International Comparison of Surface Roughness and Step Height (Depth) Standards, SIML-S2 (SIM 4.8)

K. Doytchinov, Institute for National Measurement Standards, National Research Council (NRC), Ottawa, Canada
F. Kornblit, Instituto Nacional de Technologia Industrial (INTI), Buenos Aires, Argentina
C. Colin Castellanos, Centro Nacional de Metrologia (CENAM), Queretaro, Mexico
J. C. V. Oliveira, National Institute of Metrology Standardization and Industrial Quality (INMETRO), Duque de Cacias, Brazil
T.B. Renegar and T.V. Vorburger, National Institute of Standards and Technology (NIST), Gaithersburg, USA

Abstract: Calibration services of five countries from the Sistema Interamericano de Metrología (SIM) region are compared through measurements of surface roughness and step height standards. A surface roughness standard with a nominal roughness average (*Ra)* value of 0.2 µm, a surface roughness standard with a nominal *Ra* value of 3 µm and a nominal spatial wavelength of 99 µm, and three step height standards with nominal values of 2.55 µm, 0.38 µm and 0.03 µm are compared. Special attention is paid to the influence of the long wavelength cut-off ratio of the measurements. Results are reported for *Ra*, maximum height of profile *Rz*, mean width of profile elements *RSm*, and step height *d*, depending on the sample measured. The initial reported results show that the laboratories agree on all of the measurements within their stated and published uncertainties. Observations are then discussed about the definition of *Rz*, the effect of instrument noise on *Rz*, the different step height parameters *d* and *Pt*, differences between the laboratories in reporting Type A statistical uncertainties, the method for calculating the uncertainty of the reference value, and the importance of accounting for correlations between the reference value and individual lab values when calculating the degrees of equivalence. After corrections and reanalysis the laboratories still agree well considering their stated uncertainties.

1. Introduction

Under an international Mutual Recognition Arrangement (MRA) [1] , 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 Committee International des Poids et Mesures (CIPM) and regional comparisons by the regional metrology organizations in collaboration with the Consultative Committees. Five national measurement institutes (NMIs) from the SIM (Sistema Interamericano de Metrología) region have carried out a surface roughness regional comparison, with the National Research Council (NRC) as the pilot laboratory. Participating laboratories as shown above were: NRC, NIST, INMETRO, INTI, and CENAM. Results of this international comparison are intended for inclusion in the agreement for establishing the metrological equivalence between the NMIs.

The measurements were carried out according to the schedule below:

Laboratory	Country	Date
NRC	Canada	October 2000
NIST	USA	December 2000
INMETRO	Brazil	February 2001
INTI	Argentina	March 2001
CENAM	Mexico	April 2001

Participants used the equipment and procedures normally used when calibrating customer artifacts.

2. Description of the Standards

The following physical standards were used in the study:

- Step height specimen, Model 112/964, serial number 18235 manufactured by Taylor Hobson Pneumo Ltd.* (UK) with three patches with 2.55 µm, 0.38 µm and 0.030 µm nominal step height, respectively (shown in Fig. 1).

Figure 1: Step height specimen, serial number 18235

- Surface roughness specimen, serial number Special 3144, having nominal 0.20 µm roughness average (*Ra*), manufactured by Taylor Hobson Pneumo Ltd. (UK), shown in Fig. 2. Both standards belong to NRC and were supplied in special cases as shown in Fig. 3 below.

Figure 2: Surface roughness specimen serial number "Special 3144"

* Certain equipment and materials are described in this article in order to specify clearly the experimental procedures used. In no case does this imply a recommendation by NRC, INTI, CENAM, INMETRO, or NIST that the equipment or materials are necessarily the best available for the purpose.

Figure 3: NRC step height and roughness specimens in their cases

- NIST Standard Reference Material (SRM) 2073a Prototype, Serial No. 1292 (Fig. 4), calibrated for *Ra* and surface spatial wavelength D, a parameter nearly identical to the standard parameter *RSm,* the mean width of profile elements. The standard is intended for calibration checks of stylus instruments used to measure both *Ra* and *RSm*. The SRM is a steel block of nominal Knoop hardness 500, which has been coated by the electroless nickel deposition process. A sinusoidal roughness profile has been machined onto the top surface of the specimen. The surface profile is highly sinusoidal as shown in Fig. 5. This standard belongs to NIST.

Figure 4: Standard Reference Material (SRM) 2073a Prototype, Serial No. 1292

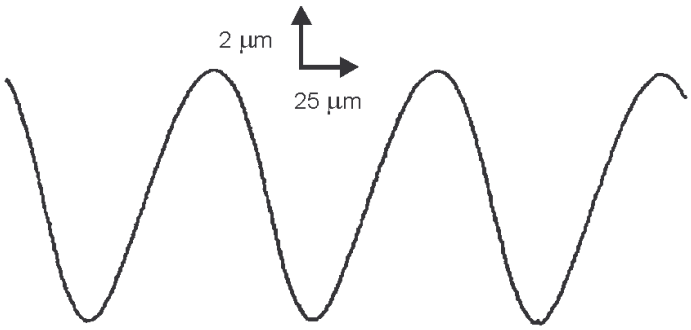


Figure 5: Representative surface profile trace of SRM 2073a. Dimensions are approximate.

3. Measuring equipment.

The instruments used in each laboratory to measure the surfaces are described below.

Organization	Description	Traceability
NRC	Form Talysurf 120L with a laser pick-up. Nominal stylus tip radius – 2 µm	Calibration ball, 22 mm radius. Special laser interferometer set-up for generating step heights. Both traceable to SI unit of length through NRC.
NIST	Roughness: Federal Surfanalyzer 2000 with LVDT stylus pickup and nominal 2 µm tip. Laser interferometer system for calibration of X-displacement. Step height: Taylor Hobson Talystep with LVDT stylus pickup and nominal 2 µm tip.	Z-Calibration: Both instruments calibrated by measurement of an interferometrically calibrated step height standard, of similar height to the Z-scale being used. Step standards traceable to the SI unit of length. X-Calibration (Federal 2000): Displacement measured directly with HeNe Laser Interferometer.
INMETRO	PERTHOMETER "S8P" (MAHR/PERTHEN) with the "PRK" reference unit and a "pick-up" type "RFHTB-50" with a nominal stylus radius of 5 µm.	Standards traceable to SI unit of length through PTB
INTI	Federal 5000 Surfanalyzer, different stylus tips with nominal radii between 2 µm and 5 µm.	Height standard, H=5,76 µm, traceable to SI unit of length through PTB
CENAM	Form Talysurf 120 mm series 2 with inductive pick-up. Nominal stylus tip radius - 2 µm.	Calibration ball, 12,5 mm radius traceable to SI unit of length through CENAM

4. Measurement Conditions

Although the participants were asked to use the equipment and procedures normally used when calibrating clients' standards, it was agreed that the cut-off filter for the *Ra* measurements should be the ISO Gaussian type [2] for which the specifications are the long-wavelength cut-off λc, the short-wavelength cut-off λs, and the resulting cut-off ratio between them $\lambda c/\lambda s$. The participants agreed to use a λc of 0.8 mm for which the default λs value is 2.5 µm, resulting in a cut-off ratio of 300:1, as specified in the ISO 3274-1996 standard [3]. However, many companies in industry use instruments with a roughness cut-off ratio of 100:1. These companies are using the services of the calibration labs to get their standards calibrated. If a calibration standard is calibrated using a cut-off ratio of 300:1 and then is used to calibrate an instrument with a 100:1 capability, there is a danger of a bias being introduced in the traceability chain because roughness measured with a 300:1 cut-off ratio should be larger than roughness measured with a 100:1 cut-off ratio. That is because roughness is basically a noise property and its measured value depends on the bandwidth of the measuring instrument. To estimate the magnitude of the problem, the pilot laboratory specifically asked the participants to evaluate the same raw roughness data using both 100:1 and 300:1 roughness cut-off ratio. The goal was to show if the differences in the results would be significant.

