Inter-comparison Example

Speaker: Maritoni Litorja[2]

Authors: Adriana Hornikova[1], Maritoni Litorja[2], Stefan D. Leigh[1], Joel B. Fowler[2]
[1] Statistical Engineering Div., [2] Optical Technology Div. of
National Institute of Standards and Technology, DoC
100 Bureau Dr. Gaithersburg, MD 20899

This paper overviews results from the Supplementary Comparison S2 organized by Consultative Committee on Photometry and Radiometry (CCPR). It details measurement and analysis of aperture areas involving 9 laboratories using 8 distinct artifacts, each having a different nominal value.

Individual laboratory estimates are compared with a comparison reference value (RV). The evaluation also includes the investigation of effects attributable to different measurement methods. Three artifacts were measured by several laboratories using both contact and non-contact methods. The pilot laboratory proposed the use of an optional second method to compare the bias of these methods (validation of methods) of the two different methods. More complex experimental designs would enable the study of other influential factors such as differing manufactures or measurement positioning schemes.

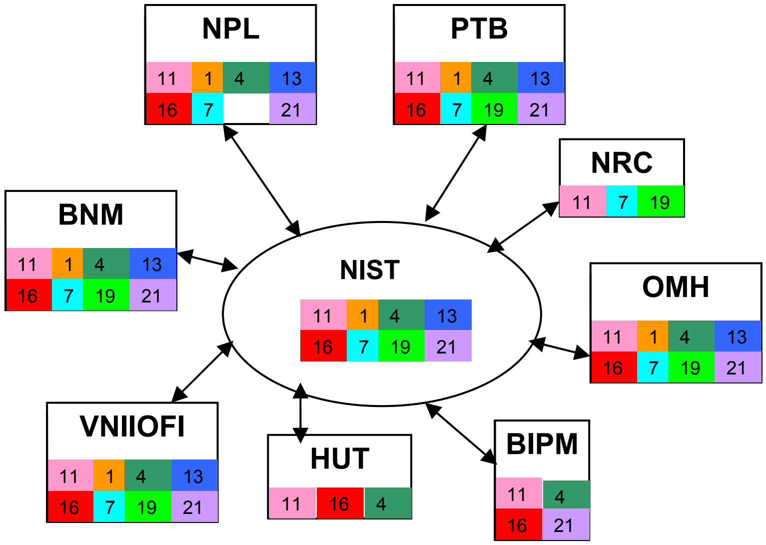
1. Introduction

The Consultative Committee on Photometry and Radiometry in 1999 decided to undertake an international comparison of the capabilities of member laboratories to measure the geometric areas of apertures used for radiometry. The accuracy of aperture areas affects the accuracy of radiometric and photometric standards and is vital to many radiometric and photometric measurements. The supplementary comparison detailed here was carried out within the framework of the Mutual Recognition Arrangement for national measurement standards and follows the *Guidelines for CIPM key comparisons*.

The National Institute of Standards and Technology, NIST, the national metrological institute (NMI) of the United States, was chosen as the Pilot Laboratory for this supplementary comparison S2. NIST was responsible for the fabrication, initial and periodic measurements during the comparison, and circulation of the transfer apertures used in the supplementary comparison.

2. Organization of the comparison

The Supplementary Comparison S2 was designed to determine laboratory differences among area measurements of apertures commonly used in radiometry and photometry. The apertures circulated for comparison are referred here to as transfer apertures. Seven apertures were fabricated at NIST while another was supplied by the Physikalische Technische Bundesanstalt (PTB), the NMI of Germany. The comparison was conducted in a star pattern (A-B-A-C-A-D-A…) where A represents the Pilot Laboratory.



