CCL Key Comparison

The Calibration of Internal and External Diameter Standards
CCL-K4

FINAL REPORT

Contents

January 2007

1. Introduction

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

At its meeting in September 1997, the Consultative Committee for the Definition of the Metre, CCDM, (today called the Consultative Committee for Length, CCL) identified key comparisons in the field of dimensional metrology and decided upon the general content and the proposed pilot laboratory and time-frames of each key comparison. In particular, it was decided that a key comparison on diameter standards shall be performed with the National Institute of Standards and Technology (NIST), as the pilot laboratory. Subsequent meetings of the CCL broadened the scope of this first key comparison to include internal diameter standards and external diameter standards. The CCL recommended the other measurement operations that may fall inside this general field, such as roundness, spherical diameter, spherical and cylindrical form, be included in future activity.

The results of this international comparison will contribute and be included in the agreement for establishing metrological equivalence. The interregional CCL key comparison will be combined with regional comparisons following similar protocol. Laboratories participating in both the interregional and the regional comparisons establish the link between the comparisons and assure their equivalence. The measurement results outlined in this report followed the guidelines established by the BIPM[1] and provide clear and unequivocal comparison of the participating laboratories' measurement performance in the area of internal and external diameter standards. This report is patterned after similar CCDM and CCL comparisons carried out in other fields from 1993 to 2002[2,3].

2. Organization

The preliminary list of participants was drafted by the pilot laboratory and was approved at the CCL meeting of 18 July 1998. The general requirement for the participating laboratories was the ability to measure, by any primary means, provided it was a measurement service to clients, the diameter of external diameter standards within the range 2 mm to 100 mm and the diameter of internal diameter standards within the range 5 mm to 100 mm. The uncertainty requirements for the diameter measurements was set at approximately 200 nm at $k = 1$.

Contact Person	**National Metrology Institute Address**	**Tel: / Fax: Email:**
L. Abramova	VNIIM 19, perspective de Moskovsky 198005 St. Petersburg, Russia	Tel. +7-812-251-8638 Fax +7-812-113-0114 e-mail: l.ju.abramova@vniim.ru
Miguel Viliesid	CENAM A.P. 1-100, Centro Queretaro, Qro. 76900, Mexico	Tel. +(52-42) 11-05-74 Fax +(52-42) 11-05-77 e-mail: mviliesi@cenam.mx
Chu-Shik Kang	KRISS 1 Doryong-dong Yusong-gu, Daejeon 305-340 Korea	Tel. +82-42-868-5103 Fax +82-42-868-5608 e-mail: cskang@kriss.re.kr
Zhang Heng	NIM No. 18 Bei San Huan Dong Lu Beijing 100013, China	Tel. +86-1-64218627 Fax +86-1-64218703 e-mail: zhangh@nim.ac.cn
John Miles	National Measurement Institute (NMI) 71 Normanby Road Clayton, Melbourne, Victoria, Australia	Tel. +613-9542-4010 Fax +613-9542-4001 e-mail: john.miles@measurement.gov.au
Oelof Kruger	CSIR Meiring Naude Road Brummeria, Pretoria 0001, Republic of South Africa	Tel. + 27-12-841-3005 Fax + 27-12-841-4458 e-mail: oakruger@csir.co.za
Otto Jusko	PTB Bundesallee 100 D-38114 Braunschweig, Germany	Tel. +49-531-592-5310 Fax +49-531-592-9292 e-mail: otto.jusko@ptb.de
Rudolf Thalmann	METAS Lindenweg 50 CH-3003 Bern – Wabern, Switzerland	Tel. +41-31-323-3385 Fax +41-31-323-3210 e-mail: rudolf.thalmann@metas.ch
David Flack	NPL Centre for Length Metrology NPL, Hampton Road, Teddington, Middlesex TW11 0LW, UK	Tel. +44-208-943-6347 Fax +44-208-943-2945 e-mail: david.flack@npl.co.uk
Attilio Sacconi	IMGC Strada delle Cacce, 73, 10135- Torino, Italy	Tel. +39-11-3977-466 Fax +39-11-3977-7459 e-mail: A.Sacconi@imgc.to.cnr.it
Coordinator:		
John Stoup	NIST Building 220, Rm B113 Gaithersburg, MD 20899 USA	Tel. +1-301-975-3476 Fax +1-301-975-8291 e-mail: john.stoup@nist.gov

Table 1. Participating laboratories.

The schedule for the comparison was carried out in a mixed form, star type circulation. The artifacts were circulated within a region and then returned to the pilot laboratory for re-measurement before circulation in the next region. Each laboratory had approximately one month to complete the measurements and provide transportation to the next laboratory on the schedule. Although a great attempt was made to keep the intercomparison moving quickly, there were some unforeseen difficulties relating to equipment failures, customs, and transportation. These problems did delay the circulation somewhat, but the intercomparison was completed within a reasonable time period.

Region	Laboratory	Country	Measurement Date
EUROMET	METAS	Switzerland	Nov. 2000
	NPL	U.K.	Jan. 2001
	PTB	Germany	Feb. 2001
	IMGC	Italy	Mar. 2001
Pilot Lab	NIST	USA	May 2001
	CENAM	Mexico	Aug. 2001
Pilot Lab	NIST	USA	Sept. 2001
APMP	NIM	China	Nov. 2002
	CSIRO	Australia	Oct. 2001
	KRISS	Korea	Dec. 2001
Pilot Lab	NIST	USA	Feb. 2002
SADCMET	CSIR	South Africa	Apr. 2002
COOMET	VNIIM	Russia	Jul. 2002
Pilot Lab	NIST	USA	Dec. 2002

Table 2. The actual time schedule of the comparison.

3. Description of the Artifacts

The artifacts used for this intercomparison were a combination of internal and external diameter standards spanning a fairly large measurement range and manufacturing style. The package contained 4 ring gages made of steel and 5 cylinders made of steel. The thermal expansion coefficient of the artifacts was supplied by the manufacturer and was not independently verified. It was reported to be $11.5 \pm 0.5\ 10^{-6}\ K^{-1}$. The artifacts are identified in Table 3. Roundness traces for each of the artifacts are shown in Appendix C.

Ring Gages

Identification	Nominal diameter (mm)	Expansion coeff. (10^{-6} K^{-1})	Manufacturer
R1	5.17	11.5 ± 0.5 (*k*=1)	Glastonbury Gage
R2	11.95	11.5 ± 0.5 (*k*=1)	Glastonbury Gage
R3	49.3	11.5 ± 0.5 (*k*=1)	Glastonbury Gage
NIST-7	100	11.5 ± 0.5 (*k*=1)	Glastonbury Gage

Cylinders

Identification	Nominal diameter (mm)	Expansion coeff. (10^{-6} K^{-1})	Manufacturer
D1	2.0	11.5 ± 0.5 (*k*=1)	Glastonbury Gage
D2	3.465	11.5 ± 0.5 (*k*=1)	Glastonbury Gage
D3	24.0	11.5 ± 0.5 (*k*=1)	Glastonbury Gage
42198	50.0	11.5 ± 0.5 (*k*=1)	SIP
D5	98.5	11.5 ± 0.5 (*k*=1)	Glastonbury Gage

Table 3. Standards used in the intercomparison.

4. Measurement instructions and data reporting

Before measurement, the artifacts were inspected for damage of the measurement surfaces, particularly at the gaging points. Any damage was recorded using appropriate diagrams. Although some damage near the gauging points was reported on some of the artifacts, none of the damage appeared to affect the quality of the measurement results.

The measurement quantity of interest was the diametrical distance between the nominal gauge points, defined as mid-elevation along the gauge cylinder and in the diameter direction specified by the engraved marks on the gauge. Although the directional markings were supplied by the manufacturer, in some cases the etchings did not align with the true diametrical axis. In these instances, the measurements were performed along the true diametrical diameter parallel to the plane indicated by the etchings. These offsets were quite small and since the geometry of the artifacts was good, it appears as though any discrepancy of the actual gaging position did not affect the quality of the measurements.

The measurement results were appropriately corrected to the reference temperature of 20° C using the thermal expansion coefficient given in the protocols. Additional corrections were applied according to the equipment and procedures used by each laboratory. Any artifacts found to have a magnetic condition had that condition removed per individual laboratory practices before the diameter measurements were performed.

One important feature of this comparison was that a laboratory was allowed to submit measurements from more than one measurement system as long as the timetable was adhered to and that each measurement system was available to general clients for measurement services. Unlike some measurement areas, many laboratories use multiple diameter-measuring systems to span the full range of sizes required by clients. Many of these systems have very different designs and measurement styles. In some cases, a laboratory may offer several levels of service depending on measurement cost and uncertainty requirements. It was initially decided to allow laboratories with these multiple systems to submit multiple sets of results in the interest of sampling the wide variety of techniques and equipment used for diameter measurement. Full measurement uncertainty analyses were required for each submitted measurement system.

5. Measurement Methods and Instruments

The artifact measurements were performed by the participants using a variety of instruments and techniques. Probe diameter values and applied force conditions are of particular interest in determining how these variables affect the measurement results. In some cases, a participant used several pieces of equipment to span the range of artifact sizes in the intercomparisons. Below are brief descriptions of the equipment and the measurement variables used by the participants.

CENAM: SIP-305m single axis horizontal comparator. Sphere feelers used to measure external and internal diameters compared with gage blocks calibrated by interferometry. For small external diameters, flat feelers were used. Resolution of 0.1 µm. Measurements performed at 0.5 N of applied force then corrected to undeformed conditions.

METAS: Length-based measuring machine designed by SIP and METAS. Internal and external measurements performed through displacement measurements using a plane mirror interferometer. Spherical probe calibrated using gauge blocks. Applied measurement force extrapolated to zero.

NIM: Mahr 828 CIM - single axis comparator. Spherical contacts compared with a reference ring, traceable to the NIM internal diameter instrument, for internal measurements and a 10 mm gage block for external measurements. Measurement force is 0.2 N to 2.0 N corrected to undeformed conditions for internal measurements and 1.0 N to 2.0 N corrected to undeformed conditions for external measurements.

NIM: Internal diameter instrument with a laser interferometer designed by NIM (denoted as "NIM Interf." in the graphs and tables). Absolute measurement using a fused quartz, 10 mm box measured by interferometry as the master artifact. 5 mm diameter probe at 0.2 N of applied force, corrected to undeformed conditions.

VNIIM: 1D laser interference comparator. Non-contact measurement using symmetrical images from a perflectometer with a scanning slit. Distances between these images measured using an interferometer.

KRISS: Federal Model 136B-3 internal and external comparator. Spherical contacts from 0.8 mm to 3.2 mm in diameter, depending on artifact size. Comparison to gage block stacks, measured by interferometry, by gauge substitution method.

IMGC: Modified M3 Moore Measuring Machine, equipped with a laser interferometer and LVDT probes. For internal and external measurements, the probe diameters range from 1.0 mm to 4.0 mm and are calibrated using a 10 mm gage block as the master. The probe force is 3 mN.

CSIRO: An NPL/Matrix internal diameter measuring machine. Instrument uses reference box standards, two-point spherical contacts with tip radii ranging from 2.4 mm to 4.8 mm, and a Hilger and Watts autocollimator reading to 0.1 arcseconds as the differential sensor. An applied force of 0.2 N was used and corrected to undeformed conditions. This instrument used for all but the 100 mm ring.

CSIRO: MU214B universal measuring machine. Instrument uses reference gage blocks aligned to the motion axis to determine feeler diameters. Comparison done through gauge substitution method. Probe diameter is 6 mm and the applied force is 0.02 N, corrected to undeformed conditions.

CSIRO: Tesa Modul comparator. For external diameter measurements, unit consists of mutually opposed inductive probe heads, pneumatically retracted, mounted in a Tesa comparator stand. Gage blocks were used as reference standards. Measurements were done using gauge substitution. The probe tip diameters were 40 mm and the applied force was 1 N and 0.63 N for each of the two probes, corrected to undeformed conditions.

CSIR: Federal internal and external comparator. For internal diameter measurements, spherical contacts are used. For external diameter measurements, a SIP 305 with an external laser and a wedge-type comparator were used. All measurements performed through gauge substitution using gauge blocks as the transfer standards.

NIST: Moore M48 coordinate measuring machine (denoted as “NIST CMM” in the graphs and tables.). Internal and external diameters measured on laser-based CMM using gage blocks or precision spheres to characterize the 3 mm diameter probe. Measurement force is 0.1 N.

NIST: Federal comparator and 1D micrometer designed by NIST. External diameters measured by displacement interferometry using a laser-based, air-bearing micrometer. Contacts are a 3 mm flat, 10 mm cylinder combination. Measurement force extrapolated to zero. Federal comparator, using gage block stacks measured by interferometry, also used for internal measurements. Variable probe diameters using an applied force of 2 N corrected to undeformed through equations. The comparator is denoted as “NIST Comp” in the graphs and tables, and the laser micrometer is denoted as “NIST micro”. Because of limited space, in some tables “NIST Comp.” is used to denote both instruments.