Measurement of *Ra* and *Rz* (ISO) on the Roughness Specimen, Serial number SPECIAL 3144

NRC, INTI, CENAM, and INMETRO performed 20 measurements on this roughness patch according to the pattern of Fig. 6. NIST followed its customary calibration procedure of nine measurements as shown schematically in Fig. 7. All labs used a sufficient traversing length so that after filtering, the evaluation length amounted to five cut-offs. The cut-off filter was the ISO Gaussian type, discussed in ISO Standard 11562[2] and ASME B46.1[4] with a cut-off length of 0.8 mm. CENAM and NRC evaluated their raw data using both 100:1 and 300:1 cut-off ratios. INTI and NIST evaluated the data using the 300:1 cut-off ratio only. The measuring equipment of INMETRO did not have a λs filter built in. NRC, CENAM, NIST, and INTI used styli with a nominal tip radius of 2 µm while INMETRO used a stylus with a nominal tip radius of 5 µm. Because the tip radius acts as a mechanical filter that attenuates small wavelengths of the surface, the participants decided to put the INMETRO results in the 100:1 cut-off ratio category.

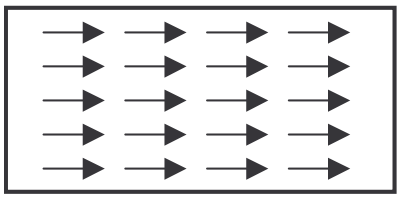


Figure 6: Schematic diagram of 20 positions used by NRC, INTI, INMETRO, and CENAM for measurement of *Ra* and *Rz* (ISO) on the Roughness Specimen, Serial number SPECIAL 3144

Both the roughness average *Ra* and the maximum height of profile *Rz* were calculated from the profile data. The calculation of *Rz* can be ambiguous. *Rz* is presently defined in ISO Standard 4287 (1997)[5] as "the sum of the largest profile peak height *Zp* and the largest profile valley depth *Zv* within a sampling length." The method of calculation is illustrated in Fig. 8. When this *Rz* (ISO-1997) is evaluated by averaging over five sampling lengths as was the case here, the result is identical to the earlier parameter *Rz* (DIN)[6]. Therefore, software for *Rz*(DIN) may be used to calculate *Rz*(ISO-1997). The value for both is larger than that derived from a previous parameter *Rz*(ISO 4287-1984), now withdrawn, which was defined as the average of the heights of the five highest peaks and depths of the five lowest valleys within the sampling length.

Measurement of *Ra* and *RSm* on the 3 µm Roughness Specimen SRM 2073a

The scans were performed according to Fig. 7. A total of nine measurements were performed on this roughness patch using a sufficient traversing length so that after filtering, the evaluation length amounts to 5 cut-offs. The cut-off filter was the ISO Gaussian type, discussed in ISO Standard 11562[2] and ASME B46.1[4] with a cut-off length of 0.8 mm. CENAM and NRC evaluated the same raw data using both 100:1 and 300:1 cut-off ratios. INTI and NIST evaluated the data using a 300:1 cut-off ratio only. The styli were the same as those used for the SPECIAL 3144 specimen. It is known that different manufacturers have different approaches when calculating the *RSm* values. Some algorithms count all of the peaks while others count only the peaks above a pre-defined threshold. The participants were advised to reduce these thresholds to a minimum if the software allowed.

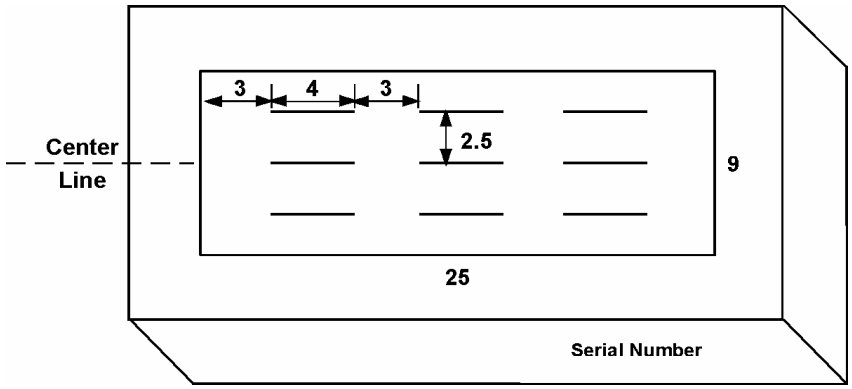


Figure 7: Measurement positions for a prototype of NIST roughness specimen SRM 2073a. All dimensions are in mm. A similar pattern with smaller dimensions was used by NIST for the SPECIAL 3144 specimen.

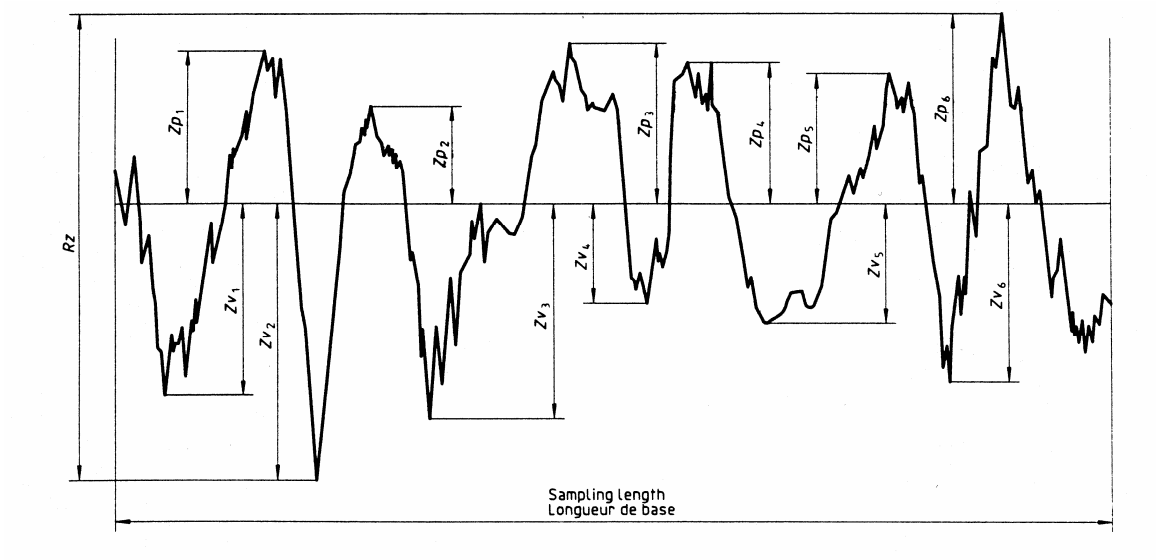


Figure 8: Schematic diagram, reprinted from ISO 4287, Ref. 5, of the method of calculation of the maximum height of profile parameter *Rz*

Avoidance of flaws

SRM 2073a has been manufactured with very few visible flaws. The participants were advised to avoid measuring paths on SRM 2073a that intersect flaws visible to the naked eye. In addition and more specifically, even single flaws in surface profiles, such as those shown in Fig. 9, should be avoided because they can change the measured *RSm* value significantly (by 2.5 % if the surface profile contains 40 peaks and valleys). That is because the customary algorithms for calculating *RSm* rely on counting the crossings of the mean line (shown dashed here) by the surface profile. Single flaws such as these do not change the *Ra* values as significantly as they can change *RSm*.

