



**Supplementary Comparison EURAMET.L-S2.2.n01**  
**Calibration of gauge blocks by mechanical**  
**comparison**  
**(EURAMET project 1581)**

**Final Report**

AS Metrosert  
Central Office of Metrology

Jako Siim Eensalu, Metrosert  
Coordinator: Allar Pärn, Metrosert  
Tallinn, September 2026

---

## Contents

|      |                                                                                    |    |
|------|------------------------------------------------------------------------------------|----|
| 1    | Document control .....                                                             | 3  |
| 2    | Introduction .....                                                                 | 5  |
| 3    | Organization.....                                                                  | 6  |
| 3.1. | Participants.....                                                                  | 6  |
| 3.2. | Schedule .....                                                                     | 7  |
| 3.3. | Standards .....                                                                    | 7  |
| 4    | Measurement instructions and reporting of results .....                            | 8  |
| 5    | Stability and condition of the gauge blocks.....                                   | 10 |
| 6    | Laboratory measurements.....                                                       | 12 |
| 7    | Measurement results.....                                                           | 12 |
| 7.1. | Results and standard uncertainties as reported by participants .....               | 12 |
| 7.2. | Measurement uncertainties.....                                                     | 19 |
| 8    | Analysis.....                                                                      | 21 |
| 8.1. | Calculation of the SCR <sub>V</sub> , Degrees of Equivalence, and Chi-squared..... | 21 |
| 8.2. | Participant responses .....                                                        | 34 |
| 8.3. | Analysis of revised DPM results .....                                              | 42 |
| 8.4. | Discussion of results.....                                                         | 43 |
| 9    | Conclusions.....                                                                   | 44 |
|      | Appendix A – Equipment and measuring processes of the participants .....           | 45 |
| 10   | References.....                                                                    | 51 |
| 11   | Source code for calculating SCR <sub>V</sub> and other statistics .....            | 52 |

# 1 Document control

| Version   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Draft A.1 | <i>Issued on May 29, 2025.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Draft A.2 | <i>Issued on January 28, 2026.</i><br><br>Revised Tables Table 13, Table 14, Table 15, Table 17.<br><br>Added missing decimals to Metrosert results in Table 4 through Table 9.<br><br>Added “Participant responses” chapter.<br><br>Corrected absolute expanded uncertainty CMC entry equation for BMM from quadratic to linear in Table 10.<br><br>Added the pilot’s measured and time-averaged deviation from nominal central length values to Table 11 and variation values to Table 12 per A. Lewis’ request.<br><br>DPM reported on Sept. 17, 2025 that due to a technical error in Excel processing, i.e. $L_c = L_n + R_2 - R_1$ instead of $L_c = L_{ref} + R_2 - R_1$ , all calibration results from DPM were stated to not be traceable to the EIM reference values. Hence, per the recommendation from the TC-L chair citing T. Coveney, DPM was excluded from the RV calculation in deviation from nominal central length, and the corrected data was assessed separately.<br><br>Added Table 20, Figure 12. |
| Draft A.3 | <i>Issued on March 12, 2026</i><br><br>The standard uncertainty values reported by BMM were previously mistakenly divided by 2 (to $k = 0,5$ ) in the report, as reminded by T. Božović. Thus, SCR <sub>V</sub> of deviation from nominal central length and variation was recalculated for all gauge blocks with the corrected BMM uncertainty values.<br><br>All graphs and tables containing results by BMM were revised.<br><br>Added Table 19 to summarise the statistical analysis.<br><br>Revised Table 20, Figure 12 to reflect the updated SCR <sub>V</sub> .<br><br>Corrections to minor style and omission mistakes, as noted by A. Lewis.<br><br>Added ( $k = 1$ ) to chapter 8.1 text to emphasize standard uncertainties were used as input in all analysis calculations.<br><br>Revised the analysis description to include chi-squared analysis for participants who failed the $E_n$ evaluation as per recommendation by A.                                                                              |

|            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|            | <p>Lewis. NMI was assigned <math>u(k = 1)</math> of <math>Q[24,3;0,42 L]</math> based on CMC entry data for chi-squared analysis, as length dependent uncertainty components were not provided by NMI in the supplied uncertainty budget.</p> <p>Birge ratio analysis was corrected to use the dynamic 95 % confidence upper limit for the Birge ratio cutoff instead of 1,0.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Draft B.1  | <p><i>Issued on April 8, 2026</i></p> <p>No changes since A.3.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Draft B.2. | <p><i>Issued on June 11, 2026</i></p> <p>Corrected spelling mistakes on pages 23, 25, and 43.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Draft B.3. | <p><i>Issued on September 14, 2026</i></p> <p>The cross-correlation of shared uncertainty sources was mistakenly grouping all participants with non-zero off-diagonal elements into one correlated group and one non-correlated group instead of two separate correlation groups and one non-correlated group. Python scripts were revised.</p> <p>Added Python source code to chapter 11 for reproducibility and traceability of statistics.</p> <p>In addition, several corrections were made in response to reviewer comments.</p> <p>Corrected minor spelling mistakes.</p> <p>The time-averages of the pilot were replaced with the first set of results to ensure equivalent input parameters for all participants, and to reflect the standard service provided by the pilot laboratory, as advised during peer review.</p> <p>Correlated uncertainty contributions are now color-coded per group in Table 10.</p> <p>Added units in nm and standardized uncertainty budgets to Table 10.</p> <p>Some presumably incorrect values and partial entries were corrected on a best effort basis in Table 10, denoted in curly brackets, without changing the uncertainty claims of participants.</p> <p>Revised Figure 12 due to updated SCRV.</p> <p>Revised Table 11 through Table 20 to accommodate feedback from reviewer comments and corrected co-variance, chi-squared calculations. Excluded participants are outlined per gauge block and measurand in gray shading.</p> |

## 2 Introduction

The metrological equivalence of national measurement standards and of calibration certificates issued by national metrology institutes is established by a set of key and supplementary comparisons chosen and organized by the Consultative Committees of the CIPM or by the regional metrology organizations in collaboration with the Consultative Committees.

In 2022, at the annual EURAMET TC-L meeting it was proposed and agreed upon to have a supplementary comparison on gauge block calibration by mechanical comparison. The purpose of this supplementary comparison is to support new and existing CMC claims for this method.

For this supplementary comparison numbered EURAMET.L-S2.2.n01 Metroser Ltd acted as the pilot laboratory and NPL acted as the co-pilot. Metroser was responsible for providing the travelling standards, monitoring the stability of travelling standards, gathering the measurement data and evaluation and reporting of the measurement results. NPL was responsible for performing additional interferometric measurements to provide a traceability link to a gauge block key comparison, although it was confirmed later on during EURAMET TC-L 2025 (see e-mail by A. Lewis, 14 October, 2025, page 42) that this requirement was unnecessary for a supplementary comparison. Thus, the interferometric results were never used. The results of this international comparison contribute to the Mutual Recognition Arrangement (MRA) between the national metrology institutes of the Meter Convention [1].

## 3 Organization

### 3.1. Participants

All members of EURAMET TC-L were invited to participate. A total of 11 institutions decided to take part. The list of participants is given in Table 1.

Table 1. List of participant laboratories and their contacts.

| COUNTRY           | CONTACT               | ADDRESS                                                  | PHONE, FAX, E-MAIL                                                          |
|-------------------|-----------------------|----------------------------------------------------------|-----------------------------------------------------------------------------|
| ALBANIA           | Defrim Bulku          | DPM<br>Autostrada Tirane-Durres Km 8<br>Tirana           | +355 4 2233 174<br>defrim.bulku@dpm.gov.al                                  |
| AUSTRIA           | Michael Matus         | BEV<br>Arltgasse 35<br>1160 Wien                         | +43 1 21 110 6540<br>michael.matus@bev.gv.at                                |
| BULGARIA          | Denita<br>Tamakyarska | BIM<br>52B, G.M. Dimitrov blvd.<br>1040 Sofia            | +359 2 970 27 19<br>d.tamakjarska@bim.government.bg                         |
| ESTONIA           | Jako Siim<br>Eensalu  | METROSERT<br>Teaduspargi 8, 12618 Tallinn                | +372 5340 7617<br>jako-siim.eensalu@metrosert.ee                            |
| HUNGARY           | Gábor<br>Szikszai     | BFKH<br>Németvölgyi út 37-39<br>1534 Budapest            | +36 1 4585854<br>szikszai.gabor@bfkh.gov.hu                                 |
| IRELAND           | Rory<br>Hanrahan      | NSAI NML<br>Claremont Avenue, Glasnevin<br>Dublin 9      | +351 1 8082611<br>rory.hanrahan@nsai.ie                                     |
| LATVIA            | Larisa<br>Svedova     | LATMB<br>Kr. Valdemara street 157<br>1013 Riga           | +371 67 51 77 27<br>larisa.svedova@latmb.lv                                 |
| LITHUANIA         | Lilija<br>Chaleckiene | FTMC<br>Sauletekio av. 3,<br>LT-10257 Vilnius            | +370 61216739<br>+370 5 262 7123<br>lilija.chaleckiene@ftmc.lt              |
| MOLDOVA           | Alexandru<br>Braguta  | NMI (MD)<br>Eugen Coca street 28<br>2064 Chisinau        | +373 79 004 982<br>alexandru.braguta@inm.gov.md                             |
| MONTENEGRO        | Tomo Božović          | BMM<br>Arsenija Boljevića bb<br>81000 Podgorica          | +382 20 601 360<br>tomo.bozovic@metrologija.gov.me                          |
| SLOVENIA          | Bojan Acko            | MIRS/UM-FS/LTM<br>Smetanova ulica 17<br>2000 Maribor     | +386 2 220 7581<br>bojan.acko@uni-mb.si                                     |
| UNITED<br>KINGDOM | Andrew Lewis          | NPL<br>Hampton Road<br>TW11 0LW Teddington,<br>Middlesex | +44 208 943 6074<br>andrew.lewis@npl.co.uk                                  |
| COORDINATOR       | Allar Pärn            | METROSERT<br>Teaduspargi 8, 12618 Tallinn                | +372 5388 9335<br>allar.parn@metrosert.ee<br>tc-length.contact@metrosert.ee |

### 3.2. Schedule

The comparison was initially organized in two loops with each laboratory allowed approximately one month to make their measurements and to prepare for transportation to the next participant. The schedule was tailored to satisfy the laboratories for scheduling the measurements. Any changes to the schedule, after the start of the circulation, were discussed and agreed among the participants and the TC-L chairman. A third loop was added to allow BMM and NSAI NML to repeat their measurements after problems with experimental setups and facility HVAC had been solved. The co-pilot laboratory, NPL, provided the traceability link of the artefacts to a gauge block key comparison by interferometric measurements. Most delays in the schedule are attributed to improper packaging and expiration of the ATA Carnet, and the addition of two unplanned measurement slots.

Table 2. Schedule of the comparison.

| Institute      | Country | Original schedule | Date of measurement | Results received |
|----------------|---------|-------------------|---------------------|------------------|
| METROSERT      | EE      | July 2023         | July 2023           |                  |
| BEV            | AT      | August 2023       | August 2023         | September 2023   |
| BIM            | BG      | October 2023      | October 2023        | December 2023    |
| FTMC           | LT      | November 2023     | November 2023       | January 2024     |
| METROSERT      | EE      | January 2024      | January 2024        |                  |
| MIRS/UM-FS/LTM | SI      | January 2024      | February 2024       | February 2024    |
| NMI (MD)       | MD      | February 2024     | March 2024          | May 2024         |
| DPM            | AL      | March 2024        | April 2024          | April 2024       |
| LATMB          | LT      | April 2024        | June 2024           | July 2024        |
| NPL            | GB      | May 2024          | July 2024           | July 2024        |
| BFKH           | HU      | June 2024         | July 2024           | October 2024     |
| METROSERT      | EE      | July 2024         | September 2024      |                  |
| NSAI NML       | IE      | September 2023    | October 2024        | April 2025       |
| BMM            | ME      | December 2023     | January 2025        | April 2025       |
| METROSERT      | EE      | -                 | April 2025          |                  |

### 3.3. Standards

The circulated artefacts were 6 steel gauge blocks of grade K of rectangular cross section, nominally according to the standard ISO 3650. As the gauge blocks were brand new, the artefact instability due to unknown long-term drift had to be measured during the comparison.

Table 3. Steel gauge block artefacts.

| Identification | Nominal length (mm) | Manufacturer |
|----------------|---------------------|--------------|
| 218004         | 0.5                 | Mitutoyo     |
| 217026         | 4                   | Mitutoyo     |
| 222541         | 10                  | Mitutoyo     |
| 221877         | 25                  | Mitutoyo     |
| 221469         | 50                  | Mitutoyo     |
| 221625         | 100                 | Mitutoyo     |

## 4 Measurement instructions and reporting of results

Before calibration, the measurement surfaces of the gauge blocks had to be inspected for damage. Any damage, rust or scratches had to be documented in the drawing in appendix B.

The gauge blocks had to be measured by mechanical comparison with the participant laboratory's reference gauge blocks, using the normal calibration procedure. Interferometric measurements of the artefacts were prohibited.

The "A" surface was the marked measuring face for gauge blocks with nominal length < 6 mm and the right-hand measuring face for gauge blocks with nominal length  $\geq 6$  mm, respectively (see Figure 1). Measurements had to be done in only one orientation: face "A" at the top in horizontal orientation, and face "B" opposite to face "A" on the platen, the marked face of gauge blocks with nominal length  $\geq 6$  mm facing the operator, marking reading direction toward the comparator or up, respectively.

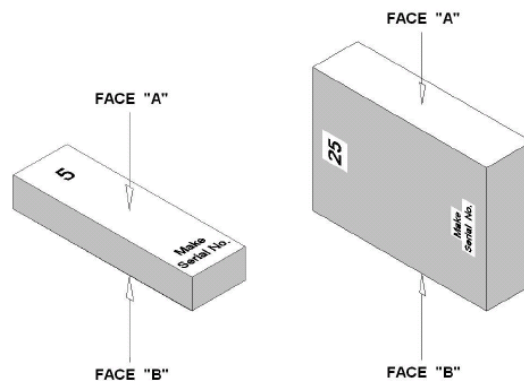


Figure 1. Gauge measurement position.

The mechanical calibration was to be performed as a five-point measurement (Figure 2) and the measurands are as stated in ISO 3650.

The following were the measurands for each gauge block: the deviation of the central length,  $l_c$ , from the nominal length,  $l_n$ , thus,  $\Delta l_c = l_c - l_n$ , and variation in length,  $v$  (Figure 3).

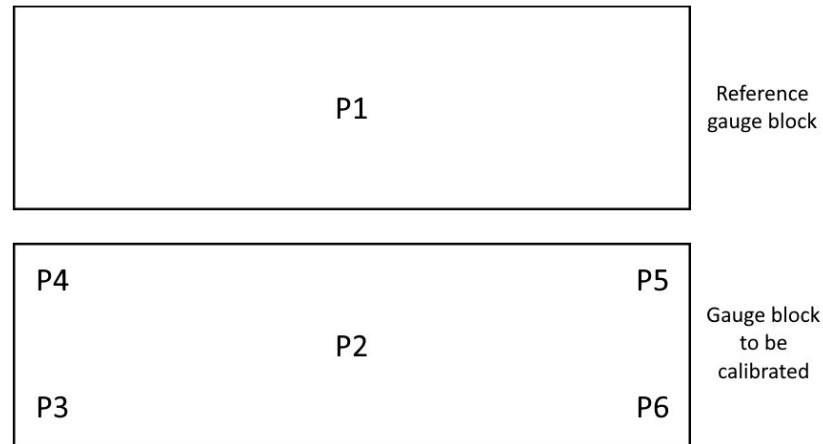


Figure 2. Gauge block measurement points.