Apt 1:	**NIST – PTB(5) – NIST - NPL(5) – NIST – BNM(5) – NIST – VNIIMOFI(5) – NIST – OMH(5) – NIST (15)**
Apt 4:	**NIST – PTB(5) – NIST - NPL(5) – NIST – BNM(5) – NIST – VNIIMOFI(5) – NIST– HUT(5) – NIST – BIPM(5) –NIST – OMH(5) – NIST (17)**
Apt 7:	**NIST –PTB(5) – NIST - NPL(5) – NIST – BNM (5) – NIST – VNIIMOFI (5) – NIST– OMH(5) – NIST (14)**
Apt 7c:	**NIST- PTB (10) – NIST- NPL (2) – NIST– OMH(5) – NIST– NRC (5) – NIST (5)**
Apt 11:	**NIST –PTB(5) – NIST - NPL(5) – NIST – BNM(5) – NIST – VNIIMOFI (5) – NIST -HUT (5) – NIST – BIPM (5) – NIST – OMH (5) – NIST(11)**
Apt 11c:	**NIST - PTB(10) – NIST - NPL (2) – NIST () –NRC (4) –NIST (5)**
Apt 13:	**NIST –PTB (5) – NIST - NPL (5) – NIST – BNM(5) – NIST – VNIIMOFI (5) – NIST – OMH (5) – NIST (11)**
Apt 16:	**NIST –PTB(5) – NIST - NPL(5) – NIST – BNM(5) – NIST – VNIIMOFI (5) –NIST – HUT(5) – NIST – BIPM (5)– NIST– OMH(5) – NIST(14)**
Apt 19:	**NIST–BNM(5) – NIST – VNIIMOFI(5) –NIST – OMH(5) – NIST (13)**
Apt 19c:	**NIST - PTB(5) – NIST - OMH (5) – NIST –NRC (4) –NIST (5)**
Apt 21:	**NIST –PTB(5) – NIST - NPL(5) – NIST – BNM(5) – NIST – VNIIMOFI (5) – NIST- BIPM(5) – NIST – OMH(5) – NIST(12)**

Figure 1. Comparison Scheme and timeline: The pilot laboratory, here NIST, is responsible for the statistical evaluation of the supplementary comparison.

The areas of the circulated apertures were measured at NIST prior to shipping to the participant laboratory, and again after they were received from the participant laboratory. A complementary set of control apertures kept at the pilot laboratory were also re-measured periodically during the comparison.

The comparison began in January 1999 and the last measurement at NIST was taken in November 2003. The time taken for each round (pilot-participant-pilot) was not uniform throughout the comparison for a variety of reasons, principally the unavailability of measuring instruments at the participant laboratory, or at NIST, or both. Due to instrumental limitations, not all participant laboratories measured all transfer apertures.

3. Physical principle of measurement

The measurement methods used by the participant laboratories were generally either (1) non contact methods where the edges are located via optical contrast and (2) by contact method where a mechanical probe touches the aperture edge to determine its location. The edge locations are then used to determine radius or diameter of the circular aperture. Two laboratories, HUT and BIPM, determined the effective area by measuring radiation throughput, rather than the geometric area which the other laboratories, either through contact or non-contact methods, used. Not every participant laboratory could measure all the transfer apertures due to certain instrumental limitations. Some laboratories were able to measure the apertures using both non-contact and contact methods.