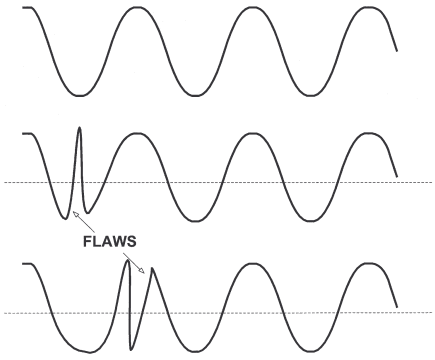


Figure 9: Avoidance of flaws in surface measurements

Measurement of the step heights on the step height specimen, serial number 18235

The step height samples contain three grooves. Only the depth of the middle one was measured. The scans were performed in a direction perpendicular to the groove. The scanned data were to be "levelled" using a least squares (LS) best-fit line on the two portions marked A and B in Fig. 10 below. The step height was defined in accordance with ISO standard 5436-1[7] as the average depth (*d*) from the best-fitted LS line to the middle third of the bottom of the groove (C in Fig. 10). A total of 9 scans over the central 2 mm of each step were performed. NRC, INTI, CENAM, and NIST calculated step heights in this way. However, INMETRO calculated the peak-valley height *Pt*[5] from the measured step profiles.

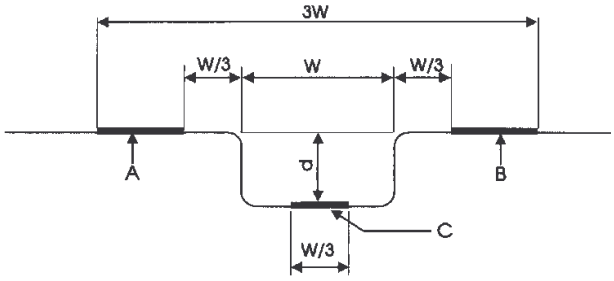


Figure 10: Assessment of the step height samples (Ref. 7)

5. Initial Results and Observations

Results for *Ra*, *Rz*, *RSm*, *d*, and *Pt* obtained by each laboratory are shown in Figs. 11-14 along with their estimated uncertainties (k=2) calculated according to the GUM [8]. The sources of uncertainty include both Type A and Type B components. The Type A components were calculated from statistical methods. The Type B components were calculated by non-statistical methods such as estimates based on physical models.

The principal source of type A uncertainty is variation in results measured at different positions across the specimen.

Other components of uncertainty may be calculated by Type A or Type B methods depending on the data or models used to derive the estimates. For the z-parameters, *Ra*, *Rz*, and *d*, major sources of uncertainty include

- The form error and surface finish of a master artifact, such as a step height or sphere, used to calibrate the z-scale of the stylus instrument,
- Instrumental variations during measurements taken to calibrate the z-scale of the instrument. These can arise from instrument noise occurring during the measurement of a master calibration artifact and lead to variations in the z-scale of the instrument from day-to-day,
- Uncertainty in the calibrated value, such as step height or radius, of the master artifact,
- Uncertainty in the horizontal resolution of the instrument, due primarily to uncertainty in the short-wavelength cutoff,
- Instrumental noise, which tends to increase measured roughness values.
- Uncertainty due to nonlinearity of the instrument transducer,
- Uncertainty (for measurement of *d*) due to the mechanical or electrical ringing that occurs when the stylus encounters the trailing edge of a step.

In addition to the Type A variation of measured values, Type B sources of uncertainty for *RSm* include

- Uncertainty in the calibration of the x-axis displacement.
- Variation due to the choice of algorithm to calculate the spacings of profile irregularities.

According to a protocol agreed upon by the five laboratories before the measurements began, the reference value (x_{ref}) for each measured quantity was the arithmetic mean of the values x_{i} from the individual NMIs, given by

$$x_{\mathrm{ref}} = \overline{x} = (1/n)\sum_i x_i , \tag{1}$$

and the standard uncertainty $u(x_{\mathrm{ref}})$ of the reference value was taken as the standard deviation of the mean of the set of reported x_{i} values, that is,

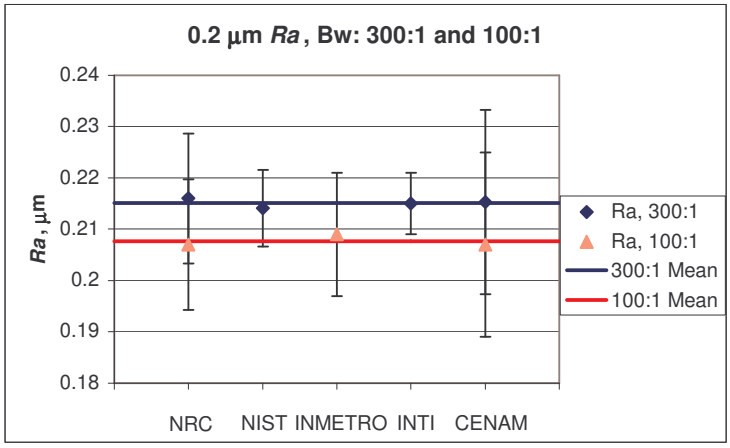
$$u\left(x_{\mathrm{ref}}\right) = \sqrt{\frac{\sum \left(x_i - x_{\mathrm{ref}}\right)^2}{n(n-1)}}. \tag{2}$$

In order to evaluate the degree of equivalence between each NMI and the mean x_{ref}, the E_n statistic was used. It was calculated as

$$E_n = \frac{x_i - x_{\mathrm{ref}}}{2 \cdot \sqrt{u_i^{\,2} + u^2(x_{\mathrm{ref}})}}, \tag{3}$$

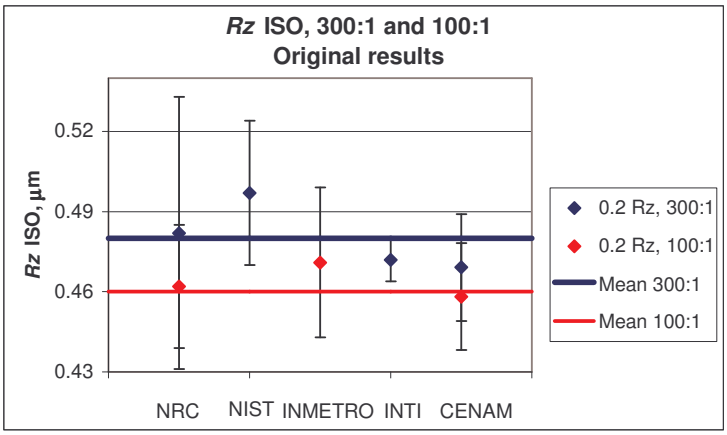
where u_{i} is the standard uncertainty for the measured value x_{i}. The formula of Eqn. 3 for degree of equivalence does not account for the correlation between x_{ref} and x_{i}. This correlation is especially significant because of the small number of contributions, four at most, for each average. Therefore, although permitted by the MRA, the above formula for equivalence is not strictly correct. This subject is addressed later in Section 6, when we describe the reanalysis of the results.

Another point is that the method of Eqn. 2 for calculating the reference value uncertainty emphasizes the consistency between the results of the different laboratories, rather than the overall uncertainty of the results. The same bias could be present in all of the results for a particular parameter, and the uncertainty of the reference value for that parameter as calculated here would not account for that bias. Such Type B sources of bias should be accounted for in the quoted uncertainties u_{i} of the individual laboratory values x_{i}. These sources of bias include stylus tip radius and uncertainties in the filter characteristics of the instruments. From Eqn. 1, because the input quantities for the reference value x_{ref} are the individual laboratory values x_{i}, the uncertainty of the reference value should be calculated from the uncertainties of the individual laboratory values. We take this approach for the re-analysis shown later below.



