DOI: 10.15199/17.2024.3.1

Intercomparison in the field of high-pressure calibration of gas meters

Porównania międzylaboratoryjne w zakresie wzorcowania gazomierzy na wysokim ciśnieniu.

Yuriy Brodyn, Aneta Korda-Burza, Daniel Wysokiński*

Keywords: *interlaboratory comparisons, gas meter calibration.*

Abstract
One of activities meant to confirm the validity of laboratory's results is comparison with results obtained by other laboratories. Laboratory's ability to analyse results of its participation in intercomparisons and to utilise them for effective oversight of its calibrations is an important element of assessing the competence of laboratories in accreditation and supervision processes. The paper presents issues related to the organization of comparisons, as well as the calculation methods used and the analysis of the results of interlaboratory comparisons organized by the Gas Meter Calibration Laboratory at the European level. The Gas Meter Calibration Laboratory carries out its tasks as part of Testing and Certification Division of the Gas Transmission Operator GAZ-SYSTEM S.A., which plays a strategic role in the Polish economy and is responsible for natural gas transmission.

Słowa kluczowe: *Porównania międzylaboratoryjne, wzorcowanie gazomierzy.*

Streszczenie
Jednym z działań, na podstawie którego laboratoria powinny wykazać ważność swoich wyników, jest porównanie z wynikami uzyskanymi przez inne laboratoria. Umiejętność analizy wyników uczestnictwa i ich stosowanie przez laboratoria do kontroli swoich wyników jest istotnym elementem oceny kompetencji laboratoriów w procesach akredytacji i nadzoru. W artykule przedstawiono zarówno kwestie związane z organizacją porównań jak również zastosowane metody obliczeniowe oraz analizę wyników porównań międzylaboratoryjnych, jakie zostały zorganizowane przez Laboratorium Wzorcowania Gazomierzy na poziomie europejskim. Laboratorium Wzorcowania Gazomierzy realizuje swoje zadania w ramach Pionu Badań i Certyfikacji Operatora Gazociągów Przesyłowych GAZ-SYSTEM S.A., który pełni strategiczną rolę w polskiej gospodarce i odpowiada za przesył gazu ziemnego.

1. Introduction

In accordance with the standard PN-EN ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories, laboratories should monitor their activities by comparing them with the results of other laboratories, taking into account participation in proficiency tests and/or participation in interlaboratory comparisons other than proficiency tests. The standard requires laboratories to plan their activities in this area and then review them for analysis, use for oversight and improvement of laboratory activities [7]. In 2021-2022, an international intercomparison in the field of high-pressure calibration of gas meters was performed between five European calibration laboratories with the purpose of both proving the competence in conducting calibrations of gas-meters and to check the Lab-to-Lab traceability on a lower level, where calibration routines of metrologically subordinate laboratories were compared. While giving short description of the participating laboratories, test objects and the program, the presentation addresses some ILC superintendence issues and mathematical apparatus utilized, the major focus being into the results obtained. In the field of high pressure recalibrations of industrial-size gas meters we currently have explicit harmonization between the NMI's. Harmonization intercomparisons are scheduled and performed and its results are published by EuReGa consortium (European References for Gas) on regular basis. But according to the rules of EuReGa [9] only laboratories that has realized the national high-pressure cubic meter entirely independent of other laboratories can participate in these comparisons (hereinafter "**primary-level labs**"). Meanwhile, the rest of the calibration labs (hereinafter "**secondary-level labs**") that receive its traceability from one of the primary-level labs do not have adequate data to directly prove mutual traceability with similar secondary-level labs. Never the less, these labs still must prove their proficiency to local accreditation bodies by conducting PT/ILC on regular basis, being usually bound solely to bilateral comparisons with the primary-level labs being their source of traceability (hereinafter "**source labs**"). Although there are multiple reports of inter-comparisons conducted on atmospheric air under the auspices of EURAMET [6,10], no comparisons for pressurized natural gas were performed previously between secondary-level labs. In 2021 this situation encouraged Gas Meter Calibration Laboratory "Laboratorium Wzorcowania Gazomierzy LWG" to initiate first

---

*) **Yuriy Brodyn, Dr.Ing.** – Operator Gazociągów Przesyłowych GAZ-SYSTEM. S.A. Expert of the Gas Meter Calibration Laboratory, yuriy.brodyn@gaz-system.pl; **Aneta Korda-Burza** – ORCID:0000-0002-2910-1011, Operator Gazociągów Przesyłowych GAZ-SYSTEM. S.A., Director of the Research and Certification Division; **Daniel Wysokiński** – Operator Gazociągów Przesyłowych GAZ-SYSTEM. S.A., Chief of the Gas Meter Calibration Laboratory

horizontal interlaboratory comparisons between European secondary-level labs in the field of high-pressure calibration of gas meters (hereinafter “**ILC**”).

Three major targets was set for the ILC:

1. Finding partners for ILC among secondary-level EU (and non-EU) labs that derive its traceability from distinct source primary-level labs;
2. Organizing ILC according to PN-EN ISO/IEC 17043:2011 [8] and EA-4/21 [3] in maximum possible range, suitable for all participants (working flow-rates and pressures);
3. Processing the results based on different principals:
    - comparing participants results to the reference value obtained by a recognized primary-level lab
    - comparing participants results to so called “Comparison Reference Value” (hereinafter “**CRV**”) derived from results of all participants;
    - cross-comparing participants results between each other.

2. Organization of interlaboratory comparisons

2.1. Participants

At the initial stage, invitations were sent to 43 laboratories worldwide, including EU, UK, eastern Europe, USA and Asia, including 17 independent commercial labs, 7 manufacturer’s labs and 19 metrological institutes. 10 labs expressed preliminary interest. Finally ILC was conducted between 6 participants, including FORCE Technology, Denmark as a reference laboratory and 5 second-level labs: pilot lab – Gas Meter Calibration Laboratory “Laboratorium Wzorcowania Gazomierzy LWG” by Gaz-System S.A., Poland, Bishop Auckland Research and Testing Facility by DNV, Great Britain, Gas Meter Calibration Laboratory LACAP by Enagás S.A., Spain, qbig GmbH, Germany and RMA Mess – u. Regeltechnik GmbH & Co.KG, Germany.

Concerning organizational matters participating laboratories in general terms hereinafter jointly are referred to as "**Participants**", and each individually as a "**Participant**". The organizer of ILC – Gas Meter Calibration Laboratory “Laboratorium Wzorcowania Gazomierzy LWG” by Gaz-System S.A. – hereinafter is referred to as “**Organizer**”, and the reference laboratory FORCE Technology hereinafter is referred to as “**Reference lab**”. Following the requirements of ISO 17043, for the purpose of this publication results are anonymized, participating laboratories (except Reference Lab) are referred to as “**Lab.1**”, “**Lab.2**” etc.

2.2. Test objects

Upon discussing of Participants’ harmonization needs and calibration ranges, Organizer has supplied the test objects for ILC – travel meters of turbine and ultrasonic types (hereinafter “**TRM**” and “**USM**” respectively): one 8-path ultrasonic DN300 meter with spool pieces and a flow conditioner, and two turbine meters – DN150 and DN80. All three meters have demonstrated long-term stability, that was evaluated based on their calibration history. Utilised formulas and criteria’s border values are described below.

The explicit description of the travel meters, including dimensioning and packing specifications, handling, conditioning, connection instructions etc. were shared between the participants (see figure 1 for example).

2.3. Calibration scope

After analyzing CMC of all Participants, the scope of calibration was set for each travel meter. Fig. 2 below presents the example for USM. The pressure of 38 bar was set for comparison with the Reference Laboratory. Another three pressures were agreed for comparison to mutually established CRV and cross-comparisons between the Participants, namely: 50, 16 and 8 bar.

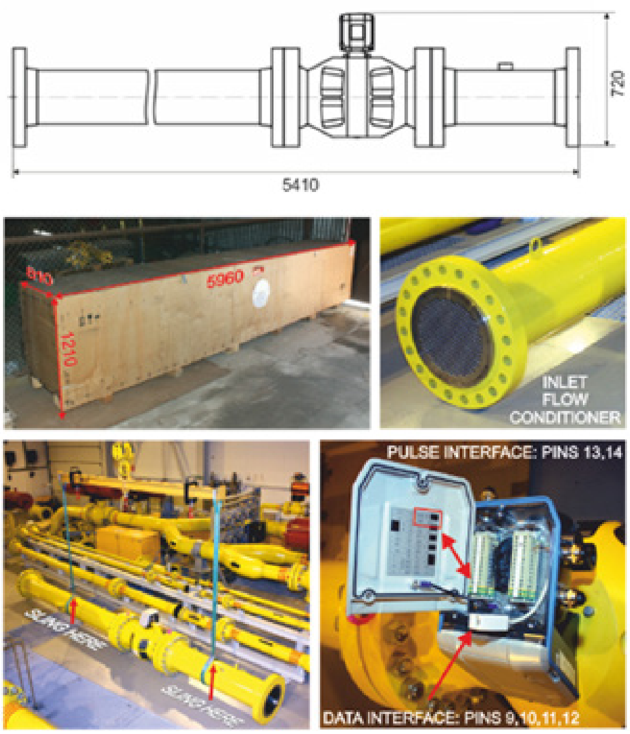


Fig. 1 Test objects description (example for USM)

Rys.1 Opis obiektów testowych

Note: hereinafter the gauge pressure is used. The common flow-points were selected according to the meter flow-range and the working range of each Participant.

2.4. Time schedule and logistics

Considering significant geographical dispersion of the Participants (see fig. 3) one of the most demanding organizational activities considering the requirements of ISO 17043 was arranging the logistics while ensuring constant supervision over the test objects.