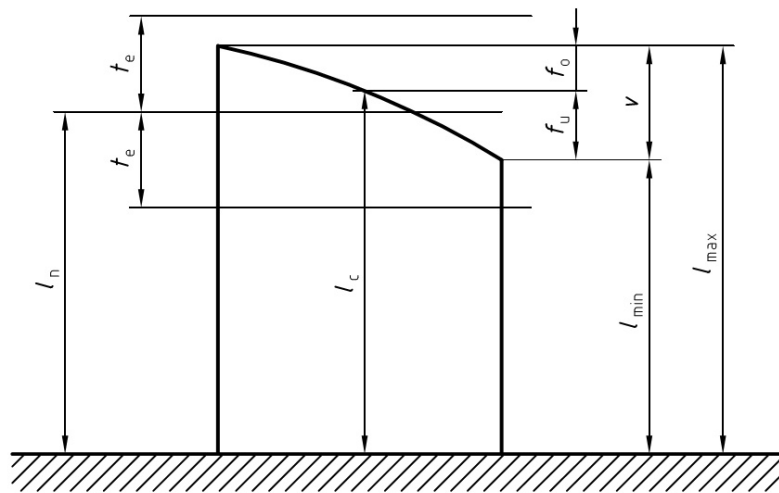


Figure 3. Nominal length ( $l_n$ ), central length ( $l_c$ ), variation ( $v$ ) of gauge blocks.

The uncertainty of measurement was to be estimated according to the *ISO Guide to the Expression of Uncertainty in Measurement*. The participating laboratories were encouraged to use their usual model for the uncertainty calculation.

The participants were asked to report the standard uncertainty for each deviation from nominal central length  $u(\Delta l_c)$  and standard uncertainty for each variation in length  $u(v)$ , and the uncertainty budget of the resulting combined uncertainty.

In case of relevant CMC-s the expanded measurement uncertainty for deviation from nominal central length from the nominal length  $U(\Delta l_c)$  had to be expressed in the usual length-dependent form:

$$U(\Delta l_c) = Q[a, b \times l_n] = \sqrt{a^2 + (b \times l_n)^2}$$

Measurement results had to be reported for the reference conditions as per the standard ISO 3650 with uncertainties calculated according to GUM (JCGM 100:2008) [2].

## 5 Stability and condition of the gauge blocks

The central length of the gauge blocks was stable during the 625-day journey. The central length and variation were consistent with zero (Figure 4, Figure 5), remaining within the uncertainty of the pilot laboratory for all gauge blocks, indicating a stable condition. Variation may have increased somewhat over time due to wear of face “B” against the platen. Thus, the standard evaluation procedure was not altered and the drift of the reference values was evaluated by least squares linear regression. Only the first results of the pilot were used in the analysis to replicate the standard measurement service, and to ensure fairness of pilot vs participant results in this single-blind supplementary comparison.

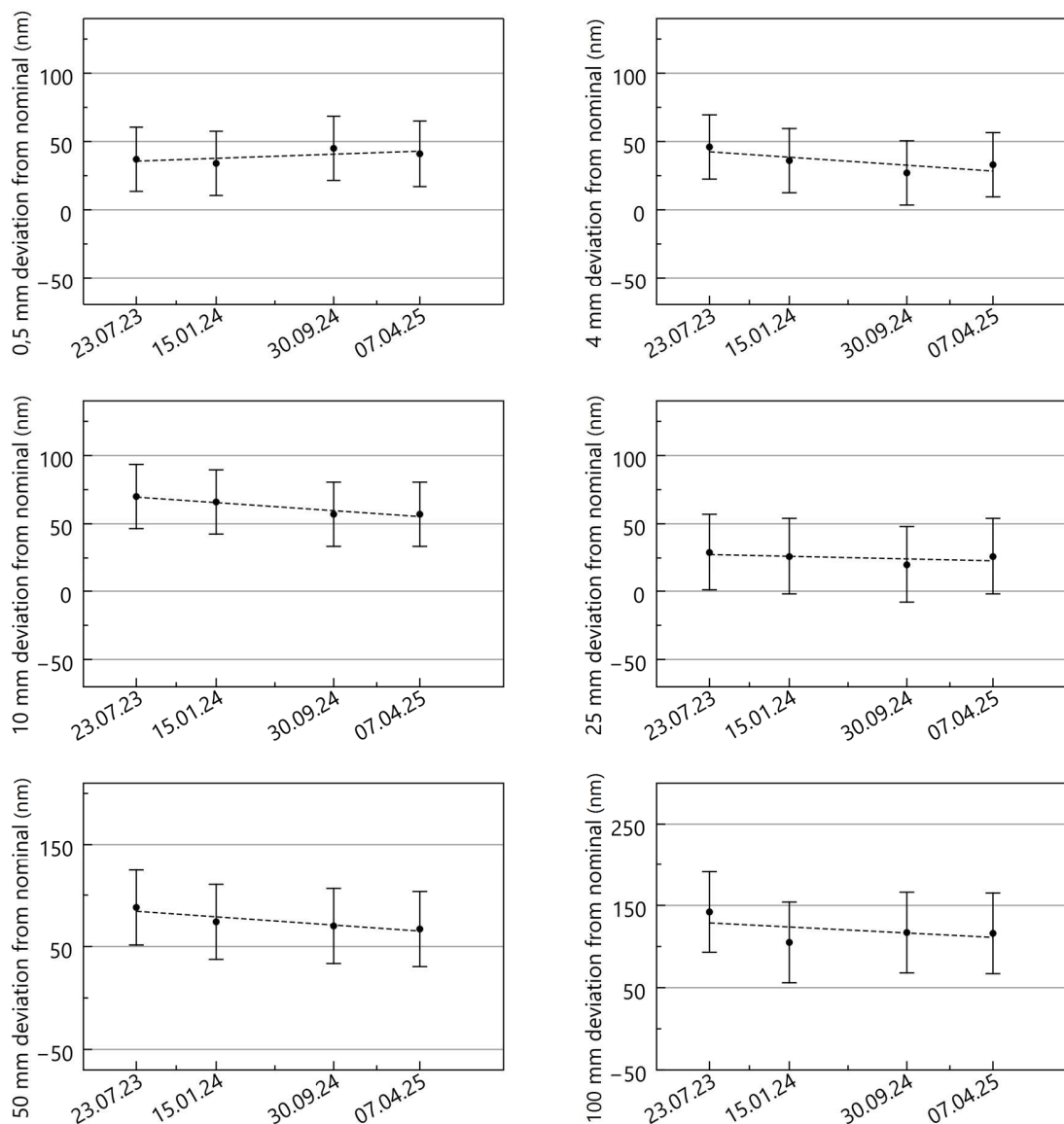


Figure 4. Drift of the deviation from nominal central length of the gauge blocks ( $k = 1$ ).

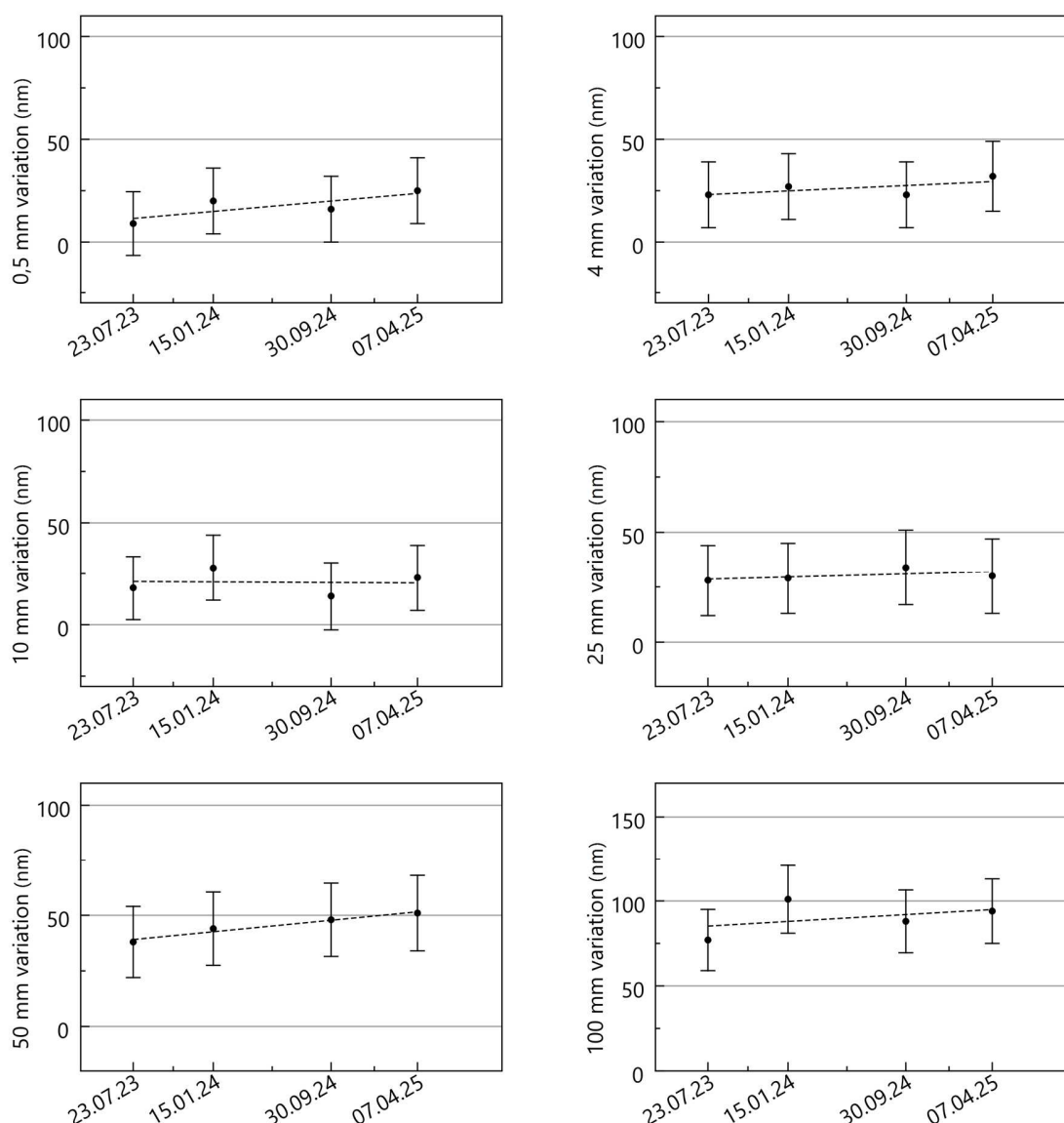


Figure 5. Variation of the gauge blocks ( $k = 1$ ) over time.

At the beginning of the comparison, most of the gauge blocks were in pristine condition, yet some showed signs of use: the center of face A on the 4 mm gauge block was scratched, and face B of the 100 mm gauge block had four scratches running parallel to the longer edge of the face. During the comparison, all but one participant noted after visual inspection in Appendix A that there were no apparent damages, whereas BMM stated to have detected severe damages putting the measurement results at risk.

All six gauge blocks were reported to have visible scratches on faces A and B, some of which participants attributed to wringing, as some of the streaks ran parallel to the long edge of the measuring faces. Based on observations by multiple participants, evidence provided in Appendix B, and subsequent visual inspection by the pilot laboratory at the end of the

comparison, suggests the 100 mm gauge block had been slid across the comparator surface at some point during the first year of the comparison while the lower mechanical probe was touching the gauge block, leaving a straight line across the center of face B.

More detailed inspection by less arbitrary methods, e.g. RMS surface roughness mapping, by the pilot laboratory may be prudent in future supplementary gauge block comparisons to mitigate the inherent ambiguity of visual inspection of the measurement surfaces, as visual acuity and availability of inspection tools of participants is presumed to differ.

## **6 Laboratory measurements**

The principal equipment used by the participants was a selection of gauge block comparators manufactured by Mahr (5 participants), TESA (4 participants), Cary (2 participants), and Feinmess Suhl (1 participant). The largest reported temperature deviation from the reference 20 °C was +0,6 °C.

## **7 Measurement results**

### **7.1. Results and standard uncertainties as reported by participants**

The results reported by the participants are given in tables Table 4 through Table 9 and figures Figure 6 through Figure 11 show the reported deviations from nominal length, with standard uncertainties ( $k = 1$ ).

Table 4. Deviation from nominal central length, variation, and standard uncertainties ( $k = 1$ ), as reported by laboratories in chronological order.

| $l_n$ 0,5 mm   | $\Delta l_c$ , nm | $u(\Delta l_c)$ , nm | $v$ , nm | $u(v)$ , nm |
|----------------|-------------------|----------------------|----------|-------------|
| METROSERT      | 37                | 23,5                 | 9        | 15,5        |
| BEV            | 21                | 25                   | 14       | 15          |
| BIM            | 33                | 50                   | 40       | 55          |
| FTMC           | 45                | 22                   | 38       | 22          |
| METROSERT      | 34                | 235                  | 20       | 16          |
| MIRS/UM-FS/LTM | 30                | 18                   | 30       | 20          |
| NMI            | 92                | 30                   | 10       | 24          |
| DPM            | -13               | 27                   | 8        | 24          |
| LATMB          | 10                | 26                   | 20       | 26          |
| NPL            | 53                | 16                   | 30       | 20          |
| BFKH           | 22                | 20                   | 49       | 31          |
| METROSERT      | 45                | 23,5                 | 16       | 16          |
| NSAI NML       | 18                | 10                   | 20       | 10          |
| BMM            | 40                | 26                   | 20       | 26          |
| METROSERT      | 41                | 24                   | 25       | 16          |

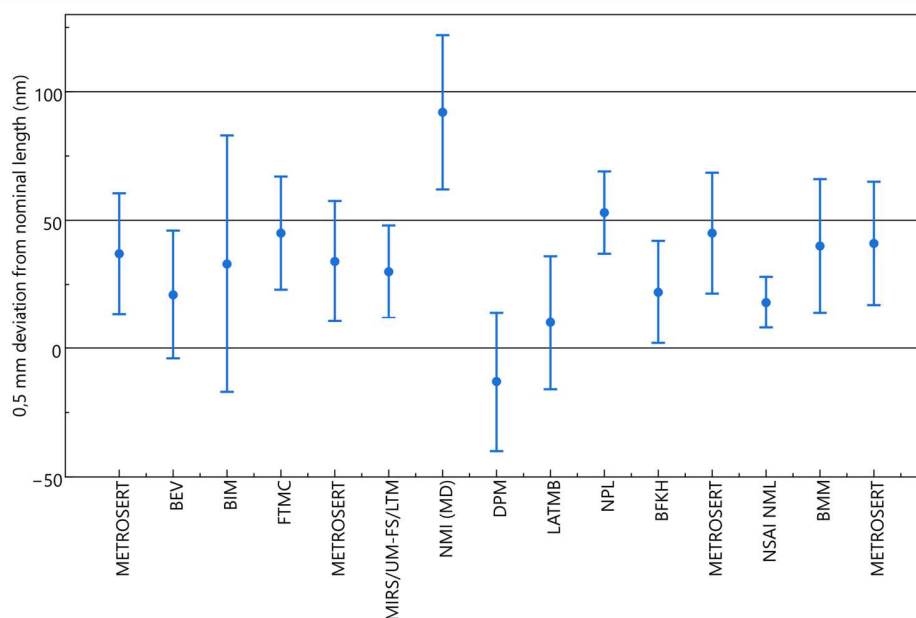


Figure 6. Deviation from nominal central length of the 0,5 mm steel gauge block ( $k = 1$ ).

