Cox methodology applied to interlaboratory comparison program for flow measurement

N M Taira¹, O S Yoshida¹, M R Avelino²,³
¹ Institute for Technological Research of São Paulo, Av. Prof. Almeida Prado, 532, São Paulo, SP, 05508-901, Brazil
² National Institute of Metrology, Quality and Technology - INMETRO, Xerêm, Brazil
³ State University of Rio de Janeiro, UERJ, Rio de Janeiro, Brazil

E-mail: nmtaira@ipt.br, olga@ipt.br, mravelino@inmetro.gov.br, Mila@UERJ.br

Abstract. This paper presents an application of the methodology developed by Cox in Interlaboratory Comparison on water flow measurement. As no reference laboratory wouldn’t be chosen, the comparison was carried out using a reference value by consensus estimated from measurements reported by participating laboratories.

1. Introduction

In Brazil, 2008, the Technical Committee in Flow Measurements – TC-13 establishes the first Interlaboratory Comparison (ILC) Program for water flow measurement. Due to the lack of proficiency of laboratories, on flow meter calibration, to meet the accreditation requirements as in NIT-DICLA-026 and in ISO 17025, this intercomparison program was the one of the key actions of the Committee.

In accordance with the General Coordination for Accreditation (CGRE) – NIT-DICLA-042, a technical subcommittee was created involving representatives of the seven laboratories participating of the Interlaboratory Comparison Program. The laboratories were accredited by CGRE in the field of liquid flow and applications. One of the first actions of the Subcommittee was to produce the Protocol of the ILC, whose main propose was to harmonize the information and establish the requirements and procedures to be complied with by the participating laboratories. The Protocol, following the instructions of DOQ-CGCRE-005 – Guidelines for Interlaboratory Comparisons, review of February 1st, 2010, recommended the nomination of a reference laboratory to provide the reference value to be compared with declared values by each participant laboratory. The ILC selected two meters type mass Coriolis as travelling standards, and fortunately, one of them presented reasonable stability during the entire program, which lasted over 18 months. However, the ILC results were not well accepted because of unexpected rejection of two laboratories, which raised the possibility of not reliability of the reference value.

After the document DOQ-CGCRE-005 was withdrawn in 2012, the TC-13 held a second edition of ILC, this time without the nomination of a reference laboratory to indicate the reference value, but with the reference value estimated from the values of each participant laboratory. The reasons for this decision were due to low maturity and experience of participant’s laboratories, besides, the Cox [1] procedures for the statistical analysis of comparison measurements could overcome the difficulties on...
comparing data with large range on uncertainties and from laboratories with high discrepancies in some flow rates.

The section 2 of this paper presents the Cox methodology (Procedure A) for calculating the reference value as the weighted mean together with tests to check the consistency of a comparison data of four laboratories, that fails because of discrepancies within laboratories at low flow rates that could not be removed from data set. As the consistency test failed, the Cox-methodology (Procedure B) was applied using as a reference value the median since it is robust for the inconsistencies not removed from the data set. Section 3 presents merits of the Cox methodology applied on the comparison data of seven participating laboratories in ILC since the first edition of the program.

2. Calculation methodology to flowrate interlaboratorial comparison programs

The most preferred performance statistic for ILC is the $E_N$ number, as recommended by the extinct DOQ-CGCRE-005-2005 and by the standard ABNT NBR ISO/IEC 17043:2011 [2]. The definition of $E_N$ is presented in equation (1),

$$E_{N_i} = \left(\text{value declared by laboratory } i - \text{reference value}\right) \frac{1}{U_{\text{lab}_i}^2 + U_{\text{reference}}^2}$$

where

$U_{\text{lab}_i}$ Expanded uncertainty of laboratory $i$,

$U_{\text{reference}}$ Expanded uncertainty of the reference value.

The criteria limit is 1, so the measurement is considered consistent if $E_N<1$ or inconsistent for $E_N>1$. Cox [1] presents procedures developed to obtain the key reference value (KCRV) for comparisons involving key laboratories of National Institutes of Metrology (NMI). The procedure proposed by Cox has been applied on many Interlaboratory Comparison Programs, and it is well documented in metrology literature: as in Mikan [3] and in Arias et al. [4]. Although the Cox methodology was developed for comparison of key laboratories it is also very suitable for common laboratory comparisons through efficient use of more sophisticated statistical tools.