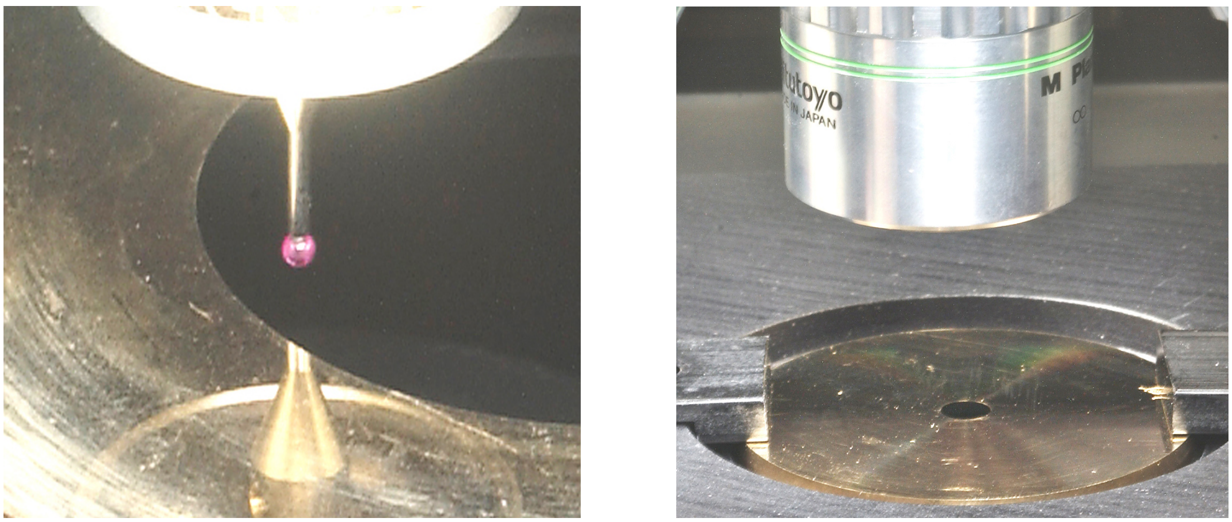


Figure 2: Two general methods of aperture area measurement. **Left:** contact probe locates the edge **Right:** Optical contrast is used to locate edges. A microscope is shown here as an example for edge detection

3.1 Description of apertures

The apertures used in the comparison, varied in size (small, medium, large), fabrication method (diamond turned vs. conventionally turned), material (copper, aluminum bronze, aluminum), and edge type (knife edged vs. cylindrical (with land)).

There were three sets of aperture containing eight apertures each, one set for circulation (transfer), one set kept at NIST as a control and one spare. The transfer set consisted of four diamond-turned apertures, three conventionally machine turned, ground, and polished apertures and one diamond turned aperture fabricated by PTB which has a small (10 µm) radius land. The apertures fabricated at NIST (APT 01 to 19) were all nominally 50 mm outer diameter by 6 mm thick. Two inside diameters (IDs), nominally 5.0 mm and 25.0 mm, two edge types made with two types of material were manufactured using each fabrication method. The edge is either sharp 'knife-edged' (only a few micrometers thick or thinner) or cylinder type 'land' (thickness of 50 µm -100 µm). The knife-edged apertures could only be measured using a non-contact technique while apertures with a land could be measured using either non-contact or contact method. The PTB aperture could also only be measured by the non-contact method.

Table 3.1 summarizes the apertures used for the comparison. There are three nominal sizes, three types of material, two edge types, and two fabrication methods. There are nine participating laboratories and two general methods of measurement, either non-contact or contact. Apt 01 to APT 06 are made from UBAC[1,2] copper-plated oxygen free high-conductivity (OFHC) copper inserts.

Table 3.1 List of apertures used for the S2 inter-comparison

ID number	Diameter [mm]	Material	Edge type	Fabrication
1-3	25	Cu	sharp	Diamond turned
4-6	5	Cu	sharp	Diamond turned
7-9	25	Al bronze	cylinder	Diamond turned
10-12	5	Al bronze	cylinder	Diamond turned
13-15	25	Al bronze	sharp	conventional
16-18	5	Al bronze	sharp	conventional
19-21	25	Al bronze	cylinder	conventional
P21-P23	20	Al	sharp	Diamond-turned

Table 3.2 lists the participating laboratories, the apertures measured and the measurement method. Not every laboratory could measure all the transfer apertures due to certain instrumental limitations. Some laboratories were able to measure the apertures using both methods. The laboratories that made contact measurements are affixed with a lower case c after the laboratory name to distinguish them from the non-contact methods.

Table 3.2 List of participating laboratories in the S2 comparison

Lab	Apertures Measured							
	1	4	7	10/11	13	16	19	P21
PTB	•	•	•	•	•	•		•
PTBc			•	•			•	
NPL	•	•	•	•	•	•	•	•
NPLc			•	•				
BNM	•	•	•	•	•	•	•	•
VNIIOFI	•	•	•	•	•	•	•	•
HUT		•		•		•		
BIPM		•		•		•		•
OMH	•	•		•	•	•	•	•
OMHc			•				•	
NISTc			•	•			•	
NRCc			•	•			•	
NIST	•	•	•	•	•	•	•	•

The first laboratory, PTB, reported edge damage to Apt 10 after their contact measurement. Even though the NIST measurement of Apt 10 upon return did not show significant difference

[1] UBAC is a registered trademark for UDYLITE Bright Acid Copper by the Enthone-OMI, Inc. subsidiary of ASARCO. Their address is 21441 HOOVER Road, Warren, Michigan 48089 USA. The phone number is 313-497-9100.

