Standard Deviation or Standard Deviation of the Mean - How to Report Statistical Variation in Surface Calibrations?

J. F. Song and T. V. Vorburger
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899

Keywords: surface roughness, surface calibration, uncertainty, geometrical non-uniformity, statistical variation.

1. Introduction

Statistical variation plays an important role in the uncertainty budget of surface calibration reports. Statistical variation mainly comes from two sources: the random variation of the measurement system (the measurement repeatability) and the random variation over the calibrated surface (the geometrical non-uniformity). Geometrical non-uniformity is an important quality factor for a surface calibration specimen. It usually contributes a dominant part of the statistical variation. In a surface calibration report, statistical variation might be represented by either the standard deviation σ_x or the standard deviation of the mean $\sigma_{\bar{x}}$. Both are calculated from a set of multiple measurements distributed at different positions within the measurement area. If the standard deviation is used for representing the statistical variation, the measurement uncertainty is reported for a single measurement. Otherwise, the measurement uncertainty is reported for the mean value of multiple measurements. From the point of view of statistics, there is no difference if this information is provided sufficiently and used correctly. If not, however, an over- or under-estimated calibration uncertainty might result.

2. Uncertainty Procedure in Surface Calibrations

The measurement uncertainty in surface calibrations mainly comes from two sources: the measurement system and the measured surface. The quoted expanded uncertainty U (with a coverage factor $k = 2$) is equal to twice the combined standard uncertainty u_c calculated from the quadratic sum of the instrument (or the measurement system) standard uncertainty $u_{(i)}$ and the statistical variation of the measurements $u_{(s)}$ [1,2]. The first part, $u_{(i)}$, mainly represents the quality of the measurement system, while the second part, $u_{(s)}$, mainly represents the geometrical non-uniformity of the measured surface combined with the repeatability of the measurement system.

For the uncertainty budget of NIST surface calibrations, performed using a stylus instrument, the instrument standard uncertainty $u_{(i)}$ is a quadratic sum of six uncertainty components arising from: 1) the geometric non-uniformity and the surface finish of the step-height master calibration standard; 2) the calibration uncertainty of the step-height master standard calibrated by interferometric and stylus measurements; 3) the variation of the measured calibration constant; 4) the non-linearity of the instrument transducer; 5) the instrument noise; and 6) the tip size of the contact stylus. A detailed description of these uncertainty components may be found in Ref. 1.

The statistical variation of the measurements $u_{(s)}$ is mainly derived from the geometrical non-uniformity of the specimen under evaluation, but it also includes instrumental random variation, or measurement repeatability which usually takes a small part. The statistical variation may be represented by the standard deviation σ_x or the standard deviation of the mean $\sigma_{\bar{x}}$, and included as a Type A uncertainty [3, 4] in the uncertainty budget. Then the statistical variation $u_{(s)}$ is quadratically added with the instrument standard uncertainty $u_{(i)}$, resulting in the combined standard uncertainty u_c, from which the expanded measurement uncertainty U is calculated with a coverage factor of $k = 2$.

3. For Calibration Laboratories

How to report statistical variation in surface calibrations, standard deviation σ_x or standard deviation of the mean $\sigma_{\bar{x}}$? Currently there is not a standardized procedure. The procedure should depend on how customers use the calibrated specimen. In many cases, customers use the specimen only for a single-trace calibration, therefore, the measurement uncertainty should be reported for a single measurement. If the customer uses the specimen for a multi-trace calibration, and takes the mean value as the calibration result, then the measurement uncertainty for the mean value of the multiple measurements should be reported. The number of measurements, n, with the distribution of the measurement positions must be included in the calibration report. In either case, as long as sufficient information is shown in the calibration report, customers may convert calibration uncertainties from one approach to the other, or re-calculate uncertainties when the number of measurements is different from that shown in the calibration report.

For example, Table 1 shows a calibration result from a NIST surface calibration report for the roughness average *Ra* value of a roughness specimen. Unless the customer specifically states otherwise, it is assumed that the specimen is used for single-trace calibrations and checks. Therefore, the calibration uncertainty is reported for the single measurement. The calibration is performed at nine different positions of the roughness specimen. The calibration result, 2.958 µm ± 0.070 µm, is reported for the single measurement that includes a statistical variation of one standard deviation calculated from nine measurements at different positions. In the Appendix attached to the Calibration Report, it is useful to include the following information:

- Each *Ra* value measured at nine different positions;
- A drawing showing the distribution of the nine measurement positions;
- A detailed description of uncertainty components from the measurement system, based on which the instrument standard uncertainty $u_{(i)}$ is calculated [1];
- A measurement summary (see lower part of Table 1), which includes the average *Ra*, the statistical variation $u_{(s)}$ represented by the standard deviation σ_x, the instrument standard uncertainty $u_{(i)}$, the combined standard uncertainty u_c and/or the expanded uncertainty U ($k = 2$).

Table 1. A calibration result intended for the single measurement of *Ra*

Specimen Patch ID	Roughness Average (*Ra*) (µm)	Expanded Uncertainty (*k* = 2, µm)
3.0 µm	2.958	0.070

Appendix

Average *Ra*	Statistical Variation (1 Standard Deviation)	Instrument Standard Uncertainty	Combined Standard Uncertainty	Expanded Uncertainty (*k* = 2)
2.958 µm	0.027 µm	0.022 µm	0.035 µm	0.070 µm

If the customer wants to report the calibration uncertainty for the mean value of nine measurements using the same data shown in Table 1, the stated measurement uncertainty for the single measurement can be easily converted to the measurement uncertainty for the mean value of nine measurements as shown below (see Table 2):

- The statistical variation $u_{(s)}$ is converted from the standard deviation σ_x to the standard deviation of the mean: $\sigma_{\bar{x}} = \sigma_x / \sqrt{n} = 0.027 / \sqrt{9} = 0.009$ µm;
- The instrument standard uncertainty $u_{(i)}$ is the same, 0.022 µm;
- The combined standard uncertainty u_c is calculated as 0.024 µm, the expanded uncertainty is $U = 0.048$ µm.