NPL: Modified Zeiss Metroscope (denoted as “NPL Metro.” in the graphs and tables). External diameters measured between two parallel 2 mm flat measuring faces operating under a contact force of 2.5 N, whose separation is measured by laser interferometry. Results corrected by formulas to undeformed conditions. Also, a Meseltron Movotelit comparator using gauge blocks as the reference was used for some external diameter measurements (denoted as “NPL Mesel.” in the graphs and tables). Instrument uses capacitive sensor with flat and parallel faces under 0.8 N of force.

NPL: Internal diameter measuring machine designed by NPL (denoted as “NPL IDM” in the graphs and tables) . Displacement measured by interferometer. Probe diameter calibrated by using a transparent, fused silica box with a 10 mm length. Measurement force is 0.06 N.

PTB: PTB-developed, Abbe-type length comparator MFU8. Used for both internal and external diameters and has laser scales and an inductive probe. Probes are ruby spheres of various diameters under a 2 mN/µm contact force extrapolated to zero. A gauge block is the reference artifact. Measurements also performed using a Lako laser comparator (denoted as “PTB Lako” in the tables) but this data was later withdrawn.

PTB: Comparator of diameter and form designed by PTB (denoted as “PTB KOMF” in the graphs and tables). Also used for both internal and external measurements. Laser interferometer Abbe-type length comparator incorporating two probing systems with inductive transducers using 5 mm ruby spheres under a contacting force of 2 mN/µm. Probe diameter determined using 3-body method.

6. Stability of the Artifacts

The pilot laboratory measured the artifacts five times: at the beginning of the comparison, after each regional loop, and at the end of the artifact circulation. Figures 1 and 2 show the results. The artifacts were measured using both the M48 CMM and the 1D comparator or laser micrometer at each re-measurement interval. No relevant damage was observed on the artifacts during the circulation. The artifacts did show only light wear, consistent with this level of intercomparison.

The dimensional stability was more difficult to determine based on the pilot measurements. The data was generally good but did show some potential artifact drift for some of the large ring and plug gauges. The pilot laboratory remeasured the gauges in late 2004 to determine if any statistically significant drift was occurring with the artifacts. From these measurements, it was determined that the 98.5 mm plug was the only artifact that did have a statistically significant drift at the k=2 level of +19 nm per year. The reported data for this plug gauge will be adjusted for the artifact drift in the measurement results section of this report.

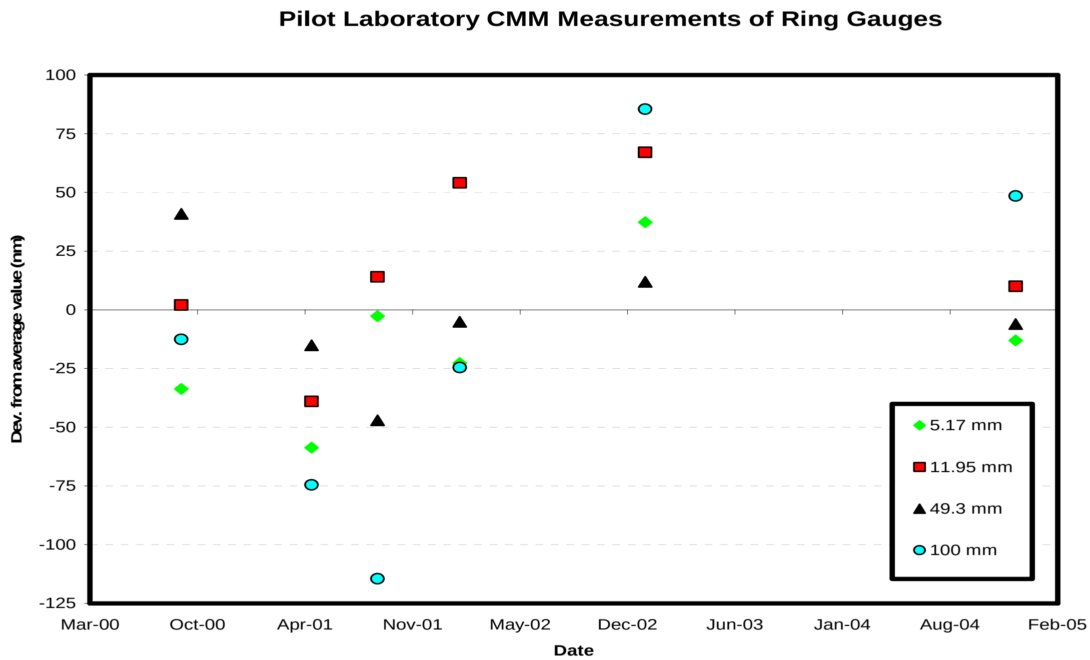


Figure 1. Pilot laboratory measurements of internal diameter standards.

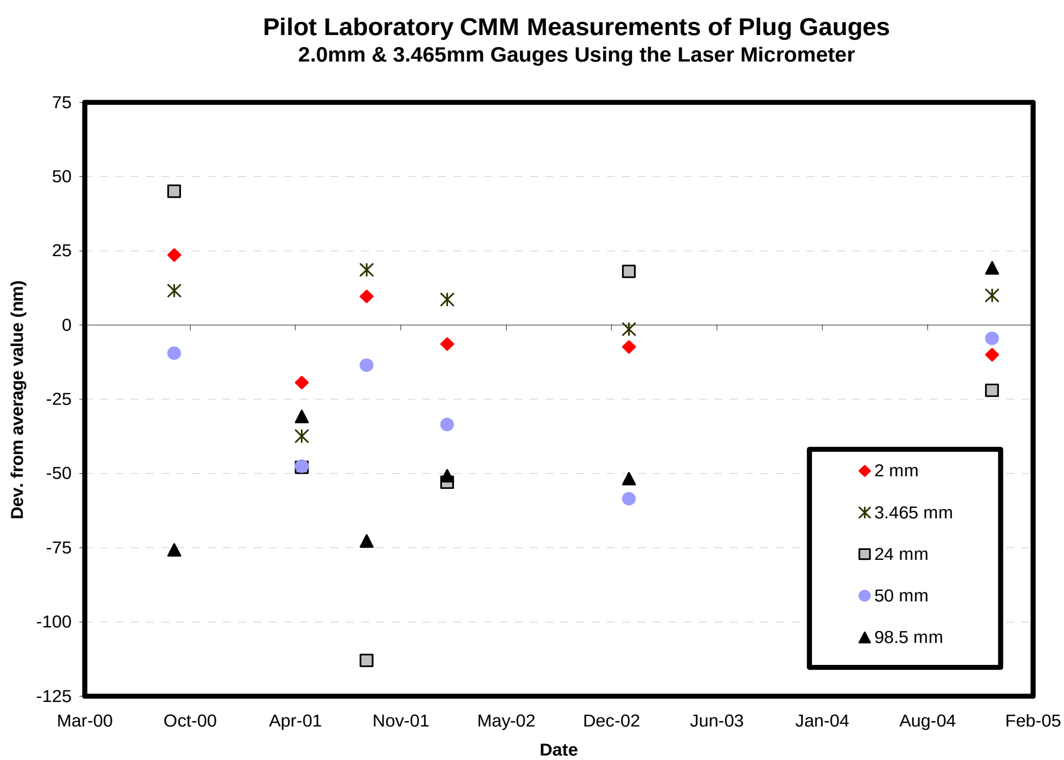


Figure 2. Pilot laboratory measurements of external diameter standards.

7. Measurement Results and Uncertainty Components

Table 4 gives the reported results for each ring and plug gage at the defined measurement location. Table 5 gives the reported results with the 98.5 mm plug data adjusted using an artifact instability term of -19 nm per year. Table 6 gives the standard uncertainty values for each of these measurements. The pilot laboratory measurements are reported once, at the time of their regional circulation.

DATE	Dec 2000	Jan 2001	Jan 2001	Jan 2001	Feb 2001	Feb 2001	Feb 2001	Mar 2001	May 2001	May 2001	Aug 2001	Oct 2001	Dec 2001	Apr 2002	Jul 2002	Nov 2002	Nov 2002
SIZE (mm)	**METAS**	**NPL IDM**	**NPL Metro.**	**NPL Mesel.**	**PTB Lako**	**PTB MFU8**	**PTB KOMF**	**IMGC**	**NIST Comp./micro.**	**NIST CMM**	**CENAM**	**CSIRO**	**KRISS**	**CSIR (SA)**	**VNIIM**	**NIM Mahr**	**NIM Interf.**
RING																	
5.17	-70	-140			-111	-114		-21	-50	-125	-100	60	660	-400	25	-50	
11.95	-230	-250			-356	-181	-269	-280	-196	-219	-210	-70	-130	-150	-270	20	
49.3	60	0			185	78	80	156	68	51	20	30	30	-100	-350	140	210
100	10	-120			-80		-97	-106	42	-67	30	230	170	-50	-500	230	270
PLUG																	
2	90		120	40	120	38	85	-34	67		100	110	60	-100	-330	-30	
3.465	560		630	560	588	541	548	543	530		470	650	420	200	65	200	
24	1400		1370		1377	1328	1376	1314	1406	1355	1180	1400	1160	1300	740	1120	
50	1130		1090		1141		1159	1090	1134	1077	1110	940	1230	980	815	1010	
98.5	2300		2500		2335	2321	2262	2215	2296	2151	1790	2210	2130	2100		2050	

Table 4. Measurement Results: deviation from nominal value, in nanometers. See Section 5 for an explanation of the abbreviations denoting various measuring instruments.

DATE	Dec 2000	Jan 2001	Jan 2001	Jan 2001	Feb 2001	Feb 2001	Feb 2001	Mar 2001	May 2001	May 2001	Aug 2001	Oct 2001	Dec 2001	Apr 2002	Jul 2002	Nov 2002	Nov 2002
SIZE (mm)	**METAS**	**NPL IDM**	**NPL Metro.**	**NPL Mesel.**	**PTB Lako**	**PTB MFU8**	**PTB KOMF**	**IMGC**	**NIST Comp./micro.**	**NIST CMM**	**CENAM**	**CSIRO**	**KRISS**	**CSIR (SA)**	**VNIIM**	**NIM Mahr**	**NIM Interf.**
RING																	
5.17	-70	-140			-111	-114		-21	-50	-125	-100	60	660	-400	25	-50	
11.95	-230	-250			-356	-181	-269	-280	-196	-219	-210	-70	-130	-150	-270	20	
49.3	60	0			185	78	80	156	68	51	20	30	30	-100	-350	140	210
100	10	-120			-80		-97	-106	42	-67	30	230	170	-50	-500	230	270
PLUG																	
2	90		120	40	120	38	85	-34	67		100	110	60	-100	-330	-30	
3.465	560		630	560	588	541	548	543	530		470	650	420	200	65	200	
24	1400		1370		1377	1328	1376	1314	1406	1355	1180	1400	1160	1300	740	1120	
50	1130		1090		1141		1159	1090	1134	1077	1110	940	1230	980	815	1010	
98.5	2300		2498		2332	2318	2259	2210	2288	2143	1777	2194	2111	2075		2014	

Table 5. Adjusted Measurement Results: 98.5 mm plug data corrected using the artifact instability term of -19 nm per year.

DATE	Dec 2000	Jan 2001	Jan 2001	Jan 2001	Feb 2001	Feb 2001	Feb 2001	Mar 2001	May 2001	May 2001	Aug 2001	Oct 2001	Dec 2001	Apr 2002	Jul 2002	Nov 2002	Nov 2002
SIZE (mm)	**METAS**	**NPL IDM**	**NPL Metro.**	**NPL Mesel.**	**PTB Lako**	**PTB MFU8**	**PTB KOMF**	**IMGC**	**NIST comp.**	**NIST CMM**	**CENAM**	**CSIRO**	**KRISS**	**CSIR (SA)**	**VNIIM**	**NIM Mahr**	**NIM Interf.**
RING																	
5.17	44	43			30	50		31	54	50	200	80	100	130	60	100	
11.95	66	43			19	36	41	51	45	50	110	80	100	130	60	100	
49.3	67	48			23	30	18	40	51	50	130	90	160	100	70	130	70
100	68	60			31		37	60	63	50	200	160	260	130	80	160	120
PLUG																	
2	37		41	38	18	29	6	34	19		100	40	100	130	60	70	
3.465	29		40	41	19	32	6	29	19		100	50	100	130	70	70	
24	38		49		20	27	26	35	29	50	120	50	120	100	70	80	
50	33		68		23		48	48	53	50	140	80	170	80	70	100	
98.5	120		115		36	42	20	68	85	50	190	140	270	130	80	130	

Table 6. Reported Standard Uncertainty (*k*=1). Values in nanometers.