[2] Certain commercial materials and equipment are identified in this paper to specify adequately the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply necessarily that the product is the best available for the purpose.

from prior to shipping to PTB, Apt 10 was replaced by Apt 11 in the subsequent measurements. Hence, only PTB measured Apt 10, while the rest of the participants measured Apt 11.

4. Measurement results

Each laboratory measured each aperture five times according to the laboratory's measurement protocol. The reported areas (*A*) are the means of the replicate measurements, expressed in mm^2 and corrected for thermal expansion if necessary to 20 °C. Uncertainties of area measurement ($u(A)$) are expressed in mm^2 in the form of a combined standard uncertainty.

Table 4.1 Supplementary Comparison Reference Values (abr. SCRV) and associated standard uncertainty for each artifact based on the maximum likelihood estimator.

Artifact	SCRV	u (SCRV)
1	550.720	0.042
4	34.019	0.005
7	505.126	0.028
11	23.295	0.005
13	525.676	0.113
16	21.290	0.014
19	518.164	0.053
21	314.097	0.026

As inputs to the comparison, each laboratory provided the pilot lab with the mean value of several measurements taken over a short time period, accompanied by laboratory specific estimates of type A and type B uncertainties. The pilot laboratory was responsible for the ongoing stability checking of artifacts. The artifacts were deemed stable for the lifetime of the study. The evaluation assumes all artifacts to be different with distinct true values and therefore all artifacts were analyzed as separate items.

Table 4.2 lists the main uncertainty components that each laboratory had to supply. The uncertainty budget from a participating laboratory includes several different components contributing to the overall uncertainty.

Table 4.2 Components of the uncertainty budget. The ratio between the largest and smallest reported combined uncertainties across artifacts went from 4 to 22.

Source of uncertainty	Type
Measurement, random	A
Measuring chain as interferometer, cosine, Abbe offset	B
Image (coherence, magnification, pixel size, stage/camera angle, focus)	B
Temperature correction to measurement conditions	B
Artifact geometry (edge quality)	B

Comparison among laboratories is based on the pair-wise degrees of equivalence (differences between estimates of aperture area between two laboratories) for each artifact separately (not provided here, but available on Database B of the MRA BIPM web-side). Individual laboratory mean estimates were also compared with a supplementary comparison reference value (SCRV).

Each aperture was assigned a reference value and associated uncertainty based on data from all the laboratories which participated in the measurement of the aperture. Reference values take into account both variation internal to each laboratory as well as across laboratories. Degrees of equivalence were computed, but are not displayed here.[1]

The reference values for this comparison were chosen as the maximum likelihood estimators.[2]