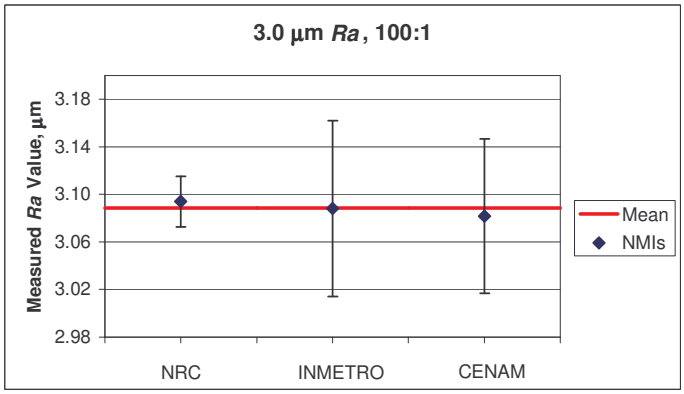
NMI	CR, 300:1			CR, 100:1		
	Ra, µm	*U*, µm	*En*	*Ra*, µm	*U*, µm	*En*
NRC	0.216	0.013	0.07	0.207	0.013	-0.05
NIST	0.2141	0.0075	-0.13	N/A	N/A	N/A
INMETRO	N/A	N/A	N/A	0.209	0.012	0.11
INTI	0.215	0.006	-0.02	N/A	N/A	N/A
CENAM	0.2153	0.018	0.01	0.2070	0.018	-0.04
Std. Dev.	0.0008			0.0012		
Ref. Value	0.2151			0.2077		
$U(X_{ref})$, $k=2$	0.0008			0.0013		

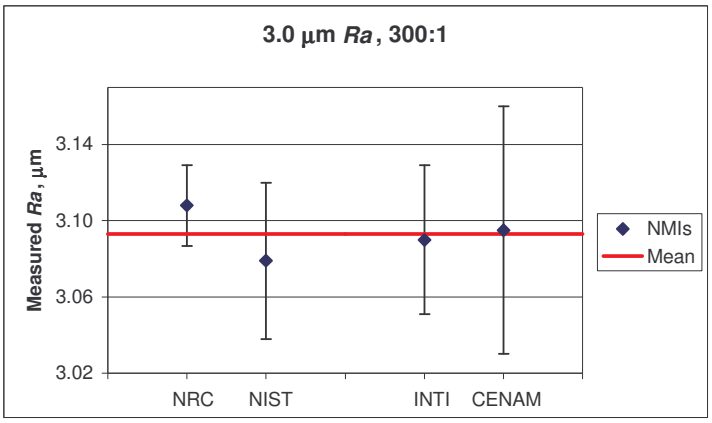
Figure 11: Initial results for roughness specimen "SPECIAL 3144" – *Ra* with band ratios of 300:1 and 100:1.

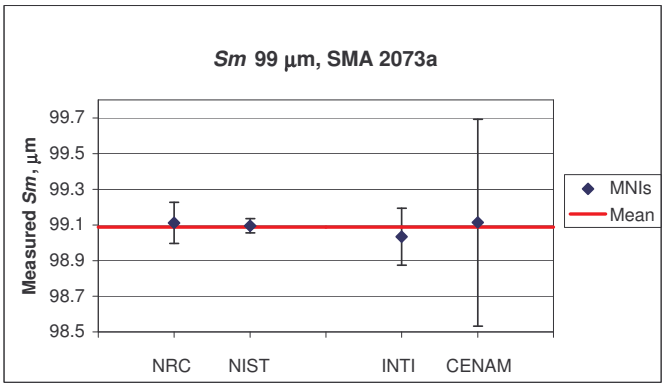


NMI	CR, 300:1			CR, 100:1		
	Rz, µm	*U*, µm	*En*, 300:1	*Rz*, µm	*U*, µm	*En*, 100:1
NRC	0.482	0.051	0.04	0.462	0.023	-0.07
NIST	0.497	0.027	0.57	N/A	N/A	N/A
INMETRO	N/A	N/A	N/A	0.471	0.028	0.25
INTI	0.472	0.008	-0.54	N/A	N/A	N/A
CENAM	0.4691	0.020	-0.46	0.4582	0.020	-0.26
Std. Dev.	0.013			0.007		
Ref. Value	0.480			0.464		
$U(X_{ref})$, $k=2$	0.013			0.008		

Figure 12: Initial results for roughness specimen "SPECIAL 3144" – *Rz* with cut-off ratios (CR) of 300:1 and 100:1.

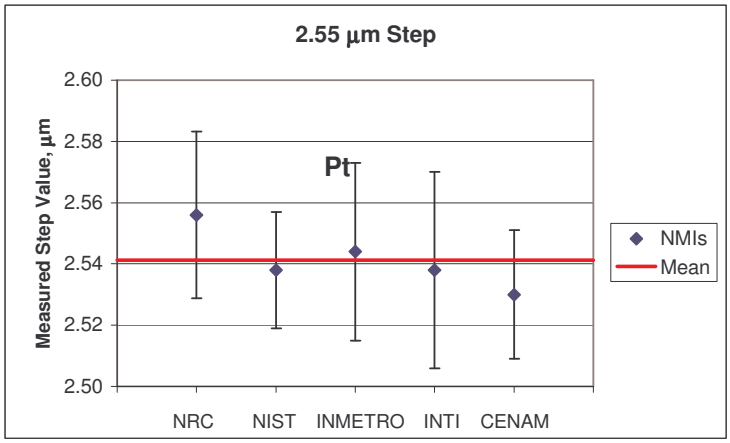


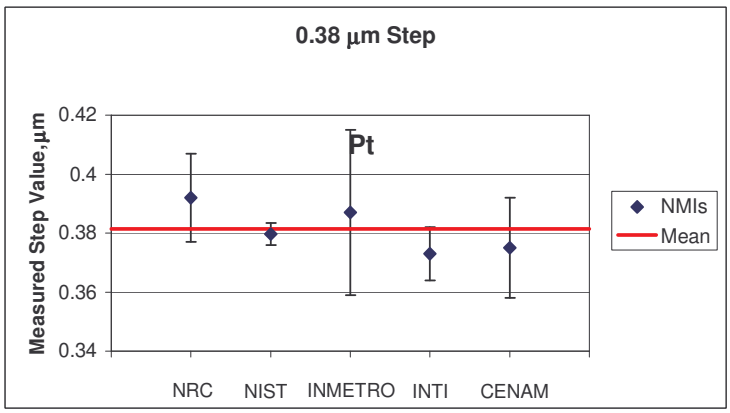


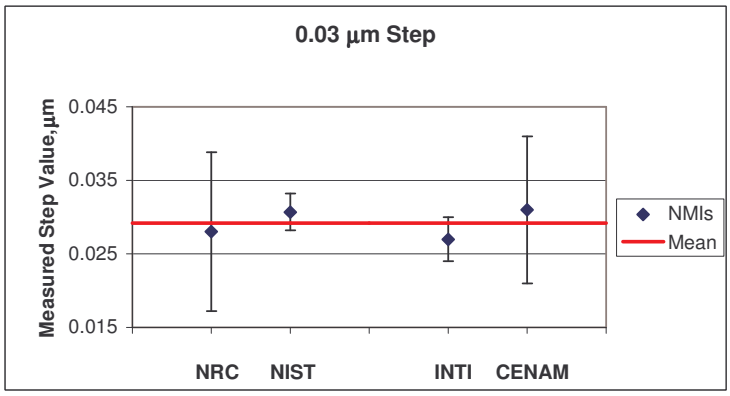


NMI	*Ra*						*RSm*		
	CR, 300:1			CR, 100:1			CR, 300:1		
	Ra, µm	*U*, µm	*En*	*Ra*, µm	*U*, µm	*En*	*RSm*, µm	*U*, µm	*En*
NRC	3.108	0.021	0.61	3.094	0.021	0.27	99.112	0.115	0.20
NIST	3.079	0.041	-0.33	N/A	N/A		99.094	0.040	0.11
INMETRO	N/A	N/A	N/A	3.088	0.074	0.00	N/A	N/A	N/A
INTI	3.090	0.039	-0.07	N/A	N/A		99.033	0.16	-0.33
CENAM	3.0951	0.065	0.03	3.0817	0.065	-0.09	99.113	0.58	0.04
Std. Dev.	0.012			0.006			0.038		
Ref. Value	3.093			3.088			99.088		
$U(X_{ref})$, $k=2$	0.012			0.007			0.038		