The participants used a wide range of techniques to break down and classify sources of uncertainty. The many different instruments and techniques used in the intercomparison make it difficult to compare individual uncertainty components. The following tables attempt to group the uncertainty components into relatively similar classifications of error for comparison purposes. These classifications separate the uncertainty components into influences of: alignment, mastering, environmental, temperature, CTE, repeatability, deformation and contact effects, and artifact geometry effects. These classifications are subject to differing interpretations and consequently there was some ambiguity in assigning values to the entries.

DATE	Dec 2000	Jan 2001	Feb 2001	Feb 2001	Feb 2001	Mar 2001	May 2001	May 2001	Aug 2001	Oct 2001	Dec 2001	Apr 2002	Jul 2002	Nov 2002
Uncertainty Component	**METAS**	**NPL Metro.**	**PTB Lako**	**PTB MFU8**	**PTB KOMF**	**IMGC**	**NIST comp.**	**NIST CMM**	**CENAM**	**CSIRO**	**KRISS**	**CSIR (SA)**	**VNIIM**	**NIM Mahr**
Alignment	12.1	29	1.2	13.4	1.2	6.2	-	-	109	28.9	88.1	-	23.5	62
Mastering Technique	18	27.3	15.9	2.9	2.8	21.3	7.2	30	10.9	13.4	14.2	-	0.7	25
Contact/deformation/gain corrections	-	6	2.8	3.2	2.5	2	6	2	20	21.6	47.2	-	85	-
Environmental effects	1.8	-	2.2	1.2	0.7	-	1.6	1.6	-	-	-	-	4.3	-
Temperature related effects	2.2	17	8	7.5	1.5	2.8	6	1.2	6.5	31.9	54	-	2.6	1.6
Artifact geometry/form	30.4	-	6.9	6.9	6.9	23.7	12	4	-	14.5	-	-	-	-
Repeatability/reproducibility	6	10	3.5	6	25.6	11.3	23.9	41	44	5.8	31.9	-	56	13
CTE	2.5	18	-	-	-	0.2	-	1.2	-	8.1	6	-	-	2.1

Table 7. Standard uncertainty components (in nanometers) for the 24 mm plug gauge.

DATE	Dec 2000	Jan 2001	Feb 2001	Feb 2001	Feb 2001	Mar 2001	May 2001	May 2001	Aug 2001	Oct 2001	Dec 2001	Apr 2002	Jul 2002	Nov 2002	Nov 2002
Uncertainty Component	METAS	NPL IDM	PTB Lako	PTB MFU8	PTB KOMF	IMGC	NIST comp.	NIST CMM	CENAM	CSIRO	KRISS	CSIR (SA)	VNIIM	NIM Mahr	NIM Interf.
Alignment	12.3	32.2	1.2	11.8	0.8	6.2	12.5	-	109	-	96.4	-	71	50	50
Mastering Technique	36	26	15.9	20.2	15.3	21.3	45	30	14.6	38.1	14.8	-	1.5	25	25
Contact/deformation/gain corrections	-	9.4	2.8	3.2	2.5	2	2.5	2	20	74.8	47	-	85	-	-
Environmental effects	3.5	5	4.5	2.4	1.5	-	-	3.1	-	-	-	-	8.8	-	-
Temperature related effects	4.4	3	12.7	14.8	3.1	5.7	12.5	2.5	13	32.7	111	-	5.4	32.9	32.9
Artifact geometry/form	54.3	-	7.5	7.5	8.4	31	5	4	-	14.6	-	-	-	-	-
Repeatability/reproducibility	6	21	5.3	5.1	1.1	11.5	14	42	65	-	34.7	-	56	6	6
CTE	5	7	-	-	-	0.2	-	2.5	-	6.4	5.5	-	-	4.3	4.3

Table 8. Standard uncertainty components (in nanometers) for the 49.3 mm ring gauge.

The 24 mm plug gauge and the 49.3 mm ring gauge were used as examples of internal and external artifacts. A dash in the tables indicates that no value could be interpreted for this uncertainty classification.

7.1 Changes in Submitted Results.

After completion of the draft B1 report of the comparison, PTB requested the following changes in their submitted results: (1) The results from the PTB laser comparator were withdrawn from consideration due to decommissioning of the instrument. These results are not included in the calculation of the reference value. They also will not be included in the BIPM database of NMI capabilities. (2) Following a suggestion from the WGDM, the uncertainties for some of the results of the PTB diameter and form instrument were increased to account for uncertainty associated with poor artifact geometry. The gauges identified below had larger local form deviations due to local roundness and surface finish conditions. The identifying marks on the gauges also allowed for a small amount of freedom in interpretation of the measurement locations, which sampled the local form deviations These facts led to the changes in standard uncertainty shown in Table 9.

Artifact	Original Uncertainty (nm)	New Uncertainty (nm)
11.95 mm ring	16	41
100 mm ring	22	37
24 mm plug	6	26
50 mm plug	8	48

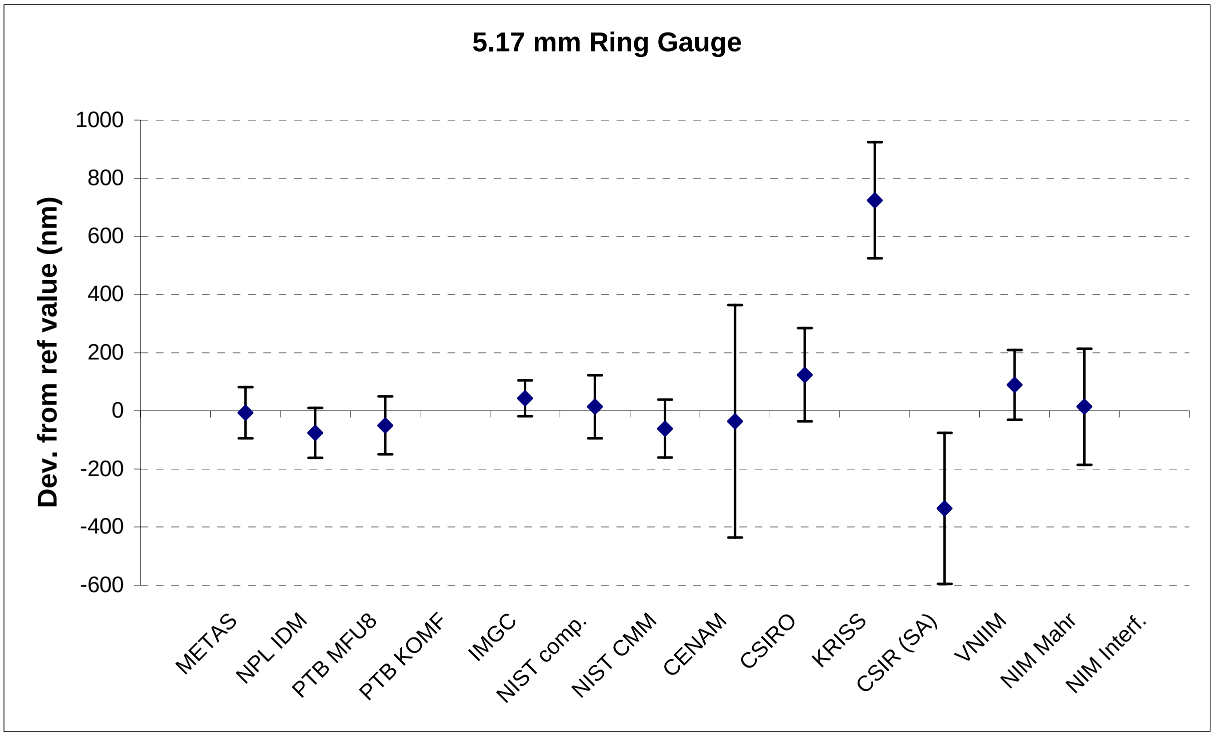
Table 9. Changes in standard uncertainty for PTB diameter and form measurement instrument (KOMF).