Table 5. Deviation from nominal central length, variation, and standard uncertainties ( $k = 1$ ), as reported by laboratories in chronological order.

| $l_n$ 4 mm     | $\Delta l_c$ , nm | $u(\Delta l_c)$ , nm | $v$ , nm | $u(v)$ , nm |
|----------------|-------------------|----------------------|----------|-------------|
| METROSERT      | 46                | 23,5                 | 23       | 16          |
| BEV            | 13                | 25                   | 32       | 15          |
| BIM            | -2                | 50                   | 30       | 55          |
| FTMC           | 15                | 22                   | 20       | 22          |
| METROSERT      | 36                | 23,5                 | 27       | 16          |
| MIRS/UM-FS/LTM | 30                | 18                   | 5        | 20          |
| NMI            | 52                | 30                   | 2        | 24          |
| DPM            | 41                | 27                   | 28       | 24          |
| LATMB          | 10                | 26                   | 20       | 26          |
| NPL            | 72                | 17                   | 43       | 20          |
| BFKH           | 6                 | 19                   | 37       | 29          |
| METROSERT      | 27                | 23,5                 | 23       | 16          |
| NSAI NML       | 42                | 10                   | 20       | 10          |
| BMM            | 30                | 26                   | 40       | 26          |
| METROSERT      | 33                | 23,5                 | 32       | 17          |

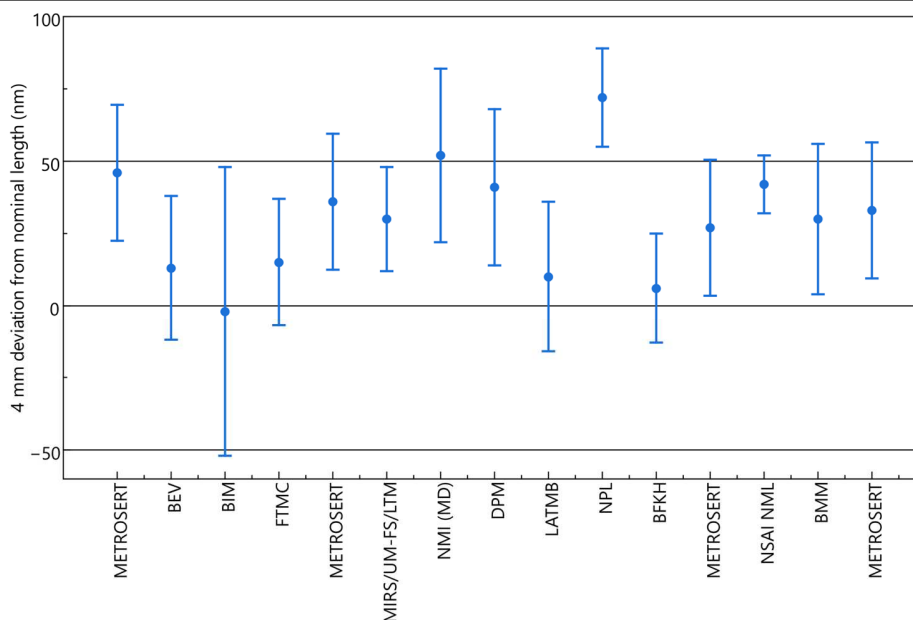


Figure 7. Deviation from nominal central length of the 4 mm steel gauge block ( $k = 1$ ).

Table 6. Deviation from nominal central length, variation, and standard uncertainties ( $k = 1$ ), as reported by laboratories in chronological order.

| $l_n$ 10 mm    | $\Delta l_c$ , nm | $u(\Delta l_c)$ , nm | $v$ , nm | $u(v)$ , nm |
|----------------|-------------------|----------------------|----------|-------------|
| METROSERT      | 70                | 26,5                 | 18       | 15,5        |
| BEV            | 30                | 25                   | 37       | 15          |
| BIM            | -17               | 50                   | 30       | 55          |
| FTMC           | 26                | 22                   | 10       | 22          |
| METROSERT      | 66                | 26,5                 | 28       | 16          |
| MIRS/UM-FS/LTM | 70                | 18                   | 50       | 20          |
| NMI            | 248               | 31                   | 20       | 24          |
| DPM            | 32                | 28                   | 12       | 24          |
| LATMB          | 2                 | 27                   | 20       | 27          |
| NPL            | 43                | 17                   | 40       | 20          |
| BFKH           | 43                | 20                   | 40       | 30          |
| METROSERT      | 57                | 26,5                 | 14       | 16,5        |
| NSAI NML       | 38                | 11                   | 10       | 11          |
| BMM            | 40                | 27                   | 30       | 27          |
| METROSERT      | 57                | 26,5                 | 23       | 16          |

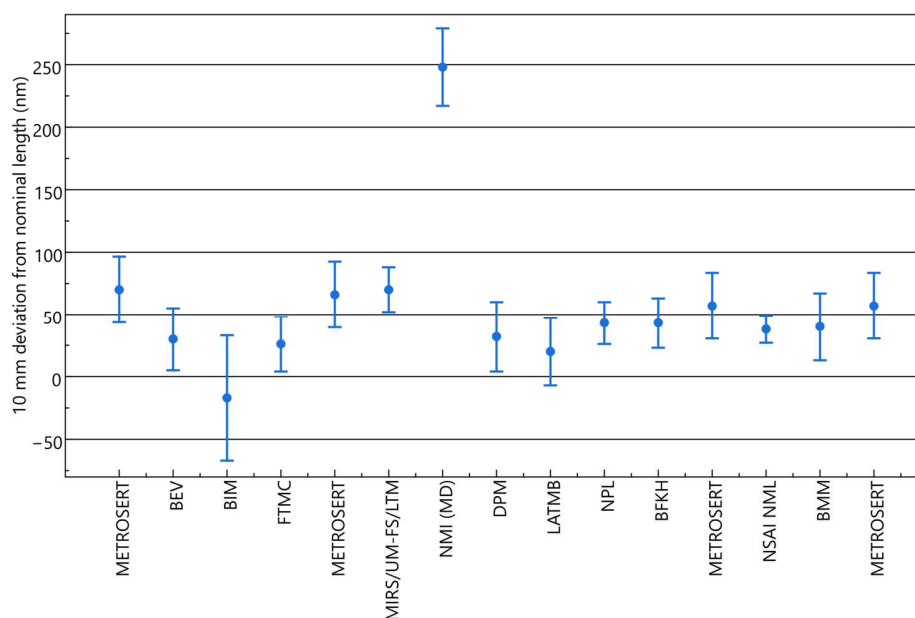


Figure 8. Deviation from nominal central length of the 10 mm steel gauge block ( $k = 1$ ).

Table 7. Deviation from nominal central length, variation, and standard uncertainties ( $k = 1$ ), as reported by laboratories in chronological order.

| $l_n$ 25 mm    | $\Delta l_c$ , nm | $u(\Delta l_c)$ , nm | $v$ , nm | $u(v)$ , nm |
|----------------|-------------------|----------------------|----------|-------------|
| METROSERT      | 29                | 28                   | 28       | 16          |
| BEV            | -23               | 26                   | 28       | 15          |
| BIM            | -17               | 52                   | 50       | 57          |
| FTMC           | 17                | 24                   | 14       | 24          |
| METROSERT      | 26                | 28                   | 29       | 16          |
| MIRS/UM-FS/LTM | 10                | 19                   | 40       | 20          |
| NMI            | 220               | 32                   | 30       | 24          |
| DPM            | -57               | 34                   | 28       | 24          |
| LATMB          | 10                | 29                   | 30       | 29          |
| NPL            | 46                | 19                   | 27       | 20          |
| BFKH           | 31                | 23                   | 52       | 33          |
| METROSERT      | 20                | 28                   | 34       | 17          |
| NSAI NML       | 6                 | 14                   | 30       | 14          |
| BMM            | 10                | 33                   | 30       | 33          |
| METROSERT      | 26                | 28                   | 30       | 17          |

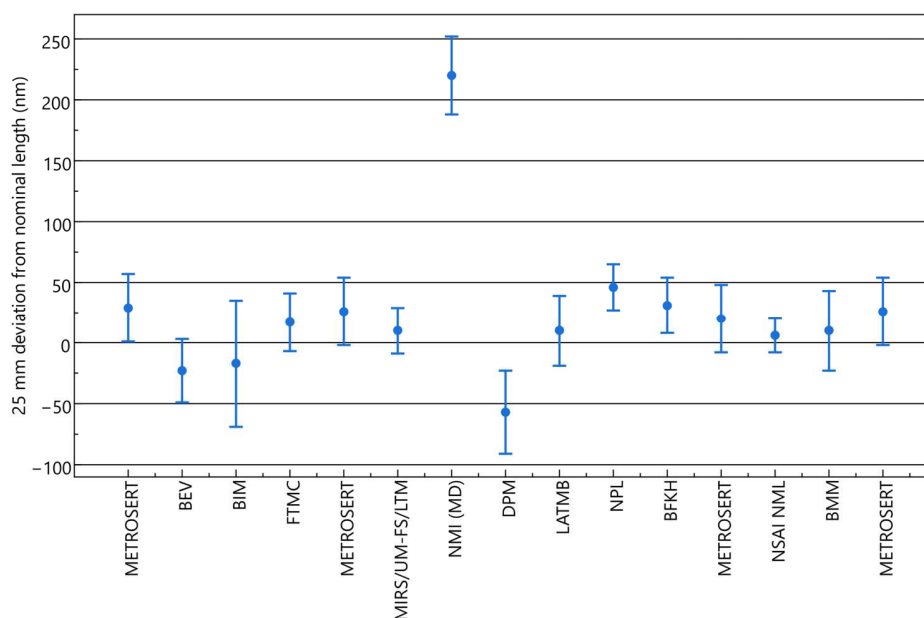


Figure 9. Deviation from nominal central length of the 25 mm steel gauge block ( $k = 1$ ).

Table 8. Deviation from nominal central length, variation, and standard uncertainties ( $k = 1$ ), as reported by laboratories in chronological order.

| $l_n$ 50 mm    | $\Delta l_c$ , nm | $u(\Delta l_c)$ , nm | $v$ , nm | $u(v)$ , nm |
|----------------|-------------------|----------------------|----------|-------------|
| METROSERT      | 88                | 36,5                 | 38       | 16          |
| BEV            | 37                | 27                   | 54       | 15          |
| BIM            | 24                | 56                   | 40       | 61          |
| FTMC           | 39                | 29                   | 31       | 29          |
| METROSERT      | 74                | 36,5                 | 44       | 16,5        |
| MIRS/UM-FS/LTM | 70                | 22                   | 60       | 20          |
| NMI            | 63                | 33                   | 30       | 24          |
| DPM            | 9                 | 47                   | 47       | 24          |
| LATMB          | -30               | 35                   | 50       | 35          |
| NPL            | 120               | 25                   | 47       | 20          |
| BFKH           | 24                | 31                   | 79       | 44          |
| METROSERT      | 70                | 36,5                 | 48       | 16,5        |
| NSAI NML       | 50                | 21                   | 20       | 21          |
| BMM            | 50                | 48                   | 60       | 48          |
| METROSERT      | 67                | 36,5                 | 51       | 17          |

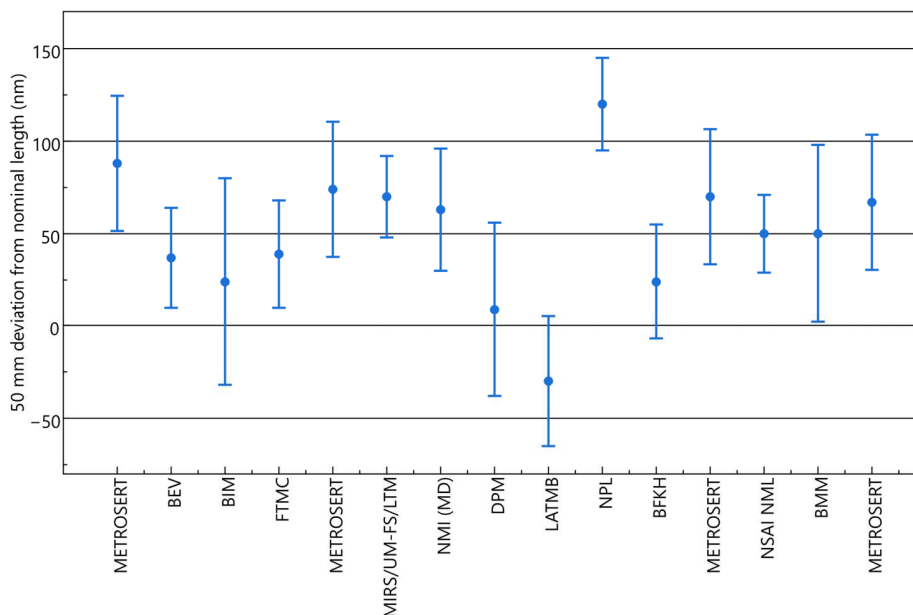


Figure 10. Deviation from nominal central length of the 50 mm steel gauge block ( $k = 1$ ).

Table 9. Deviation from nominal central length, variation, and standard uncertainties ( $k = 1$ ), as reported by laboratories in chronological order.

| $l_n$ 100 mm   | $\Delta l_c$ , nm | $u(\Delta l_c)$ , nm | $v$ , nm | $u(v)$ , nm |
|----------------|-------------------|----------------------|----------|-------------|
| METROSERT      | 142               | 49                   | 77       | 18          |
| BEV            | 72                | 32                   | 108      | 15          |
| BIM            | 62                | 71                   | 70       | 76          |
| FTMC           | 116               | 44                   | 67       | 44          |
| METROSERT      | 105               | 49                   | 101      | 2           |
| MIRS/UM-FS/LTM | 90                | 33                   | 90       | 20          |
| NMI            | 347               | 36                   | 60       | 24          |
| DPM            | -245              | 81                   | 46       | 24          |
| LATMB          | 60                | 54                   | 120      | 54          |
| NPL            | 148               | 42                   | 113      | 20          |
| BFKH           | 100               | 60                   | 121      | 95          |
| METROSERT      | 117               | 49                   | 88       | 185         |
| NSAI NML       | 86                | 38                   | 50       | 38          |
| BMM            | 100               | 84                   | 120      | 84          |
| METROSERT      | 116               | 49                   | 94       | 19          |

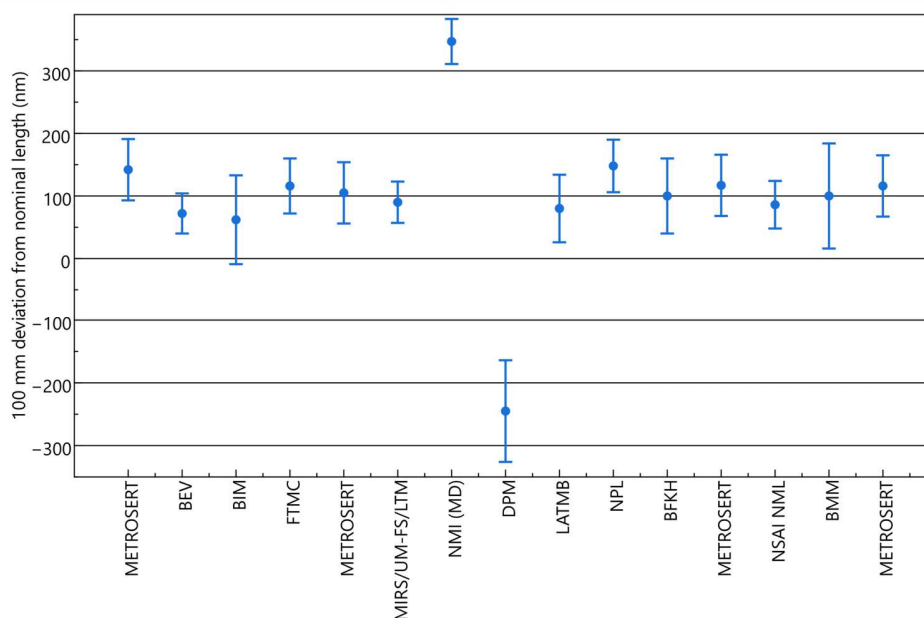


Figure 11. Deviation from nominal central length of the 100 mm steel gauge block ( $k = 1$ ).