The meters were transported in specially constructed transportation boxes to ensure no damage during shipment. The boxes were furnished with specially designed lodgements. Boxes for turbine meters were equipped with the anti-vibration feet. All boxes were equipped with shock indicator, corresponding to cargo size and weight. The condition of the indicators was checked each time prior to and after the loading/unloading operations (see fig. 4).

After calibration were finished, the inlet and outlet of the travel meters was to be sealed with the appropriate stickers (the set of spare stickers was supplied by the Organizer and included in the transportation boxes), then meters were to be sheathed into the transportation boxes, prepared to shipping and stored according to the Participant’s procedures until dispatching.

To minimise risks of damage during the transportation only dedicated transport was utilised (no transhipment of cargo was allowed, only direct shipment between labs). To ensure maximum control over the travel meters during their transportation, the online GPS monitoring of cargo was introduced, allowing the Organiser to track cargo status during the transportation.

To document that no damage has occurred during the transport and handling the Record of transportation and handling of the travel meters (hereinafter “**RTH record**”) was introduced, describing proper actions that each of the Participants performed: upon arrival and prior to departure the visual check of the transportation boxes, including lock mechanisms, was performed, condition of the shock indicators was checked and documented in the RTH record, travel meters were photographed. Any observed damages or nonconformities was to be

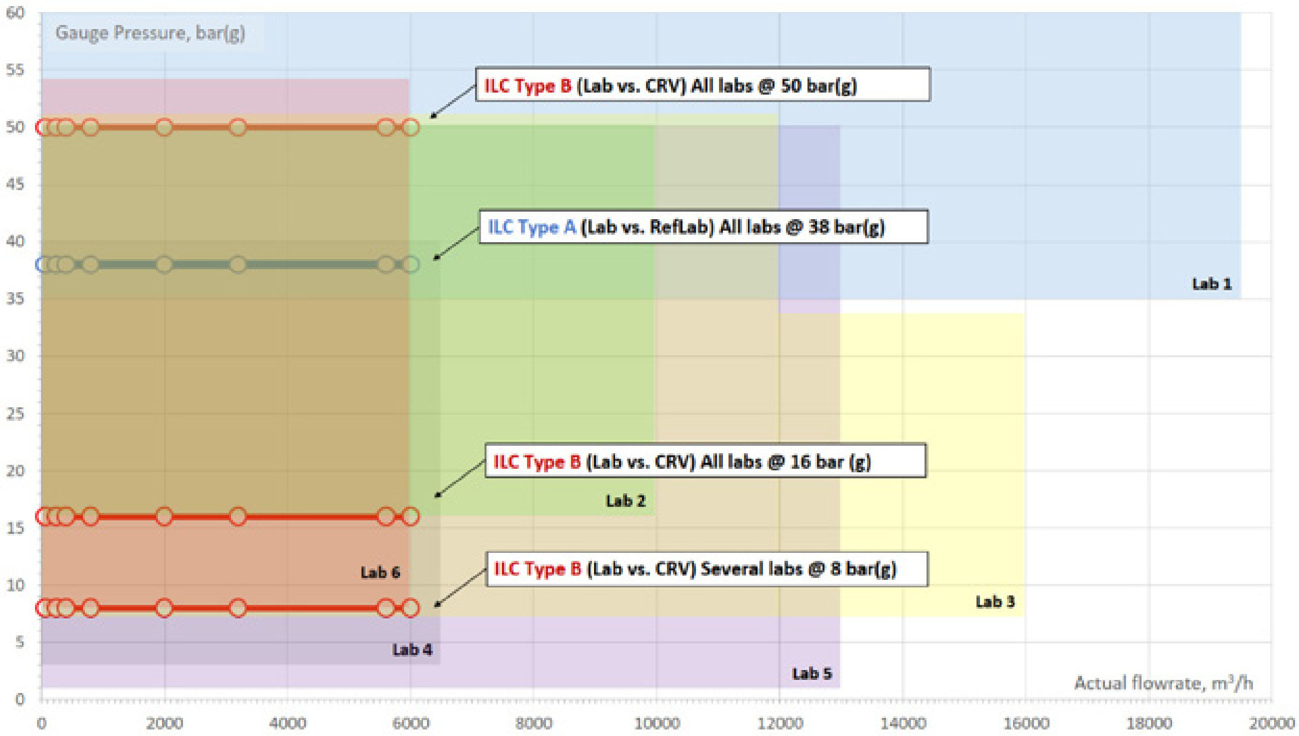
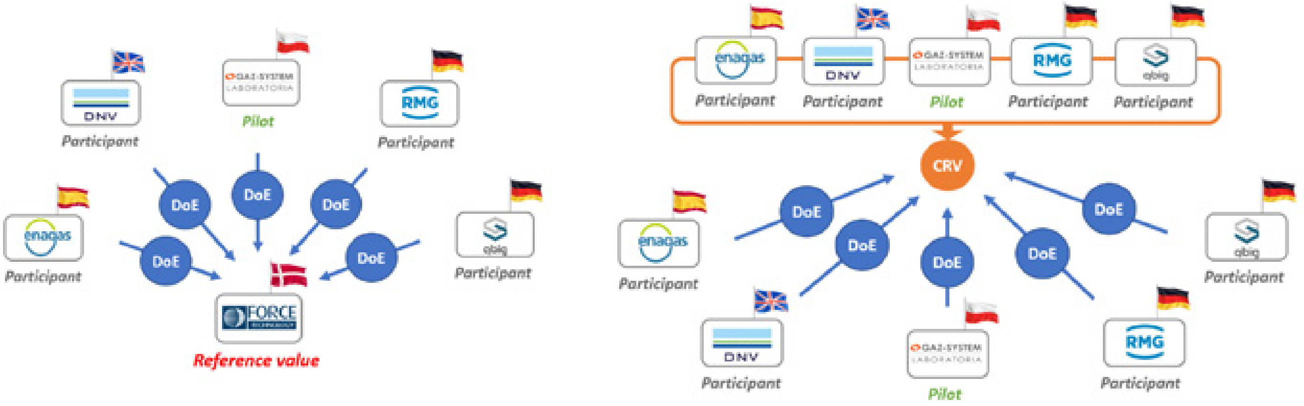


Fig. 2 Example of setting the mutual calibration scope for USM DN300,
Rys.2 Przykład ustawienia zakresu wzajemnej kalibracji dla USM DN300

Fig. 3 ILC logistics

Fig. 4 Handling of the travel meters.
Rys.4 Obsługa liczników podróży



a) ILC Type A *b) ILC Type B*

Fig. 5 Test methods

Rys.5 Metody testowe

documented, RTH record was filled in, signed, scanned and sent to the Organiser together with the photographs the same day the meters were sent to the next laboratory. No damages or nonconformities was documented during the ILC, RTH records were filled in, sent to and stored by the Organizer.

3. Test methods

Intercomparisons included two major type of tests further referred to as "**ILC Type A**" and "**ILC Type B**". As a measure of equivalence, the **D**egree **o**f **E**quivalence was introduced, similar to E_n value calculated in accordance with [8], Appendix B, equation (B.5). Data reporting and processing as well as evaluation of the equivalence was performed based on the provisions for the DoE limits as in effect in best known EURAMET practices [6, 10]. For the purpose of this document the DoE hereinafter will be represented by the ***En*** value.

ILC Type A were bilateral comparisons where the result of each Participant were compared to the assigned value determined by calibration in the external well recognized laboratory, in our case – Force Technology (see figure 5, *a*). Within a scope of Type B comparisons, the assigned value was determined based on the calibration results of all the participating laboratories, weighed by means of their uncertainties, considering possible correlations (see fig. 5, *b*). The **C**omparisons' **R**eference **V**alue **CRV** was established. The Chi-Squared test was applied to confirm the consistency of each result and whether it has a contribution to CRV. Utilised formulas and criteria border values are described below.

Calibration certificates was sent to the Organizer immediately after the calibration were finished. Minimum data set reported per travel meter by each of the Participants consisted of: nominal calibration pressure P_n, nominal flow point Q_{nom}, real flowrate at the travel meter Q_{MUT} (measured by the reference meter(s) and adjusted for the conditions at the travel meter), gas temperature at the travel meter T_{MUT}, averaged for a given flow point Q_{nom}, errors of the travel meter calculated for each of five successive measurements $E_1 \dots E_5$, average error E_{av} and total uncertainty U_{tot} of calibration at given flow point Q_{nom} as well as gas chemical composition.

Participants were not allowed to share their results other than with the Organizer. In order to assure the impartiality, the Organizer was not allowed to share Participant's results end of ILC. In order to fulfil the requirements of [110], after calibrations were finished the results were kept confidential, code IDs was introduced (Lab.1, Lab.2 etc.), calibration results provided by each Participant to the Organizer was treated as confidential.

The error of the meter was determined for each measurement as follows:

$$e = \frac{V_{indicated} - V_{real}}{V_{real}} \times 100\ (\%) \tag{1}$$

where e is the error of the meter determined for each of $n = 5$ successive measurements; $V_{indicated}$ is the indicated volume i.e volume, measured by the travel meter; V_{real} is the real volume i.e volume, measured by the reference meter(s) and adjusted for the conditions at the travel meter.

After the calibration the average error $\bar{e}$ at a given static pressure and flowrate point was calculated.