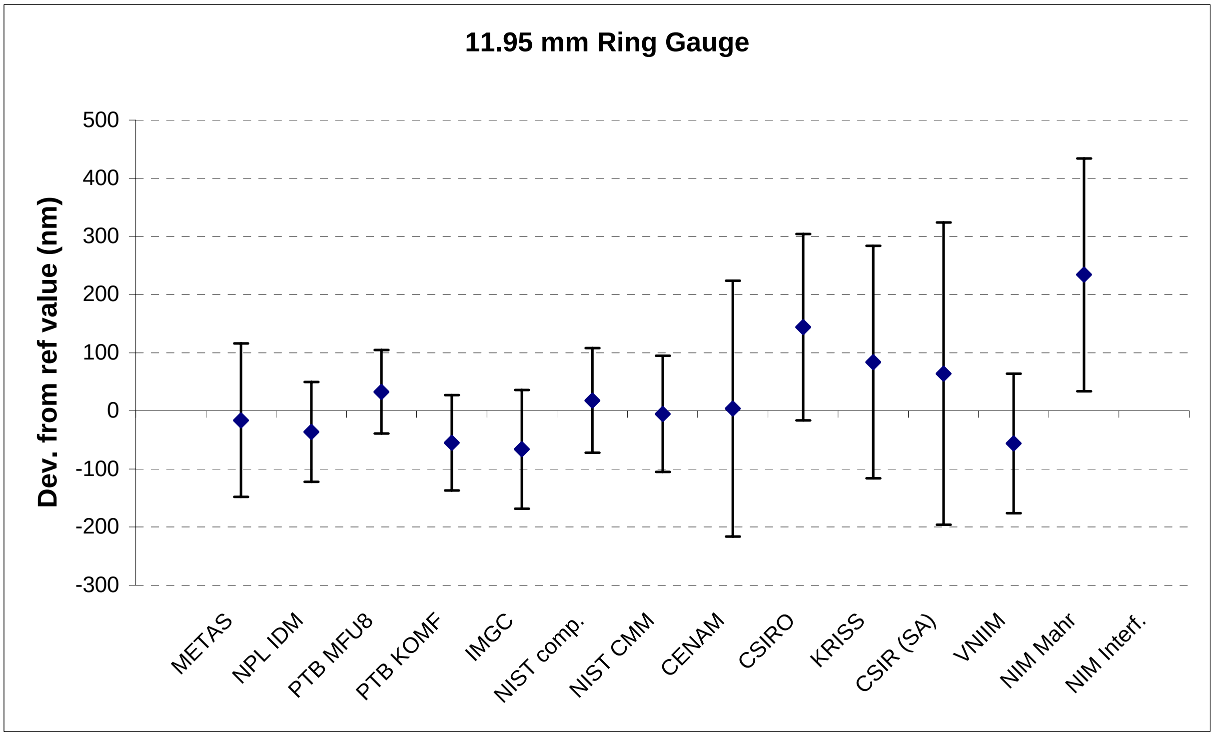
8. Conclusions

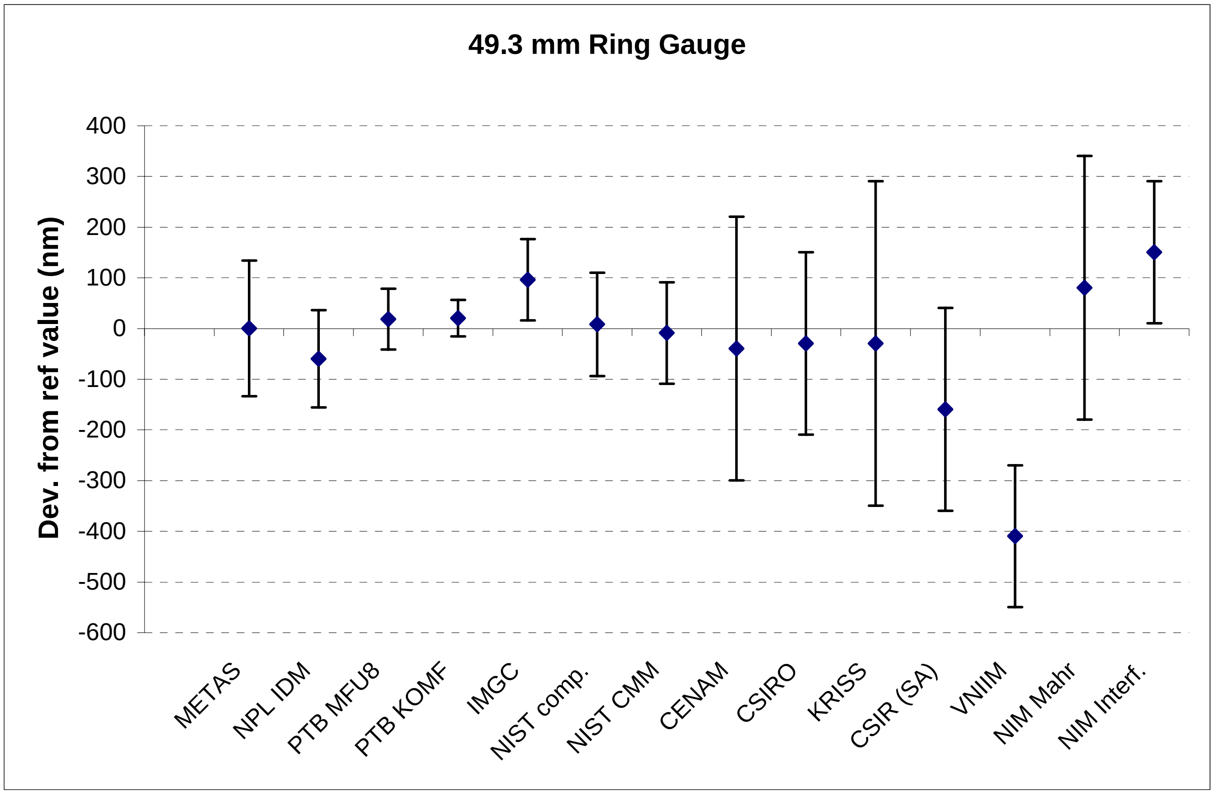
The results of this comparison underscore the limitations that exist in the area of internal and external diameter standards and how these limitations affect the determination of a reference value. The diversity of the measurement equipment used for these measurements range from simple 1D length measuring machines to state-of-the-art and unique systems indicating that the field of diameter measurement has allowed development of very different approaches to achieve the required accuracies. Individual national laboratories have set their own acceptable levels of accuracy, likely dependent on the requirements from their country's particular industries, and have not pushed for a universally accepted approach to diameter measurement. A single universal approach to diameter measurement, similar in style to what has evolved in the measurements of gage blocks, may be difficult for a number of reasons including the large variety of cylindrical artifact designs and the generally poorer surface quality of cylindrical diameter artifacts. The relatively small number of laboratories with reasonably low uncertainties for diameter measurements is an issue that may need to be addressed in future intercomparisons. Future intercomparisons may also want to incorporate the measurement of precision spheres as artifacts since they are not susceptible to the same types of limitations as cylindrical standards.

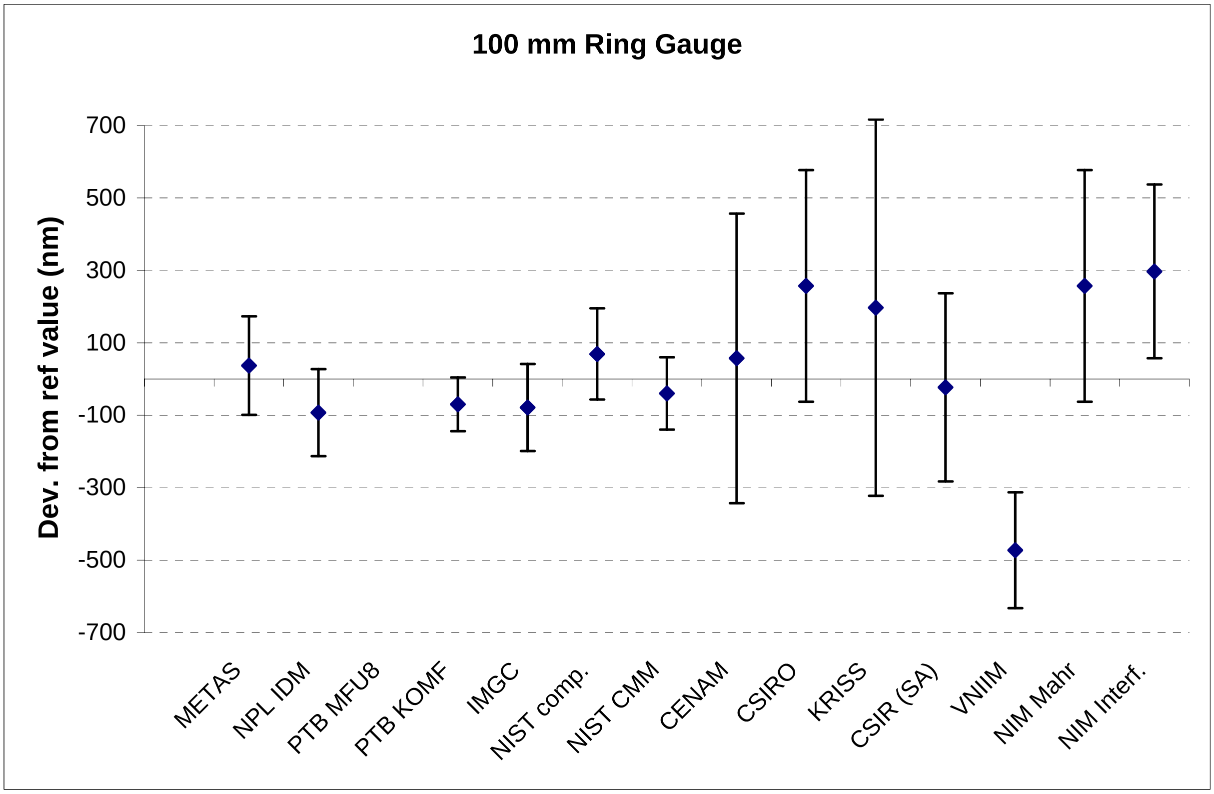
9. Measurement Results – Charts

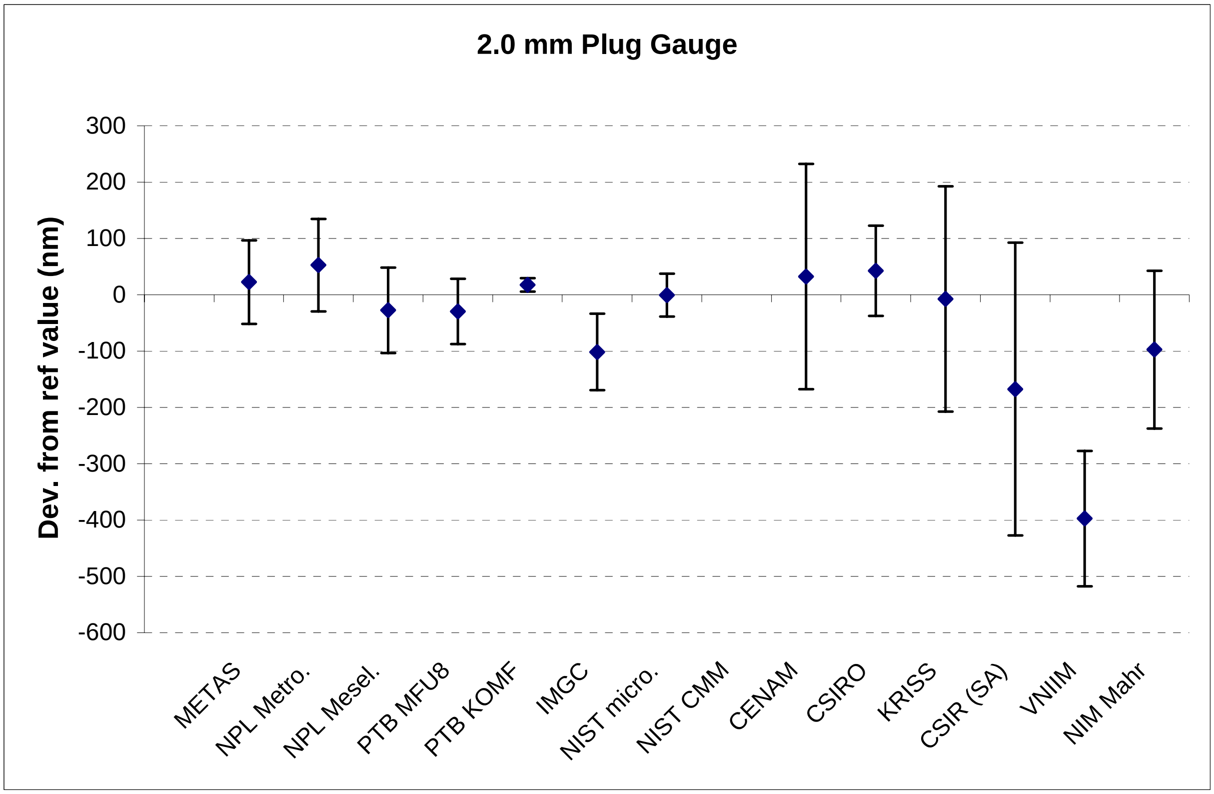
The following charts show the measurement results and the associated expanded (k = 2) uncertainty in graphical form. The reference value is derived from the average of the modified weighted mean and the total bootstrapped median and is used as the zero of the Y axis.