2.1. “Procedure A” proposed by Cox

The Procedure A estimates the reference value by the mean of the values declared by the laboratories weighted by the their respective standard uncertainties, according to equation (2),

$$y = \frac{x_1}{u_{x_1}^2} + \frac{x_2}{u_{x_2}^2} + \cdots + \frac{x_n}{u_{x_n}^2}$$

where $x_1, x_2, \cdots, x_n$ are the errors of the meter at a nominal flow rate, as declared by the laboratories 1, 2, ⋯, n, and $u_{x_1}, u_{x_2}, \cdots, u_{x_n}$ are the uncertainties of the errors reported by the respective laboratories 1, 2, ⋯, n.

The uncertainty of the reference value is obtained by equation(3)

$$\frac{1}{u_{y}^2} = \frac{1}{u_{x_1}^2} + \frac{1}{u_{x_2}^2} + \cdots + \frac{1}{u_{x_n}^2}$$

The data consistency test, equation (4-5), or chi-square test, verifies the validity of the dataset to produce a reliable comparison reference value (CRV),

$$Pr\{\chi^2(\nu) > \chi^2_{\text{obs}}\} < 0.05$$
For the test it is necessary that all measurements are valid.

Having validated the data set, the next step is to determine the discrepancies between the error stated by the laboratories relative to the reference error, \( y = x_{\text{ref}} \), i.e., \( d_i = x_i - x_{\text{ref}} \). The acceptance criteria is presented in equation (6) or equation (7),

\[
|d_i| \leq 2u(d_i) \quad \text{or} \quad E_N = \left| \frac{d_i}{2u(d_i)} \right| \leq 1
\]

where \( u(d_i) = \sqrt{u_{x_i}^2 - u_{x_{\text{ref}}}^2} \).

The four participating laboratories calibrated the standard meter in ten points in the range 15 to 3000 L/h. Figure 1 shows the flow meters calibration curves and the weighted mean curve as the reference value (CRV) curve.

![Flow meters calibration curves by four laboratories and the reference curve (CRV)](image)

**Figure 1.** Flow meters calibration curves by four laboratories and the reference curve (CRV)

Figure 2 presents the curves \( (E_N) \). It is clear from the Figure 2 that, according to the criteria defined by the equation 7, results obtained for L1 laboratory are not consistent for flow rates above 750 l/h, and the laboratories L3 and L4 are not consistent on others points.

Table 1 shows, for each flow rate, the reference values (CRV) with their uncertainties, and the results of data consistency tests in (4). It can be observed in Table 1 (a) that only four points could be used for comparison of the results. However, as recommended in Cox (2002), the data which could have generated the inconsistencies may be withdrawn. Thus, Table 1 (b) shows the results obtained without taking into account the laboratory dataset L1, which had the highest amount of \( (E_N) \) points above the acceptable limit (Fig.2).
Figure 2. $E_N$ values of four participating laboratories obtained by Procedure A.

Table 1. Reference values and data consistency test, according to procedure A.

| Flowrate (L/h) | CRV (%) | standard uncertainty of CRV (%) | $\chi^2_{obs}$ | Flowrate (L/h) | CRV (%) | standard uncertainty of CRV (%) | $\chi^2_{obs}$ |
|----------------|---------|---------------------------------|----------------|----------------|---------|---------------------------------|----------------|
| 15             | -0.122  | 0.071                           | inconsistent   | 15             | -0.152  | 0.133                           | inconsistent   |
| 30             | 0.988   | 0.098                           | inconsistent   | 30             | 1.085   | 0.113                           | inconsistent   |
| 60             | 1.288   | 0.079                           | ok            | 60             | 1.329   | 0.095                           | ok            |
| 120            | 1.289   | 0.075                           | ok            | 120            | 1.274   | 0.088                           | ok            |
| 350            | 0.613   | 0.034                           | ok            | 350            | 0.604   | 0.035                           | ok            |
| 750            | 0.027   | 0.033                           | inconsistent   | 750            | -0.012  | 0.035                           | ok            |
| 1000           | -0.357  | 0.033                           | inconsistent   | 1000           | -0.384  | 0.034                           | ok            |
| 1500           | -0.598  | 0.058                           | ok            | 1500           | -0.688  | 0.071                           | ok            |
| 2500           | -0.982  | 0.056                           | inconsistent   | 2500           | -1.094  | 0.068                           | ok            |

(a) All four laboratories  
(b) Without laboratory L1

Although discrepancies still remain at the two lowest flowrates of the calibration range, neglect them is not a reasonable decision, once it is expected that flow meters have low performance at low flow rates. For this case, Procedure A is not made viable, even by following the recommendations in Cox (2002), and one may proceed to the application Procedure B, which is robust for datasets with some discrepant measurements.