Total uncertainty of the calibration is calculated based on EA-4/02 M [7] by using 95% confidence level for n-1 (number of successive measurements – 1) degrees of freedom, and assuming k_{95}=2. The two equations below shows how these calculations were done.

$$u_{MUT.S} = SD = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sum(e-\bar{e})^2}{n(n-1)}} \tag{2}$$

where $U_{MUT.S}$ is the standard uncertainty of the travel meter calculated for the given probability and n successive measurements (the uncertainty component due to the short-term stability of the meter, i.e. repeatability ***SD***); σ is the standard deviation of $n = 5$ successive measurements; $\bar{e}$ is the average error of the meter calculated from the errors e reported the for $n = 5$ successive measurements.

$$U_{tot} = 2 \times \sqrt{\left(\frac{U_{CMC}}{2}\right)^2 + u_{MUT.S}^2} \tag{3}$$

where U_{tot} is the total expanded uncertainty of the meter calibration in a Participant's laboratory, facilitating calibration and measurement capability (hereinafter **CMC**) and the standard uncertainty contribution due to the short-term stability of the meter (i.e. repeatability); $U_{CMC} = k_{95} \cdot u_{CMC}$ represents Participant's lab CMC i.e. expanded uncertainty of the average of n calibration results for the "Best Existing Device" using the laboratory's reference standard, for the 95 % confidence level assuming Gaussian distribution; u_{CMC} is the combined CMC standard uncertainty (the uncertainty component due to the laboratory's reference standard(s), measuring equipment, applied methodology etc.); $k_{95} = 2$ is the coverage factor used to calculate expanded uncertainty of the output estimate from its combined standard uncertainty (95 % confidence level assuming Gaussian distribution).

In addition to the short-term uncertainty contribution $u_{MUT.S}$ given by equation (2), the long-term uncertainty contribution from the travel meter in this intercomparisons was evaluated based on numerous historical calibrations performed by LWG in preceding years at different calibration pressures. The long-term uncertainty contribution is calculated based on uniform distribution between minimal and maximum errors, historically registered for the given travel meter per calibration point at different calibration pressures:

$$u_{MUT.L} = \frac{\bar{e}_{max}-\bar{e}_{min}}{2\sqrt{3}} \tag{4}$$

where $u_{MUT.L}$ is the average value of the long-term uncertainty contribution (i.e. reproducibility) based on historical calibration data; $\bar{e}_{max}$ and $\bar{e}_{min}$ are maximum and minimum average errors of the meter registered at given flowrate and static pressure, respectively.

Based on historical calibration data, the estimated long-term standard uncertainty contribution of USN DN300 travel meter equals to $u_{MUT.S}$ = 0,06% for flowrates below Qt (310 m^3/h) and $u_{MUT.S}$ = 0,03% for flowrates higher then Qt. For both TM DN150 and DN80 the estimated long-term standard uncertainty contribution equals to $u_{MUT.S}$ = 0,03% for whole working range.

The difference between the chemical composition of gases used for calibration must also be accounted for. Based on historical experience the uncertainty contribution caused by (10÷20)% change of methane fraction in gas composition may roughly be estimated as the u_{GC} = 0,05% for both ultrasonic and turbine travel meters. This value was achieved during previously conducted Join industry project (JIP) on stability of flow meters for renewable gases and was agreed among the participants.

Standard uncertainty of the intercomparisons are facilitating laboratory's CMC, both short-term and long-term stabilities of the meter as well as the gas composition contribution and thus shall be calculated as follows:

$$u_{ILC.LAB_k} = \sqrt{\left(\frac{U_{tot.LAB_k}}{2}\right)^2 + u_{MUT.L}^2 + \{u_{GC}^2\}} \tag{5}$$

where $u_{ILC.LAB_k}$ is the standard uncertainty calculated for the given Participant's laboratory *k*, referred to calibrations in this ILC; $u_{tot.LAB_k}$ is the total expanded uncertainty by the given Participant's laboratory *k*, facilitating CMC and $u_{MUT.S}$ and calculated according to equation (3); $u_{MUT.L}$ is the standard uncertainty contribution due to the long-term stability of the meter (i.e. reproducibility) as described above; u_{GC} is the standard uncertainty contribution due to the difference in gas composition as described above (must be taken into account in case the significant differences in gas compositions are reported).

Equation (5) is valid for all calibrations performed during the ILC, both at the Participant's labs and at the Reference laboratory. The expanded uncertainty referred to the ILC was calculated by multiplying the respective standard uncertainty value by coverage factor k_{95} = 2 (95 % confidence level assuming Gaussian distribution):

$$U_{ILC.LAB_k} = 2 \cdot u_{ILC.LAB_k} \tag{6}$$

3.1. Reproducibility checks

In order to fulfil the demands of [10] considering supervision over the test objects during ILC, the Organiser introduced systematic reproducibility checks of the travel meters. Two types of reproducibility check were suggested: 1) based on comparing of the current Participant's result to the Reference laboratory result (hereinafter "**Participant-to-Reference reproducibility check**"), and 2) based on comparing of two Organizer's results between each other (hereinafter "**Organizer-to-Organizer reproducibility check**").

The idea of Participant-to-Reference reproducibility check was to make sure that: 1) no unexpected accidental change of the meter's metrological behaviour has occurred while transporting to and handling in the Participant's lab, and 2) the Participant's results are reliable and may be taken for further processing.

The Organizer-to-Organizer reproducibility check was to be performed in case Participant-to-Reference reproducibility check showed possible change of the reproducibility – in this case meter was to be delivered back to the Organizer where re-calibration was to be performed and the Organizer's result received for given travel meter on the initial stage was to be compared to the result of re-calibration.

The Organizer-to-Organizer reproducibility check was also performed on the final stage of the ILC to prove no significant changes to the metrological performance has occurred during the comparisons and thus the ILC results are reliable – in this case the Organizer's result received on the initial stage was compared to the Organizer's result received on the final stage of the ILC.

Participant-to-Reference reproducibility check was performed at 38 bar. Since first scheduled calibration it he scope of ILC was calibration in Reference lab at 38 bar (see fig. 4), it was decided to start calibration in sequent Participant's lab from this pressure. After Participant has finished the 38 bar calibration and sent the results to Organizer, the Participant-to-Reference reproducibility check was performed and based on that Organizer decided whether no change of the reproducibility was registered and thus calibration may be continued on other pressures.

For this check typical E_n criterion as described in [10] was utilised as DoE (formulas used for its calculation will be described in the following).

Considerations presented in [7] proves that the values of $E_n \leq 1{,}2$ reflects the dominance of non-stochastic uncertainty components compared to the stochastic influences thus these values are usually chosen for comparisons for gas flow. Values of $E_n \leq 1{,}5$ reflects the dominance of stochastic influences in the uncertainty budget, but still acceptable for some other types of comparisons. Otherwise, values of $E_n > 1{,}5$ are indicating that the reproducibility of the meter becomes inferior to the possible stochastic influences and thus should be investigated. Taking into account these considerations the upper threshold of the reproducibility criterion here was set to $E_n = 1{,}5$.

This means that if Participant-to-Reference reproducibility check gives $E_n \leq 1{,}5$ in all calibration points, the meter reproducibility is considered **acceptable** (the travel meter characteristic is stable, Participant's result may be non-equivalent to the assigned value, but still is reliable and thus may be taken for further processing). Otherwise, if Participant-to-Reference reproducibility check_gives $E_n > 1{,}5$ in any calibration point, the meter reproducibility is considered being in the **warning level** (some unexpected change of the travel meter characteristic may occurred, or otherwise Participant's result is either not reliable and thus may not be taken for further processing, or is reliable but considerably divergent and thus non-equivalent to the reference value).

In later case, if no damages or nonconformities were recorded during the transportation/handling (lack of corresponding entries in the RTH record), the organiser was to require the Participant to analyse the adequacy of its results. If the Participant confirmed that no mistakes or nonconformities were made during the calibration and the results are adequate, then the Organiser was to take a decision regarding the practical implications and the necessity of the re-calibration of the travel meter(s) in its laboratory to check whether no unexpected changes in its characteristic(s) has occurred.

In case the decision stating necessity of the re-calibration would be taken, the Organizer was to arrange transportation of the travel meters from the Participant's lab back to the Pilot lab (LWG), where the meter-under-question was to be re-calibrated and its reproducibility was to be checked utilising the Organizer-to-Organizer reproducibility check. No such situation occurred during the ILC. The Organizer-to-Organizer reproducibility check was performed on the final stage of the ILC proving no significant changes to the metrological performance has occurred during the comparisons and thus the ILC results were reliable.

Determination of reproducibility in the Pilot lab (Organizer-to-Organizer reproducibility check) was performed utilising the method implemented in the accredited quality system of LWG, based on the variance analysis. As a measure of reproducibility, the standard deviation is taken as an expression of the variability of the results within each series, and the variability of the results between series (caused by a specific factor i.e. measurement conditions, etc.).

The standard deviation of the reproducibility S_R is calculated as:

$$U_{ILC.LAB_k} = 2 \cdot u_{ILC.LAB_k} \tag{7}$$

where S_r is the standard deviation of the repeatability for all results (reflecting short-term stability $u_{TM.S}$ given by equation (2) above), assuming no significant differences between the standard deviation of each series, and S_L is the standard deviation between the series (reflecting long-term stability).