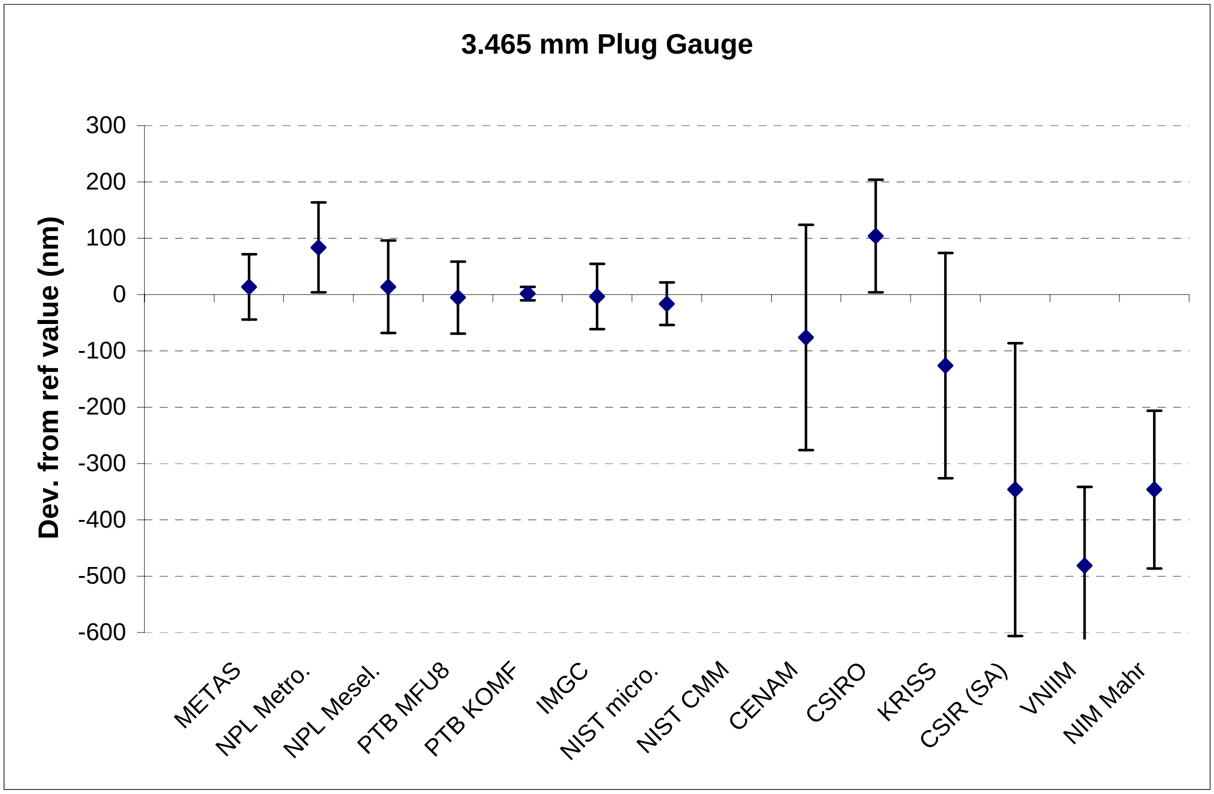


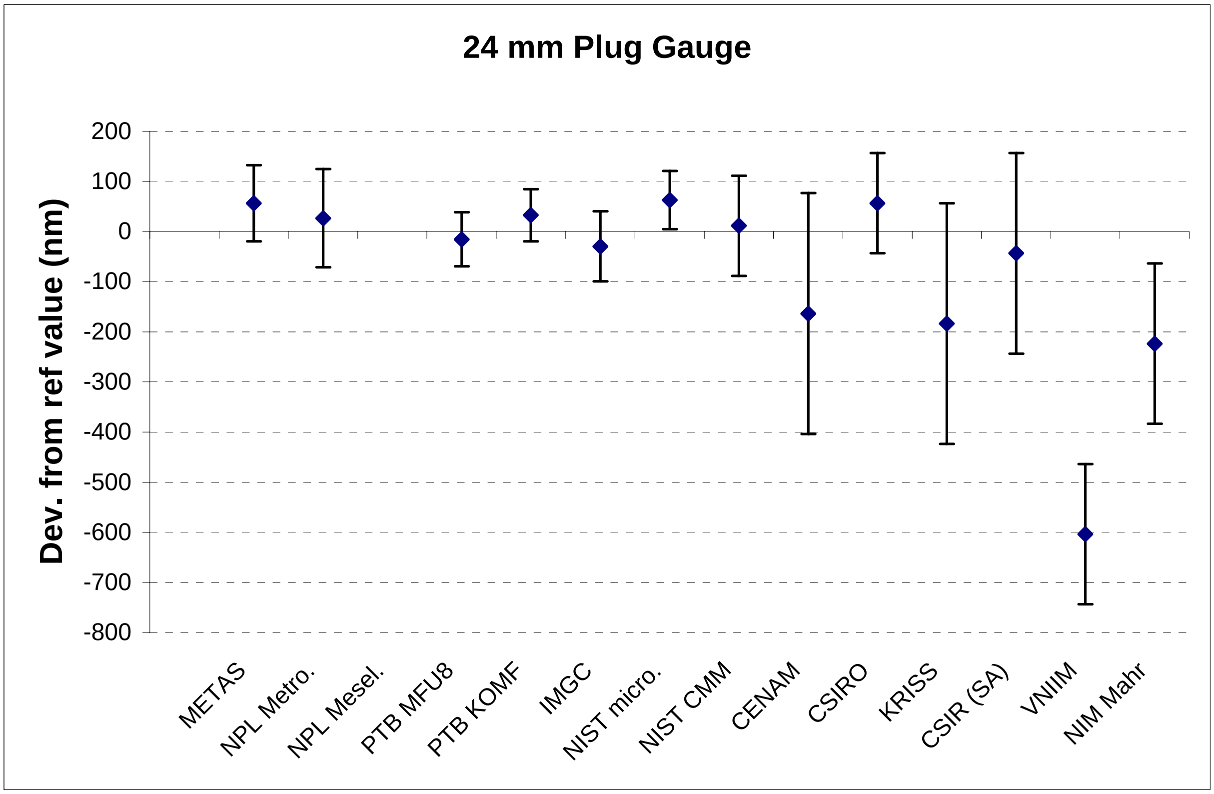


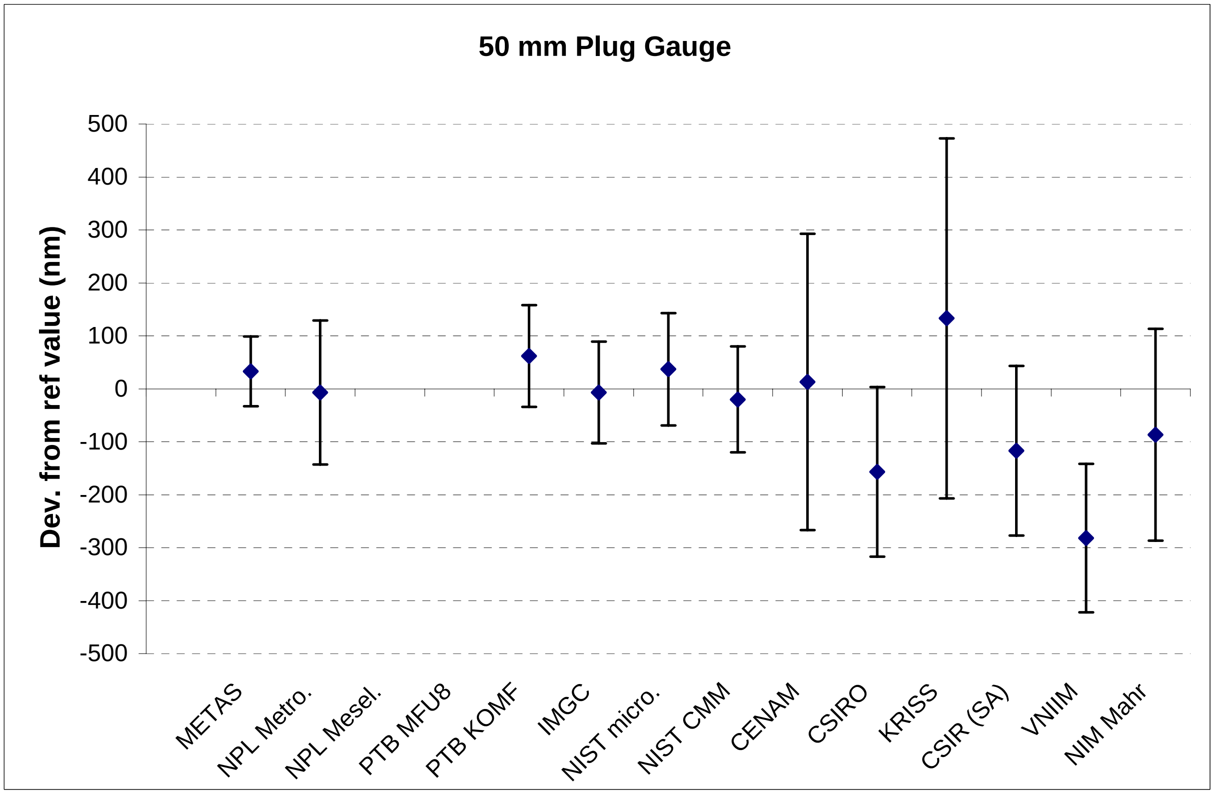

49.3 mm Ring Gauge
Dev. from ref value (nm)
400
300
200
100
0
-100
-200
-300
-400
-500
-600
METAS
NPL IDM
PTB MFU8
PTB KOMF
IMGC
NIST comp.
NIST CMM
CENAM
CSIRO
KRISS
CSIR (SA)
VNIIM
NIM Mahr
NIM Interf.

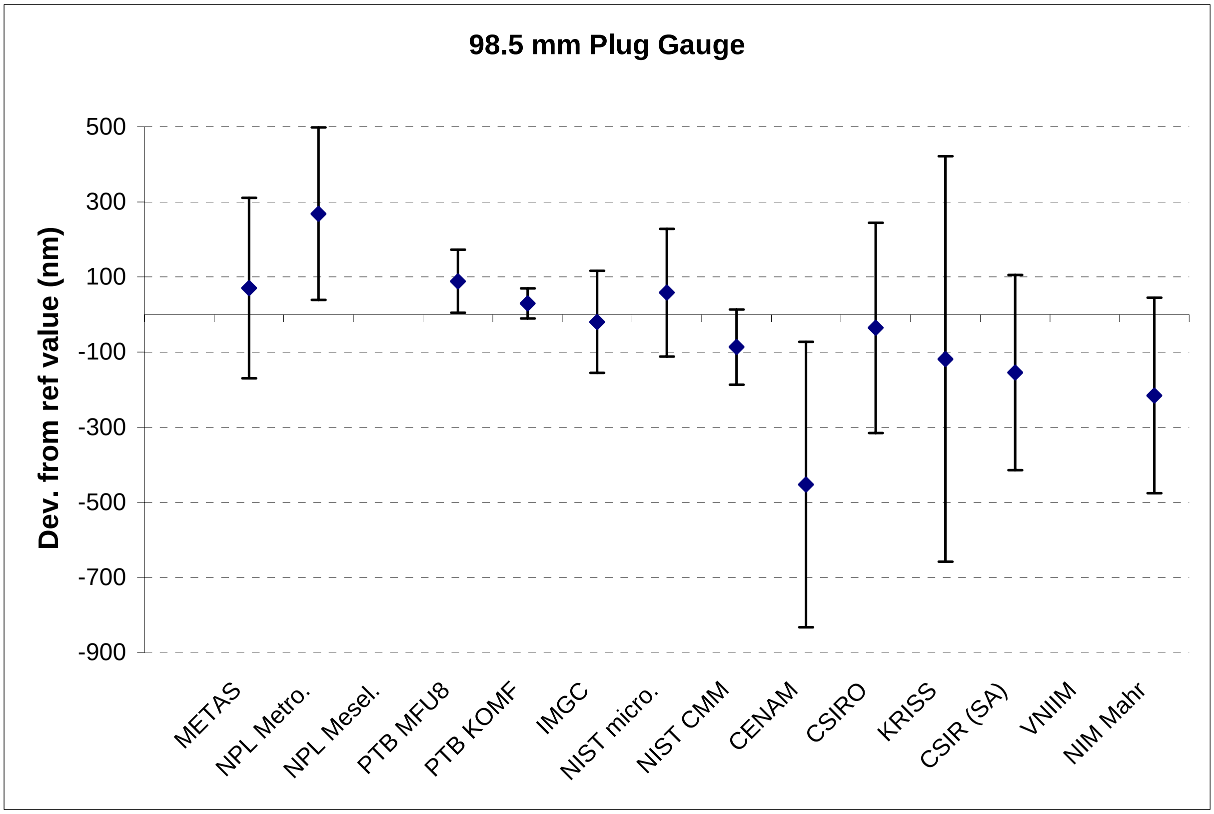

100 mm Ring Gauge
Dev. from ref value (nm)
700
500
300
100
-100
-300
-500
-700
METAS
NPL IDM
PTB MFU8
PTB KOMF
IMGC
NIST comp.
NIST CMM
CENAM
CSIRO
KRISS
CSIR (SA)
VNIIM
NIM Mahr
NIM Interf.


2.0 mm Plug Gauge
Dev. from ref value (nm)
300
200
100
0
-100
-200
-300
-400
-500
-600
METAS
NPL Metro.
NPL Mesel.
PTB MFU8
PTB KOMF
IMGC
NIST micro.
NIST CMM
CENAM
CSIRO
KRISS
CSIR (SA)
VNIIM
NIM Mahr


3.465 mm Plug Gauge
Dev. from ref value (nm)
300
200
100
0
-100
-200
-300
-400
-500
-600
METAS
NPL Metro.
NPL Mesel.
PTB MFU8
PTB KOMF
IMGC
NIST micro.
NIST CMM
CENAM
CSIRO
KRISS
CSIR (SA)
VNIIM
NIM Mahr


24 mm Plug Gauge
Dev. from ref value (nm)
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
METAS
NPL Metro.
NPL Mesel.
PTB MFU8
PTB KOMF
IMGC
NIST micro.
NIST CMM
CENAM
CSIRO
KRISS
CSIR (SA)
VNIIM
NIM Mahr


50 mm Plug Gauge
Dev. from ref value (nm)
500
400
300
200
100
0
-100
-200
-300
-400
-500
METAS
NPL Metro.
NPL Mesel.
PTB MFU8
PTB KOMF
IMGC
NIST micro.
NIST CMM
CENAM
CSIRO
KRISS
CSIR (SA)
VNIIM
NIM Mahr



10. References

1. Quinn T. J. 1999 *Guidelines for key comparisons carried out by Consultative Committees*, BIPM, Paris (available at http://www.bipm.fr).

2. Vaucher G. and Thalmann R. 1993 *CCDM Gauge Block Intercomparison,* OFMET, Wabern

3. Thalmann R., Decker J., Brown N. 1998 *CCL Key Comparison: Calibration of Gauge Blocks by Interferometry*, OFMET, Wabern

4. Thalmann R. 2000 *Metrologia* **37** 253-260

5. Lewis A. 2003 *Metrologia* **40** 04004

6. Cox M and Iguzquiza E.P. 2001 *The Total Median and its Uncertainty*, in A*dvanced Mathematical and Computational Tools in Metrology V*, World Scientific Publishing Company, 106-117

Appendix A. Reference Value Determinations

The calculation of the reference values for this intercomparison was complicated due to many factors. First, the widely varying instrumentation and methods used by the participants may have resulted in inconsistent realizations of the measurand. This was the anticipated consequence. The ensuing reference value calculations were required to arrive at some reasonable and statistically relevant solution taking this into consideration. Additionally, as noted in previous intercomparisons of ring and plug gages [4], the data displayed an asymmetric distribution with opposing tendency between the internal and external diameter results. Many of the primary uncertainty sources involved with these measurements would produce one-sided, biased results. At some laboratories, the same equipment was used for both internal and external diameter measurements. These one-sided errors would be propagated in opposite directions depending on which artifact was being measured.

The artifacts themselves also contributed to the analysis difficulty due to the fact that these gauges were of typical geometry and surface finish. These characteristics complicated the fact that the participants used a wide range of methods to arrive at an undeformed measurement result. In some cases, applied forces of greater than 2 N were used and then corrected using elastic formulae back to an undeformed result, while other participants used forces of only a few milli-Newtons. Based on these facts, it would be assumed that the gauges with the smoothest surfaces would produce the most consistent results, even between very different techniques. It is difficult to determine if this assumption is proven from the actual data. It is likely that in some cases the potential systematic errors from the correction for deformation were small compared to other larger instrument related error sources.