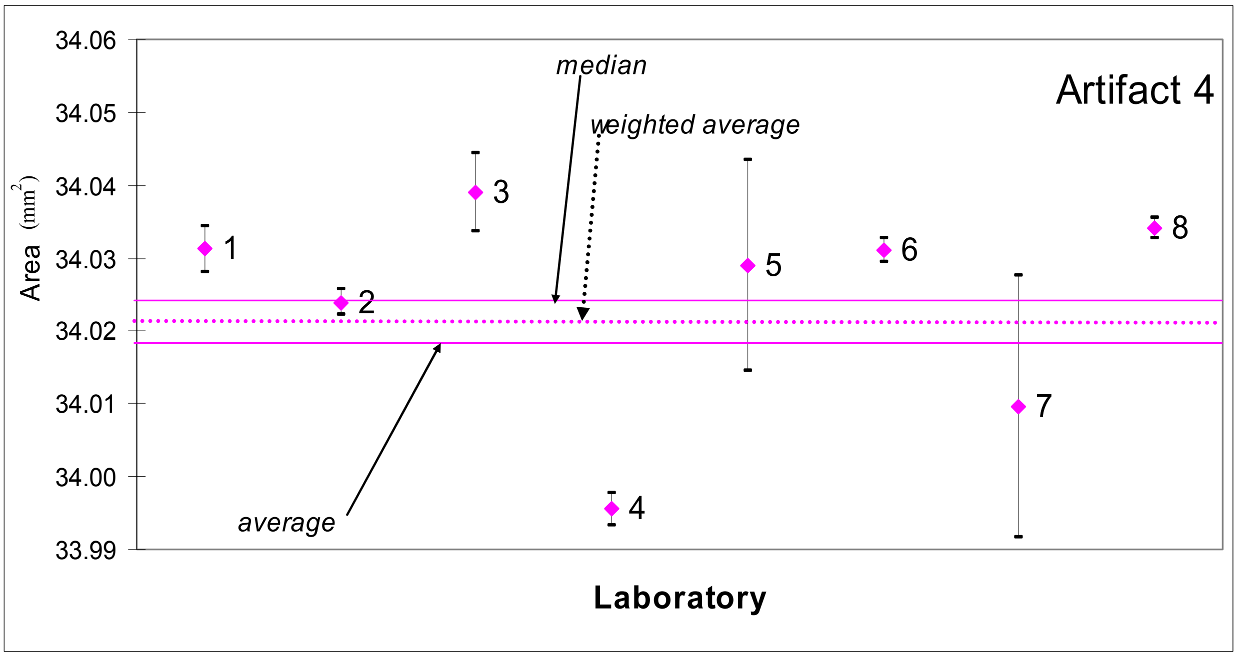


Figure 3. A graph showing the results from measuring artifact 4 across participating labs. Diamonds mark the mean values and uncertainties are represented by vertical lines. Purple horizontal lines denote average, weighted average (with weights equal to laboratories' uncertainties) and median.

4.1 Outliers

Outliers were tested for graphically using Mandel's h and k statistics. K-statistics are ratio's of an artifact's within laboratory standard deviation divided by its pooled standard deviations across laboratories. H-statistics divide an artifact's laboratory specific average deviation from across-laboratory average by its standard deviation of all the laboratories specific numbers. Both k- and h-values can be examined across artifacts and across laboratories to ascertain the origin of extreme values. Outlying values are determined by comparison to tabulated critical values.[3] Participating laboratories were notified of potential outliers. For the final analysis, no points were discarded.