The standard deviation S_L is formed by the standard deviation characterizing the dispersion of the means around the overall mean, corrected for the standard deviation of the repeatability. It is calculated according to the equation below:

$$S_L = \sqrt{\frac{|\bar{S}^2 - S_r^2|}{n_0}} \tag{8}$$

where $\bar{S}$ is the standard deviation of the mean of each series, and n_0 is the average number of series of m total series. These parameters are calculated according to the following formulas:

$$\bar{S} = \sqrt{\frac{\sum_1^m n \cdot (\bar{e} - \bar{\bar{e}})^2}{m-1}} \tag{9}$$

$$n_0 = \frac{1}{m-1}\left(\sum_1^m n - \frac{\sum_1^m n^2}{\sum_1^m n}\right) \tag{10}$$

where $n = 5$ is the number of successive measurements during the given calibration in $m = 2$ total series (series no 1 being initial calibration and series no 2 being check calibration). The variable $\bar{\bar{e}}$ is the overall mean of all results, defined by the following equation:

$$\bar{\bar{e}} = \frac{\sum_1^m n \cdot \bar{e}}{\sum_1^m n} \tag{11}$$

Based on the standard deviation of reproducibility, the reproducibility limit is determined, i.e. the upper threshold value which, with a certain probability, should not be exceeded by the absolute value of the difference of two consecutive calibration series performed under reproducibility conditions. The threshold of the reproducibility criterion here is given by the following formula:

$$r_R = t(P, f) \cdot \sqrt{2} \cdot S_R \tag{12}$$

were $t(P, f)$ is the Student's t-value for the $f = n - 1$ degrees of freedom and the confidence level of $P = 95\%$. Usually for practical applications $t(P, f) = 2$ is sufficient. The $\sqrt{2}$ multiplier reflects the fact that the difference of two results is being compared.

The criterion for the reproducibility condition for the travel meters is established based on the historical calibration data. This criterion depends from the type of the travel meter. For both turbine meters TM DN80 and TM DN150 the upper threshold of the reproducibility criterion is $r_R = 0{,}2\%$ for the whole working range. For the ultrasonic meter USM DN300 the upper threshold of the reproducibility criterion is $r_R = 0{,}6\%$ for flowrates Q ≤ Qt (Qt= 310 m^3/h), and $r_R = 0{,}3\%$ for flowrates Q > Qt.

It means that if Organizer-to-Organizer reproducibility check shows $r_R \leq 0{,}2\%$ for TM DN80 or TM DN150, or $r_R \leq 0{,}6\%$ at the flowrates lower or equal to 310 m^3/h for USM DN300, or $r_R \leq 0{,}3\%$ at the flowrates higher that 310 m^3/h for USM DN300 — then the meter reproducibility is considered **acceptable**, acknowledging that no unexpected change of the meter characteristic has occurred and thus calibration results received earlier are reliable**.**

Otherwise if Organizer-to-Organizer reproducibility check shows $r_R > 0{,}2\%$ for TM DN80 or TM DN150, or $r_R > 0{,}6\%$ at the flowrates lower or equal to 310 m^3/h for USM DN300, or $r_R > 0{,}3\%$ at the flowrates higher that 310 m^3/h for USM DN300 — then the meter reproducibility is considered being in the **warning level**, meaning that the additional investigation shall be undertaken to establish the cause-and-effect relationship and decide whether or not given meter may still be used for the ILC.

The above-described method was to be utilised both to check the reproducibility of the travel meter(s) during the ILC (optionally, in case the decision of re-calibration was taken), and to perform the final reproducibility check at the end of the comparisons (mandatory), when the travel meters returned to LWG after all calibrations were finished.

3.2. Data processing – ILC type A

For the ILC Type A results of each of the Participants were considered as the Participant's results, and results of the Reference Laboratory were considered as the reference value and thus the assigned value. Evaluation of the equivalence between the Participant's laboratory k and the Reference Laboratory, i. e. the "**Equivalence Lab-to-RefLab** " was performed as following.

First the DoE values were calculated according to equation below:

$$En_{LAB_k.vs.RefLab} = \left|\frac{\bar{e}_{LAB_k} - \bar{e}_{RefLab}}{U_{LAB_k.vs.RefLab}}\right| \tag{13}$$

where $En_{LAB_k.vs.RefLab}$ is the DoE in the ILC type A, used to evaluate the equivalence between the results of Participant's laboratory k and the results the Reference Laboratory; $\bar{e}_{LAB_k}$ is the Participant's result i.e. average error in a calibration point determined by the Participant's laboratory k; $\bar{e}_{RefLab}$ is the reference value i.e. average error in the same calibration point determined by the Reference Laboratory; $U_{LAB_k.vs.RefLab}$ is the expanded uncertainty of the offset between the average result $\bar{e}_{LAB_k}$ of Participant's laboratory k and average result of the Reference Laboratory.

Uncertainty of the offset between two results is depending on the possible correlation between these input values. Based on the equation describing the pure propagation of (standard) uncertainty of the offset between two variables x_1 and x_2, the (standard) uncertainty of the offset is the quadratic sum of the uncertainties of the inputs (u_1 and u_2) subtracting twice the covariance (cov) between the two input values [6,10]:

$$u_{x_1-x_2}^2 = \begin{bmatrix} \frac{\partial(x_1-x_2)}{\partial x_1} & \frac{\partial(x_1-x_2)}{\partial x_2} \end{bmatrix} \begin{bmatrix} u_1^2 & cov \\ cov & u_2^2 \end{bmatrix} \begin{bmatrix} \frac{\partial(x_1-x_2)}{\partial x_1} \\ \frac{\partial(x_1-x_2)}{\partial x_2} \end{bmatrix} = u_1^2 + u_2^2 - 2 \cdot cov \tag{14}$$

As far as the ILC type A is concerned, there are two particular cases describing possible correlation between input values, they are as follows.

Case A1: ILC type A, offset "LAB-to-RefLab", independent laboratories

If the Participant's laboratory k is not directly traceable to the Reference Laboratory i.e. does not calibrate its reference standards in the Reference Laboratory, then these two labs are considered as independent and there is no covariance between their results. It this case standard uncertainty of the offset between the results $\bar{e}_{LAB_k}$ and $\bar{e}_{RefLab}$ is calculated as follows:

$$u_{LAB_k.vs.RefLab} = \sqrt{u_{ILC.LAB_k}^2 + u_{ILC.RefLab}^2} \tag{15}$$

where $u_{ILC.LAB_k}$ and $u_{ILC.RefLab}$ are the expanded uncertainties referred to the ILC calculated correspondingly for the Participant's laboratory k and the Reference Laboratory according to equation (5).

Case A2: ILC type A, offset "LAB-to-RefLab", dependent laboratories

If the Participant's laboratory k is directly traceable to the Reference Laboratory i.e. if its reference standards were calibrated in the Reference Laboratory, then Participant's laboratory is metrologically depended from the Reference Laboratory and thus the results from these two labs are correlated – the covariance between these results is associated with the uncertainty due to the shared traceability. For practical purposes it is appropriate to determine the conservative estimate of an upper limit of this covariance. The upper limit is determined for the theoretical case if there are no additional stochastic influence in the traceability of the laboratory from its source. Then the results of this Participant's laboratory *k* would be strongly correlated with results of the Reference Laboratory (correlation coefficient = 1). Based on the practical experience gained from the historical bilateral intercomparisons with the Reference Laboratory being the participant of EuReGa group, the Organizer has set this shared uncertainty to be $u_{Shared} = 0{,}05\%$.

Note: Prior to ILC this value was consulted with FORCE Technology and was set to 0,05% based on the experience gained from the harmonization between the laboratories being active members of the EuReGa Consortium.

It this case the covariance is expressed as and considering equation (14) the standard uncertainty of the offset between the results and is calculated as follows:

$$u_{LAB_k.vs.RefLab} = \sqrt{u^2_{ILC.LAB_k} + u^2_{ILC.RefLab} - 2 \cdot u^2_{Shared}}, \tag{16}$$

with = 0,05%.

Equations (15) and (16) use the standard uncertainties. The expanded uncertainty $U_{LAB_k.vs.RefLab}$ used in equation (13) is determined by:

$$U_{LAB_k.vs.RefLab} = 2 \cdot u_{LAB_k.vs.RefLab}. \tag{17}$$

After calculating the *En* values, evaluation of the equivalence "Lab-to-RefLab" was performed based on the provisions for the DoE limits as in effect in best known EURAMET practices [3, 4]:

- absolute values ***En*** **≤ 1.0** were considered **acceptable** (results subject to comparisons are equivalent);
- absolute values **1.0 <** ***En*** **≤ 1.2** would be considered as being in the **warning level** (results subject to comparisons may still be considered as equivalent, but check actions are recommended);
- absolute values ***En*** **> 1,2** would be considered as **unacceptable** (results of the laboratories subject to comparisons are not equivalent and thus should be investigated).

3.3. Data processing – ILC type B

For the ILC Type B results of each of the Participants were considered as the Participant's results, and the assigned value i.e. the CRV was determined based on the calibration results of all the participating laboratories. Evaluation of the equivalence between the Participant's laboratory *k* and the Participant's laboratory *l* (with $k \neq l$) i. e. the "**Equivalence Lab-to-Lab**", as well as the equivalence between the Participant's laboratory *k* and the CRV, i. e. the "**Equivalence Lab-to-CRV**" were performed as follows.

Determination of the CRV

First the reference value was calculated as weighted mean error (weighed by means of Participant's uncertainty):

$$y = \frac{\frac{\bar{e}_{LAB_1}}{u^2_{ILC.LAB_1}} + \frac{\bar{e}_{LAB_2}}{u^2_{ILC.LAB_2}} + \dots\dots\dots \frac{\bar{e}_{LAB_k}}{u^2_{ILC.LAB_k}}}{\frac{1}{u^2_{ILC.LAB_1}} + \frac{1}{u^2_{ILC.LAB_2}} + \dots\dots\dots \frac{1}{u^2_{ILC.LAB_k}}}, \tag{18}$$

where $\bar{e}_{LAB_1}$, $\bar{e}_{LAB_2}$, … $\bar{e}_{LAB_k}$ are errors of the meter in a given calibration point determined by the given Participant's laboratory *(1, 2, ... k)*; $u_{ILC.LAB_1}$, $u_{ILC.LAB_2}$, … $u_{ILC.LAB_k}$ are the standard uncertainties calculated for the given Participant's laboratory *(1, 2, ... k)*, referred to calibrations in this ILC. These uncertainties were to be determined according to the equation (5).