## 7.2. Measurement uncertainties

Table 10. Standard uncertainty components of deviation from nominal central length uncertainty budgets ( $u_c$ ,  $k = 1$ ;  $U_{CMC}$ ,  $k = 2$ ) provided by participants.  $L$  is  $l_n$  in mm. Entries indexed with superscript  $a$  signify an uncertainty estimation for a 100 mm nominal length gauge block. Entries indexed with superscript  $b$  in curly brackets signify reviewed corrected values derived from Appendix C. Correlated uncertainty sources between groups of participants with shared traceability are outlined in shaded colors (light blue and light orange).

|                              | METROSERT            | BEV                  | BIM             | FTMC                 | MIRS/UM-FS/LTM     | NMI                   | DPM              | LATMB              | NPL              | BFKH               | NSAI NML          | BMM                |
|------------------------------|----------------------|----------------------|-----------------|----------------------|--------------------|-----------------------|------------------|--------------------|------------------|--------------------|-------------------|--------------------|
| $u(l_s)$                     | Q[10,5 nm; 0,10 L]   | 15 nm                | 0,15 L          | 14,5 nm              | Q[10 nm; 0,09 L]   | 50 nm                 | Q[13; 0,21 L]    | 11,5 nm            | Q[13 nm; 0,21 L] | 15,0 nm            | 5,25 nm           | Q[17,5 nm; 0,25 L] |
| $u(\theta)$                  | 0,29 L               | 1,0 nm               | 0,20 L          | 21,2 nm              | 0,054 L            |                       | 0,34 L           |                    | 0,132 L          | 16,7 nm            | 0,35 nm           | 2,92E-4 L          |
| $u(\theta_{a,m})$            | 0,087 L              |                      |                 |                      |                    |                       |                  |                    |                  |                    |                   |                    |
| $u(\delta\theta)$            | 0,20 L               | 17 nm                | 0,25 L          | 33,2 nm              | 0,132 L            | 0,33 {L} <sup>b</sup> | 0,40 L           |                    | 0,135 L          | 33,2 nm            | 0,29 nm           | 3,32E-4 L          |
| $u(\alpha_s)$                |                      | 0,12 L               | 0,17 L          |                      | 0,030 L            |                       | 0,035 L          |                    | 0,116 L          | 2,89 nm            |                   | 5,77E-5 L          |
| $u(\delta\alpha)$            |                      |                      |                 |                      | 0,021 L            | 0                     |                  |                    | 0,231 L          |                    |                   | 4,08E-8 L          |
| $u(\alpha_a)$                |                      | 0,12 L               | 0,13 L          |                      |                    |                       | 0,047 L          |                    |                  | 2,89 nm            |                   |                    |
| $u(\alpha_s)u(\theta)$       |                      |                      |                 |                      |                    |                       | 0,117 L          |                    |                  |                    |                   |                    |
| $u(\alpha_a)u(\theta)$       |                      |                      |                 |                      |                    |                       | 0,115 L          | 0,24E-6 nm         |                  |                    |                   |                    |
| $u(\alpha_s)u(\delta\theta)$ |                      |                      |                 |                      |                    | 0,24 {L} <sup>b</sup> | 0,027 L          | 0,24 nm            |                  |                    |                   |                    |
| $u(c)$                       | 15,1 nm              |                      | 30 nm           | 15 nm                |                    | 21 nm                 |                  | 17,3 nm            |                  |                    |                   |                    |
| $u(d)$                       |                      | 1,5 nm               | 20 nm           | 4,02 nm              | Q[10,5 nm; 0,19 L] | 7,4 nm                | 24 nm            |                    | 8,2 nm           |                    |                   | 5,48 nm            |
| $u(d_s)$                     |                      |                      | 12 nm           | 2,25 nm              |                    |                       |                  | 6,9 nm             |                  |                    | 3,8 nm            | 16 nm + 6E-4 L     |
| $u(\beta_{\text{drift}})$    |                      | 0,144 L              | 0,25 L          |                      | 0,082 L            |                       | 0,115 L          |                    |                  |                    |                   | 0,102 L            |
| $u(e_{\text{drift}})$        | 8,7 nm               | 11,5 nm              |                 | 18,5 nm              |                    |                       |                  | 8,2 nm             |                  | 28,9 nm            | 11,5 nm           | 8,16 nm            |
| $u(\Delta p)$                | 5,0 nm               | 5,0 nm               |                 |                      |                    | 9,62 nm               |                  |                    |                  |                    |                   | 4,04 nm            |
| $u(\varepsilon)$             | 0,192 L <sup>2</sup> |                      |                 |                      |                    |                       |                  | 2,9 nm             |                  | 2,89 nm            |                   |                    |
| $u(\kappa)$                  | 2 nm                 |                      |                 |                      | ≅0                 |                       |                  | 0,12E-6 nm         |                  | 7,22 nm            | 5,7 nm            |                    |
| $u(\rho)$                    | 20 nm                |                      | 20 nm           |                      | 4 nm               |                       |                  | 10 nm              |                  | 32,7 nm            | 3,0 nm            |                    |
| $u(\gamma)$                  | 2,9 nm               |                      |                 |                      |                    |                       |                  |                    |                  | 4,08 nm            |                   |                    |
| $u_c$                        | Q[24 nm; 0,43 L]     | 31,3 nm <sup>a</sup> | Q[50 nm; 0,5 L] | 43,7 nm <sup>a</sup> | Q[15 nm; 0,25 L]   | Q[24,3 nm; 0,645 L]   | Q[27 nm; 0,76 L] | Q[25,6 nm; 0,47 L] | Q[16 nm; 0,38 L] | 60 nm <sup>a</sup> | 17,8 nm           | Q[26 nm; 0,8 L]    |
| $U_{CMC}$                    | Q[54 nm; 0,91 L]     | Q[50 nm; 0,4 L]      | Q[100 nm; 1 L]  | Q[48 nm; 0,9 L]      | Q[35 nm; 0,5 L]    | Q[50 nm; 0,84 L]      | N/A              | Q[80 nm; 1,3 L]    | Q[32 nm; 0,76 L] | Q[40 nm; 1,3 L]    | Q[83 nm; 22,32 L] | [54 nm + 1,2 L]    |

Uncertainty component designations:

$u(l_s)$  – standard gauge block uncertainty

$u(\theta)$  – temperature deviation of standard gauge block from reference temperature uncertainty

$u(\theta_{a,m})$  – air or material temperature measurement uncertainty

$u(\delta\theta)$  – temperature difference between gauge blocks uncertainty

$u(\alpha_s)$  – coefficient of thermal expansion of standard uncertainty

$u(\alpha_a)$  – coefficient of thermal expansion of artifact uncertainty

$u(\delta\alpha)$  – difference of coefficient of thermal expansion uncertainty

$u(d)$  – length difference of gauge blocks uncertainty

$u(d_s)$  – surface roughness of artefact gauge block uncertainty

$u(e_{\text{drift}})$  – constant component of drift of standard gauge block deviation from nominal central length uncertainty

$u(\beta_{\text{drift}})$  – length dependent component of drift of standard gauge block deviation from nominal central length uncertainty

$u(\Delta p)$  – flatness difference between standard and artifact gauge block uncertainty

$u(\varepsilon)$  – deformation uncertainty

$u(c)$  – comparator uncertainty

$u(\kappa)$  – comparator linearity uncertainty

$u(\rho)$  – comparator repeatability uncertainty

$u(\gamma)$  – resolution of consecutive readings uncertainty

## 8 Analysis

### 8.1. Calculation of the SCRv, Degrees of Equivalence, and Chi-squared

Regarding correlation, the source of traceability for interferometric calibration of reference gauge blocks was GUM for three participants (Metroserf, FTMC, NSAI NML), and VTT MIKES for two participants (BIM, LATMB). The traceability source was unique for all other participants. Thus, the reported standard gauge block uncertainty components for those participants were used for covariance matrix calculations when calculating the reference value (RV) for the Deviation from nominal central length.

Outliers were eliminated iteratively on a per-gauge and measurand basis before calculating the reference value by eliminating the participant with the largest  $E_n$  magnitude until the Birge ratio

satisfied a value of  $R_B < \sqrt{1 + \sqrt{\frac{8}{I-1}}}$ , where  $I$  is the number of contributing laboratories, when considering standard uncertainties ( $k = 1$ ).

The Birge ratio  $R_B$  is derived from the application of least squares algorithms and the  $\chi^2$ -test:

$$R_B = \frac{u_{\text{ext}}(\bar{x}_w)}{u(\bar{x}_w)}$$

where  $u_{\text{ext}}(\bar{x}_w)$  is the external standard deviation

$$u_{\text{ext}}(\bar{x}_w) = \sqrt{\frac{1}{(I-1)} \cdot \frac{\sum_{i=1}^I w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^I w_i}},$$

$w_i$  is the normalised weight for the result  $x_i$

$$w_i = C \cdot \frac{1}{[u(x_i)]^2},$$

and the normalising factor,  $C$ , is given by

$$C = \frac{1}{\sum_{i=1}^I \left[ \frac{1}{u(x_i)} \right]^2}.$$

The weighted mean,  $\bar{x}_w$ , is calculated by:

$$\bar{x}_w = \sum_{i=1}^I w_i \cdot x_i$$

The uncertainty of the weighted mean is calculated by:

$$u(\bar{x}_w) = \sqrt{\frac{1}{\sum_{i=1}^I \left[ \frac{1}{u(x_i)} \right]^2}} = \sqrt{C}$$

The equation for the  $R_B$  cutoff is derived from the Birge ratio squared, which is effectively the reduced chi-squared statistic ( $\chi^2_v/v$ ), where the degrees of freedom is  $v = I - 1$  with  $I$  being the number of contributing laboratories.

From standard statistics, the properties of a chi-square distribution dictate:

Expected value:  $E[\chi^2] = v$

Variance:  $Var(\chi^2) = 2v$

To find the properties of the reduced chi-squared ( $R_B^2$ ):

Expected value of  $R_B^2$ :  $\frac{v}{v} = 1$

Variance of  $R_B^2$ :  $\frac{2v}{v^2} = \frac{2}{v} = \frac{2}{I-1}$

Standard deviation of  $R_B^2$ :  $\sigma = \sqrt{\frac{2}{I-1}}$

To establish a threshold for 95 % confidence ( $k = 2$ ), two standard deviations are added to the expected value:

$$R_B^2 < 1 + 2\sigma \Leftrightarrow R_B^2 < 1 + 2\sqrt{\frac{2}{I-1}} \Leftrightarrow R_B^2 < 1 + \sqrt{\frac{8}{I-1}} \Leftrightarrow R_B < \sqrt{1 + \sqrt{\frac{8}{I-1}}}$$

After deriving the weighted mean and its associated standard uncertainty, the deviation of each laboratory's result from the weighted mean is determined as  $x_i - \bar{x}_w$ . The uncertainty of this deviation is calculated as a combination of the uncertainties of the result,  $u(x_i)$ , and the uncertainty of the weighted mean  $u(\bar{x}_w)$ .

The degree of equivalence (DoE) is calculated as:

$$DoE_i = x_i - \bar{x}_w$$

The uncertainty of DoE is calculated in the generalized form to account for correlated uncertainty contributions between relevant participants:

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 + [u_{int}(\bar{x}_w)]^2 - 2 \cdot cov(x_i; \bar{x}_w)}$$

$E_n$ , which is used to check for statistical consistency of the results with their associated uncertainties, is defined as the ratio of the deviation from the weighted mean, divided by the expanded uncertainty of this deviation:

$$E_n = \frac{x_i - \bar{x}_w}{2 \cdot u_{DoE,i}}$$

The covariance of  $x_i$  with  $\bar{x}_w$  is calculated by:

$$\text{cov}(x_i; \bar{x}_w) = \sum_{i=1}^I w_i V_{ij}$$

where the covariance matrix  $V$  is constructed as:

$$V_{ij} = \begin{cases} u_i^2 & \text{if } i = j \\ u_{\text{common},i} \cdot u_{\text{common},j} & \text{if } i \neq j \end{cases}$$

$u_{\text{common},i}$  being the uncertainty contribution within  $u_i$  originating from the same source of traceability as  $u_{\text{common},j}$  in  $u_j$ .

While the Birge ratio evaluates the consistency of the consensus group for a single measurand, judging whether an individual participant's results support their overall uncertainty claim requires evaluating their performance across all measured artifacts. Because comparisons involving several measurands increase the statistical risk that a participant with correctly evaluated uncertainties may still observe an isolated large normalized error, a systematic acceptance procedure is required.

Consider a single participant reporting results for different measurands relating to a particular CMC claim. The DoE for the  $i$ -th measurand represents the difference between the participant's result and the SCR, alongside its associated standard uncertainty  $u_{\text{DoE},i}$ . The uncertainty accounts for the uncertainty of the SCR, the participant's claimed uncertainty, and the mathematical correlation between the two if the participant contributed to the reference value.

The assessment proceeds in two stages:

### Normalized Error Criterion

For a participant reporting results for  $m$  different measurands, indexed  $i \in \{1, \dots, m\}$ , let  $\text{DoE}_i$  denote the Degree of Equivalence, i.e., the difference between the participant's result and the corresponding SCR, and  $u_{\text{DoE},i}$  its standard uncertainty, which already accounts for the uncertainty of the SCR, the uncertainty claimed by the participant, and the correlation between the two arising from the participant's own contribution to the weighted-mean SCR.

First, the normalized errors are computed as

$$E_{n,i} = \frac{\text{DoE}_i}{k \cdot u_{\text{DoE},i}}$$

utilizing a coverage factor of  $k = 2$  for approximately 95% confidence. If  $|E_{n,i}| \leq 1$  for all of the participant's measurements, indicating the difference from the SCR is within the expanded uncertainty of that difference for every measurand, the participant's uncertainty claim is immediately accepted.

### Multivariate Chi-Squared Assessment

If the participant has  $|E_{n,i}| > 1$  for one or more measurements, a chi-squared test statistic is computed

$$\chi_{obs}^2 = \sum_i (k \cdot E_{n,i})^2 = \sum_i \frac{DoE_i^2}{u_{DoE,i}^2}$$

summed over all  $m$  measurands for that participant.  $\chi_{obs}^2$  is compared against  $\chi_{95,m}^2$ , the 95<sup>th</sup> percentile of the chi-square distribution with  $m$  degrees of freedom (obtainable from statistical tables, or e.g. `CHISQ.INV(0.95, M)` in Excel or `scipy.stats.chi2.ppf(0.95, M)` in Python).

Briefly, the Chi-squared test statistic was calculated for participants that failed the  $E_n$  and Birge ratio test per CCL/WG-MRA/GD-3 (v1.7, July 2024), which itself follows Beissner, 2002. [3] As CCL/WG-MRA/GD-3 Appendix, Section 7 notes,  $Q$  strictly follows a chi-square distribution only if the normalized errors across the  $M$  measurands are independent. Because the results for one participant typically come from a single measurement campaign, some correlation between measurands (shared systematic effects) is likely in practice.

The two-step procedure below consisting of an immediate-acceptance rule for the fully-correlated case combined with the independent-sum-of-squares  $Q$  test is a deliberate, simple approximation that keeps the acceptance probability close to 95 % without requiring a full covariance model between measurands.

Attempting to calculate the full covariance with partial data via Mahalanobis distance yielded negative eigenvalues in an eigenvalue test. Thus, in order to attempt to calculate a full covariance model, participants in future comparisons would have to supply the necessary input data, denoting length-dependent and independent correlated and uncorrelated uncertainty sources in their uncertainty budget. That would, in theory, permit to calculate a correlated chi-squared statistic that accounts for both the systematic error between measurands of a participant and the SCR<sub>V</sub> of both measurands via groups of shared traceability chains.

**Decision Rule** The resulting test statistic  $\chi_{obs}^2$  is compared to  $\chi_{95,m}^2$ , which is the 95<sup>th</sup> percentile of the chi-square distribution with  $m$  degrees of freedom.

- If  $\chi_{obs}^2 \leq \chi_{95,m}^2$ , the participant's overall uncertainty claim is accepted, and the isolated values are deemed statistical anomalies consistent with the participant's stated uncertainty.
- If  $\chi_{obs}^2 > \chi_{95,m}^2$ , the threshold is exceeded and there is a possible systemic problem; the participant's results and uncertainty claim must be investigated.

The measured and reference values for the two measurands are listed in Table 11 and Table 12 with the associated standard uncertainty and Birge ratio. The gauge specific values for  $E_n$ ,  $w$ , DoE and  $u_{DoE}$  for all participants are listed in tables Table 13 through Table 18.