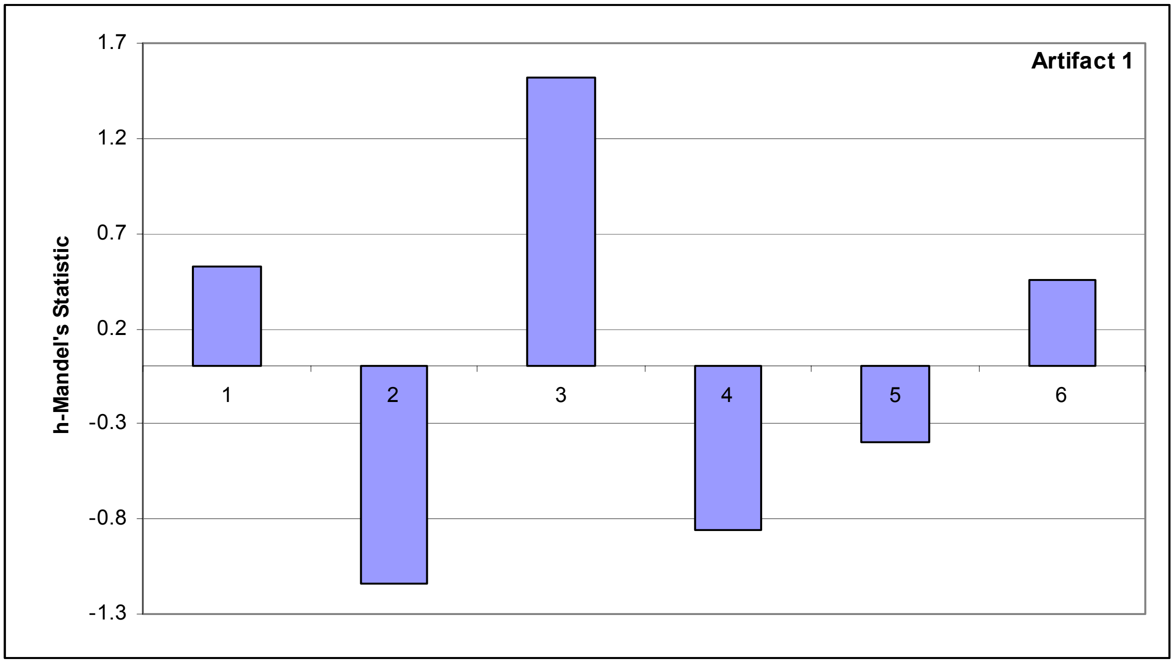


Figure 4. Outlier check for artifact 1, Mandel's h-statistic
(critical value for six laboratories is 1.87).

4.2 Stability check

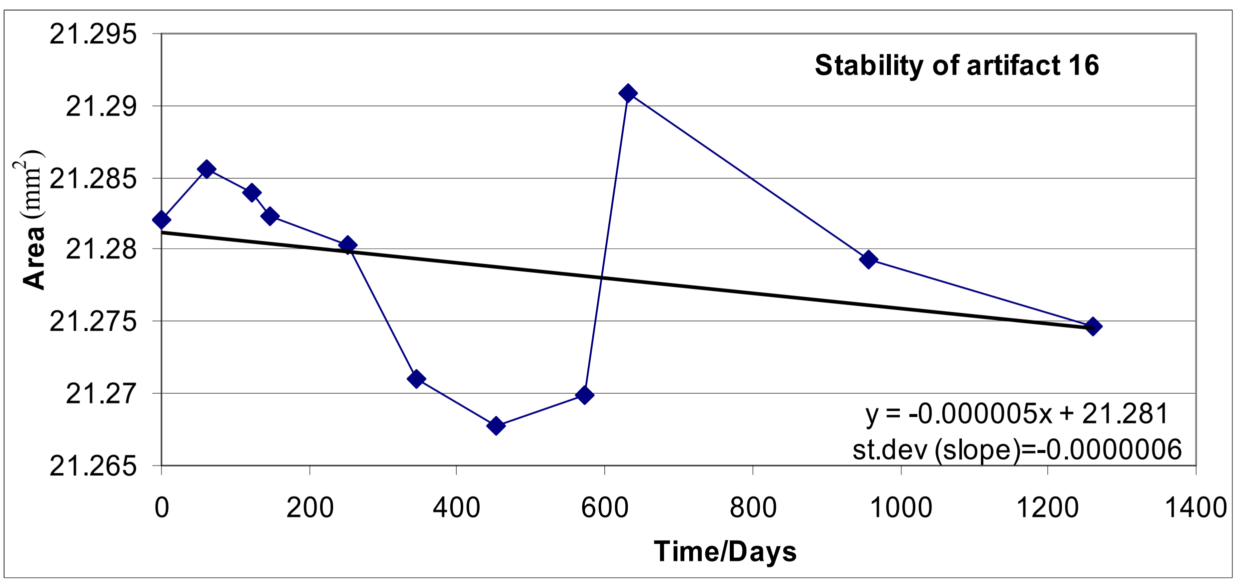


Figure 5a. Stability checks are an inevitable part of a comparison evaluation. This graph shows measurements taken by NIST on the traveling artifact number 16 at different times of the comparison, precisely before and after each laboratory as well as before and after the entire comparison. In conjunction with graphical methods, quantitative methods were used to analyze the negligibility of aperture's instability.