2.2. “Procedure B” proposed by Cox  
Procedure B associates each value declared by the participating laboratory to the normal distribution and to the mean $\mu$ and standard deviation. Samples are generated from the distribution by simulation and their median, the mean of the medians is the reference value.
Figure 3 shows examples of histograms of the values of the median estimated from the simulation (10^6 samples) of results of four laboratories and their distributions reference value calculated according to the Procedure B.

Thus, one can define a new equation for the normalized/standard error $E_N$ according to the equation (7), defined as the degree of equivalence according to procedure B. Cox, which is calculated according to ISO GUM [5] also adopted by Arias et al. [4], and also according to Appendix B of Cox(2002). In his work, Cox suggests to estimate the lower confidence interval for each value of $d_i$.

Table 2 presents the results of the ILC in a volume of 100 mL artifact conducted within the CIPM, where are presented the DoE calculated values and the calculated values according to the computational tools called $\text{Cox}_B$ and $\text{Cox}_B\text{ _full}$, which are described in item 2.3.

In Table 2 it is possible to notice the adherence of the DoE declared values and the ones calculated with the developed computational tools. Recalling that those are not exact values once they are calculated by using the simulation date.

### Table 2. Application of Cox methodology for CIPM key comparisons.

| NMI    | Declared values $x_i$(mL) | $x_i$(mL) | DoE (Degree of Equivalence) |
|--------|---------------------------|-----------|-----------------------------|
|        |                           |           | Cox Arias* | Cox B | Cox B full |
| CENAM  | 99.89350                  | 0.00077   | 0.44        | 0.44  | 0.45       |
| NRC    | 99.89780                  | 0.00080   | 1.80        | 1.81  | 1.87       |
| SP     | 99.89500                  | 0.00160   | 0.31        | 0.32  | 0.32       |
| IMGC   | 99.89300                  | 0.00083   | 0.67        | 0.65  | 0.70       |
| NMIA   | 99.89550                  | 0.00110   | 0.65        | 0.62  | 0.66       |
| INMETRO| 99.89290                  | 0.00061   | 0.81        | 0.81  | 0.87       |
| KCRV   | 99.8942                   | 0.00060   | 99.8942     | 0.00060 |

*Adjusted from de Arias et al. [4]

2.3. Computational Tools

For the purposes of the Cox Procedure B in an ILC of flow calibrations hundreds of millions of data are computed. For this reason, the use of basic computer tools has increasingly become inappropriate. Due to the considerable processing time and computational efforts the use of spreadsheet has shown
itself inconvenient when over $10^4$ samples were simulated. Thus, both procedures A and B, as proposed by Cox (2002), are developed for Mathcad® e VBA Excel® programming environments. Additionally, a Supplement CSharp Excel (Add-Ins) was specifically developed to include functions of Cox_A(); Cox_B() e Cox_B_full(). This implementation is compatible for Excel® Office® package, in versions 97, 2003, 2007 and 2010.

The Cox_A function calculates the comparison reference value (CRV), mean differences ($d_i$) and the degrees of equivalence (DoE) from the values reported by participating laboratories of a ILC according to the procedure proposed by Cox, and described in Section 2.1.

The Cox_B function calculates the comparison reference value of CRV by the "median method" calculated from the Monte Carlo simulation of M random values of the inverse of the normal cumulative probability distribution. The parameters taken into account are the declared value (mean) and the uncertainty specified by the each of the participating laboratory. According to Cox (2002), it is recommended to use $M = 10^6$ random elements to characterize the values in a population with a distribution similar to the normal.

The Cox_B function produces the following result:

- $u_{CRV}$: Standard uncertainty of the value of CRV;
- $d_i$: calculated average difference for laboratory $i$ relative to the reference value;
- $u(d_i)$: uncertainty of the mean difference $d_i$ calculated according to ISO GUM [5];
- DoE: equivalence degree defined as the magnitude of the ratio $d_i/(2u(d_i))$.

![Figure 4](image)

**Figure 4.** $E_N$ values of four participating laboratories obtained by Cox_B function.

Figure 4 shows the result of ILC with four participating laboratories, by applying the Procedure B to the same data of the Figures 1 and 2. It is observed the same trend of $E_N$ for L1 and L3 laboratories, but the laboratory L4 no longer presents results higher than recommended for ILC, as if it was possible to apply Procedure A.

The Cox_B_full function calculates the reference value in the same way that Cox_B function, but $u_{d_i}$ is calculated in accordance with Appendix B of Cox(2002) that estimates the lower confidence interval for each value.
Figure 5. $E_N$ values of four participating laboratories obtained by Cox_B_full function.

Figure 5 shows the results obtained by applying the function to Cox_B_full to the same set of data relating to Figures 3 and 4.