Figure 13: Initial results for roughness specimen SRM 2073a - *Ra* and *RSm*







NMI	*d* = 2.55 µm			*d* = 0.38 µm			*d* = 0.03 µm		
	d, µm	***U***, µm	***En***	***d***, µm	***U***, µm	***En***	***d***, µm	***U***, µm	***En***
NRC	2.556	0.027	0.52	0.392	0.015	0.64	0.028	0.011	-0.11
NIST	2.538	0.019	-0.15	0.3797	0.0038	-0.20	0.0307	0.0025	0.48
INMETRO (*Pt*)	2.544	0.029	0.09	0.387	0.028	0.20	N/A	N/A	N/A
INTI	2.538	0.032	-0.10	0.373	0.009	-0.72	0.027	0.003	-0.61
CENAM	2.530	0.021	-0.49	0.375	0.017	-0.34	0.031	0.010	0.18
Std. Dev.	0.010			0.008			0.0020		
Ref. Value	2.541			0.381			0.0292		
$U(X_{ref})$, $k=2$	0.009			0.007			0.0020		

Figure 14: Initial results for step height specimen, Serial Number 18235

Overall, the initial agreement between the five laboratories was very good, as indicated by the generally small values of $|E_n|$. A satisfactory degree of agreement is indicated by $|E_n| < 1$. The average $|E_n|$ value for all results here was 0.24 with a standard deviation of 0.21. The maximum $|E_n|$ value for all measurements was 0.72. This suggests that at least some of the uncertainty budgets may have been on the conservative side. For individual parameters we observed the following.

The *Ra* measurements of the specimen SPECIAL 3144 (See Fig.11) have excellent agreement. The parameter $|E_n|$(max) is equal to 0.13. For the *Rz* measurement of the same specimen, the results (Fig. 12) are somewhat more varied. Explanations for this behavior are explored below.

For the measurement of *Ra* of the SRM 2073a Sinusoidal Prototype (See Fig. 13), the results are in very good agreement with $|E_n|$(max) = 0.61.

The step height measurements (See Fig. 14) are also in good agreement with $|E_n|$(max) = 0.72.

This comparison was also used to study the effect of roughness cut-off ratio (bandwidth)[2]. Two of the laboratories (NRC and CENAM) measured *Ra* and *Rz* with a cut-off ratio ($\lambda c/\lambda s$) of both 300:1 and 100:1. The differences between results for both *Ra* and *Rz* at the two cut-off ratios are significant for the rectangular profile specimen with nominal *Ra* of 0.2 µm, which has a significant fraction of short spatial wavelengths on its surface (See Fig. 11 and 12). Clearly, the shortest spatial wavelength components of this surface are more attenuated by the filter with cut-off ratio of 100:1, whose short cut-off value is 8 µm, than by the filter with 300:1 cut-off ratio. For the sinusoidal specimen SRM 2073a, which has a fundamental spatial wavelength of 100 µm and only a very small fraction of spatial wavelengths below 8 µm, the differences in results using the different bandwidth instruments are not significant (See Fig. 13).

For the measurement of *RSm*, the quoted uncertainties vary widely. This is because this parameter is not as commonly used as *Ra*. Hence, the measurement algorithms and the uncertainty budgets for this parameter are likely to vary widely. Best practices for *RSm* are not as firmly established as they are for other parameters, such as *Ra* and *Rz*. NIST's low uncertainty is based on direct use of laser interferometry to measure the lateral displacement of the stylus. This is not common practice for stylus type measurements of *RSm*. Also the method of calculation of *RSm* [5] can vary significantly depending on what discrimination criteria are applied to ignoring peaks that are small or close together. By contrast, the definition for a similar parameter *Sm* defined in the 1995 revision of the U.S. national standard [4] does not apply such criteria and counts all of the profile peaks and valleys. However, these issues are not significant in the case of a sinusoidal specimen with 100 µm *RSm,* because very short spatial wavelengths are not a significant component of the surface roughness. In spite of the lack of a common practice, the agreement between the laboratories is well within the quoted uncertainties.

For step height measurements, the well known algorithm from ISO 5436-1(2000) [7] (Fig. 10) was used by NRC, NIST, CENAM, and INTI. INMETRO measured a different quantity, known as *Pt*, the height difference between the highest and lowest points on the profile. For a leveled, flat profile, the value of *Pt* should be very close to the step height itself. All results agree with the reference value within the quoted uncertainties.

6. Further Observations and Reanalysis

After the initial submission and analysis of the results, a number of differences between labs were recognized. We therefore applied a number of corrections to the analysis and recalculated some of the parameters and uncertainty values. The changes took place in several categories as described below.

6.1 Definition of *Rz*

During the evaluation process the participants noticed that both NRC and CENAM had *Rz* (ISO) values smaller than *Rz* (DIN) if the same raw data is used. Further investigation revealed that the software algorithm of both instruments was based on the older ISO standard. Both laboratories used their raw data and recalculated the Rz (ISO) values using the *Rz* (DIN) algorithm, which is identical to the new *Rz* (ISO) algorithm when the evaluation length consists of five sampling lengths, as was the case here.

6.2 Instrument Noise and *Rz*

Because *Rz* is an extreme value parameter, its value is more sensitive to noise than averaged parameters such as *Ra*. Hence, instrument noise is a more important cause of variation for *Rz* measurements than for *Ra* measurements. Because the instruments likely had different noise levels, it is important to consider the effects carefully.

During a summary coordination meeting held at NRC, the participants discussed the measurement uncertainties. It was noted that NRC's expanded uncertainty for *Rz* (51 nm) was very large in comparison with the other labs. The main reason for this large uncertainty was that NRC estimated the effect of the instrument noise by measuring *Rz* on an optical flat using the same system and filter conditions that were used to measure the 3144 specimen. This experiment revealed significant noise and therefore influenced the NRC uncertainty estimate. Because it is a "range" parameter, the *Rz* parameter is extremely sensitive to noise. Furthermore, the noise can only increase the *Rz* values. The *Rz*-noise, which is clearly evident during a smooth-surface measurement, is also present during an artifact measurement and introduces a significant systematic increase into measured *Rz* results, effectively causing a bias.

The participants agreed to conduct the same experiment as NRC, and the results in most of the cases have shown significant instrument noise. It was proposed that each participant correct its *Rz* result and uncertainty by the following to account for the *Rz* noise bias and its large uncertainty:

$$Rz \text{ Correction} = -(0.5\ Rz\text{-noise}) \pm (0.5\ Rz\text{-noise}/\sqrt{3}) \quad (k=1). \tag{4}$$

The uncertainty conservatively assumes a rectangular distribution for the correction because the size of the correction would depend on the shape of the surface profile and is difficult to calculate with precision. In fact, we expect the correction to be larger for a periodic surface profile such as the SPECIAL 3144 specimen than for a random surface profile.