The calculation of the reference values for this key comparison needed to take all of these issues into consideration. The calculations evolved into the investigation of several methods. The first was the well-documented weighted mean calculation as used regularly in other completed key comparisons. Within the context of this method, the Birge Ratio, R_b, is used as a test of the overall statistical consistency of the data, and the consistency of an individual result with the reference value is measured by E_n, the normalized deviation of a result from the reference value. In the analysis used here we define E_n using a coverage factor $k = 1$, as is done in [5]. See reference [5] for details of the calculation of E_n, R_B, and the reference value.

The PTB requested that their submitted laser comparator data be excluded from the reference value calculation since the instrument is no longer in service. When the results from all laboratories except the PTB laser comparator submission are included in a weighted mean, the results are as shown in Table A1. Note that the Birge Ratio R_B and the normalized deviation E_n are often much larger than 1, indicating that the uncertainty estimates are not all reliable. Most artifacts were measured 14 times. For 14 measurements and k = 2, the Birge ratio should be less than 1.34 for consistency. By contrast, what we actually see is that for all but one of the artifacts R_B >1.34. For this data a “typical” Birge ratio is R_B = 2.15, where by “typical” we mean the root mean square (RMS) value of the Birge ratio for all artifacts. This result clearly indicates that the data are inconsistent and that the weighted mean will not be reliable without some modification.

In some past comparisons it has been possible to identify one or two laboratories as outliers who encountered some problem with the measurements. It may be possible to eliminate these outlier results from further consideration, leaving a subset of laboratories with mutually consistent data. In the current comparison this might also be possible, but only by excluding a large amount of data in addition to accounting for withdrawal and updated uncertainty statements of PTB. That is, there are indications that, in general, a number of laboratories had difficulty in estimating the uncertainty. In light of this difficulty, we explored a method for computing the reference value as described by Thalmann [4]. Rather than look at laboratory performance for all artifacts in the aggregate, we look at each artifact individually. We form the weighted mean and throw out the laboratory with the largest E_n value, where E_n is a measure of consistency. Then the weighted mean is recomputed and the process iterated until all remaining E_n values are less than 2. This procedure resulted in the exclusion of the measurements of between two and five laboratories in computing the reference values of each artifact. The laboratories whose measurements were excluded differed from one artifact to the next.

This method is not perfectly efficient in a statistical sense, because some "good" data at the edges of the probability distribution will be excluded, but this is a small price to pay in return for the significant immunity of the reference value to undue influence from outliers. It is also worth noting, as a caution, that when two E_n values are nearly equal, it is not clear which of the corresponding measurements should be excluded to improve the consistency of the remaining values. For simplicity we ignore this complication (we always exclude the measurement corresponding to the largest E_n), but we must recognize that more elaborate procedures might be slightly more efficient in finding an appropriate reference value.

	Rings				Plugs				
	5.17 mm	11.95 mm	49.3 mm	100 mm	2 mm	3.465 mm	24 mm	50 mm	98.5 mm
Participant	E_n	E_n	E_n	E_n	E_n	E_n	E_n	E_n	E_n
METAS	-0.50	-0.23	-0.09	1.29	0.40	0.58	1.68	1.64	0.43
NPL IDM	-2.27	-0.86	-1.42	-0.82					
NPL Metro.					1.10	2.18	0.65	0.10	2.19
NPL Mesel.					-0.94	0.40			
PTB MFU8	-1.36	1.06	0.43		-1.31	-0.08	-0.46		1.77
PTB KOMF		-1.41	1.02	-0.74	3.40	1.53	1.59	1.68	0.81
IMGC	1.06	-1.33	2.35	-0.57	-3.25	-0.02	-0.76	0.14	-0.59
NIST Comp.	-0.01	0.46	0.04	1.94	-0.45	-0.74	2.52	1.00	0.47
NIST CMM	-1.60	-0.07	-0.31	0.15			0.33	-0.14	-2.22
CENAM	-0.25	0.05	-0.36	0.52	0.25	-0.74	-1.33	0.19	-2.49
CSIRO	1.39	1.85	-0.41	1.91	0.88	2.14	1.25	-1.84	-0.39
KRISS	7.18	0.86	-0.23	0.94	-0.15	-1.24	-1.50	0.86	-0.51
CSIR	-2.72	0.51	-1.67	0.19	-1.35	-2.65	-0.39	-1.33	-1.35
VNIIM	1.28	-0.94	-6.03	-5.50	-6.78	-6.86	-8.68	-3.96	
NIM Mahr	-0.01	2.38	0.57	1.91	-1.51	-4.92	-2.77	-0.75	-1.82
NIM Interf.			2.08	2.90					
Birge Ratio	**2.53**	**1.15**	**1.98**	**2.07**	**2.37**	**2.75**	**2.85**	**1.55**	**1.49**

Table A1. E_n values and Birge ratio calculations for the full dataset (except the PTB Laser data).

In any event, the result as calculated here should provide an unbiased estimate of the reference value, and is obtained in a straightforward, well-defined manner.

	Rings				Plugs				
	5.17 mm	11.95 mm	49.3 mm	100 mm	2 mm	3.465 mm	24 mm	50 mm	98.5 mm
Participant	E_n	E_n	E_n	E_n	E_n	E_n	E_n	E_n	E_n
METAS	-0.18	-0.14	-0.09	1.01	0.24	0.48	1.07	0.77	0.35
NPL IDM	-1.95	-0.72	-1.43	-1.15					
NPL Metro.					0.95	2.08	0.18	-0.28	2.06
NPL Mesel.					-1.10	0.34			
PTB MFU8	-1.09	1.24	0.44		-1.52	-0.17	-1.38		1.53
PTB KOMF		-1.26	1.11	-1.35	1.40	0.67	0.63	1.13	0.02
IMGC	1.57	-1.21	2.14	-0.90	-3.35	-0.11	-1.44	-0.42	-0.74
NIST Comp.	0.24	0.60	0.04	1.64	-0.78	-0.89	1.68	0.51	0.35
NIST CMM	-1.32	0.05	-0.31	-0.25			-0.13	-0.68	-2.20
CENAM	-0.19	0.10	-0.36	0.43	0.19	-0.76	-1.52	0.01	-2.53
CSIRO	1.56	1.93	-0.41	1.80	0.73	2.06	0.79	-2.06	-0.47
KRISS	7.13	0.92	-0.23	0.87	-0.21	-1.26	-1.69	0.72	-0.55
CSIR	-2.58	0.55	-1.67	0.04	-1.40	-2.66	-0.62	-1.65	-1.42
VNIIM	1.52	-0.84	-5.85	-5.38	-6.83	-6.85	-8.75	-4.06	
NIM Mahr	0.13	2.38	0.57	1.80	-1.59	-4.93	-2.99	-1.00	-1.90
NIM Interf.			2.02	2.67					
Birge Ratio	**1.17**	**0.96**	**0.79**	**1.17**	**1.02**	**0.69**	**1.15**	**0.85**	**1.05**

Table A2. Birge ratios are within acceptable levels. The measurements corresponding to all E_n values greater than 2 were excluded from the reference value and from calculation of the Birge ratio.

Table A2 shows results after excluding enough data to achieve consistency. The Birge ratios calculated from the remaining data are reasonable. The E_n values in the table were calculated in accordance with the following explanation. Normally E_n is calculated using:

$$E_n = \frac{x_i - \overline{x}_w}{\sqrt{[u(x_i)]^2 - [u(x_{\text{int}})]^2}}$$

where x_i is the measurement result for laboratory i, $\overline{x}_w$ is the weighted mean, $u(x_i)$ is the laboratory's stated standard uncertainty, and $u(x_{\text{int}})$ is the internal standard deviation. Here the subtraction of internal standard deviation accounts for correlations between the reference value and the measurement in question. However, when a measurement has been excluded from the calculation of the reference value and is not correlated with the reference, E_n is given by

$$E_n = \frac{x_i - \overline{x}_w}{\sqrt{[u(x_i)]^2 + [u(x_{\text{int}})]^2}}$$

In principle, a given result might be excluded from consideration because E_n as calculated from the first formula has magnitude greater than 2, but the result could then appear in the table with E_n <2 when it is recalculated according to the second formula. In practice, this did not occur for this data set.

Another method for calculation of the reference value was also performed. Median calculations are well known for their robustness in the face of widely dispersed data. The total median or total bootstrap technique [6] is also an appealing estimator for reference values because it retains the robust property of the median but has a smaller mean-squared error. The total median, variance of the total median, and coverage intervals for the total median can be readily obtained using bootstrap re-sampling techniques.

Bootstrap re-sampling is interpreted as given any numerical population, a sample of size n can be drawn from it and an estimator calculated from that sample. Total bootstrap is defined as drawing, with replacement, all distinct bootstrap samples from the given population. For a population of size p, there are p^p distinct bootstrap samples. For large populations it is prohibitive to perform the enumerations, requiring some form of simulation. However, for the median estimator, the consequences of analyzing the results of full enumeration can be obtained without carrying out the full enumeration process. The total median, T, is defined as the mathematical expectation of the median according to the bootstrap. For details on the properties of and the implementation of the total median, refer to the literature [6].

To explore differences in various methods for finding the reference value, we calculated the reference values using the total median bootstrap (1000 bootstrap samples), the mean (unweighted), the median, the weighted mean, and the modified weighted mean. Results are shown in Table A3. (These calculations were carried out prior to several later revisions in the analysis are given here only to illustrate typical differences in the methods.) The graph in Figure 3 shows the varying results of the different methods. It is offered that averaging the most robust methods with reasonably low calculated uncertainty of the reference value is one way to arrive at a fair estimation of the reference value due to the data inconsistency, varied laboratory techniques, wide range of reported uncertainty values, and one-sided bias errors in the data set.

	Mean	**Std. dev. (nm)**	**Median**	**Std. dev. (nm)**	**Weighted Mean**	**Std. dev. (nm)**	**Total Median**	**Std. dev. (nm)**	**Modified Weighted Mean**	**Std. dev. (nm)**
Rings										
5.17 mm	5.169966	236	5.169930	38	5.169937	38	5.169936	23	5.169937	16
11.95 mm	11.949801	95	11.949786	35	11.949729	22	11.949794	30	11.949779	16
49.3 mm	49.300044	134	49.300060	39	49.300091	26	49.300053	30	49.300066	13
100.0 mm	99.999997	198	99.999980	84	99.999919	32	100.000001	63	99.999944	21
Plugs										
2.0 mm	2.000024	121	2.000064	36	2.000079	13	2.000054	33	2.000081	6
3.465 mm	3.465465	180	3.465542	97	3.465547	15	3.465521	35	3.465546	5
24.0 mm	24.001273	180	24.001342	60	24.001369	16	24.001325	49	24.001362	12
50.0 mm	50.001070	109	50.001090	40	50.001145	15	50.001094	31	50.001109	18
98.5 mm	98.502194	172	98.502210	55	98.502266	26	98.502217	43	98.502259	17

Table A3. Reference values obtained using multiple techniques.

Under this qualifier, the average of the modified weighted mean and the total bootstrap median is offered as a good estimate: If x_w is the modified weighted mean and if x_t is the total median, then the reference value x_{ref} is:

$$x_{ref} = \frac{x_w + x_t}{2}$$

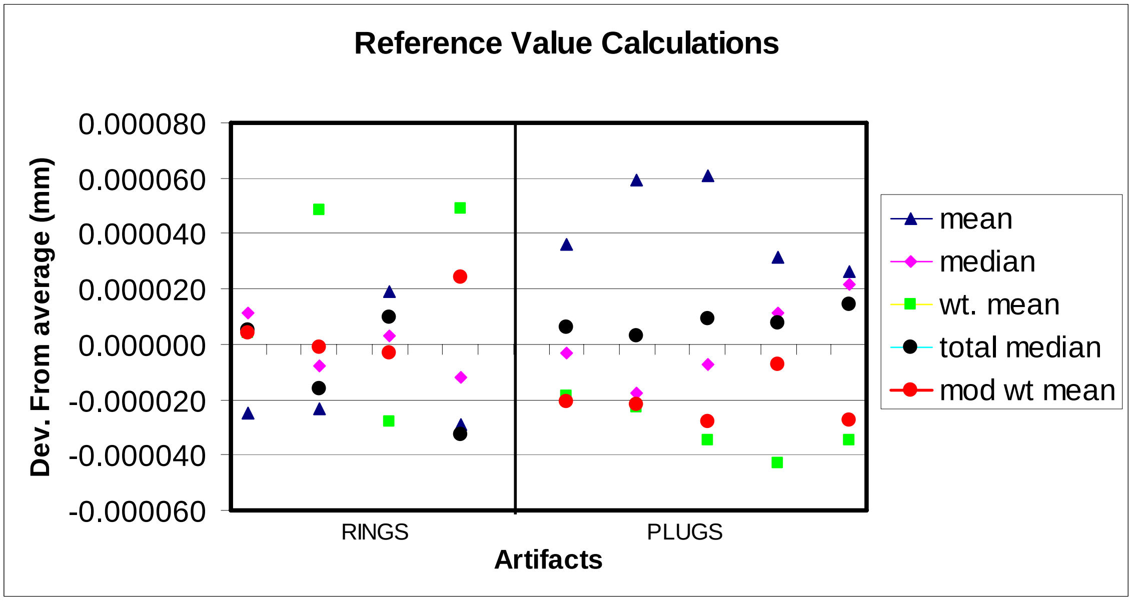


Figure A1. Graphical spread of reference values using multiple methods of calculation.