Comparing the results generated by the three functions described, for the same set of data, it is observed that the Cox_A generates more "conservative" results and that the Cox_B function generates more "optimistic" results, that is, according to the understanding of the participating laboratories. On the other hand, the Cox_A_full function may be considered more "careful", as shown in the DoE values, as shown in Table 2.

3. Application of the tools in the flowrate interlaboratorial comparison program

Figure 6 shows the result of ILC in mass flowrate with seven participating laboratories and only one laboratory. Results of Laboratory E are considered to be unsatisfactory. For this ILC one laboratory was designed to be the reference laboratory, as described in item 1.

Figure 6. Example of $E_N$ values in ILC mass flow with seven participating laboratories
The application of Cox_A function results in inconsistencies in various flowrates. By applying the Cox_B_full function it is possible to notice the existence of two groups of laboratories whose artifact calibration curves have different shapes in relation to the CRV reference value, as shown in Figure 7.

Figure 7. Mass flowrate meter calibration curves relative to the CRV obtained thought Cox_B_full function in ILC with seven participating laboratories

The application of Cox_B_full function generates the (DoE's) where all the laboratories showed satisfactory results, including the laboratory E. The same applies to the reference laboratory, as shown in Figure 8.

For the case shown in Figure 8 expanded uncertainties reported by participating laboratories ranged from 0.06% to 0.22%, ie, the uncertainties are not "harmonized" and these can affect laboratory results that declared the highest value and to promote the laboratory that declared lower uncertainty values, as can be deducted from equation 2 for the case of Procedure A. The same applies to Figure 3 for the case of Procedure B.

Figure 8. Curves of $E_n$ calculated by Cox_B_full function in ILC mass flow with seven participating laboratories
Assuming that all laboratories reported an expanded uncertainty uniform and equal to 0.06%, it is found that there will be no change in the form of the error curve as shown in Figure 9. The $E_N$ curves, shown in Figure 10, indicate that four laboratories had unsatisfactory results.

![Figure 9. Mass flow meter calibration curves relative to the CRV calculated as Cox_B_full function in ILC with seven participating laboratories and expanded and uniform uncertainty equal to 0.06%.

Figure 9.](image)

Assuming that all laboratories reported an expanded uncertainty uniform and equal to 0.22%, it is verified from Figure 11 that there are no changes in the form of the error curves. The $E_N$ curves indicate that all laboratories present satisfactory results, as shown in Figure 11, as was expected.

![Figure 10. Curves of $E_N$ calculated as Cox_B_full function in ILC mass flow with seven participating laboratories and expanded and uniform uncertainty equal to 0.06%.

Figure 10.](image)
Figure 11. Curves of $E_N$ calculated as $Cox_B\_full$ function in ILC mass flow with seven participating laboratories and expanded and uniform uncertainty equal to 0.22%.

4. Conclusion

Although there are several acceptance criteria and calculation methodologies for the ILC could be established by consensus or to be defined by the laboratory accreditation manager, as described in the standard ISO/IEC 17043, there are, on the field of flow measurement in Brazil, a considerable lack of knowledge and common understanding on the measurement uncertainties likely to be reached by the flow meter calibration laboratories.

The Cox methodology was also applied in ILC on water meters with the participation of manufacturer’s laboratories, laboratories of water companies and independent laboratories. During the first edition of ILC, several unsatisfactory laboratory performances occurred, resulting largely from the lack of harmonization of the declared uncertainties, which ranged from 0.01% to 0.64%. During the second edition of the ILC, there were a significant progress and an improvement in ensuring the metrological reliability in calibration activities and the water meter verification activities in Brazil, as well as the dissemination of best practices and harmonization of metrological concepts. The application of Cox procedures had a fundamental role on this improvement process.

The application of methodologies Cox proved to be flexible and efficient to promote harmonization of the laboratories performances as the application of procedure B would keep on comparison data set the results of discrepant laboratories or laboratories with uncertainties that could not cover their measurements, allowing the learning of the not mature laboratories.

References

[1] Cox M G, 2002 The evaluation of key comparison data Metrologia 39, p.589-595

[2] Associação Brasileira de Normas Técnicas, 2011 ABNT NBR ISO/IEC 17043: Avaliação de conformidade – Requisitos gerais para ensaios de proficiência. (Rio de Janeiro: ABNT)

[3] Mikan B and Valenta T, 2009 Final Report - Draft B, interlaboratory comparison calibration of the turbine gas meter G6500 EURAMET Project No. 1006, March

[4] Arias R et al., 2005 CIPM Key Intercomparison volume at 20 L and 100 mL - CCM.FF-K4, March

[5] International Organization for Standardization, ISO, 1993 Guide to the expression uncertainty in measurement.