The standard uncertainty of the reference value is given by

$$\frac{1}{u_y^2} = \frac{1}{u^2_{ILC.LAB_1}} + \frac{1}{u^2_{ILC.LAB_2}} + \dots\dots\dots \frac{1}{u^2_{ILC.LAB_k}}. \tag{19}$$

The expanded uncertainty of the reference value U_y is calculated by multiplying the aforementioned standard uncertainty value by coverage the factor $k_{95} = 2$:

$$U_y = 2 \cdot u_y \tag{20}$$

After the calculation of the reference value the chi-squared test for consistency check was performed using the values of errors of the meter in each flow rate. At first iteration the chi-squared value $\chi^2$was calculated by

$$\chi^2 = \frac{(\bar{e}_{LAB_1} - y)^2}{u^2_{ILC.LAB_1}} + \frac{(\bar{e}_{LAB_2} - y)^2}{u^2_{ILC.LAB_2}} + \dots\dots\dots\dots \frac{(\bar{e}_{LAB_k} - y)^2}{u^2_{ILC.LAB_k}} \tag{21}$$

The degrees of freedom *ν* were calculated by

$$\nu = k - 1 = 4 \tag{22}$$

where *k* is number of the Participants taking part in calibrations. The consistency check would be failing if

$$\Pr\{\chi^2_\nu > \chi^2\} < 0{,}05 \tag{23}$$

Note: Pr denotes "probability of".

To calculate χ^2_ν the function ***CHIINV(a;n)*** in MS Excel was used with the following parameters: alpha level *α* = 0,05 (5%), degrees of freedom *ν = 4*.

The consistency check was considered failed if ***CHIINV(0,05; 4)***$<\chi^2$

If the consistency check did not fail then *y* was accepted as the **comparison reference value** $\bar{e}_{CRV}$, and U_y was accepted as the **expanded uncertainty of the comparison reference value** U_{CRV}. The standard uncertainty of the CRV for 95% probability was calculated as u_{CRV}=1/2 U_{CRV}.

In case the consistency check failed the laboratory with the highest value of $\frac{(\bar{e}_{LAB_k} - y)^2}{u^2_{ILC.LAB_k}}$ was to be excluded for the next iteration of evaluation, and the new reference value *y* was to be re-calculated without the values of excluded laboratory. Then the new standard uncertainty of the reference value u_y and the new chi-squared value χ^2 had to be re-calculated, and a new iteration of the consistency check performed. This procedure was to be repeated until the consistency check will pass.

Evaluation of the equivalence "Lab-to-CRV" and "Lab-to-Lab"

After the CRV was determined, equivalence "Lab-to-Lab" and "Lab-to-CRV" were evaluated. The DoE in these cases is calculated as follows:

$$En_{LAB_k.vs.CRV} = \left|\frac{\bar{e}_{LAB_k} - \bar{e}_{CRV}}{U_{LAB_k.vs.CRV}}\right|, \tag{24}$$

$$En_{LAB_k.vs.LAB_l} = \left|\frac{\bar{e}_{LAB_k} - \bar{e}_{LAB_l}}{U_{LAB_k.vs.LAB_l}}\right|, \tag{25}$$

where $U_{LAB_k.vs.CRV}$ is the expanded uncertainty of the offset between the the average result $\bar{e}_{LAB_k}$ of the Participant's laboratory *k* and the CRV $\bar{e}_{CRV}$; $U_{LAB_k.vs.LAB_l}$ is the expanded uncertainty of the offset between the the average result $\bar{e}_{LAB_k}$ of the Participant's laboratory *k* and average result $\bar{e}_{LAB_l}$ of Participant's laboratory *l* in a given calibration point, with $k \neq l$.

As far as the ILC type B is concerned, following cases describing possible correlation between input values were considered.

Case B1: ILC Type B, offset "LAB-to-CRV", independent laboratories with contribution to the CRV

The covariance between the result of the given Participant's lab (with contribution to the CRV) and the CRV is the variance of the

CRV itself [1]. It this case the covariance is expressed as $cov = u^2_{CRV}$ and the standard uncertainty of the offset between the results $\bar{e}_{LAB_k}$ and $\bar{e}_{CRV}$ therefor is calculated as follows:

$$u_{LAB_k.vs.CRV} = \sqrt{u^2_{ILC.LAB_k} + u^2_{CRV} - 2 \cdot u^2_{CRV}} = \sqrt{u^2_{ILC.LAB_k} - u^2_{CRV}}, \qquad (26)$$

Case B2: ILC Type B, offset "LAB-to-CRV", independent laboratories without contribution to the CRV

If the laboratory does not contribute to the CRV (as if chi-squared test has eliminated the laboratory's result from calculation of the CRV in a given calibration point) there is no covariance between the result and the CRV. In this case the offset between this result and the CRV will have the uncertainty given by:

$$u_{LAB_k.vs.CRV} = \sqrt{u^2_{ILC.LAB_k} + u^2_{CRV}} \qquad (27)$$

Case B3: ILC Type B, offset "LAB-to-CRV", dependent laboratories with a contribution to the CRV

If two or more laboratories have common traceability source i.e. they calibrate their reference standards in the same source laboratory, then their contributions to CRV are subject to correlation. The covariance between such laboratory's result and the CRV is associated with the common source of traceability. It this case, similar to case A2, it is reasonable to determine a conservative estimation of an upper limit of this covariance. And similar to case A2, the results considered here would be strongly correlated (correlation coefficient = 1) and there would be the same covariance to the CRV expressed as $cov = u^2_{CRV}$, and the standard uncertainty of the offset between the results $\bar{e}_{LAB_k}$ and $\bar{e}_{CRV}$ therefor is calculated as follows:

$$u_{LAB_k.vs.CRV} = \sqrt{u^2_{ILC.LAB_k} + u^2_{CRV} - 2 \cdot u^2_{CRV}} = \sqrt{u^2_{ILC.LAB_k} - u^2_{CRV}} \qquad (28)$$

Case B4: ILC Type B, offset "LAB-to-LAB", independent laboratories

There is no covariance between the results of two independent Participant's laboratories *k* and *l* :

$$u_{LAB_k.vs.LAB_l} = \sqrt{u^2_{ILC.LAB_k} + u^2_{ILC.LAB_l}} \qquad (29)$$

Case B5: ILC Type B, offset "LAB-to-LAB", dependent laboratories with common source of traceability

If laboratories *k* and *l* have common traceability source i.e. calibrate their reference standards in the same source laboratory then their results are corelated. In this case again a covariance between these labs is associated with the common source of traceability and again, for the same reason as in case B3 and similar to the case of Participant-to-Reference reproducibility check, a conservative upper limit of the covariance can be determined as $cov = u^2_{CMC.SourceLab}$:

$$u_{LAB_k.vs.LAB_l} = \sqrt{u^2_{ILC.LAB_k} + u^2_{ILC.LAB_l} - 2 \cdot u^2_{CMC.SourceLab}} \qquad (30)$$

where $u_{CMC.SourceLab}$ represents the standard CMC uncertainty of the laboratory being the common source of traceability. It can be derived from the expanded uncertainty stated in the calibration certificate by $u_{CMC.SourceLab} = (1/k_{95})\ U_{CMC.SourceLab.}$

Note: Based on the practical experience gained from the historical bilateral intercomparisons with the Reference Laboratory being the participant of EuReGa group, the Organizer has set this uncertainty to be $u_{CMC.SourceLab} = u_{Shared} = 0{,}05\%$[].

The equations from (30) to (34) use standard uncertainties. The expanded uncertainties $U_{LAB_k.vs.LAB_l}$ and $U_{LAB_k.vs.CRV}$ to be substituted into the equations (27) and (28) were determined by:

$$U_{LAB_k.vs.CRV} = 2 \cdot u_{LAB_k.vs.CRV} \qquad (31)$$
$$U_{LAB_k.vs.LAB_l} = 2 \cdot u_{LAB_k.vs.LAB_l} \qquad (32)$$

After calculating the DoE values, the evaluation of the equivalence "Lab-to-CRV" and "Lab-to-LAB" was performed based on the provisions for the DoE limits as in effect in best known EURAMET practices [3, 4]:

- absolute values ***En* ≤ 1.0** was considered **acceptable** (results subject to comparisons are equivalent);