Assigning an uncertainty to the reference value is difficult. We can expect that errors in x_w and x_t will have significant correlation, so that the uncertainty of the reference value, u_{ref}, could be as large as

$$u_{ref} = \frac{u_w + u_t}{2}$$

where u_w and u_t are the uncertainties of the modified weighted mean and total median. However, it has been pointed out that the uncertainty derived from the modified weighted mean calculations is almost assuredly too small, because the data chosen for the analysis was selected for consistency. A conservative estimate for the uncertainty of the reference value should probably be closer to the uncertainty of the total median than to the (much smaller) uncertainty that has been assigned to the weighted mean. For the uncertainty of the reference value, we use an *ad hoc* combination of the two uncertainties that achieves this desired result:

$$u_{ref} = \sqrt{(u_w^2 + u_t^2)/2}$$

To this must be added an additional uncertainty to account for possible artifact instability. This additional uncertainty, u_a, is on the order of 25 nm for rings and plugs greater than 25 mm in diameter. (Some details of estimating u_a, u_t, and u_w are discussed in Appendix B.) Including this artifact uncertainty, the final uncertainty of the reference value is

$$u_{ref} = \sqrt{\frac{u_w^2 + u_t^2}{2} + u_a^2}$$

Table A4 shows the final reference values and their associated calculated uncertainties.

Artifact	Rings				Plugs				
	5.17 mm	11.95 mm	49.3 mm	100.0 mm	2.0 mm	3.465 mm	24.0 mm	50.0 mm	98.5 mm
reference value (mm)	5.169936	11.949786	49.300060	99.999973	2.000067	3.465546	24.001343	50.001097	98.502229
std unc. (nm)	29	21	19	51	22	25	36	32	39

Table A4. Final reference value calculations with standard uncertainties.

For this intercomparison, it was decided to let laboratories submit multiple measurements using different pieces of equipment. This seems to have been advantageous since the different submissions within each laboratory appear uncorrelated and independent, allowing us to treat each submission essentially as an independent measurement. This resulted in populating the sample with more valid measurements using well understood measurement techniques than would have been possible otherwise. This resulted in a better estimate of the reference value and a more meaningful intercomparison than would have occurred otherwise if each laboratory was limited to only one submission.

As mentioned at the start of this section, alignment errors can be expected to cause correlated errors and to bias the results. In principle, these errors should be estimated and the bias should be removed as described in the GUM. In practice, such adjustments are almost never made in dimensional metrology, and it appears that none of the participants in this comparison removed the bias in the recommended manner. Without complete knowledge of how alignment errors were estimated, it would be inappropriate for the pilot laboratory to carry out this adjustment; we will make no corrections for the bias. However, for completeness we should at least estimate the possible bias that alignment errors might cause in the reference value. As an example, consider the case of the 49 mm ring. For each laboratory, we assign an error due to misalignment that is equal to the uncertainty for alignment given in Table 8. When no estimate of this uncertainty component is interpreted, we assign an error comparable to what other laboratories with the same overall uncertainty estimate for their alignment uncertainty. Using these assigned errors, we estimate that misalignment will shift the reference value by about 18 nm. This is not insignificant relative to the 24 nm standard uncertainty of the reference value. In future comparisons it would be desirable to more carefully consider the effect of correlated alignment errors.

Appendix B: Some details of the calculation of reference values, uncertainty, and degrees of equivalence

As mentioned in the text, we calculate the reference value from the average of the modified weighted mean and the total median.

B.1 The total median: The total median may be somewhat less familiar than the weighted mean. The total median can be found either by using a formula based on calculated weights or by bootstrap simulation. The bootstrap simulation involves sampling with replacement from the N results obtained by the participants. N samples are picked at random from the N results, where one particular result can be picked more than one time (even as many as N times). A median is determined for the N samples. The process is repeated many times, and the medians are averaged to obtain the bootstrap median, which is equal to the total median within a small statistical uncertainty. This process has the effect of assigning weights to the results from each laboratory, and these weights can be used to calculate the reference value. The weight for a given laboratory depends on (1) the total number of laboratories participating and (2) where the laboratory lies in a list of numerically ordered results from the participants. Laboratories that lie near the center of the ordered list are given high weight, and laboratories that achieved extreme results, at the two ends of the list, are given low weights.

For a given number of participants, the weighing factors are always the same, and they can be determined analytically as has been shown by Cox [6]. They can also be obtained through numerical simulation of artificial data. For example, if the bootstrap median is calculated for a set of 5 results x_i= 0, 2, 4, 6, and 8, and then recalculated for results x_i=0, 2, 5, 6, 8 (with the third value increased by 1, which does not change the order), then the difference between the first bootstrap median and the second gives the weighing factor for the third participant in an ordered set of 5 results.

The weighing factors needed to evaluate our data (N=12, 13, or 14) are given in table B.1. These values have been determined through bootstrap simulation as described above.

N=12	*N*=13	*N*=14
0.00011	0.00002	0.00001
0.0045	0.00146	0.00093
0.0297	0.0142	0.00945
0.0878	0.0549	0.0385
0.1624	0.1243	0.0929
0.2155	0.1936	0.1571
0.2155	0.2232	0.2011
0.1624	0.1936	0.2011
0.0878	0.1243	0.1571
0.0297	0.0549	0.0929
0.0045	0.0142	0.0385
0.00011	0.00146	0.00945
	0.00002	0.00093
		0.00001

Table B.1: Weighting factors for N=12, N=13, and N=14. The last digit in each entry is uncertain.

B.2 Uncertainty of the Reference Value. As has already been discussed, we assign an uncertainty to the reference value that is a combination of the uncertainty of the total median, the uncertainty of the modified weighted mean, and the uncertainty due to possible artifact instability:

$$u_{ref}^2 = \frac{u_w^2 + u_t^2}{2} + u_a^2$$

Because the uncertainty of the weighted mean is arguably too small, the calculated uncertainty is more heavily weighted toward the uncertainty of the total median than would be calculated according to a standard propagation of uncertainty. In his paper, Cox gives results for the 95% uncertainty interval of the total median as calculated according to his mathematical model. According to his paper, the endpoints of the 95% confidence interval are x_k and x_{N-k+1} where x_k is the kth result in an ordered list of the N measurement results, and k=3 for N=12, k= 4 for N=13, and k=4 for N=14. The uncertainty of the weighted mean is calculated from the artifact uncertainties as shown in reference 6.

The artifact uncertainties due to drift were estimated in the following manner:

(1) Fit a line to the pilot lab stability data and find the uncertainty of the slope.
(2) Estimate the maximum possible rate of artifact drift by adding the magnitude of the best fit slope to twice the uncertainty in the best fit slope.
(3) Assign a k=2 artifact uncertainty equal to this drift times half the duration of the comparison. Assuming a linear drift, this will correctly estimate the uncertainty at the beginning and the end of the comparison, although it will overestimate the uncertainty for labs participating near the mid-point in time of the comparison.

In some cases, the procedure above yields unrealistically large uncertainties. Our experience has been that dimensional instability of steel artifacts is always $\Delta l/l<3\text{x}10^{-6}$ per year. When the rate as estimated above exceeds this value, we determine the uncertainty assuming a rate of 3x10^{-6} per year. We note that at the k=2 level all of our results are consistent with the possibility that there is no measurable instability, with the exception of the 98.5 mm plug gauge.

B.3 Degrees of equivalence: In order to determine the degree of equivalence, we must calculate the differences d_i=x_i-x_{ref} between each laboratory's result x_i and the reference value x_{ref}, and we must compute the uncertainty of this result in a manner that properly account for correlations between x_i and x_{ref}. For the total median, Cox does not discuss these correlations or the uncertainties of d_i. However, we note that the weighing factors in Table B.1 never exceed 22%, suggesting that correlations are most likely small. When the number of participants is 12 or greater, correlations of individual laboratories with the total median are probably of only minor importance.

Correlations with the weighted mean are much more important. Before PTB expanded their uncertainty, the greatest concern was the 2 mm plug, where the PTB form machine contributed 80% of the total weight. Ignoring possible small correlations with the total median, the PTB result now contributes 40% in the final reference value. Below we show that even in this most extreme case, correlations have only minor importance.

In the following discussion we ignore the uncertainty due to artifact instability.

Suppose that we could write an analytic expression

$$x_{ref} = w_i x_i + f \qquad (1)$$

where w_i is a weight for some result x_i and f is a function that is independent of x_i: $f = f(x_1, x_2, ...x_{i-1}, x_{i+1}, x_N)$. For example, equation (1) can be written when the reference value is the mean or the weighted mean. The situation is much less clear for the total median.

For our particular case, averaging the weighted mean and the total median, we can write

$$x_{ref} = (w_i^{'} x_i + f_1 + f_{tm}) / 2 \qquad (2)$$

where $w_i^{'} x_i + f_1$ is the usual weighted mean, $w_i^{'}$ the weight for laboratory i calculated according to the standard prescription, and f_{tm} is the total median. When more than 12 laboratories participate in a comparison, exclusion of a single laboratory does not change the total median by a great deal, and we can say that (at least approximately) f_{tm} is independent of x_i. If (a) we accept this approximation, (b) we identify f in equation (1) with (f_1 +f_{tm})/2, and (c) we set $w_i = w_i^{'} / 2$, then equation (1) will correctly describe the situation for our reference value.

Now find the uncertainty in the deviation d = $x_i - x_{ref}$:

$$d = (1 - w_i) x_i - f \qquad (3)$$

and consequently

$$u_d^2 = (1 - w_i)^2 u_i^2 + u_f^2 \qquad (4)$$

Note that the overall uncertainty of the reference value is

$$u_{ref}^2 = w_i^{\,2} u_i^2 + u_f^2 \qquad (5)$$

Combining (4) and (5) give

$$u_d^2 = (1 - 2w_i) u_i^2 + u_{ref}^2 \qquad (6)$$

Thus, for example, for laboratories that have small weights in computing the reference value,

$$u_d^2 \approx u_i^2 + u_{ref}^2 \ . \qquad (7)$$

For the example of the PTB measurement of the 2 mm ring mentioned previously, , $w_i = 40\%$ (because $w_i^{'} = 80\%$) and

$$u_d^2 = 0.2 u_i^2 + u_{ref}^2 \qquad (8)$$

For the 2 mm ring, formula (8) yields u_d = 24.1 nm, while (7), which ignores correlations, would give u_d =24.7 nm. The difference between these two results is very small relative other uncertainties in our analysis. The smallness of the effect of correlations is a consequence of the relatively large uncertainty that we have assigned to the reference value (much larger than the uncertainty of the weighted mean as calculated by the standard method). Under these conditions, it is reasonable to ignore the effect of correlations.

Table B.3 shows the differences of measured diameters with respect to the Key Comparison Reference Values and the expanded (k=2) uncertainties of these differences for each artifact, laboratory, and method.