6.3 NIST Type A Uncertainty

NRC, INTI, INMETRO, and CENAM developed their uncertainty budgets with the principal Type A component calculated as one standard deviation of the mean of the set of measured values. This is conventional practice for laboratory comparisons. NIST, on the other hand, initially calculated its uncertainty budgets with the variation due to surface position calculated as one standard deviation of the measurements, according to its practice for measurement calibrations for customers. NIST uses this method because NIST assumes that a customer may calibrate its instrument with the NIST calibrated standard with as little as a single measurement. The method of calculation for this uncertainty component should have been agreed to beforehand but was not. For these measurements the difference in the total uncertainty for the two methods is insignificant for five of NIST's seven submitted results because the instrument uncertainty is a more dominant component in the uncertainty budget than the variation in the measured results. However, for the measurements of *Rz* and the 0.03 µm step height, the uncertainty is dominated by the variation in measurements and is therefore significantly smaller when calculated using the standard deviation of the mean. Because this method of uncertainty calculation can affect the calculated degree of equivalence between NIST and other laboratories for these two quantities, NIST recalculated its uncertainty budgets for all seven measured parameters using the standard deviation of the mean to represent the variation in measurements, thus putting its statistics in line with others.

6.4 INMETRO *Pt* Value

It is expected that the *Pt* value calculated by INMETRO should be very close to the *d* parameter for step height for quality step height surfaces, such as the 2.55 μm step and the 0.387 μm step measured by INMETRO here, which have very little roughness or flatness deviation. This is shown by the good agreement between the INMETRO results for *Pt* and the results of other laboratories for *d* in Fig. 14. However, because the parameters are different, it is better to compute the reference value for *d* only from the four measured results for *d*. Therefore, we reanalyzed the data in this way. Even though the average value changes slightly, the revised tables for *d* below still show good agreement between the INMETRO values and the other laboratories.

6.5 Summary Statistics

The summaries shown in Section 5 were done according to an agreed upon protocol before any measurements were taken. After the initial submission of results, discussion among the participants, and reviews by colleagues, we decided that the summary statistics should be changed in three ways:

1) Computing the uncertainty of the reference value $u(x_{\text{ref}})$ from the deviations of the individual lab values with respect to x_{ref} (as shown by Eqn. 3) does not take into account the estimated uncertainty u_i of each laboratory's results. The uncertainty of x_{ref} for each computed parameter should be calculated from the law of propagation of uncertainties associated with the input quantities in Eqn. 1. The propagation of uncertainty from the inputs x_i to the output x_{ref} yields a standard uncertainty $u(x_{\text{ref}})$ given by the quadratic sum

$$u^2(x_{\text{ref}}) = \frac{1}{n^2}\sum_{i=1}^{n} u_i^{\,2} \ . \tag{5}$$

The uncertainty of the reference value for each quantity was therefore recalculated in this way.

2) It was pointed out that the standard method[9] to express degree of equivalence is with the pair of values d_i, $U(d_i)$, where d_i is the difference between the lab value and reference value

$$d_i = x_i - x_{\text{ref}} \tag{6}$$

and $U(d_i)$ is the uncertainty of that difference. Therefore, the tables of final results were reformatted to show those values explicitly.

3) Taking into account the relatively small number of lab results in the comparison (for all the calculated parameters the number n of NMI´s was less than or equal to 5), the dependence of the reference value x_{ref} for any quantity with respect to a single result x_i is relatively high. So, there may be a significant correlation between x_i and x_{ref} (on the order of $1/\sqrt{n}$). Thus the square root of the denominator in Eqn. (3) may be significantly greater than the standard uncertainty of the numerator. It then follows that the parameter E_n as first calculated could be an underestimation of the normalized difference between x_i and x_{ref},. Therefore, the variance of the difference $u^2(d_i)$ was recalculated to account for the correlation between x_i and x_{ref}. This quantity is given by [9]

$$\begin{aligned} u^2(d_i) &= \text{var}(x_i) + \text{var}(x_{\text{ref}}) - 2\cdot\text{cov}(x_i, x_{\text{ref}}) = u_i^2 + \frac{1}{n^2}\cdot\sum_{i=1}^{n} u_i^2 - 2\cdot\frac{u_i^2}{n} \\ &= \left(1-\frac{2}{n}\right)u_i^2 + u^2(x_{\text{ref}}), \end{aligned} \tag{7}$$

where var is the variance of a quantity and cov is the covariance of two quantities. Alternatively, exclusive statistics can be used[10,11] to account for the correlations. In this case, different reference values would be obtained for each lab.

With this recalculation of $u^2(d_i)$ we avoid the statistical correlations between each single result and the reference value, except those arising from the possible *physical* correlations among the original results $x_1,\ldots x_n$, given by crossed traceability chains or other causes. The tables in Figs. 15 to 20 show the expanded uncertainty $U(d_i)$ equal to 2 $u(d_i)$. These tables also show the recalculated equivalence quotients *En* given by

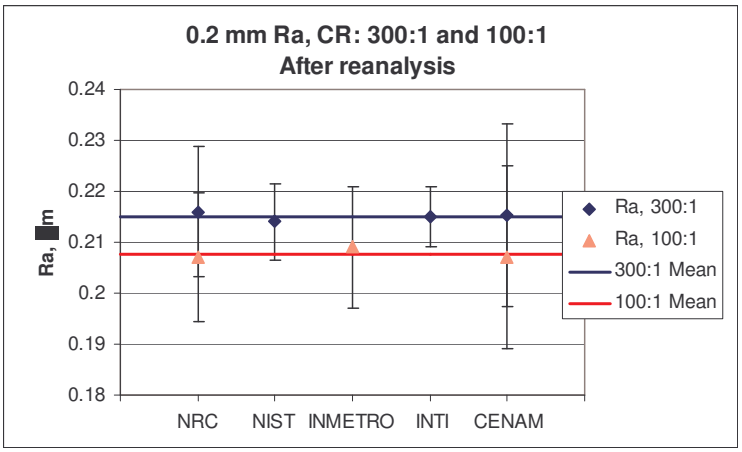
$$E_n = (x_i - x_{\text{ref}})/[2\,u(d_i)]\,. \tag{8}$$

6.6 Reanalysis Summary

Figures 16 to 20 show the results with all changes applied. Overall the agreement between the laboratories is still excellent, and only modest differences were found between E_n in the initial results and E_n in the final results. The maximum value for $|E_n|$ was 0.72 from the initial results, and one value for $|E_n|$ reaches a value of 0.98 for the final results. The modest E_n values shown in Figs. 15-20 suggest that the participants most likely assigned uncertainty values that were too conservative. For 39 measurements, we can expect that we will typically see about 2 results with E_n>1, whereas here we do not see any.

After the corrections and reanalysis, the *Rz* results of Fig. 16 all have smaller values than those of Fig. 12 and the uncertainties are larger on the average. The difference between the *Rz* reference values before and after correction is 0.018 μm for the 300:1 cut-off ratio and 0.010 μm for the 100:1 cut-off ratio. Applying the noise correction alone produces very significant changes in the values and uncertainties for *Rz*. If the noise correction were not applied, the average value of *Rz* would be higher by 0.026 μm for the 300:1data and 0.015 μm for the 100:1 data. These differences are significant with respect to the quoted uncertainties.

Because the participants used instruments based on different principles (laser, inductive) and from different manufacturers it seems clear that the surface metrology community is facing a common issue of calculation of *Rz* in the presence of noise. It is important that the proposal in Eqn. 4 for handling the noise-related component of uncertainty be discussed and that a uniform approach for reckoning noise effects for *Rz* measurement be adopted. We note that in recent comparisons of measured surface roughness parameters conducted by both the Asia-Pacific Metrology Programme [12] and the European Metrology Programme (EUROMET)[13] , there were significant differences among the laboratory results for the averaged *Rz* parameter[12] (also known as R_{tm} and *Ry5*) and for the *Rmax* parameter[12,13]. The parameter *Rmax*, defined in DIN 4768[6], is the largest *Rz* value in a set of five calculated over an evaluation length consisting of five sampling lengths. Perhaps, those differences could be due to the presence of different amounts of noise in the measured profiles and to different procedures to account for noise effects.