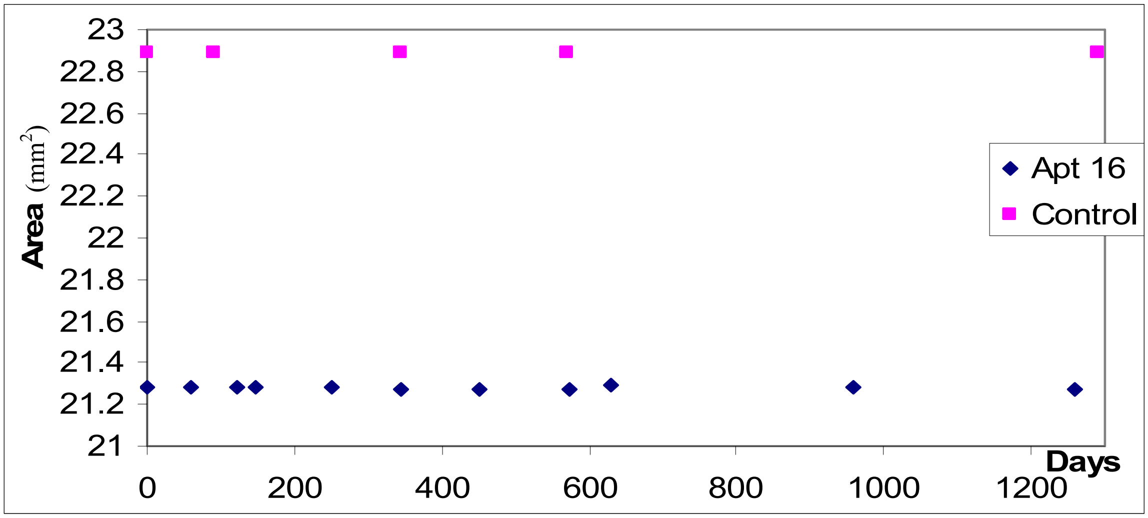


Figure 5b. For unstable artifacts, the drift has to be quantified and measurements corrected. This figure details the behavior of artifact 16 and its counterpart aperture 17, also called control, during the entire comparison period.

5. Different measurement methods

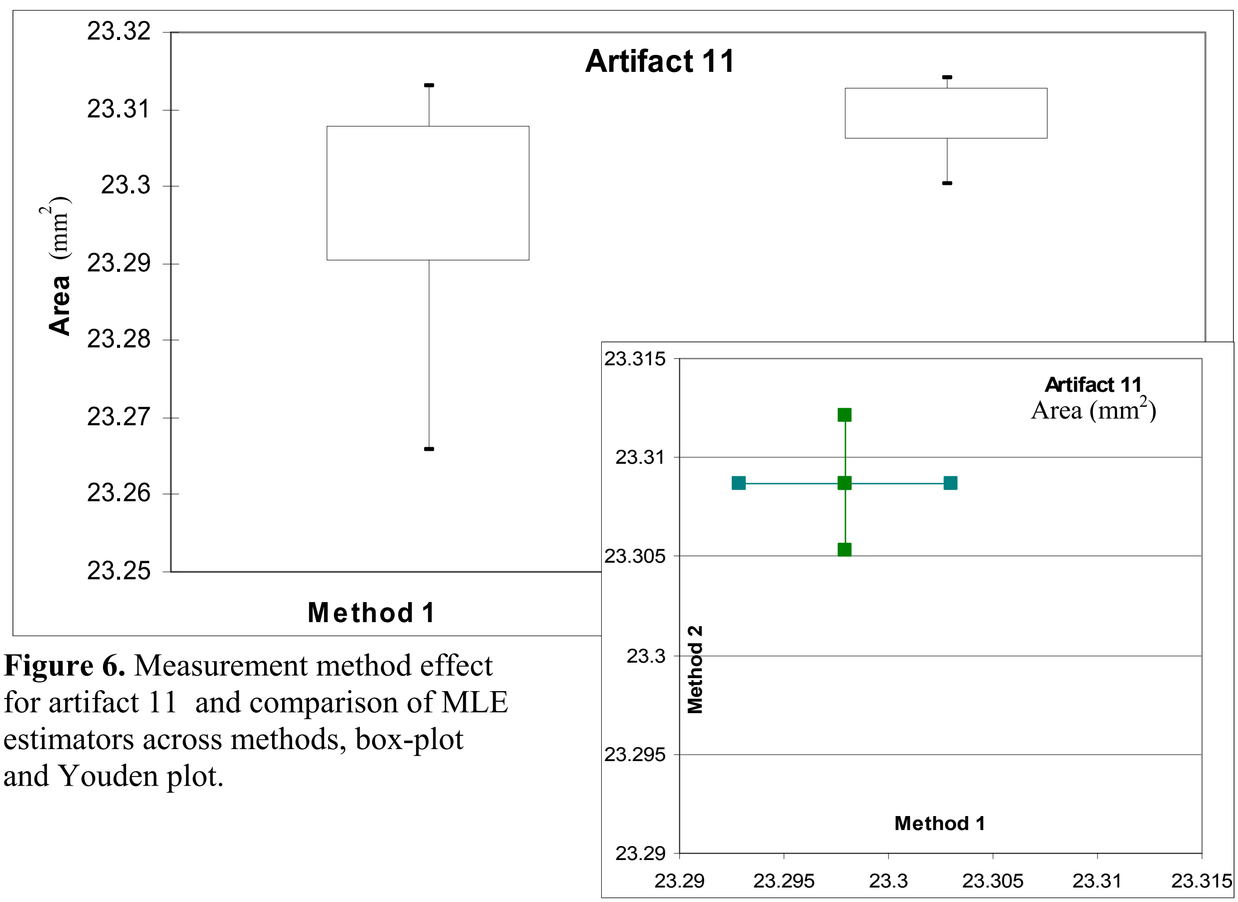
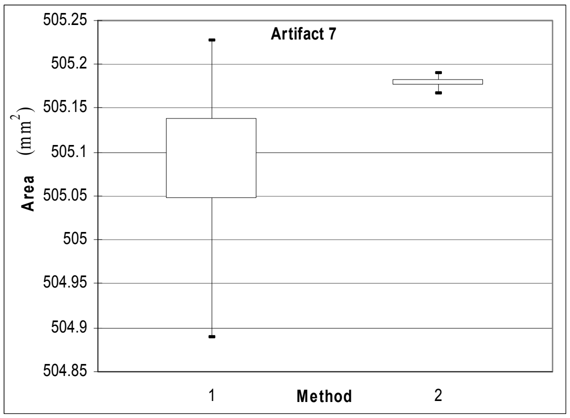