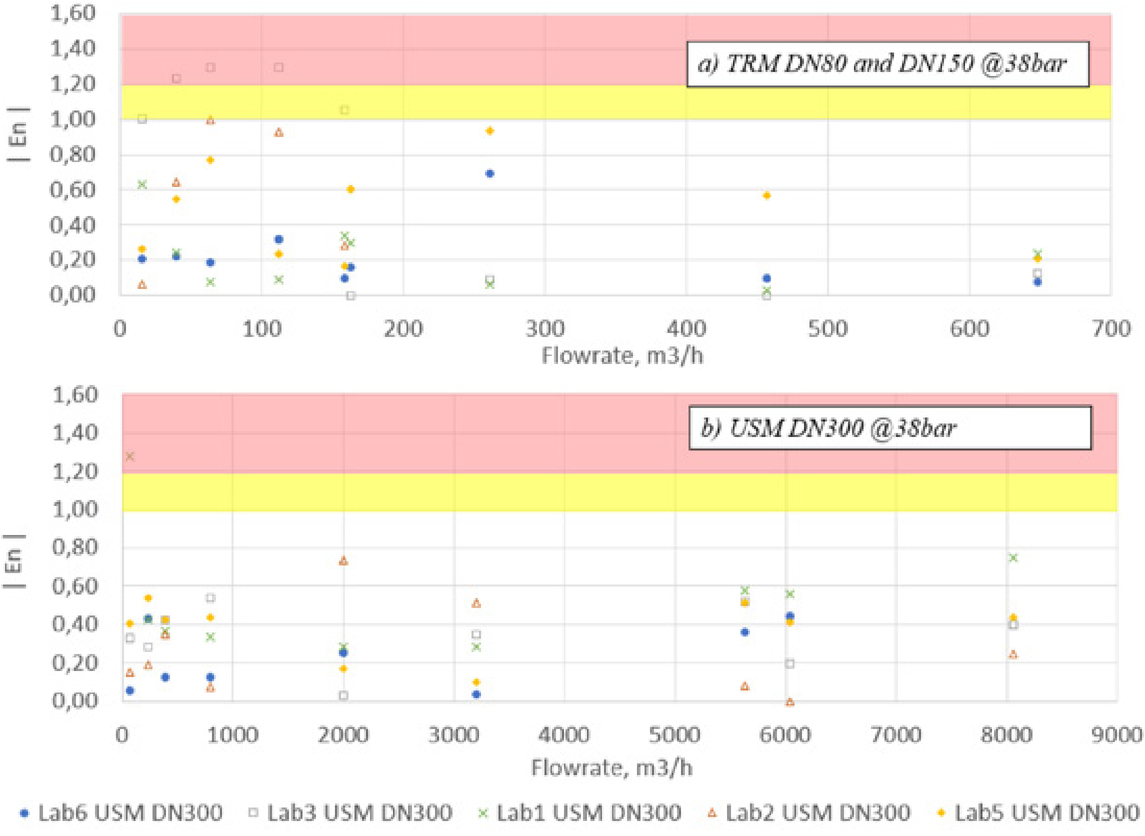


Fig. 6 ILC type A results – degree of equivalence En @38b

Rys. 6 Wyniki ILC typu A – stopień równoważności En @38bar

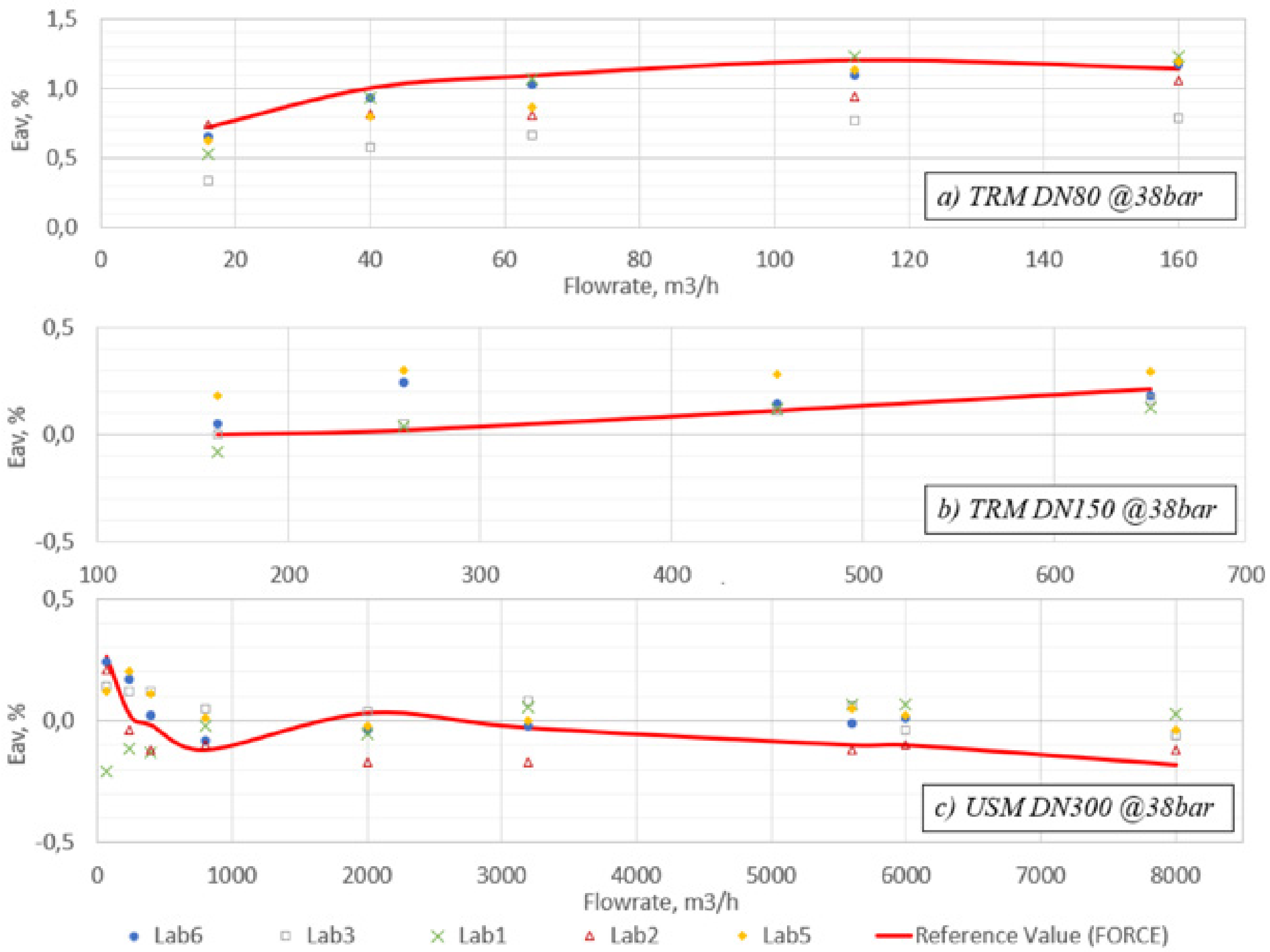


Fig. 7 ILC type A results – deviation Eav @38bar

Rys.7 Wyniki ILC typu A – odchylenie Eav @38bar

- absolute values **1.0 <** ***En*** **≤ 1.2** would be considered as being in the **warning level** (results subject to comparisons may still be considered as equivalent, but check actions were recommended);
- absolute values ***En*** **> 1,2** would be considered as **unacceptable** (results of the laboratories subject to comparisons are not equivalent and thus should be investigated).

On the whole, preliminary results of the ILC were processed by the Organizer after all Participants have finalized the calibrations and shared their results. Prior to processing the results the Organizer-to-Organizer reproducibility check was conducted at 38 barg.

A preliminary report was elaborated by the Organizer and shared with other Participants for commenting after collecting of all preliminary results and data analyses. The Participants had appropriate time for comments and/or amendments to the report, after that the Organizer shared final version of the report.

4. Summary results

4.1. ILC type A

Brief summary results for ILC type A (where the Participant's results were compared to the Reference Laboratory results at 38 bar) are presented on the graphs below. In order to satisfy the requirements of [4] the results are anonymized, for the purpose of this publication designations "**Lab.1**", "**Lab.2**" etc. are used.

The intercomparisons results at 38 bar mostly showed | En | ≤ 1,0 (meaning that given results of corresponding Participant are equivalent to the results of the reference laboratory FORCE Technology). Lab.2 has failed to install DN150 meter, rest of the Participants presented their results. Three results of Lab.3 obtained with turbine meter DN80 proved unacceptable, and two results being acceptable, but on a warning level. One result of Lab.1 for ultrasonic meter at Qmin proved unacceptable. Rest of the results proved equivalent to the assigned value.

4.2. ILC type B

Brief summary results for ILC type B (where the Participant's results were compared to the **C**omparison **R**eference **V**alue and also against each other at 50 bar, 16 bar and 8 bar) are presented on the graphs below. In order to satisfy the requirements of ISO 17043 the results are anonymized, for the purpose of this publication designations "**Lab.1**", "**Lab.2**" etc. are used.

The intercomparisons between each of the Participants and the CRV, mutually established by all Participants, showed all participating laboratories, both those having same source of traceability, and those having different source of traceability (PIGSAR, Germany or FORCE Technology, Denmark) are generally in good agreement, suggestive of good harmonization between mentioned traceability-source laboratories being active members of EuReGa consortium.

Most results of intercomparisons type B against CRV showed |En| ≤ 1,0 for pressures 50 bar, 16 bar and 8 bar, and thus considered acceptable. One result of Lab.3 for TRM DN80 showed "warning level" (|En| = 1,13 for TRM DN80 at Q = 160 m3/h and P = 50 bar). Other than this one result, all results were equivalent to the assigned value CRV.

Mutual equivalence between each pare of the Participants was also checked and proved, the results were shared between the Participants. Comparisons allowed to uncover some minor technical issues that were either corrected or accounted for.