Table 11. Measured values, and reference value  $\overline{\Delta l_{c_w}}$  of gauge block deviation from nominal central length, associated standard uncertainty  $u(\overline{\Delta l_{c_w}})$ , Birge ratio, and chi-squared. The Birge ratio represents the value obtained after excluding outlier and disqualified participants per gauge block. Excluded participants are outlined by grey shading. The results of the pilot used in calculations are outlined in blue shading.

|                                           |                                | Gauge block nominal length (mm) |             |             |             |       |             |
|-------------------------------------------|--------------------------------|---------------------------------|-------------|-------------|-------------|-------|-------------|
| Meas. date                                |                                | 0,5                             | 4           | 10          | 25          | 50    | 100         |
| 2023.07.23                                | $\Delta l_{c_w}(\text{nm})$    | 37                              | 46          | 70          | 29          | 88    | 142         |
|                                           | $u(\Delta l_{c_w})(\text{nm})$ | 23,5                            | 23,5        | 26,5        | 28          | 36,5  | 49          |
| 2024.09.30                                | $\Delta l_{c_w}(\text{nm})$    | 34                              | 36          | 66          | 26          | 74    | 105         |
|                                           | $u(\Delta l_{c_w})(\text{nm})$ | 23,5                            | 23,5        | 26,5        | 28          | 36,5  | 49          |
| 2024.01.15                                | $\Delta l_{c_w}(\text{nm})$    | 45                              | 27          | 57          | 20          | 70    | 117         |
|                                           | $u(\Delta l_{c_w})(\text{nm})$ | 23,5                            | 23,5        | 26,5        | 28          | 36,5  | 49          |
| 2025.04.08                                | $\Delta l_{c_w}(\text{nm})$    | 41                              | 33          | 57          | 26          | 67    | 116         |
|                                           | $u(\Delta l_{c_w})(\text{nm})$ | 24                              | 23,5        | 26,5        | 28          | 36,5  | 49          |
| $\overline{\Delta l_{c_w}}(\text{nm})$    |                                | 30,4                            | 30,3        | 40,7        | 14,8        | 55,0  | 96,6        |
| $u(\overline{\Delta l_{c_w}})(\text{nm})$ |                                | 6,3                             | 6,7         | 6,7         | 7,6         | 9,2   | 14,4        |
| $R_B$                                     |                                | 1,015                           | 0,876       | 0,970       | 0,854       | 1,273 | 0,658       |
| $I$                                       |                                | 11                              | 10          | 10          | 10          | 11    | 10          |
| $R_B$ cutoff                              |                                | 1,377                           | 1,394       | 1,394       | 1,394       | 1,377 | 1,394       |
| Participants excluded                     |                                | DPM                             | DPM,<br>NPL | DPM,<br>NMI | DPM,<br>NMI | DPM   | DPM,<br>NMI |

Table 12. Measured values, and reference value  $\bar{v}_w$  of gauge block variation, associated standard uncertainty  $u(\bar{v}_w)$ , Birge ratio, and chi-squared. The Birge ratio represents the value obtained after excluding outlier and disqualified participants per gauge block. Excluded participants are outlined by grey shading. The results of the pilot used in calculations are outlined in blue shading.

| Meas. date            | Gauge block nominal length (mm) |       |       |       |       |       |
|-----------------------|---------------------------------|-------|-------|-------|-------|-------|
|                       | 0,5                             | 4     | 10    | 25    | 50    | 100   |
| 2023.07.23 $v_w$ (nm) | 9                               | 23    | 18    | 28    | 38    | 77    |
| $u(v_w)$ (nm)         | 15,5                            | 16    | 15,5  | 16    | 16    | 18    |
| 2024.09.30 $v_w$ (nm) | 20                              | 27    | 28    | 29    | 44    | 101   |
| $u(v_w)$ (nm)         | 16                              | 16    | 16    | 16    | 16,5  | 20    |
| 2024.01.15 $v_w$ (nm) | 16                              | 23    | 14    | 34    | 48    | 88    |
| $u(v_w)$ (nm)         | 16                              | 16    | 16,5  | 17    | 16,5  | 18,5  |
| 2025.04.08 $v_w$ (nm) | 25                              | 32    | 23    | 30    | 51    | 94    |
| $u(v_w)$ (nm)         | 16                              | 17    | 16    | 17    | 17    | 19    |
| $\bar{v}_w$ (nm)      | 20,4                            | 23,6  | 23,6  | 29,8  | 44,4  | 86,9  |
| $u(\bar{v}_w)$ (nm)   | 5,5                             | 5,5   | 5,7   | 6,1   | 6,7   | 7,5   |
| $R_{B,v}$             | 0,545                           | 0,586 | 0,733 | 0,347 | 0,595 | 0,955 |
| $I$                   | 12                              | 12    | 12    | 12    | 12    | 12    |
| $R_B$ cutoff          | 1,361                           | 1,361 | 1,361 | 1,361 | 1,361 | 1,361 |
| Participants excluded | None                            | None  | None  | None  | None  | None  |

Table 13.  $E_n$ , weight, degree of equivalence and associated standard uncertainties ( $k = 1$ ) for deviation from nominal central length and variation for 0,5 mm gauge block. Excluded participants are outlined in grey shading.

| Participant    | Deviation from nominal central length |       |     |           | Variation |       |     |           |
|----------------|---------------------------------------|-------|-----|-----------|-----------|-------|-----|-----------|
|                | $E_n$                                 | $w$   | DoE | $u_{DoE}$ | $E_n$     | $w$   | DoE | $u_{DoE}$ |
| METROSERT      | 0,182                                 | 0,029 | 9   | 24        | 0,123     | 0,118 | -4  | 16        |
| BEV            | 0,195                                 | 0,062 | -10 | 25        | 0,270     | 0,134 | -8  | 14        |
| BIM            | 0,026                                 | 0,016 | 3   | 50        | 0,169     | 0,010 | 19  | 55        |
| FTMC           | 0,346                                 | 0,015 | 15  | 22        | 0,387     | 0,063 | 17  | 22        |
| MIRS/UM-FS/LTM | 0,013                                 | 0,119 | -1  | 17        | 0,221     | 0,076 | 9   | 20        |
| NMI            | 1,049                                 | 0,043 | 62  | 30        | 0,247     | 0,053 | -12 | 24        |
| DPM            | 0,369                                 | 0     | -21 | 28        | 0,290     | 0,053 | -14 | 24        |
| LATMB          | 0,405                                 | 0,057 | -21 | 26        | 0,031     | 0,045 | -2  | 26        |
| NPL            | 0,766                                 | 0,151 | 23  | 15        | 0,221     | 0,076 | 9   | 20        |
| BFKH           | 0,222                                 | 0,097 | -9  | 20        | 0,451     | 0,032 | 28  | 31        |
| NSAI NML       | 0,793                                 | 0,359 | -13 | 8         | 0,092     | 0,302 | -2  | 9         |
| BMM            | 0,190                                 | 0,057 | 10  | 26        | 0,031     | 0,045 | -2  | 26        |

Table 14.  $E_n$ , weight, degree of equivalence and associated standard uncertainties ( $k = 1$ ) for deviation from nominal central length and variation for 4 mm gauge block. Excluded participants are outlined in grey shading.

| Participant        | Deviation from nominal central length |       |     |           | Variation |       |     |           |
|--------------------|---------------------------------------|-------|-----|-----------|-----------|-------|-----|-----------|
|                    | $E_n$                                 | $w$   | DoE | $u_{DoE}$ | $E_n$     | $w$   | DoE | $u_{DoE}$ |
| METROSERT          | 0,005                                 | 0,029 | -1  | 24        | 0,066     | 0,106 | -3  | 17        |
| BEV                | 0,467                                 | 0,063 | -23 | 25        | 0,132     | 0,135 | 4   | 14        |
| BIM                | 0,379                                 | 0,016 | -38 | 50        | 0,016     | 0,011 | 2   | 55        |
| FTMC               | 0,488                                 | 0,015 | -21 | 22        | 0,196     | 0,063 | -9  | 22        |
| MIRS/UM-<br>FS/LTM | 0,166                                 | 0,120 | -6  | 17        | 0,564     | 0,076 | 22  | 20        |
| NMI                | 0,28                                  | 0,044 | 17  | 30        | 0,179     | 0,053 | -9  | 24        |
| DPM                | 0,336                                 | 0     | -19 | 28        | 0,008     | 0,053 | -1  | 24        |
| LATMB              | 0,507                                 | 0,058 | -26 | 26        | 0,164     | 0,045 | -9  | 26        |
| NPL                | 1,151                                 | 0,135 | 37  | 16        | 0,382     | 0,076 | 15  | 20        |
| BFKH               | 0,825                                 | 0,108 | -30 | 18        | 0,153     | 0,037 | 9   | 29        |
| NSAI NML           | 0,410                                 | 0,361 | 7   | 8         | 0,500     | 0,304 | -9  | 9         |
| BMM                | 0,111                                 | 0,058 | -6  | 26        | 0,230     | 0,045 | 12  | 26        |

Table 15.  $E_n$ , weight, degree of equivalence and associated standard uncertainties ( $k = 1$ ) for deviation from nominal central length and variation for 10 mm gauge block. Excluded participants are outlined in grey shading.

| Participant        | Deviation from nominal central length |       |     |           | Variation |       |     |           |
|--------------------|---------------------------------------|-------|-----|-----------|-----------|-------|-----|-----------|
|                    | $E_n$                                 | $w$   | DoE | $u_{DoE}$ | $E_n$     | $w$   | DoE | $u_{DoE}$ |
| METROSERT          | 0,404                                 | 0,03  | 22  | 27        | 0,079     | 0,113 | -3  | 17        |
| BEV                | 0,24                                  | 0,071 | -12 | 25        | 0,463     | 0,144 | 13  | 14        |
| BIM                | 0,591                                 | 0,018 | -59 | 50        | 0,054     | 0,011 | 6   | 55        |
| FTMC               | 0,372                                 | 0,03  | -16 | 21        | 0,333     | 0,067 | -15 | 22        |
| MIRS/UM-<br>FS/LTM | 0,850                                 | 0,137 | 29  | 17        | 0,675     | 0,081 | 26  | 20        |
| NMI                | 3,256                                 | 0     | 207 | 32        | 0,089     | 0,057 | -5  | 24        |
| DPM                | 0,375                                 | 0     | -22 | 29        | 0,261     | 0,057 | -13 | 24        |
| LATMB              | 0,412                                 | 0,061 | -22 | 27        | 0,079     | 0,045 | -5  | 27        |
| NPL                | 0,046                                 | 0,153 | 2   | 16        | 0,414     | 0,081 | 16  | 20        |
| BFKH               | 0,039                                 | 0,111 | 2   | 19        | 0,27      | 0,036 | 16  | 30        |
| NSAI NML           | 0,204                                 | 0,334 | -4  | 9         | 0,752     | 0,268 | -15 | 10        |
| BMM                | 0,030                                 | 0,061 | -2  | 27        | 0,111     | 0,045 | 6   | 27        |

Table 16.  $E_n$ , weight, degree of equivalence and associated standard uncertainties ( $k = 1$ ) for deviation from nominal central length and variation for 25 mm gauge block. Excluded participants are outlined in grey shading.

| Participant    | Deviation from nominal central length |       |     |           | Variation |       |     |           |
|----------------|---------------------------------------|-------|-----|-----------|-----------|-------|-----|-----------|
|                | $E_n$                                 | $w$   | DoE | $u_{DoE}$ | $E_n$     | $w$   | DoE | $u_{DoE}$ |
| METROSERT      | 0,197                                 | 0,044 | 11  | 27        | 0,003     | 0,131 | -1  | 16        |
| BEV            | 0,757                                 | 0,085 | -38 | 25        | 0,079     | 0,168 | -3  | 14        |
| BIM            | 0,308                                 | 0,021 | -32 | 52        | 0,176     | 0,012 | 20  | 57        |
| FTMC           | 0,052                                 | 0,054 | 3   | 23        | 0,349     | 0,066 | -17 | 24        |
| MIRS/UM-FS/LTM | 0,133                                 | 0,159 | -5  | 18        | 0,259     | 0,095 | 10  | 20        |
| NMI            | 3,123                                 | 0     | 206 | 33        | 0,004     | 0,066 | -1  | 24        |
| DPM            | 0,598                                 | 0     | -42 | 35        | 0,047     | 0,066 | -3  | 24        |
| LATMB          | 0,083                                 | 0,067 | -5  | 28        | 0,003     | 0,045 | -1  | 29        |
| NPL            | 0,900                                 | 0,159 | 32  | 18        | 0,083     | 0,095 | -4  | 20        |
| BFKH           | 0,377                                 | 0,108 | 17  | 22        | 0,337     | 0,035 | 22  | 33        |
| NSAI NML       | 0,367                                 | 0,259 | -9  | 12        | 0,006     | 0,193 | -1  | 13        |
| BMM            | 0,070                                 | 0,050 | -5  | 34        | 0,003     | 0,035 | -1  | 33        |

Table 17.  $E_n$ , weight, degree of equivalence and associated standard uncertainties ( $k = 1$ ) for deviation of central length and variation for 50 mm gauge block. Excluded participants are outlined in grey shading.

| Participant    | Deviation from nominal central length |       |     |           | Variation |       |     |           |
|----------------|---------------------------------------|-------|-----|-----------|-----------|-------|-----|-----------|
|                | $E_n$                                 | $w$   | DoE | $U_{DoE}$ | $E_n$     | $w$   | DoE | $U_{DoE}$ |
| METROSERT      | 0,285                                 | 0,045 | 21  | 36        | 0,019     | 0,176 | -1  | 15        |
| BEV            | 0,342                                 | 0,115 | -18 | 26        | 0,313     | 0,200 | 9   | 14        |
| BIM            | 0,275                                 | 0,025 | -31 | 56        | 0,047     | 0,013 | -6  | 61        |
| FTMC           | 0,279                                 | 0,075 | -16 | 28        | 0,259     | 0,054 | -15 | 29        |
| MIRS/UM-FS/LTM | 0,391                                 | 0,173 | 16  | 21        | 0,382     | 0,113 | 15  | 19        |
| NMI            | 0,137                                 | 0,077 | 9   | 32        | 0,339     | 0,078 | -16 | 24        |
| DPM            | 0,028                                 | 0     | 3   | 48        | 0,031     | 0,078 | 2   | 24        |
| LATMB          | 1,249                                 | 0,067 | -85 | 34        | 0,064     | 0,037 | 5   | 35        |
| NPL            | 1,411                                 | 0,134 | 66  | 24        | 0,037     | 0,113 | 2   | 19        |
| BFKH           | 0,513                                 | 0,087 | -31 | 30        | 0,384     | 0,024 | 34  | 44        |
| NSAI NML       | 0,116                                 | 0,171 | -5  | 19        | 0,644     | 0,102 | -26 | 20        |
| BMM            | 0,047                                 | 0,037 | -5  | 48        | 0,152     | 0,020 | 15  | 48        |

Table 18.  $E_n$ , weight, degree of equivalence and associated standard uncertainties ( $k = 1$ ) for deviation from nominal central length and variation for 100 mm gauge block. Excluded participants are outlined by grey shading.

| Participant        | Deviation from nominal central length |       |     |           | Variation |       |     |           |
|--------------------|---------------------------------------|-------|-----|-----------|-----------|-------|-----|-----------|
|                    | $E_n$                                 | $w$   | DoE | $u_{DoE}$ | $E_n$     | $w$   | DoE | $u_{DoE}$ |
| METROSERT          | 0,251                                 | 0,074 | 24  | 47        | 0,037     | 0,144 | 2   | 19        |
| BEV                | 0,4                                   | 0,201 | -23 | 29        | 0,729     | 0,256 | 19  | 13        |
| BIM                | 0,237                                 | 0,039 | -33 | 70        | 0,127     | 0,01  | -20 | 76        |
| FTMC               | 0,255                                 | 0,094 | 22  | 42        | 0,256     | 0,03  | -23 | 44        |
| MIRS/UM-<br>FS/LTM | 0,082                                 | 0,189 | -5  | 30        | 0,023     | 0,144 | 1   | 19        |
| NMI                | 3,254                                 | 0     | 253 | 39        | 0,641     | 0,1   | -30 | 23        |
| DPM                | 0,565                                 | 0     | -93 | 83        | 0,948     | 0,1   | -44 | 23        |
| LATMB              | 0,335                                 | 0,069 | -35 | 53        | 0,289     | 0,02  | 31  | 54        |
| NPL                | 0,674                                 | 0,117 | 54  | 40        | 0,645     | 0,144 | 24  | 19        |
| BFKH               | 0,045                                 | 0,058 | 6   | 59        | 0,169     | 0,007 | 32  | 95        |
| NSAI NML           | 0,126                                 | 0,134 | -9  | 36        | 0,526     | 0,04  | -40 | 38        |
| BMM                | 0,032                                 | 0,03  | 6   | 83        | 0,185     | 0,009 | 31  | 84        |