Figure 6. Measurement method effect for artifact 11 and comparison of MLE estimators across methods, box-plot and Youden plot.

The comparison also included the investigation of effects attributable to different measurement methods. ANOVA, analytical and graphical (Figure 6) was done on three artifacts that were measured by several laboratories using both contact and non-contact methods.



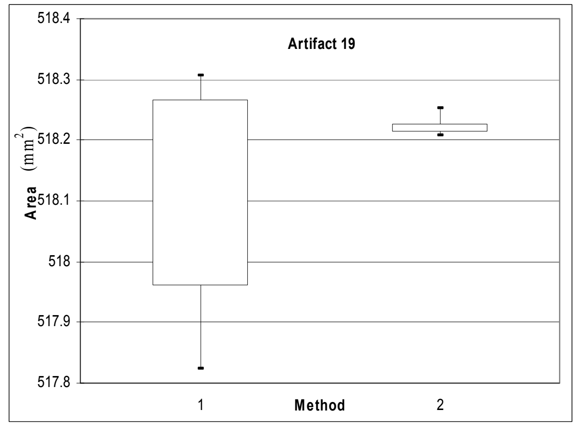


Figure 7. Different methods comparison across several artifacts. Graphical ANOVA is used to assess for bias and/or variance difference between the contact (here designated as Method 1) and non-contact (Method 2) methods. For all three artifacts the mean of method 2 is slightly above method 1 and method 2 has a smaller variability in comparison to method 1.

This comparison has attempted to accommodate the possibility of separating the influence of measurement methods used. As one of the measurement methods used within this supplementary comparison is a well established method (standardized measurement method) this enables the validation of newer, less established methods.

6. Conclusions

This supplementary comparison is one of the first in the area of radiometry. At the conclusion of the study the pilot laboratory proposed the use of a second method, in order to compare the performance of these methods. This could constitute an important step towards more complex and meaningful experimental designs in key comparisons in the future. Other influential factors could also be studied, such as differing modes of manufacture or measurement positioning schemes.

7. References

[1] BIPM Key Comparison Database: http://kcdb.bipm.org/AppendixB
[2] Rukhin, A. L., Vangel,M. G.: Estimation of a Common mean and Weighted Means Statistics, J. of the Ame. Statistical Association, March, 1998, vol. 93, no 441.
[3] Mandel, John: Evaluation and Control of Measurements, New York: Marcel Dekker, 1991.