Fig. 8 ILC type B results – degree of equivalence En @50 bar

Rys.8 Wyniki ILC typu B – stopień równoważności En przy 50 barach

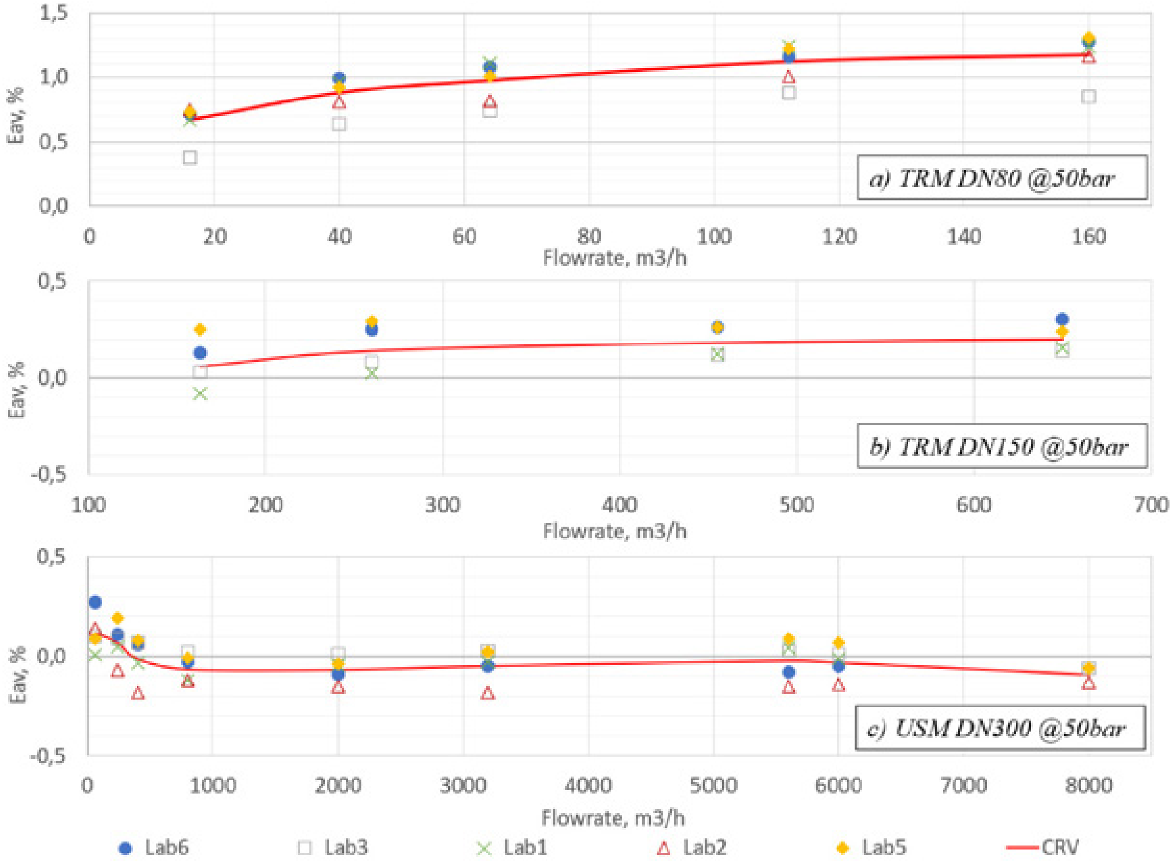


Fig. 9 ILC type B results – deviation Eav @50bar

Rys.9 Wyniki ILC typu B – odchylenie Eav @50bar

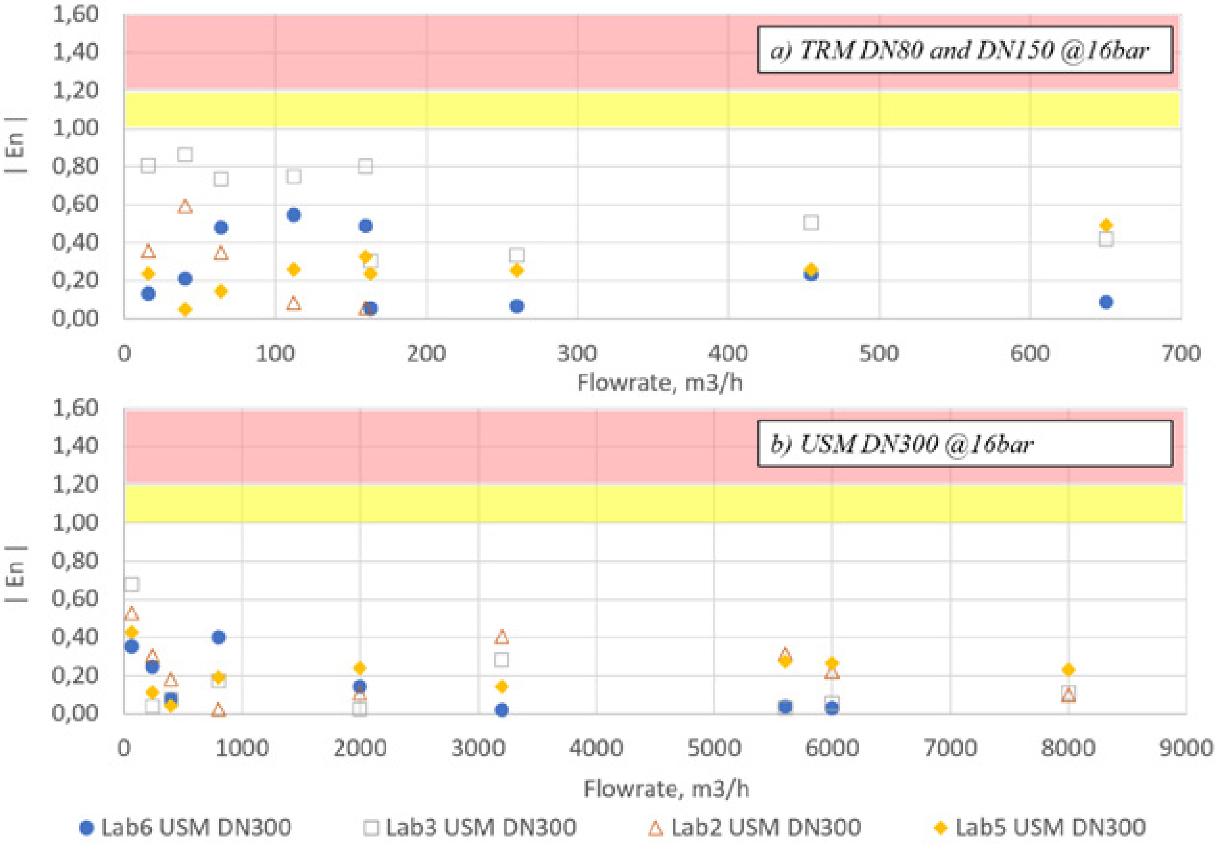


Fig. 10 ILC type B results – degree of equivalence En @16 bar

Rys.10 Wyniki ILC typu B – stopień równoważności En przy 16 barach

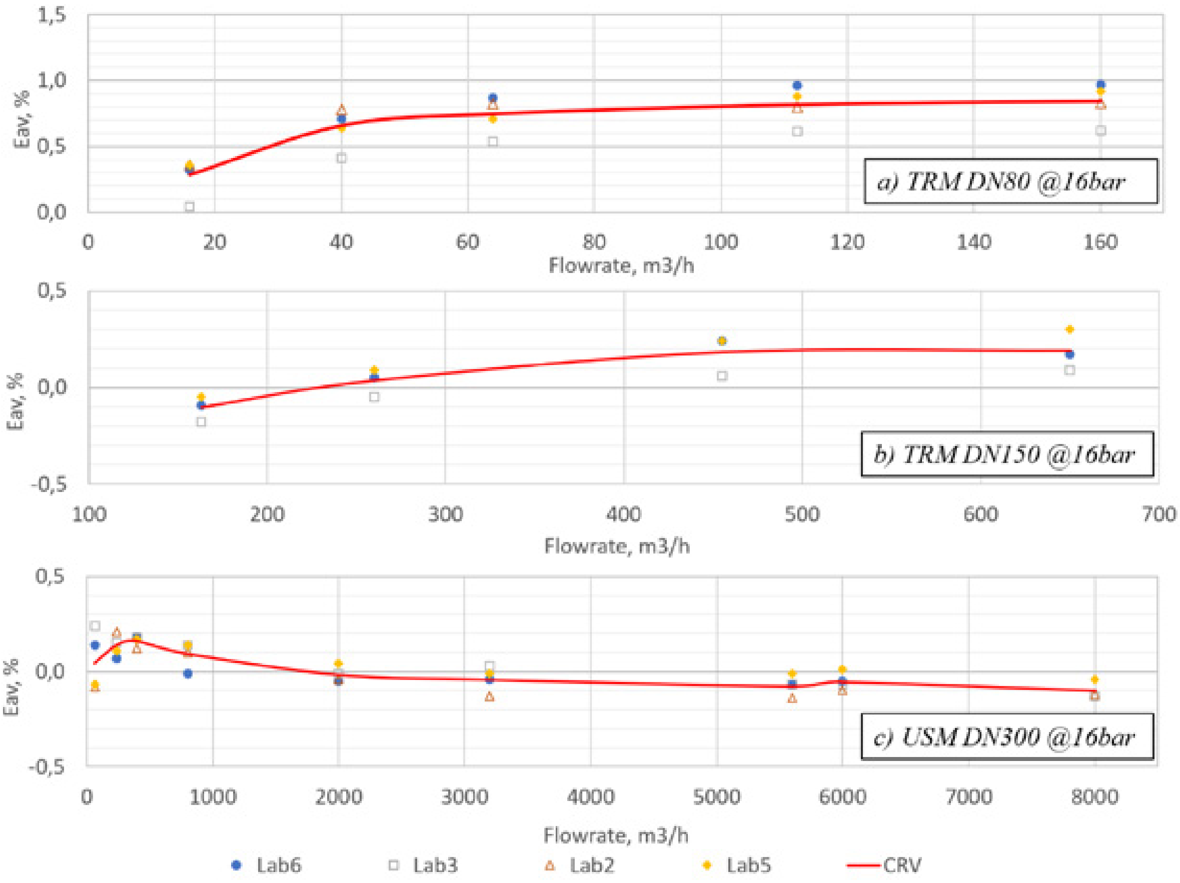


Fig. 11 ILC type B results – deviation Eav @16bar

Rys.11 ILC type B results – deviation Eav @16bar

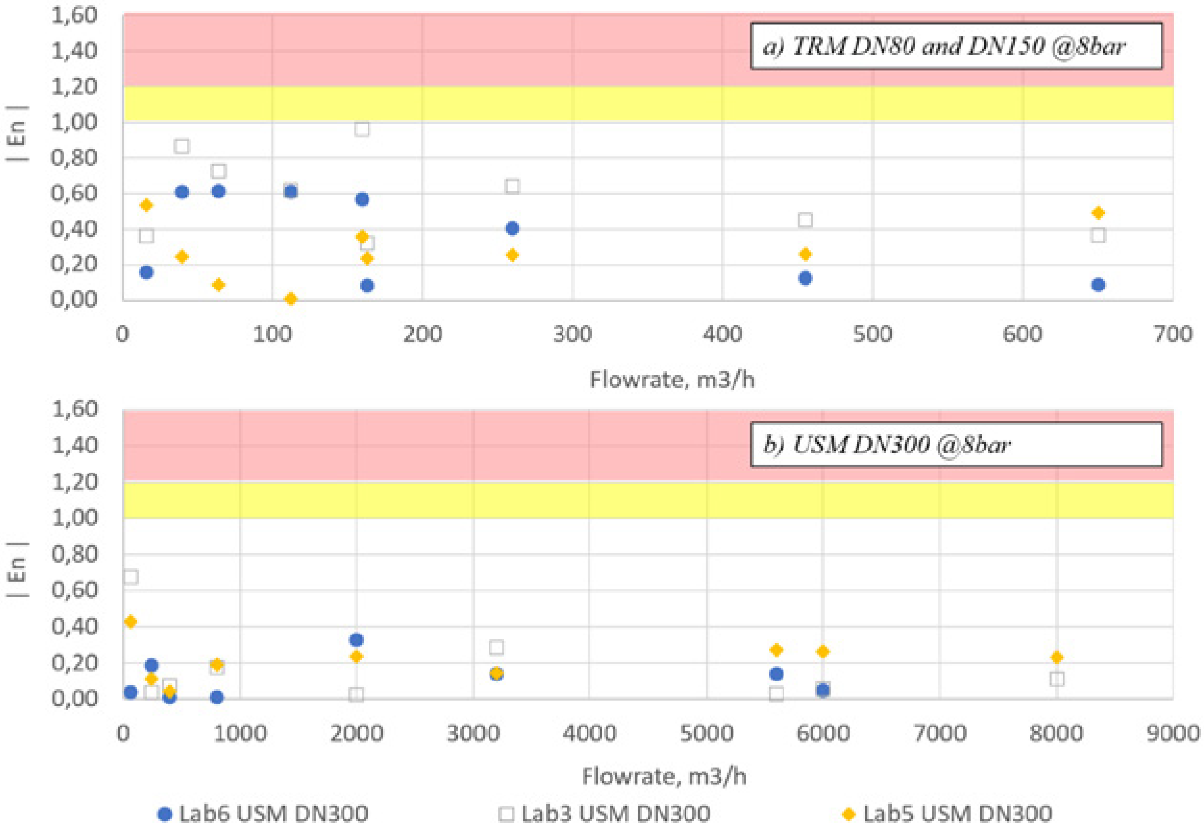


Fig. 12 ILC type B results – deviation Eav @8bar

Rys.12 Wyniki ILC typu B – odchylenie Eav @8bar

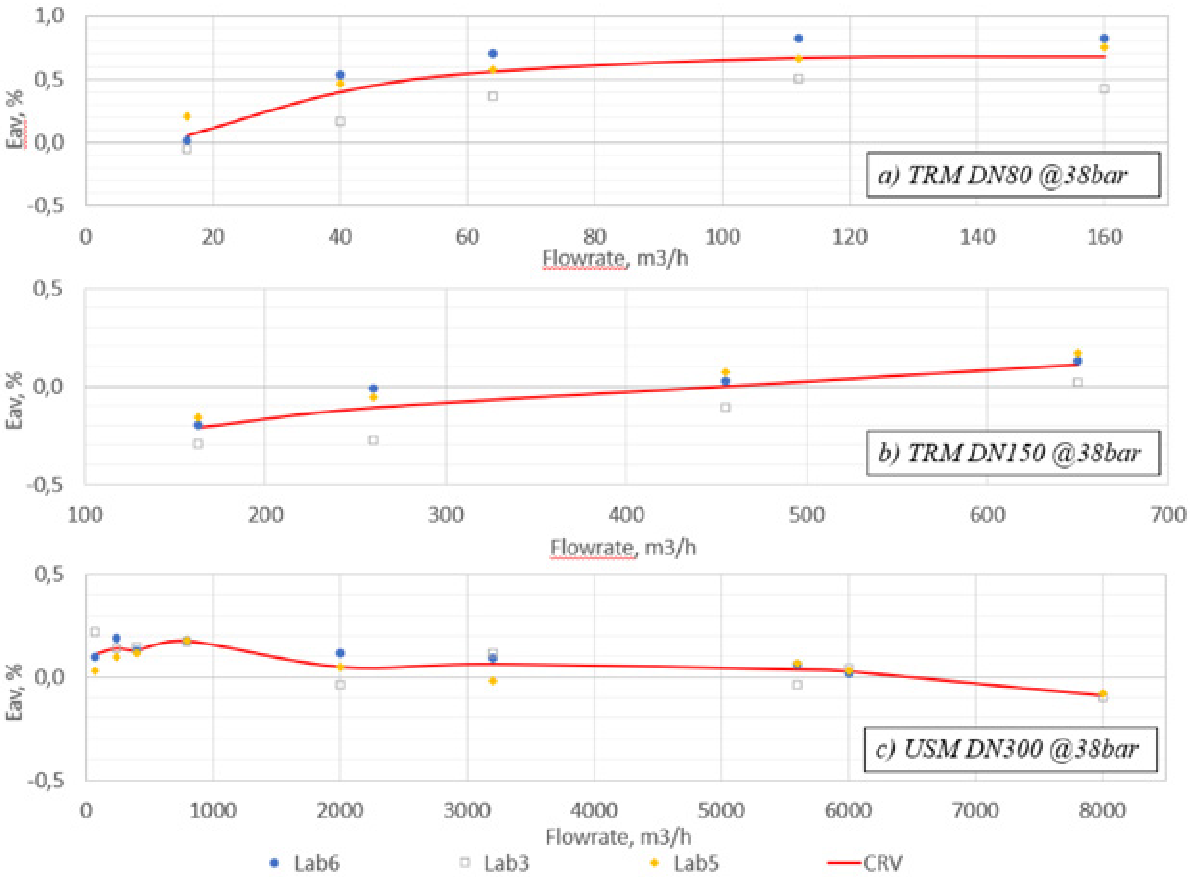


Fig. 13 ILC type B results – deviation Eav @8bar

Rys.13 Wyniki ILC typu B – odchylenie Eav @8bar

5. Conclusions

During the ILC total 591 results were obtained (see Table 1), including those for direct intercomparisons between the laboratories. The results were subject to evaluation based on the En criteria. In the field of gas meter calibrations it is common practice to use |En| ≡ 1,2 as an upper limit criterion for acceptation of the intercomparisons result [10]. Results showing 1.0 < |En| ≤ 1.2 are considered acceptable but being on warning level (check actions are recommended). In this intercomparisons four En values proved unacceptable, another eight En showed warning level. 461 results corresponded to |En| ≤0,5, and another 118 results showed 0,5 ≤ |En| ≤ 1,0. In total 98% of results were acceptable.

Table 1 Grand summary of the ILC

Tabela Całościowe podsumowanie ILC

Results received	Number	Percent
Total results	591	100%
0,0 ≤ En ≤ 0,5	461 of 591	78%
0,5 < En ≤ 1,0	118 of 581	20%
1,0 < En ≤ 1,2 i.e. warning level	8 of 591	1%
En > 1,2 i.e. unacceptable	4 of 591	1%
0,0 ≤ En ≤ 1,2 i.e. acceptable	587 of 591	98%