Table 19. Maximum  $E_n$ , chi-squared statistics, and assessment status of participants. Participants requiring further investigation are outlined by grey shading.

| Participant      | Deviation from nominal central length |          |               |                                | Variation   |          |               |                        |
|------------------|---------------------------------------|----------|---------------|--------------------------------|-------------|----------|---------------|------------------------|
|                  | $E_{n,max}$                           | $\chi^2$ | $\chi^2_{95}$ | Assessment                     | $E_{n,max}$ | $\chi^2$ | $\chi^2_{95}$ | Assessment             |
| METROSERT        | 0,57                                  | 3,96     | 12,59         | Accept<br>$E_n \leq 1$         | 0,39        | 1,34     | 12,59         | Accept<br>$E_n \leq 1$ |
| BEV              | 0,76                                  | 4,41     | 12,59         | Accept<br>$E_n \leq 1$         | 0,81        | 4,66     | 12,59         | Accept<br>$E_n \leq 1$ |
| BIM              | 0,58                                  | 2,73     | 12,59         | Accept<br>$E_n \leq 1$         | 0,18        | 0,34     | 12,59         | Accept<br>$E_n \leq 1$ |
| FTMC             | 0,37                                  | 2,07     | 12,59         | Accept<br>$E_n \leq 1$         | 0,41        | 2,02     | 12,59         | Accept<br>$E_n \leq 1$ |
| MIRS/UM-FS/LTM   | 0,87                                  | 3,75     | 12,59         | Accept<br>$E_n \leq 1$         | 0,69        | 4,08     | 12,59         | Accept<br>$E_n \leq 1$ |
| NMI              | 3,27                                  | 129      | 12,59         | Investigate<br>$\chi^2$ failed | 0,59        | 2,86     | 12,59         | Accept<br>$E_n \leq 1$ |
| DPM              | 2,08                                  | 25,1     | 12,59         | Investigate<br>$\chi^2$ failed | 0,90        | 3,79     | 12,59         | Accept<br>$E_n \leq 1$ |
| LATMB            | 1,26                                  | 10,4     | 12,59         | Accept<br>$\chi^2$ passed      | 0,31        | 0,45     | 12,59         | Accept<br>$E_n \leq 1$ |
| NPL              | 1,40                                  | 20,3     | 12,59         | Investigate<br>$\chi^2$ failed | 0,70        | 4,02     | 12,59         | Accept<br>$E_n \leq 1$ |
| BFKH             | 0,68                                  | 3,73     | 12,59         | Accept<br>$E_n \leq 1$         | 0,47        | 2,64     | 12,59         | Accept<br>$E_n \leq 1$ |
| NSAI NML         | 0,79                                  | 5,77     | 12,59         | Accept<br>$E_n \leq 1$         | 0,72        | 4,74     | 12,59         | Accept<br>$E_n \leq 1$ |
| BMM <sup>a</sup> | 0,19                                  | 0,18     | 12,59         | Accept<br>$E_n \leq 1$         | 0,32        | 0,74     | 12,59         | Accept<br>$E_n \leq 1$ |

<sup>a</sup>Updated values were assessed.

## 8.2. Participant responses

**DPM** (D. Bulku, September 17, 2025):

First of all, we would like to thank you for Report Draft A. As noted in the report, our laboratory had two outlier results at the lengths of 25 mm and 100 mm.

We would like to inform you that, following an internal review and analysis of the comparison results, we identified a **technical error in the Excel data processing**, which affected the results we initially reported. We understand that it is now too late to make formal corrections.

During our review, we clearly identified that not only were there two outlier results, but due to this technical error, **all calibration results from our laboratory were not traceable to the EIM reference values**.

Instead of using the correct traceable central length formula  $L_c = (L_{ref} + R_2 - R_1)$ , the Excel cell mistakenly used  $L_c = L_n + R_2 - R_1$ , which caused the results to fall outside of traceability.

Furthermore, we incorrectly reported the comparison measurement difference ( $R_2 - R_1$ ) as the deviation in central length, that is completely different to real traceable central length deviation.

Following the correction of the central length deviation values, according to the Report Draft A, all comparison results seem to be statistically significant.

**DPM** (D. Bulku, September 18, 2025):

I wrote you before, about the mistake in calculation of comparison results. So, for more info please find in attach file the doc:

- Appendix C 2024
- Appendix C reviewed 2025
- Excel doc. explaining the case of error.

As the results are published, Draft A, there is no possibility changing anything. Well, I stated also that, other results are incorrect, even statistically they seem to be significant. Does the report address it in a way?

**BMM** (T. Božović, August 22, 2025)

Thank you for your email and the report draft A1.

I would like to thank all participants for the opportunity to repeat the measurements and I apologize as it caused an extension of the entire process.

There is a mistake regarding our measurement uncertainty data in the Draft A1.

In our Appendix C are given  $u(l_c)$  and  $u(v)$  not  $U(l_c)$  and  $U(v)$ , so the values shouldn't be divided by 2.

Our mistake is that we reported a difference of measurement uncertainty in relation to our current CMC, as significant.

It's similar with the reported damages on the standards. Maybe we shouldn't have reported them as significant, except for the reference of 100 mm.

I apologize for that because it caused additional work in processing the measurement results.

A lesson has been learned for next time.

**NPL** (A. Lewis, September 2, 2025):

I note the two NPL results which do not agree with the reference value within the uncertainties.

We are looking into the measurement data for these and will give some possible explanations after an internal discussion. Sorry that we were such a bad linking lab for these two gauges!

I am also looking through the report in more detail but this will take some time as I have a mountain of other things to do.

I noticed a few small errors in the results tables where the METROSERT uncertainty sometimes jumps by x10 (due to missing decimal separator). I assume these are just in the display and not in the actual numbers in the analysis.

Three items to note in the analysis:

1. Only 1 METROSERT result can be included for each gauge – it should be the first measurement – the one you did before you saw any other results.
2. For the two NPL ‘bad’ results, you have excluded them from the calculations and from the results display. Our usual process is to only exclude them when the Birge ratio check is not satisfied. Then we start excluding results (set the weighting to zero) starting with the result with the highest En value, and stop excluding results from the calculation when the Birge ratio test is OK. I think this means that the NPL results for the 4 mm and 50 mm gauges should still be part of the calculation, because the Birge ratio test is OK despite the bad NPL results.
3. Even laboratories with bad results should have their En values in the tables, as well as DoE values.

There is another statistical test we can perform across the set of all NPL results to see if there is a big problem or not, but I have not yet had time to do this. It is a process proposed by Brian Eves at a recent WG-MRA meeting that we plan to implement in the guidance. When I have a chance to try it, I will share with you.

**METROSERT** reply (J. S. Eensalu, September 3, 2025)

Thank you for your feedback

1.  
In the analysis, the time-averaged Metrosert result was used for each gauge block as per chapter 7.2 in the technical protocol: Steel gauge blocks occasionally show a

growing or a shrinkage the rate of which is approximately linear with time. Since the artefacts used here are new and with no history, the instability of the blocks must be determined in course of the comparison. For this check the measurements of the pilot laboratory are used exclusively, not that of the other participants. Using these data a linear regression line is fitted and the slope together with its uncertainty is determined (per gauge block). Three cases can be foreseen: a) The linear regression line is an acceptable drift model and the absolute drift is smaller than its uncertainty. The gauge block is considered stable and no modification to the standard evaluation procedure will be applied. In fact the results of the pilot's stability measurements will not influence the numerical results in any way.

*The case A applies in this comparison based on our measurement results (see Figures 4 and 5 in the report).*

2. b) The linear regression line is an acceptable drift model and the absolute drift is larger than its uncertainty, i.e. there is a significant drift for the gauge block. In this case an analysis similar to [Nien F Z et al. 2004, Statistical analysis of key comparisons with linear trends, Metrologia 41, 231] will be followed. The pilot influences the reference value by the slope of the drift only, not by the measured absolute lengths. The data are not compatible at all with a linear drift, regarding the uncertainties of the pilot's measurements. In this case the artefact is unpredictably unstable or the pilot has problems with its measurements. TC-L has to determine the further approach.
3. Unfortunately, the results of these two gauge blocks did not satisfy the Birge ratio requirement of  $RB < 1$  when NPL was included. I followed the procedure of first including all participants, and thereafter eliminating the participant with the highest  $En$  magnitude one by one iteratively until the Birge ratio decreased below 1. The Birge ratio presented is the value obtained after eliminating the mentioned participants. Perhaps this fact should be expressed more explicitly in the report. Based on the statistical analysis all outliers by all outlined participants were significant and none were in the grey zone. It is possible that most participants are not taking into consideration something that NPL does, which would explain the somewhat positive bias of the central length deviation for NPL for all gauge blocks, but without further tests it is difficult to draw conclusions as to what this originates from.
4. I will add the  $En$  results of all participants in the tables. DoE is already included for all participants. The minus sign meant that that participant was not included in the weighting.

5. The uncertainty values for Metrosert in the tables were indeed missing decimals in a few places. I appreciate that you noticed. This affected only the table results, however, as the time-averaged values were used in the analysis.

**NPL** (A. Lewis, September 26, 2025)

Sorry for the delay in responding – too much work to do!

I agree with using the time-weighted METROSERT result for each gauge block. This should be the only METROSERT result in the En calculations. It would be good to show these time-weighted results somewhere. I suggest adding the time-weighted values to the tables of results. This will help anyone who is checking the tables perform their own En calculations using on the time-weighted result for METROSERT. I tried to do this but I get slightly different En values to the ones you report.

For example:

Using 35.5 nm as the average result for METROSERT for the 4 mm gauge, I get a Birge ratio that is OK with all participants included. The NPL En value is 1.17.

Using 62.5 nm for METROSERT for the 10 mm gauge, the Birge ratio fails until we exclude NMI (MD)

Using 25.3 nm for METROSERT for the 25 mm gauge, the Birge ratio fails until we exclude NMI (MD) (we do not need to exclude DPM).

Using 74.8 nm for METROSERT for the 50 mm gauge, the Birge ratio is OK with all included. NPL En is 1.4 and LATMB En is 1.21.

I attach my analysis Excel.

We have the following comments on the discrepant NPL results.

#### **4 mm gauge block**

The NPL master gauge was calibrated 3/2022.

Master gauge re-calibration in 09/2025 shows zero change in length.

Gauge showed pitting and light scratches which cross the measurement region.

Our 3 results had a 10 nm spread.

### **50 mm gauge block**

The NPL master gauge was calibrated 3/2022.

Master gauge re-calibration in 09/2025 shows +11 nm change in length. That would make our result longer

Gauge had some pitting but not in the measurement region.

Our 3 results had a 10 nm spread.

So, nothing obvious as to the source of the problem. Our comparator is very old (40+ years) and starting to show issues which affect its zeroing. It is possible this is the cause of the problem. Normally comparator results agree with the GBI results within the uncertainties (i.e. within 40 nm) but as the protocol did not allow interferometry measurements, we did not put the gauges in the GBI.

### **METROCERT reply (J. S. Eensalu, October 1, 2025):**

I looked at the calculations you sent. I admit not presenting the time averaged values makes it high impossible for others to reproduce the results obtained. The Python scripts I used to analyse the data are attached with the output Excel files.

The time averaged Metrocert central length deviation values used in our respective analyses were quite a bit different. Hence the differences in results. I am intrigued as to whether our results will match when using the same initial values from A1.

Note that I excluded participants if the Birge ratio was even slightly above 1.000.

Meanwhile, DPM has informed us that all of their central length deviation results were miscalculated due to using the nominal value instead of the central length deviation of their standard.

After consultation with Dariusz, all results by DPM are to be excluded from the RV calculation, and I will have to analyze their results separately in an additional table and present relevant graphs in A2, passing a decision based on their updated data, as Tim had previously suggested to Dariusz for another supplementary comparison.

Perhaps it would be prudent to suggest at the upcoming TC-L that future gauge block supplementary comparisons should make use of interferometric measurements by the pilot or co-pilot to mitigate the emergence of such complicated conundrums as much as possible? What is your opinion on this matter?

**NPL** (A. Lewis, 14 October, 2025)

Nice to see you over the last 2 days (even if only on a screen).

Just a quick note to confirm that the TC-L meeting recommended to not use interferometric references for comparisons such as this due to the measurand being slightly different. Also, Antti Lassila mentioned to Tim during lunch that supplementary comparisons do not need the linking that it used in key comparisons.

So the mechanical results should stand on their own and it looks like NPL has some issues with their comparator

**NPL** (A. Lewis, 02 February, 2026)

Thanks for the update. It's never easy being a pilot!

Tim has now taken over from me as the TC-L Delegate but he's not the technical expert for the gauge block area so I am happy to respond with the following observations.

1. Section 7.2 caption for table 10 : italicise the  $k$  in  $U_{CMC}$ ,  $k = 2$ .
2. Table 11 – for the 10 mm gauge block, NMI was also excluded (as shown in Table 15).
3. Table 11 – for the 100 mm gauge block, NMI was also excluded (as shown in Table 18).
4. For the results tables where some participants have been excluded (weight = 0), it is customary to still include their  $En$  value, (even though they are not contributing to the weighted mean). This is necessary for discussion of whether or not their results support the CMC claim. Tables 13 to 18 should include the ( $>1$ )  $En$  values for the labs that were excluded from the weighted mean.

5. In section 8.1 you state that participants are eliminated, in order, until the Birge ratio is less than 1. According to the CCL guidance document [GD-3.2](#) (the report template) the Birge ratio has an expectation value of 1 only when considering standard uncertainties. For a coverage factor of  $k = 2$  the expectation value is increased and the data in a comparison are consistent provided that

$$R_B < \sqrt{1 + \sqrt{8/(I - 1)}}$$

where  $I$  is the number of laboratories contributing data. For this comparison  $R_B$  is around 1.4 due to there being between 8 and 10 contributions to each weighted mean. I am not a statistics expert so I do not recall where this expectation value comes from but it has been used in all the comparison reports and is embedded in the spreadsheet used by several pilots. I analysed the results using this expectation value and in several cases the Birge ratio is satisfied without having to eliminate anyone. I should clarify that I was using the raw uncertainty data without taking into account correlations. My initial check suggests that the overall change is negligible – those with high  $En$  values still remain high and there is no change to the overall outcome.

6. The uncertainty formula for  $U(DoE)$  for included participants is stated as

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 + [u_{int}(\bar{x}_w)]^2 - 2 \cdot cov(x_i; \bar{x}_w)}$$

In situations where there are no correlations, the formula stated in the guidance document for results not contributing to the weighted mean is

$$u_{DoE,i} = \sqrt{[u(x_i)]^2 - [u_{int}(\bar{x}_w)]^2}$$

So I wonder if the first sign in your equation should be a '+', as when there is no correlation, the term  $2 \cdot cov(x_i; \bar{x}_w)$  is zero ? [J.S.E 12 March 2026: see updated chapter 8.1 on why I disagree]

7. I don't think the corrected analysis benefits the two NPL results with  $En > 1$ . In both cases we measure 'long'. Since the comparison, we have re-calibrated most of the gauge blocks we used as references. The results for the 4 mm gauge are unchanged. The results for the 50 mm gauge are now 11 nm shorter, but we still get  $En \sim 1.1$ .
8. Showing the NPL  $En$  values for all gauges (and for all other participants who were excluded) is useful because the CCL has now issued guidance on analysing multiple artefact comparisons, to decide if there is support for CMC claims – see section 5.2 of the [GD3](#) document (v1.7). This requires chi-squared test statistics to be performed using the calculated  $En$  values.