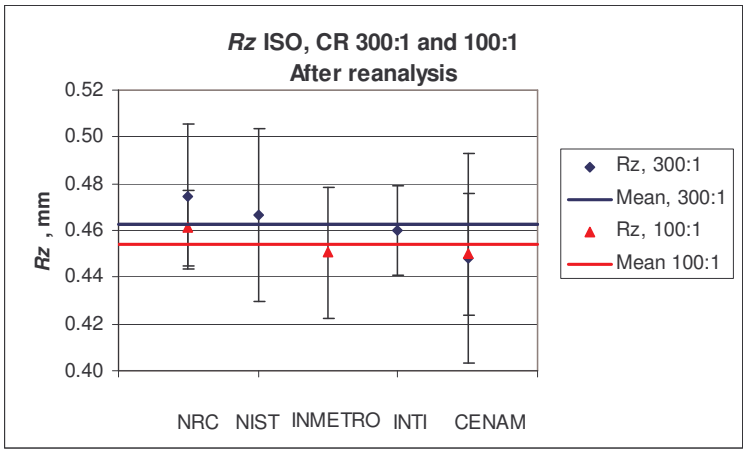
NMI	Special 3144 Ra, CR,300:1				
	Ra, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	0.216	0.013	0.0009	0.0110	0.08
NIST	0.2141	0.0074	-0.0010	0.0080	-0.13
INMETRO					
INTI	0.215	0.006	-0.0001	0.0074	-0.01
CENAM	0.2153	0.018	0.0002	0.0141	0.01

X ref	0.2151
std. unc.	0.0030
U, k=2	0.0060

NMI	Special 3144 Ra, CR,100:1				
	Ra, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	0.207	0.013	-0.0007	0.011	-0.06
NIST					
INMETRO	0.209	0.012	0.0013	0.011	0.12
INTI					
CENAM	0.207	0.018	-0.0007	0.013	-0.05

X ref	0.2077
std. unc.	0.0042
U,k=2	0.0084

Figure 15: Results for Ra measurements of roughness specimen "SPECIAL 3144" after reanalysis



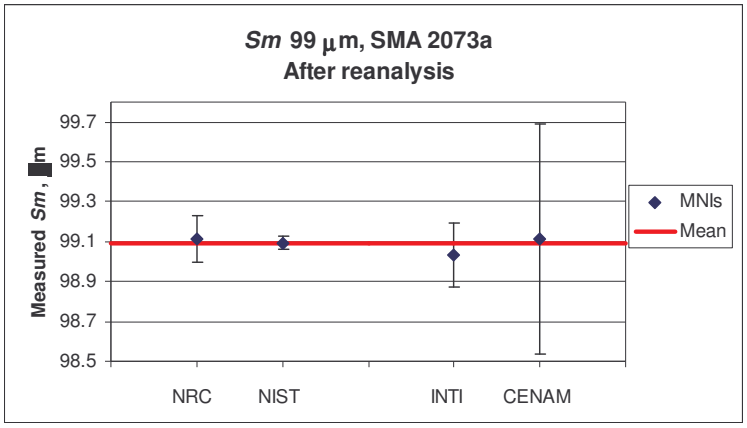
NMI	Special 3144 *Rz*, CR, 300:1				
	Rz, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	0.475	0.031	0.0125	0.028	0.45
NIST	0.467	0.037	0.0045	0.031	0.14
INMETRO					
INTI	0.460	0.018	-0.0025	0.021	-0.12
CENAM	0.448	0.044	-0.0145	0.035	-0.41

RV	0.4625
std. unc.	0.0085
U,k=2	0.0169

NMI	Special 3144 *Rz* ISO, CR, 100:1				
	Rz, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	0.461	0.016	0.0072	0.016	0.44
NIST					
INMETRO	0.4505	0.028	-0.0033	0.021	-0.16
INTI					
CENAM	0.450	0.025	-0.0038	0.020	-0.19

X ref	0.4538
std. unc.	0.0068
U, k=2	0.0136

Figure 16. Comparison of surface roughness *Rz* results from five SIM NMI laboratories after reanalysis.



NMI	SRM 2073a Ra, CR - 300:1				
	Ra, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	3.108	0.021	0.0150	0.027	0.56
NIST	3.079	0.04	-0.0140	0.036	-0.39
INMETRO					
INTI	3.09	0.039	-0.0030	0.035	-0.09
CENAM	3.0951	0.065	0.0021	0.051	0.04

RV	3.093
std. unc.	0.011
U,k=2	0.022

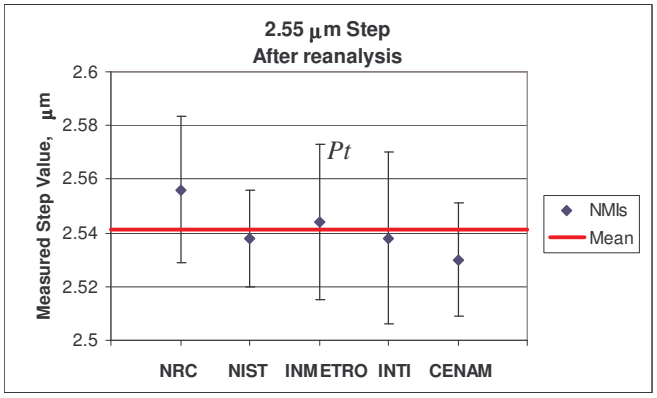
NMI	SRM 2073a *Ra*, CR - 100:1				
	Ra, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	3.094	0.021	0.0061	0.036	0.17
NIST					
INMETRO	3.088	0.074	0.0001	0.054	0.00
INTI					
CENAM	3.0817	0.065	-0.0062	0.050	-0.12

X ref	3.088
std. unc.	0.017
U, k=2	0.034

NMI	SRM 2073a RSm, CR - 300:1				
	RSm, µm	*U*, µm	*di*, µm	*U*(*di*), µm	*En*
NRC	99.112	0.115	0.024	0.174	0.14
NIST	99.094	0.035	0.006	0.155	0.04
INMETRO					
INTI	99.033	0.16	-0.055	0.191	-0.29
CENAM	99.113	0.58	0.025	0.438	0.06

X ref	99.088
std. unc.	0.077
U,k=2	0.153

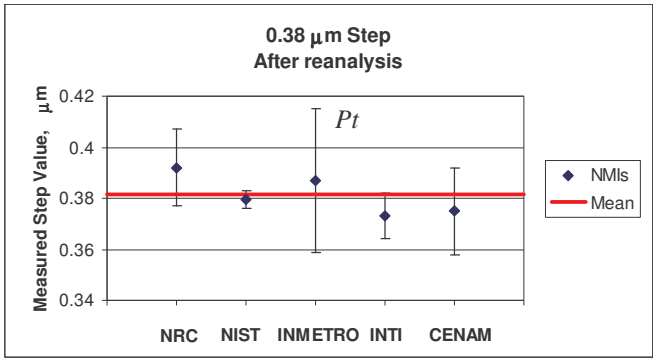
Figure 17: Comparison of surface roughness Ra and RSm results from five SIM NMI laboratories after reanalysis. The graphs for *Ra* are essentially unchanged from those of Fig. 13 and are not repeated here.