En is a test with 95% confidence and therefor agreement with 98% probability is proving ILC successful. It should be emphasised that it was first ILC of this kind. The intercomparisons allowed each of the Participants to draw its own conclusions. Certainly it was valuable experience for all Participants. For the Gas Meter Calibration Laboratory LWG, the experience gained in the organization of this ILC is a great added value that should be utilised for organizing proficiency testing and intercomparisons in future.

REFERENCES

[1] Cox M.G.2002." Evaluation of key comparison data" Metrologia (39): 589-595.

[2] Dopheide D., B.Mickan, R.Kramer, H.-J.Hotze, J.-P.Vallet, M.R.Harris, Jiunn-Haur Shaw, Kyung-Am Park,2006. CIPM Key Comparisons for Compressed Air and Nitrogen, CCM.FF-5.b – Final Report.

[3] EA-4/21 INF: 2018. Wytyczne dotyczące oceny stosowności małych porównań międzylaboratoryjnych w procesie akredytacji laboratoriów. – Marzec 2018 r. Wyd. 00. – 11 str.

[4] EA-4/02 M.2018. Evaluation of the Uncertainty of Measurement in Calibration.

[5] Gunnarsson Arnthor.2020." Bilateral Intercomparison Between Gaz-System and FORCE Technology in 2019". Final Report, March 2, 2020.

[6] Mickan Bodo, Tomáš Valenta.2007." Inter-laboratory calibration comparison of the turbine gas meter G650". EURAMET Project No. 862. Final Report Draft B. – March 8, 2007 .

[7] PN-EN ISO/IEC 17043:2011 Ocena zgodności – Ogólne wymagania dotyczące kompetencji organizatorów badania biegłości.

[8] Polityka dotycząca uczestnictwa w badaniach biegłości DA-05, wyd. 9, Warszawa, 12.04.2023r.

[9] Traceability chains of EuReGa Participants for high-pressure natural gas. EuReGa White paper. – June 15, 2018

[10] Valenta Tomáš.2015." Inter-laboratory calibration comparison of the rotary piston gas meter G650". EURAMET Project No. 1296. Final Report Draft B. – December 8, 2015 .