Sorry to not have spotted these potential issues sooner.

### 8.3. Analysis of revised DPM results

According to calculated values in Table 20, illustrated in Figure 12, the revised deviation from nominal central length of DPM supports the uncertainty claim based on chi-squared analysis. Thus, further corrective actions are not necessary. The corrected values have a systematic net negative bias vs the SCR.V.

Table 20. Comparison of SCR.V to revised DPM values for deviation from nominal central length.

|                                     | Gauge block nominal length (mm) |      |      |      |      |      |
|-------------------------------------|---------------------------------|------|------|------|------|------|
|                                     | 0,5                             | 4    | 10   | 25   | 50   | 100  |
| $\overline{\Delta l_{c_w}}$ (nm)    | 30,4                            | 30,3 | 40,7 | 14,8 | 55,0 | 96,6 |
| $u(\overline{\Delta l_{c_w}})$ (nm) | 6,3                             | 6,7  | 6,7  | 7,6  | 9,2  | 14,4 |
| $\Delta l_{c_{DPM}}$ (nm)           | 10                              | 17   | 20   | -27  | 57   | 2    |
| $u(\Delta l_{c_{DPM}})$ (nm)        | 26                              | 26   | 27   | 33   | 48   | 84   |
| $E_n$                               | 0,38                            | 0,25 | 0,37 | 0,62 | 0,02 | 0,55 |
| $\chi^2$                            | 4,14                            |      |      |      |      |      |
| $\chi^2_{95}$                       | 12,59                           |      |      |      |      |      |

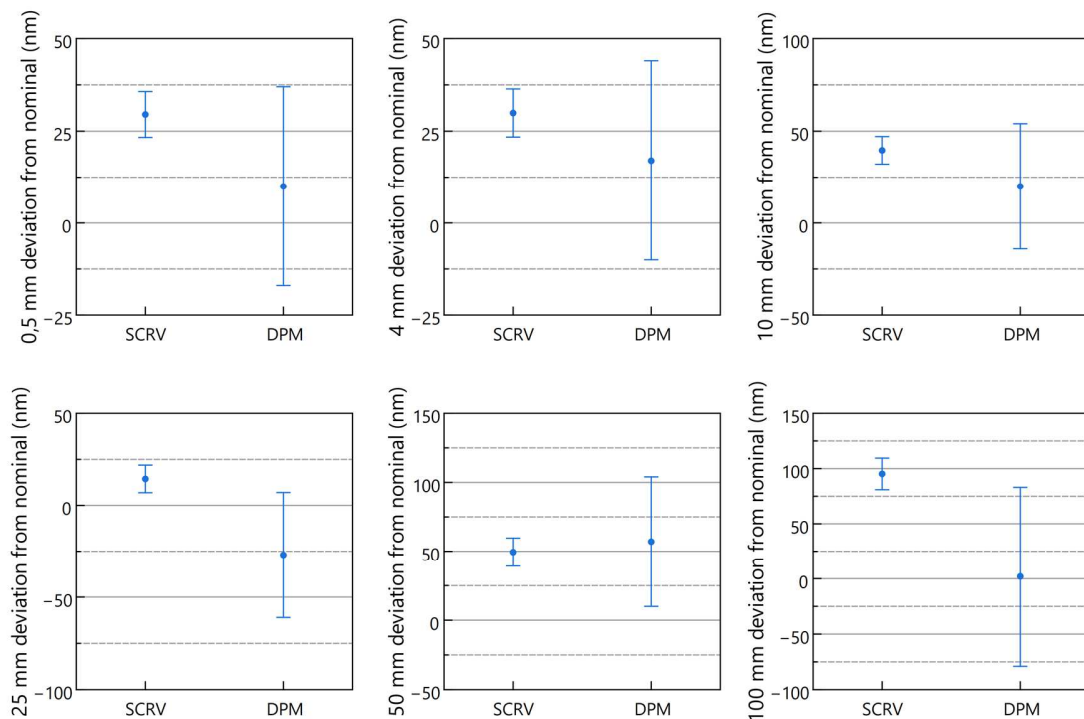


Figure 12. Comparison of SCR.V to revised DPM values for deviation from nominal central length.

#### 8.4. Discussion of results

Outliers were eliminated per measurand per gauge block until the Birge ratio was less than the specified value. Weights were not capped for analysis.

The statistical data analysis shows that NMI was an outlier for the 10 mm, 25 mm, and 100 mm gauge block deviation from nominal central length. There was a systematic net positive bias in their deviation from nominal central length measurement procedure.

NPL was an outlier for the deviation from nominal central length of the 4 mm gauge block in the  $E_n$  analysis, also exhibiting a net positive bias in the deviation from nominal central length measurement procedure. According to chi-squared analysis the CMC is not supported and further corrective actions are required.

Based on the originally submitted data, DPM was an outlier for the deviation from nominal central length of the 25 mm and the 100 mm gauge blocks with a net negative bias in general in comparison to the reference value.

Chi-squared analysis indicates the CMC claim of LATMB is supported and further corrective actions are not needed.

NMI, NPL and LATMB were asked without disclosing results whether they wanted to adjust their reported results. NMI and NPL chose not to change their results. LATMB provided updated data for the 100 mm gauge block, changing deviation from nominal central length from +80 nm to +60 nm and variation from +110 nm to +120 nm.

Regarding the principal measurand, the deviation from nominal central length, a total of 4 values out of 72 were not compatible with the final reference values per  $E_n$ , Birge ratio and chi-squared analysis, representing 5,6 % of the full set of results.

DPM did not have a relevant CMC previously, and the results originally submitted by DPM support the claimed uncertainty for gauge blocks other than 25 mm and 100 mm nominal length. However, DPM recognized after Draft A.1 that the deviation from nominal central length calculation was wrong for all 6 gauge blocks. Thus, as of Draft A.2, all deviation from nominal central length results of DPM were excluded from the SCR<sub>V</sub> calculation, and a chi-squared statistic test vs the SCR<sub>V</sub> was calculated using the corrected values submitted by DPM. The claimed uncertainty by DPM after submitting corrections is supported by chi-squared analysis.

The CMC claim is not supported for NMI (10 mm; 25 mm; 100 mm) based on outlier results in the statistical analysis, mandating corrective action. Moreover, the uncertainty budget supplied in Appendix C was inconsistent in unit conversion and did not include length dependent contributions present in the CMC budget.

Overall, the largest Birge ratio of 1,273 indicates a realistic uncertainty estimation for the remaining participants.

No outliers were found in the variation measurement results of any gauge block. The Birge ratio was largest for the variation of the 100 mm gauge block at  $R_B$  0,955.

## 9 Conclusions

The following conclusions can be drawn from the EURAMET.L-S2.2.n01 supplementary comparison:

- The time taken to perform all measurements, and transport the artefacts, was 9 months longer than originally planned. The delay was partly to allow two laboratories an opportunity to measure again, rather than being excluded from the comparison, and partly due to issues related to the ATA Carnet and Customs.
- None of the gauge blocks were damaged significantly during the circulation according to the measurement results of the pilot laboratory.
- In all, the comparison has been mostly successful. A total of 5,6 % of deviation from nominal central length results are considered outliers and 0 % of variation results are considered outliers.
- The CMC claims for all participants except NMI and NPL are supported by the results of this comparison. The CMC claim for deviation from nominal central length for NMI and NPL is not supported. Upon review, it was discovered that NMI supplied an incomplete and inconsistent uncertainty budget table in Appendix C, possibly indicating a systematically miscalculated CMC uncertainty budget. Therefore, NMI and NPL must implement corrective actions.

## Appendix A – Equipment and measuring processes of the participants

The participants were asked to supply this information in a format ready for inclusion in Appendix D of the technical protocol.

### METROSERT

Make and type of instrument(s)

TESA UPC 100 SN 8C02 for gauge blocks of 0,5 mm to 100 mm and TESA UPC 500 SN 9B 01 for gauge blocks of 0,5 mm to 500 mm. Resolution: 10 nm. Standard gauge blocks were calibrated by interferometry at GUM. Comparators are calibrated internally every 6 months.

The combined measurement sequence for central length and deviation is  $5 \times (P1-P2-P1-P3-P4-P5-P6-P2)$ . The gauge blocks were either measured by one person with both comparators on different days or by different persons on different days with undisclosed comparator and standard gauge block set in a double-blind experiment.  $E_n$  was  $<0,3$  in every case. The results of the participant contact were used in the report.

Range of gauge block temperature during measurements & description of temperature measurement method:

Temperature range: 19,5 °C to 20,5 °C taken as limits. The temperature of the gauge blocks was not measured for this comparison. The air temperature in the room in the vicinity of the gauge blocks was measured using an iButton DS1923 hygrothermochron with a resolution of 0,0625 °C and an expanded uncertainty of 0,3 °C.

An Almemo 470-1 surface temperature data logger with 3D printed PETG clamps was used to determine that the gauge block temperature of grade K 100 mm and 500 mm gauge blocks differed by less than the expanded uncertainty of 0,05 °C at about 19,74 °C after one full day of stabilisation when measured using the related Almemo NTC contact probes. The temperature of  $> 25$  mm gauge blocks was observed to be close to identical to the temperature of the comparator and the air-cushioned granite table and was not significantly influenced by the oscillations in air temperature.

### BEV

Make and type of instrument(s)

Mahr 626 + Millitron 1240, enhanced resolution (1 nm), dedicated software developed at BEV. Standard gauge blocks (steel, KOBA) are calibrated by interferometry in-house.

Our standard measurement sequence for the central length is:  $5 \times (P1-P2-P2-P1)$ , for the deviation in length:  $3 \times (P2-P3-P4-P5-P6-P6-P5-P4-P3-P2)$ . For this comparison each block was measured 5 to 15 times on different days by different persons.

Range of gauge block temperature during measurements & description of temperature measurement method:

Temperature range: 19,65 °C to 19,98 °C

Temperature of blocks were not measured actually, the air temperature (and humidity) in the near vicinity to the blocks is recorded and its average used to estimate the block temperature. Instrument used is a Vaisala HMT330 transmitter. The resolution of the sensor is 0,01 °C, the expanded uncertainty about 0,05 °C near 20 °C.

## **BIM**

Make and type of instrument(s):

Gauge block comparator TESA UPD for mechanical comparison of gauge blocks with range from 0,5 mm to 100 mm. TESA UPD allows comparative measurement of two gauge blocks with same nominal length or with variation in length up to 25 mm.

Temperature device with 4 sensors type PT-100.

TESA up software program allows all length and temperature values as measured to be further processed online and makes all the necessary corrections automatically.

Range of gauge block temperature during measurements & description of temperature measurement method:

Gauge block temperature during measurements:  $(20 \pm 0,3)^{\circ}\text{C}$ .

Temperature device - thermometer Fluke, type 1529-R with 4 sensors, type PT-100. Two sensors are used for temperature measurement of reference and tested gauge blocks, one sensor is mounted on the measuring table and one sensor is mounted on the measuring arm close to upper probe.

## **FTMC**

Make and type of instrument(s)

Mechanical comparator with two contacts Mahr 826.

Resolution -0,01  $\mu\text{m}$ .

Range of gauge block temperature during measurements & description of temperature measurement method: the temperature of laboratory environment was measured by ALMEMO 6290-7S

Range of environmental temperature:

0,5 mm -  $19,86 \div 20,01$ ; 4 mm –  $19,95 \div 20,26$  mm; 10 mm –  $19,88 \div 20,20$ ; 25 mm - $19,82 \div 20,02$ ;

50 mm –  $19,94 \div 20,12$ ; 100 mm –  $19,82 \div 20,09$  mm.

## MIRS/UM-FS/LTM

Make and type of instrument(s):

Instrument: Gauge block comparator 826, Manufacture: Mahr, Resolution: 10 nm

The traceability of the comparator was ensured by internal calibration in the laboratory (Certificate No.: 13311-LTM-23).



Range of gauge block temperature during measurements & description of temperature measurement method:

Equipment: Temperature measuring system TEMP 14, Manufacturer: JENAer Meßtechnik (Germany)

Range of gauge block temperature during measurement:

- Material sensor: (19,9262 – 20,1139) °C
- Air sensor: (19,8871 – 20,1437) °C

The temperature during the measurement was monitored on the gauge block comparator table with a material sensor, while the temperature in the measuring room was simultaneously recorded with an air sensor positioned close to device. The traceability of the measuring system was ensured by calibration in the National Temperature Laboratory - Laboratory for Metrology and Quality (LMK) of the Faculty of Electrical Engineering at the University of Ljubljana (Certificate No.: LMK0723I856-860).

## **NMI**

Make and type of instrument(s):

TESA UPC, Model 05930015, SN 5U01

Resolution of Comparator 0,01  $\mu\text{m}$

Steel gauge blocks, grade K (TESA)

Range of gauge block temperature during measurements & description of temperature measurement method:

$$T = (19,62 \dots 19,90) ^\circ\text{C}, U = 0,05 ^\circ\text{C}$$

## **DPM**

Make and type of instrument(s)

*Gauge Block comparator EMP II, Steinmeyer, Feinmess Suhl GmbH, Germany. Range 0 – 100 mm, resolution 0.01  $\mu\text{m}$ .*

*Digital thermometer AHLBORN, ALMEMO MA 25903S. Three channel with external probe, type NTC, resolution 0.01  $^\circ\text{C}$ , for gauge block temperature measurement.*

*Digital capacity hygrometer with thermometer TESTO Saveris H4D, with external probe, resolution 0.1 % and 0.1  $^\circ\text{C}$  for laboratory temperature measurement.*

Range of gauge block temperature during measurements & description of temperature measurement method:

*Temperature range for gauge blocks has been between 20.0  $^\circ\text{C}$  to 20.6  $^\circ\text{C}$ .*

*Temperature measurement method for gauge blocks is carried out by a simple contact of external probe to each gauge block surface after measurements.*

## **LATMB**

Make and type of instrument(s):

The difference measurements are carried out on a Mahr gauge blocks comparator type 826PC with a Millimar display unit, scale resolution 0,01  $\mu\text{m}$ .

Range of gauge block temperature during measurements & description of temperature measurement method:

During the measurements range of temperature were 19,8  $^\circ\text{C}$  to 20,1  $^\circ\text{C}$ .

The temperature is determined directly at the gauge blocks measuring device.

The air temperature of the laboratory is continuously recorded near the measurement device.

## NPL

Make and type of instrument(s)

Cary IVC-150 Gauge Block Comparator      S/N 4071

Cary IVC-153 X-Y table      (mounted on above item)

Carylabor display unit      S/N 13344

Grade K steel gauge blocks selected from two sets of master gauge blocks, S/N 100077 and 09136



Range of gauge block temperature during measurements & description of temperature measurement method:

The temperature is not recorded directly but is controlled such that the temperature during calibration will be in the range 19.8 °C to 20.2 °C. The indicator for the temperature is a self-built logging unit based on a Raspberry Pi computer, connected to digital DS18B20 temperature sensors. One sensor is near to the Cary, mounted on the steel surface plate where the gauge blocks are thermally stabilised. The logging unit and sensors are calibrated by the Temperature and Humidity group at NPL.

## **BFKH**

Make and type of instrument(s)

- Double side tactile gauge block comparator; Cary; IVC 154; 4141; HOS-0013/2023
- Set of standard gauge blocks; Mitutoyo; BM1-122K/YJ; 2307080; 2311471
- Standard gauge block; Mitutoyo; 0,5mm; 238563; 2400977

Combined digital multi sensor thermo, and hygrometer; TITON Bt.; AEROMES 06; 201; HÕM-0065/2024 for thermometer and GAZ-0083/2024 for hygrometer

Range of gauge block temperature during measurements & description of temperature measurement method:

20 °C ± 0,5 °C;

Indirect measurement based on the temperature of the used heat compensation table, air temperature measurement with a multi sensor resistance thermometer with a resolution of 0.01 °C.