NMI	18235, d=2.55 µm				
	d, µm	U, µm	di, µm	$U(di)$, µm	En
NRC	2.556	0.027	0.0155	0.0228	0.68
NIST	2.538	0.018	-0.0025	0.0179	-0.14
INMETRO (Pt)	2.544	0.029	0.0035	0.0316	0.11
INTI	2.538	0.032	-0.0025	0.0259	-0.10
CENAM	2.53	0.021	-0.0105	0.0194	-0.54

X ref	2.5405
std. unc.	0.0063
U,k=2	0.0125

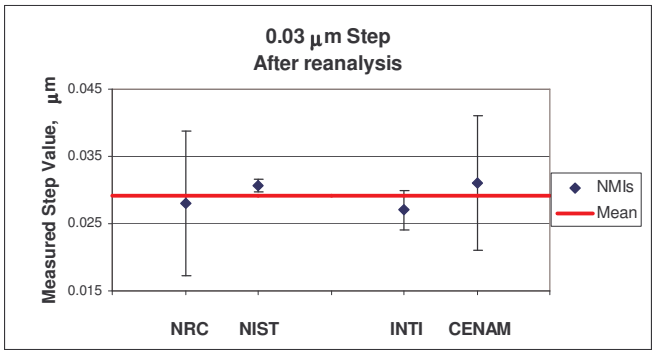
Figure 18: Comparison of step height (*d)* and *Pt* results from five SIM NMI laboratories after reanalysis for the 2.55 µm.



NMI	18235, d=0.38 µm				
	d, µm	U, µm	di, µm	$U(di)$, µm	En
NRC	0.392	0.015	0.0121	0.0123	0.98
NIST	0.3797	0.0035	-0.0002	0.0066	-0.03
INMETRO (Pt)	0.387	0.028	0.0071	0.0287	0.25
INTI	0.373	0.009	-0.0069	0.0089	-0.78
CENAM	0.375	0.017	-0.0049	0.0135	-0.36

X ref	0.3799
std. unc.	0.0031
U,k=2	0.0062

Figure 19: Comparison of step height (*d*) and *Pt* results from five SIM NMI laboratories after reanalysis for the 0.38 µm step.



NMI	18235, d=0.03 µm				
	d, µm	U, µm	di, µm	$U(di)$, µm	En
NRC	0.028	0.011	-0.0012	0.0087	-0.14
NIST	0.0307	0.0009	0.0015	0.0039	0.40
INMETRO					
INTI	0.027	0.003	-0.0022	0.0044	-0.50
CENAM	0.031	0.01	0.0018	0.0080	0.23

X ref	0.0292
std. unc.	0.0019
U,k=2	0.0038

Figure 20: Comparison of step height (*d)* and *Pt* results from four SIM NMI laboratories after reanalysis for the 0.03 µm step.

7.0 Final Observation: Bandwidth Limits

In future, the choice of the bandwidth limit at small spatial wavelengths for such comparisons will require further consideration. The short wavelength limit usually depends either on the λs filter or on the stylus tip size or perhaps both if those sizes are comparable. For the *Ra* and *Rz* results for 300:1 cut-off ratio, NRC, INTI, CENAM, and NIST all used styli with nominal radii of 2 μm and λs of 2.5 μm. However, the stylus used at NIST was measured using the method of tracing over a razor blade [14, 7], and the measured result for the stylus radius was 5.0 μm $\pm$ 0.5 μm (k=1). The styli used by the other laboratories were not measured, only specified. The comparison of NIST *Ra* results in Fig. 11 with those of NRC, INTI, and CENAM suggest that the measured 5 μm stylus tip size does not significantly attenuate measured roughness values by comparison with the results for the nominal 2 μm stylus tip sizes as much as the change from a 2.5 μm λs cut-off to a 8 μm λs cut-off. This observation is further supported by the following results from INTI for the 0.2 μm *Ra* specimen. For a nominal stylus tip radius of 5 μm, the *Ra* measured by INTI was 0.2148 μm with a standard deviation of 0.001μm, essentially the same as the average result 0.215 μm shown in Fig. 11 for nominal 2 μm tip radius. The question here is: what was the effective short wavelength cut-off of the INMETRO roughness measurements? The INMETRO data were placed in the 100:1 category and compared with the 100:1 NRC and CENAM data because the nominal stylus radius was 5 μm, higher than the nominal radius of the others. The INMETRO results agreed well with those of NRC and CENAM. However, INMETRO's nominal radius was the same as NIST's measured radius value. So a case could be made that INMETRO's results should be tabulated with the 300:1 results. Perhaps in future comparisons, the stylus radii should be evaluated from measurements or else the short wavelength limit of the measured profiles should be clearly limited by Gaussian filtering.

8.0 Acknowledgement

The participants acknowledge the support from NRC piloting this comparison and hosting an evaluation meeting and the support from SIM for some travel expenses. The NIST authors are grateful to A. Hornikova for the careful inspection of these results and suggestions concerning the analysis and to C.D. Foreman for taking surface profile measurements contributing to this report.

9.0 References

[1] *Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes* (International Committee for Weights and Measures, 1999), available at www.bipm.org/pdf/mra.pdf.

[2] ISO 11562:1996 *Geometrical Product Specifications (GPS) – Surface texture: Profile method – Metrological characteristics of phase correct filters* (International Organization for Standardization, Geneva, Switzerland, 1996).

[3] ISO 3274:1996 *Geometrical Product Specifications (GPS) – Surface texture: Profile method – Nominal characteristics of contact (stylus) instruments* (International Organization for Standardization, Geneva, Switzerland, 1996).

[4] ASME B46.1-2002 *Surface Texture* (American Society of Mechanical Engineers, New York, 2003)

[5] ISO 4287:1997 *Geometrical Product Specifications (GPS) – Surface texture: Profile method – Terms, definitions and surface texture parameters* (International Organization for Standardization, Geneva, Switzerland, 1997).

[6] Deutsche Normen DIN 4768, Part 1 *Determination of Surface Roughness Values R_a, R_z, R_{max} with Electric Stylus Instruments, Basic Data* (Beuth Verlag GmbH, Berlin, 1974).

[7] 5436-1:2000 *Geometrical Product Specifications (GPS) – Surface texture: Profile method; Measurement standards – Part 1: Material measures* (International Organization for Standardization, Geneva, Switzerland, 1999).

[8] *Guide to the Expression of Uncertainty in Measurement* (GUM), International Organization for Standardization, Geneva, 1995).

[9] M.G. Cox, "The Evaluation of Key Comparison Data," Metrologia, 2002, **39**, 589-595.

[10] A.G. Steele, B.M. Wood, R.J. Douglas, "Exclusive statistics: simple treatment of the unavoidable correlations from key comparison reference values," Metrologia, 2001, **38**, No. 6, 483-488.

[11] J.E. Decker, J. Altshuler, H. Beladie, I. Malinovsky, E. Prieto, J. Stoup, A. Titov, M. Villesid, J.R. Pekelsky, Report of SIM 4.2 Regional Comparison, Stage One: Calibration of Gauge Blocks by Optical Interferometry (in press).

[12] W.J. Giardini, C. Calderon, F. Cheng-yen, K.K. Hau, C.-S. Kang, S.M. Lee, F. Lesha, A.M. MeKawi, R. Monteros, C. Quan, V. Roonwal, N. Sakano, R.P. Singhal, S.Y. Wong, S.Z.A. Zahwi, "International Comparison of Surface-texture Parameters Organized by the Asia-Pacific Metrology Programme," Metrologia, 2002, **39,** 225-229.

[13] EUROMET Supplementary Comparison, "Surface Texture," Project No. 600, Technical Report (Physikalisch-Technische Bundesanstalt, Braunschweig, Germany, March 23, 2001).

[14] J.-F. Song, T.V. Vorburger, Stylus Profiling at High Resolution and Low Force," Applied Optics, 1991, **30**, 42.