Laboratory	RINGS				PLUGS				
	5.17 mm	11.95 mm	49.3 mm	100 mm	2.0 mm	3.465 mm	24.0 mm	50.0 mm	98.5 mm
METAS	-6 ± 105	-16 ± 139	0 ± 139	37 ± 170	23 ± 86	14 ± 77	57 ± 104	33 ± 93	71 ± 255
NPL IDM	-76 ± 104	-36 ± 96	-60 ± 103	-93 ± 157					
NPL Metro.					53 ± 93	84 ± 94	27 ± 121	-7 ± 151	269 ± 246
NPL Mesel.					-27 ± 87	14 ± 96			
PTB MFU8	-50 ± 116	33 ± 83	18 ± 71		-29 ± 72	-5 ± 81	-15 ± 90		89 ± 120
PTB KOMF		-55 ± 92	20 ± 52	-70 ± 126	18 ± 45	2 ± 51	33 ± 88	62 ± 116	30 ±95
IMGC	43 ± 85	-66 ± 110	96 ± 89	-79 ± 157	-101 ± 80	-3 ± 77	-29 ± 100	-7 ± 116	-19 ± 161
NIST Comp.	14 ± 123	18 ± 99	8 ± 109	69 ± 162	0 ± 57	-16 ± 63	63 ± 92	37 ± 124	59 ± 191
NIST CMM	-61 ± 116	-5 ± 108	-9 ± 107	-40 ± 143			12 ± 123	-20 ± 119	-86 ± 132
CENAM	-36 ± 404	4 ± 224	-40 ± 263	57 ± 413	33 ± 205	-76± 206	-163 ± 250	13 ± 287	-452 ± 390
CSIRO	124 ± 170	144 ± 165	-30 ± 184	257 ± 336	43 ± 91	104± 112	57 ± 123	-157 ± 173	-35 ± 293
KRISS	724 ± 208	84 ± 204	-30 ± 322	197 ± 530	-7 ± 205	-126 ± 206	-183 ± 250	133 ± 346	-118 ± 547
CSIR (SA)	-336 ± 266	64 ± 263	-160 ± 204	-23 ± 279	-167 ± 264	-346 ± 265	-43 ± 212	-117 ± 173	-154 ± 274
VNIIM	89 ± 133	-56 ± 127	-410 ± 145	-473 ± 190	-397 ± 127	-481 ± 149	-603 ± 157	-282 ± 154	
NIM Mahr	14 ± 208	234 ± 204	80 ± 263	257 ± 336	-97 ± 146	-346 ± 149	-223 ± 175	-87 ± 210	-215 ± 274
NIM Interf.			150 ± 145	297 ± 261					

Table B.3. Differences from the Key Comparison Reference Values with the associated expanded uncertainties (*k*=2).

Appendix C: Roundness plots of the artifacts

The roundness plots shown here were provided by METAS. Data was taken using a 2RC filter with a cutoff of 150 UPR.

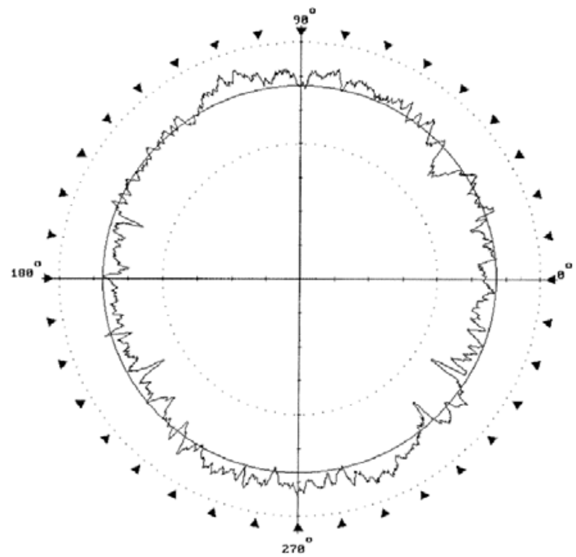


Fig. C1. 5.17 mm ring, Scale= 0.1 μm

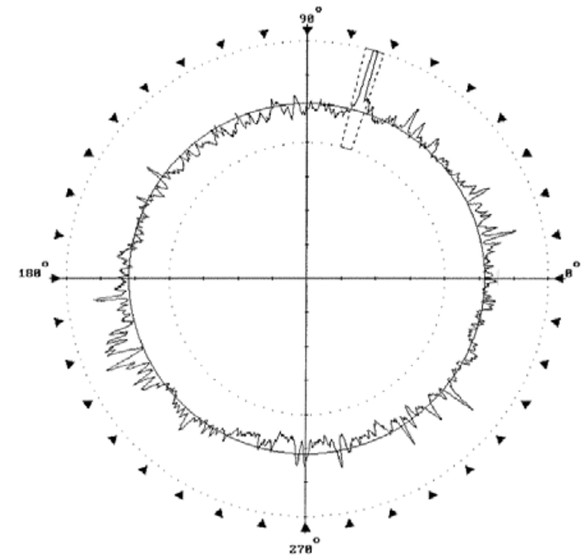


Fig. C2. 11.95 mm ring. Scale= 0.2 μm

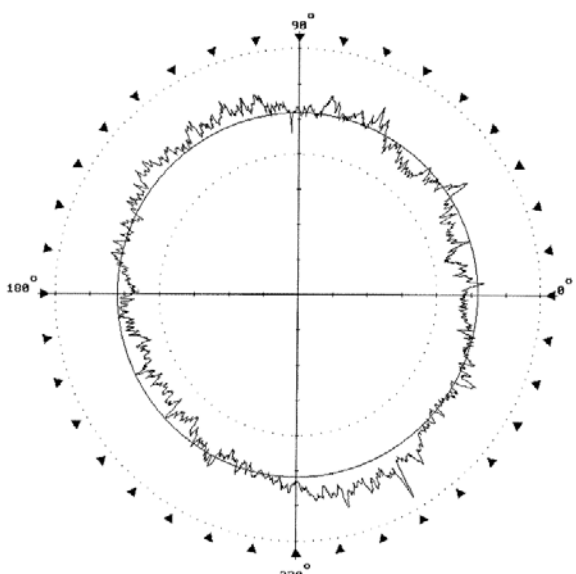


Fig. C3. 49.3 mm ring. Scale= 0.1 μm

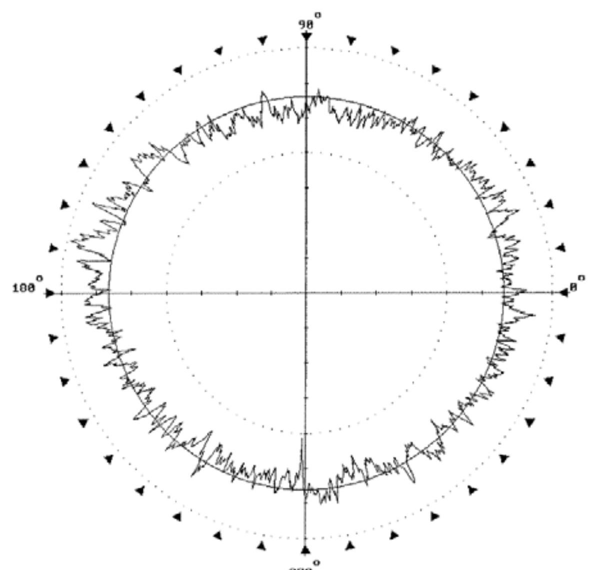


Fig. C4. 100 mm ring. Scale= 0.1 μm

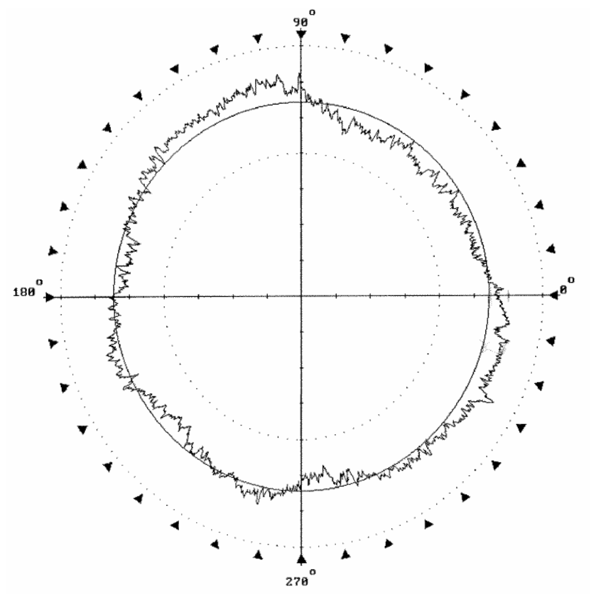


Fig. C5. 2 mm plug. Scale= 0.1 µm

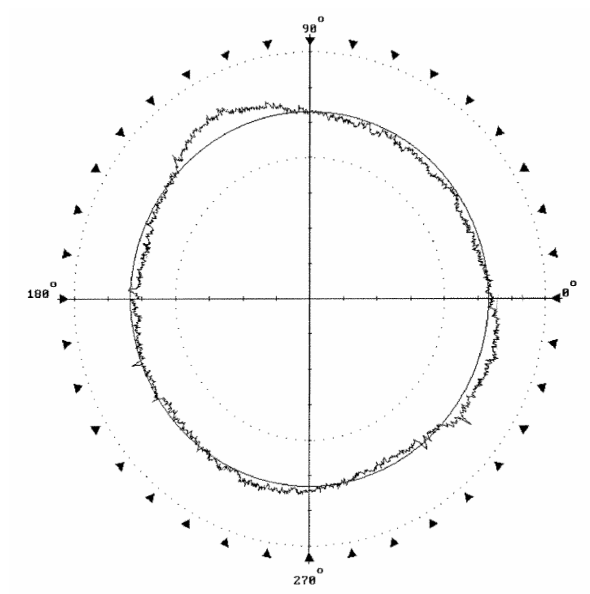


Fig. C6. 3.465 mm plug. Scale= 0.1 µm

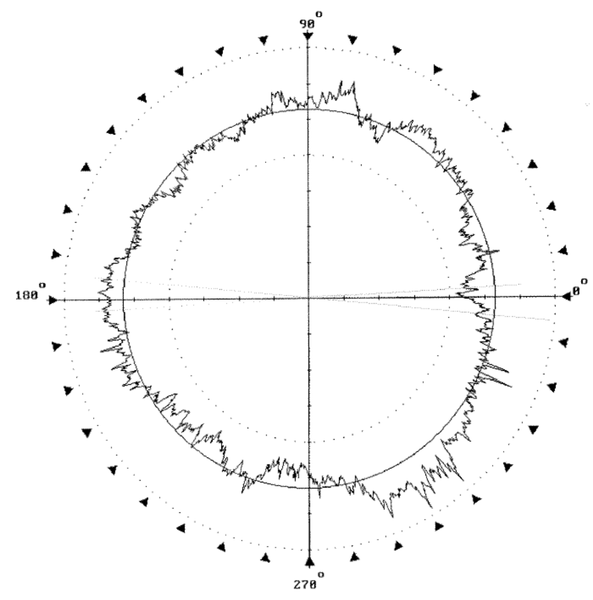


Fig. C7. 24 mm plug. Scale= 0.1 µm

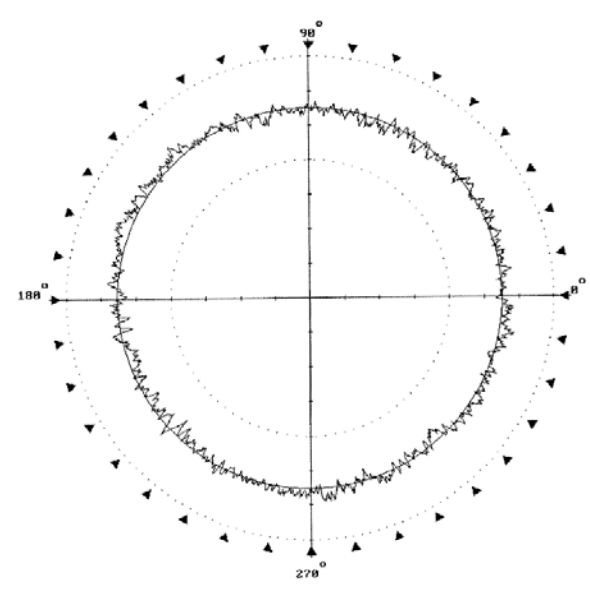


Fig. C8. 50 mm plug. Scale= 0.1 µm

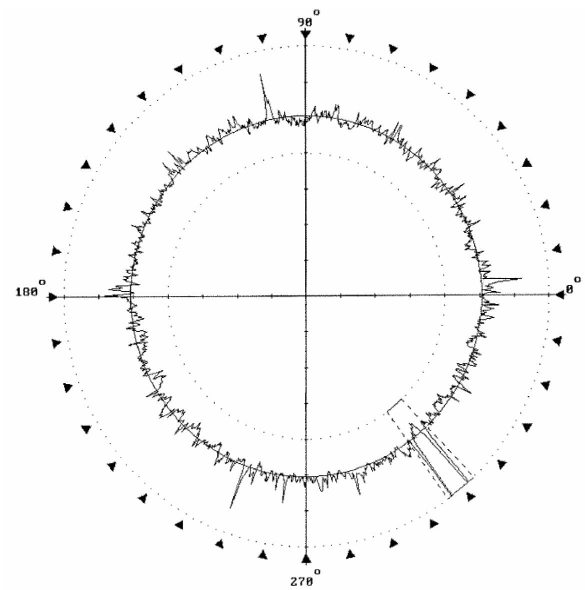


Fig. C9. 98.5 mm plug. Scale= 0.2 μm