## **NSAI NML**

Make and type of instrument(s)

TESA UPC gauge block comparator with T90 indicator

Range of gauge block temperature during measurements & description of temperature measurement method:

20.1 °C to 20.2 °C

Testo temperature system: indicator & 2 probes (1 on platten & 1 on the TESA GB comparator)

## **BMM**

Make and type of instrument(s)

Measurement is done using the mechanical gauge block comparator, PRECIMAR 826 PC (s/n 1204/23) equipped with an electronic indication device, Mahr Millimar C 1202 (s/n: 32040184), supported with QM Soft application.

Range of gauge block temperature during measurements & description of temperature measurement method:

For the establishing declared environment temperature condition  $20 \pm 0,5$  °C in our laboratory we are using Uniflier climate system.

Besides that, on the very measurement position we measure gauge blocks temperature using system for precise temperature measurements produced by JENAer Mesttechnik GmbH, typ TEMP14 consisting of three temperatures sensor s/ns: NC269, NC270, NC271 position on the antivibrate table surface and gauge blocka dn other two on the very comparator table.

The 24 hours period of stabilization of gauge blocks is prescribed before the start of calibration.

## 10 References

- [1] Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Oct. 14, 1999.
- [2] Evaluation of measurement data — Guide to the expression of uncertainty in measurement, JCGM 100:2008. doi: 10.59161/JCGM100-2008E.
- [3] K Beissner, “On a measure of consistency in comparison measurements,” Metrologia, vol. 39, no. 1, p. 59, Feb. 2002, doi: 10.1088/0026-1394/39/1/8.
- [4] Figures drawn using LabPlot, <https://labplot.kde.org/>.

## 11 Source code for calculating SCR<sub>V</sub> and other statistics

The following is the template code used for calculation of statistics for deviation from nominal central length of the 0,5 mm gauge block.

```
import math
from pathlib import Path

import numpy as np
import pandas as pd

def round_half_away_from_zero(value: float, decimals: int = 0) -> float:
    """Round to the given decimal precision, ties away from zero (GUM/BIPM convention for
    reported values)."""
    value = float(value)
    factor = 10 ** decimals
    scaled = round(value * factor, 9)
    sign = 1.0 if scaled >= 0 else -1.0
    return sign * math.floor(abs(scaled) + 0.5) / factor

def round_up(value: float, decimals: int = 0) -> float:
    """Round up (ceiling) at the given decimal precision; use for non-negative values such as
    uncertainties."""
    value = float(value)
    factor = 10 ** decimals
    return math.ceil(round(value * factor, 9)) / factor

def round_sig_away_from_zero(value: float, sig_figs: int) -> float:
    """Round to the given number of significant figures, ties away from zero."""
    value = float(value)
    if value == 0:
```

```

    return 0.0
    decimals = sig_figs - math.floor(math.log10(abs(value))) - 1
    return round_half_away_from_zero(value, decimals)

n = 12

# Participant order (index 0-11):
# 0 METROSERT, 1 BEV, 2 BIM, 3 FTMC, 4 MIRS/UM-FS/LTM, 5 NMI, 6 DPM, 7 LATMB, 8
NPL, 9 BFKH, 10 NSAI NML, 11 BMM

x_i = np.array([
    37, 21, 33, 45, 30, 92, -13, 10, 53, 22, 18, 40
])

u_i = np.array([
    23.5, 25, 50, 22, 18, 30, 27, 26, 16, 20, 10, 26
])

# Shared uncertainty components from same source, example:
u_common = np.zeros(n)
group = np.full(n, "", dtype=object) # correlation group label: participants in the same group
share correlated components; different groups are independent

length = 0.5 # nominal length in mm
label = "central length"
#label = "var"

# Group A - GUM (mutually correlated within Group A only)
u_common[0] = np.sqrt(10.5**2+(0.1*length)**2); group[0] = 'A' # METROSERT
u_common[3] = 14.5; group[3] = 'A' # FTMC
u_common[10] = 5.25; group[10] = 'A' # NSAI NML

```

```
# Group B - VTT MIKES (mutually correlated within Group B only)
```

```
u_common[2] = 0.15*length;          group[2] = 'B' # BIM
```

```
u_common[7] = 11.5;                group[7] = 'B' # LATMB
```

```
exclude_indices: list[int] = [6] # participants to exclude, e.g. 5, 6
```

```
include_mask = np.ones(n, dtype=bool)
```

```
include_mask[exclude_indices] = False
```

```
# Filter data arrays
```

```
x_i_inc = x_i[include_mask]
```

```
u_i_inc = u_i[include_mask]
```

```
u_common_inc = u_common[include_mask]
```

```
group_inc = group[include_mask]
```

```
n_inc = len(x_i_inc)
```

```
# Build covariance matrix V for included participants only
```

```
V_inc = np.zeros((n_inc, n_inc))
```

```
for i in range(n_inc):
```

```
    for j in range(n_inc):
```

```
        if i == j:
```

```
            V_inc[i, j] = u_i_inc[i]**2
```

```
        elif (u_common_inc[i] > 0 and u_common_inc[j] > 0
```

```
            and group_inc[i] != " and group_inc[i] == group_inc[j]):
```

```
            V_inc[i, j] = u_common_inc[i] * u_common_inc[j]
```

```
try:
```

```
    V_inv_inc = np.linalg.inv(V_inc)
```

```
    one_vec_inc = np.ones(n_inc)
```

```
except np.linalg.LinAlgError:
```

```

    raise ValueError("Covariance matrix is singular or ill-conditioned. Check your inputs.")

one_vec = np.ones(n)

# Compute SCRv and its uncertainty
x_scrv_inc = (one_vec_inc @ V_inv_inc @ x_i_inc) / (one_vec_inc @ V_inv_inc @
one_vec_inc)
u_scrv_inc = np.sqrt(1 / (one_vec_inc @ V_inv_inc @ one_vec_inc))

# Residuals and Birge ratio
residuals_inc = x_i_inc - x_scrv_inc
chi_sq_inc = residuals_inc @ V_inv_inc @ residuals_inc
birge_ratio_inc = np.sqrt(chi_sq_inc / (n_inc - 1))

# Calculate weights for SCRv
w = (V_inv_inc @ one_vec_inc) / (one_vec_inc @ V_inv_inc @ one_vec_inc)

# Covariance of each participant with SCRv
cov_xi_xscrsv = np.array([np.sum(w * V_inc[i, :]) for i in range(n_inc)])

# Uncertainty of difference including covariance term
u_diff = np.sqrt(u_i_inc**2 + u_scrv_inc**2 - 2 * cov_xi_xscrsv)

# Degrees of Equivalence
DoE = x_i_inc - x_scrv_inc

# Prepare output DataFrame
df = pd.DataFrame({
    'x_i': x_i,
    'u_i': u_i,
    'u_common': [round_up(v, 1) for v in u_common],

```

```

    'group': group,
})

# Create full DoE and En arrays
DoE_full = np.full(n, np.nan)
u_DoE_full = np.full(n, np.nan)

# Fill in for INCLUDED participants (with covariance subtraction)
DoE_full[include_mask] = x_i_inc - x_scrv_inc
u_DoE_full[include_mask] = np.sqrt(u_i_inc**2 + u_scrv_inc**2 - 2 * cov_xi_xscr)

# Fill in for EXCLUDED participants (no covariance subtraction - they do not contribute to the
SCRV)
excluded_indices = np.where(~include_mask)[0]
for idx in excluded_indices:
    DoE_full[idx] = x_i[idx] - x_scrv_inc
    u_DoE_full[idx] = np.sqrt(u_i[idx]**2 + u_scrv_inc**2)

# En calculated for all participants (k=2), included and excluded alike
En_full = np.abs(DoE_full / (2 * u_DoE_full))

# Weight is 0 for excluded participants (they do not contribute to the SCR)
weights_full = np.zeros(n)
weights_full[include_mask] = w

# Displayed values, rounded per GUM/BIPM reporting conventions (residual omitted - identical
to DoE)
df['En'] = [round_sig_away_from_zero(v, 2) for v in En_full]
df['weight'] = [round_sig_away_from_zero(v, 2) for v in weights_full]
df['DoE'] = [round_half_away_from_zero(v, 0) for v in DoE_full]
df['u_DoE'] = [round_up(v, 0) for v in u_DoE_full]

```

```

df['u_DoE_method'] = np.where(include_mask, 'with covariance', 'no covariance')

df.index.name = 'Participant'

# Full-precision DoE and u(DoE), kept separately for downstream statistics (e.g. the chi-squared
assessment)
raw_df = pd.DataFrame({
    'DoE': DoE_full,
    'u_DoE': u_DoE_full,
})
raw_df.index.name = 'Participant'

print("\n" * 4)
print(f"SCRV = {x_scrv_inc:.5f} ± {u_scrv_inc:.5f}")
print(f"Birge ratio = {birge_ratio_inc:.4f}")
print("\nParticipant Results:")
print(df)

# Create DataFrame for Covariance Matrix
V_inc_df = pd.DataFrame(V_inc)
V_inc_df.index.name = "Row"
V_inc_df.columns = [f"Col_{i}" for i in range(V_inc.shape[1])]

# Create DataFrame for Inverse Covariance Matrix
V_inv_df = pd.DataFrame(V_inv_inc)
V_inv_df.index.name = "Row"
V_inv_df.columns = [f"Col_{i}" for i in range(V_inv_inc.shape[1])]

# Create DataFrame for Weights
weights_df = pd.DataFrame({
    "Weight": w

```

```

})

weights_df.index.name = "Included Participant Index"

# Create summary DataFrame for excluded participants
if len(exclude_indices) > 0:
    excluded_df = pd.DataFrame({
        "x_i": x_i[exclude_indices],
        "u_i": u_i[exclude_indices],
        "u_common": u_common[exclude_indices],
        "group": group[exclude_indices]
    }, index=exclude_indices)
    excluded_df.index.name = "Excluded Participant"
else:
    excluded_df = pd.DataFrame(columns=["x_i", "u_i", "u_common", "group"])

# Round SCRV, its uncertainty and the Birge ratio for reporting
scrv_rounded = round_half_away_from_zero(float(x_scrv_inc), 1)
u_scrv_rounded = round_up(float(u_scrv_inc), 1)
birge_ratio_rounded = round_up(float(birge_ratio_inc), 3)

summary_df = pd.DataFrame({
    "Nominal length (mm)": [length],
    "Label": [label],
    "SCRV": [scrv_rounded],
    "u(SCRV)": [u_scrv_rounded],
    "Birge ratio": [birge_ratio_rounded],
})

# Export all to Excel, into a shared "stats" subfolder
output_dir = Path(__file__).resolve().parent / "stats"

```

```

output_dir.mkdir(exist_ok=True)
output_filename = output_dir / f'comparison_results_{label}_{length:.1f}_mm.xlsx"
with pd.ExcelWriter(output_filename, engine='openpyxl') as writer:
    df.to_excel(writer, sheet_name="Participant Results", index=True) # type:
    ignore[reportUnknownMemberType]
    raw_df.round(5).to_excel(writer, sheet_name="Raw", index=True) # type:
    ignore[reportUnknownMemberType]
    V_inc_df.round(5).to_excel(writer, sheet_name="Covariance Matrix", index=True) # type:
    ignore[reportUnknownMemberType]
    V_inv_df.round(5).to_excel(writer, sheet_name="Inverse Covariance", index=True) # type:
    ignore[reportUnknownMemberType]
    summary_df.to_excel(writer, sheet_name="Summary", index=False) # type:
    ignore[reportUnknownMemberType]
    ws_summary = writer.sheets["Summary"]
    ws_summary["C2"].number_format = "0.0"
    ws_summary["D2"].number_format = "0.0"
    ws_summary["E2"].number_format = "0.000"

print(f"\nAll results exported to '{output_filename}' successfully.")

```

The following is the template code used to calculate the chi-squared statistics.

```
import math
from pathlib import Path

import numpy as np
import pandas as pd
from scipy.stats import chi2 # pyright: ignore[reportMissingTypeStubs]

# Reusable script: multi-measurand consistency assessment per CCL/WG-MRA GD-3 v1.7
# (Jul 2024), section 5.2 "Assessment of KC results with several measurands"
# by I. Leroux, based on Beissner, Metrologia 39, p.59 (2002).
#
# For each participant, with M nominal lengths (measurands), i in {1..M}:
#  $En_i = d_i / (k_i * u(d_i))$ ,  $k_i = 2$  for ~95% confidence
# - if  $|En_i| \leq 1$  for all i: accept the uncertainty claim immediately.
# - otherwise, compute  $Q = \sum((k_i * En_i)^2) = \sum(d_i^2 / u(d_i)^2)$ 
# and compare it to  $\chi^2(M, 0.95)$ , the 95th percentile of the chi-square
# distribution with M degrees of freedom.
# -  $Q \leq \chi^2(M, 0.95)$ : accept the uncertainty claim.
# -  $Q > \chi^2(M, 0.95)$ : Q exceeds the threshold and  $|En_i| > 1$  for at least
# one measurement - investigate the participant's results.
#
# Q is a simple sum of squared normalized errors.
# It assumes the M measurands are independent.
# GD-3 appendix (section 7) acknowledges this is an approximation when
# the errors are actually correlated across measurands, and deliberately
# chooses this simple form over a full covariance treatment.
#
# Run all 12 ".py" scripts first so their stats/*.xlsx
# "Raw" sheets (full-precision DoE, u_DoE) are up to date before running this one.
```

```

def round_half_away_from_zero(value: float, decimals: int = 0) -> float:
    """Round to the given decimal precision, ties away from zero (GUM/BIPM convention for
    reported values)."""
    value = float(value)
    factor = 10 ** decimals
    scaled = round(value * factor, 9)
    sign = 1.0 if scaled >= 0 else -1.0
    return sign * math.floor(abs(scaled) + 0.5) / factor

PARTICIPANTS = [
    "METROCERT", "BEV", "BIM", "FTMC", "MIRS/UM-FS/LTM", "NMI",
    "DPM", "LATMB", "NPL", "BFKH", "NSAI NML", "BMM",
]
n = len(PARTICIPANTS)
lengths = [0.5, 4, 10, 25, 50, 100]
m = len(lengths)
k = 2 # coverage factor, ~95% confidence, consistent with En

folder = Path(__file__).resolve().parent / "stats"
labels = ["central length", "var"]
sheet_names = {"central length": "Central length deviation", "var": "Variation"}

chi2_95 = chi2.ppf(0.95, df=m)
chi2_95_rounded = round_half_away_from_zero(float(chi2_95), 2)

tables: dict[str, pd.DataFrame] = {}
for label in labels:
    doe = np.full((n, m), np.nan)
    u_doe = np.full((n, m), np.nan)
    for j, length in enumerate(lengths):

```

```

source_file = folder / f"comparison_results_{label}_{length:.1f}_mm.xlsx"
raw = pd.read_excel(source_file, sheet_name="Raw", index_col=0)
doe[:, j] = raw["DoE"].to_numpy()
u_doe[:, j] = raw["u_DoE"].to_numpy()

en = doe / (k * u_doe)
en_max = np.abs(en).max(axis=1)
Q = np.sum((k * en) ** 2, axis=1)

verdict = []
for p in range(n):
    if en_max[p] <= 1:
        verdict.append("Accepted (all |En| <= 1)")
    elif Q[p] <= chi2_95:
        verdict.append("Accepted (Q within limit)")
    else:
        verdict.append("Flagged - investigate (Q exceeds limit)")

tables[label] = pd.DataFrame({
    "Participant": np.arange(n),
    "En,max": [round_half_away_from_zero(v, 2) for v in en_max],
    "Q": [round_half_away_from_zero(v, 2) for v in Q],
    "chi2_0.95,M": [chi2_95_rounded] * n,
    "Participant name": PARTICIPANTS,
    "Verdict": verdict,
})

output_filename = folder / "Chi-squared total.xlsx"
with pd.ExcelWriter(output_filename, engine='openpyxl') as writer:
    for label in labels:

```

```
        tables[label].to_excel(writer, sheet_name=sheet_names[label], index=False) # type:
ignore[reportUnknownMemberType]
```

```
for label in labels:
```

```
    print(f"\n{sheet_names[label]}:")
```

```
    print(tables[label].to_string(index=False))
```

```
print(f"\nMulti-measurand consistency assessment exported to '{output_filename}')
```