Table 2. The same data from Table 1 but showing the expanded uncertainty for the mean value of *Ra* (n = 9)

Specimen Patch ID	Roughness Average (*Ra*) (µm)	Expanded Uncertainty for the Stated Mean Value of *Ra* (*k* = 2, µm)
3.0 µm	2.958	0.048

Appendix

Average *Ra*	Statistical Variation (1 Standard Deviation of the Mean, *n* = 9)	Instrument Standard Uncertainty	Combined Standard Uncertainty	Expanded Uncertainty (*k* = 2)
2.958 µm	0.009 µm	0.022 µm	0.024 µm	0.048 µm

From Tables 1 and 2, it can be seen that there are two uncertainty statements for the same set of measurement data. Table 1 shows the measurement uncertainty for a single measurement, which implies that when customers take any measurement within the measurement area, the measurement value may expected to lie within the range of

2.958 µm ± 0.070 µm with a confidence level approaching 95% for large n [3,4]. Table 2 shows the measurement uncertainty for the mean value of a set of nine measurements, which implies that when customers take nine measurements at different positions of the measurement area, the mean value could be expected within the range of 2.958 µm ± 0.048 µm with a confidence level approaching 95%.

4. For Users

If a customer received a calibration report showing that the measurement uncertainty was reported for a single measurement (see Table 1), and if the customer wanted to reduce the measurement uncertainty by measuring, say, three positions evenly distributed in the measurement area and taking the mean value as the result, the expanded measurement uncertainty for the mean value could be calculated by:

- Converting the standard deviation σ_x to the standard deviation of the mean $\sigma_{\bar{x}}$ for the statistical variation, $u_{(s)} = \sigma_{\bar{x}} = \sigma_x / \sqrt{n} = 0.027$ µm / $\sqrt{3} = 0.0156$ µm;
- Combining the new statistical variation, $u_{(s)} = 0.0156$ µm, with the instrument standard uncertainty, $u_{(i)} = 0.022$ µm, as the new combined standard uncertainty, $u_c = 0.027$ µm;
- Calculating the expanded uncertainty with a coverage factor $k = 2$, $U = 0.054$ µm (see Table 3).

Table 3. The same data from Table 1 but showing the expanded uncertainty for the mean value of *Ra* (*n = 3*)

Average *Ra*	Statistical Variation (1 Standard Deviation of the mean, $n = 3$)	Instrument Standard Uncertainty	Combined Standard Uncertainty	Expanded Uncertainty ($k = 2$)
2.958 µm	0.0156 µm	0.022 µm	0.027 µm	0.054 µm

5. For Standards Development Organizations

Different types of surface calibration specimens are specified in the International Standard ISO 5436 [5,6] and the U.S. National Standard ASME B46 standard [7]. In the previous version of ISO 5436 standard [5] and the current ASME B46 standard [7], tolerances and measurement uncertainties for some types of specimens are also specified. For example, for the Type D random profile roughness specimens [5-7], the specified tolerances are shown in Table 4 (originally Table 9 in Ref. 5 and Table 11-9 in Ref. 7). A wide tolerance range, ± (15 ~ 30) % (see the second column), was specified for the nominal value of *Ra*. This wide range is for the convenience of the specimen manufacturer and is usually not a problem for the user. The actual measured *Ra* value and the uncertainty are generally more important for the user. It must be noted that the standard deviation is used here for representing the statistical variation, $u_{(s)}$ = (3 ~ 4) % in the last column, mainly for the control of the geometrical non-uniformity of the calibration specimen. However, the tolerance for the uncertainty of measurement, or the expanded uncertainty, is specified for the stated mean value of *Ra*, U = ± (3 ~ 5) % (see the third column). This uncertainty includes a statistical variation represented by the standard deviation of the mean. In both the ISO and ASME standards [5,7], there is a note stating that the mean value is "taken from 12 evenly distributed readings" (see the bottom of Table 4). This information is important for

Table 4. Tolerances for type D specimens [5,7]

Nominal Value of *Ra*, µm	Tolerance on Nominal Value, %	Uncertainty of Measurement of Stated Mean Value of *Ra*[1)], %	Standard Deviation from Mean Value, %
0.15	± 30	± 5	4
0.5	± 20	± 3	3
1.5	± 15	± 3	3

1) Taken from 12 evenly distributed readings.

both calibration laboratories and users. It enables one to convert the standard deviation σ_x and the standard deviation of the mean $\sigma_{\bar{x}}$ when necessary.

For example, if calibration laboratories want to know whether or not their calibration capability is qualified for the calibration shown in Table 4, they can calculate the instrument standard uncertainty, $u_{(i)}$, required for this calibration. For the specimen with nominal *Ra* of 0.5 µm, the tolerance for the expanded measurement uncertainty of the stated mean *Ra* is specified as $U = \pm 3\%$ (k =2) with the measurements taken at 12 measurement positions. That means the combined standard uncertainty for the stated mean Ra is u_c= 1.5 %. Because the statistical variation $u_{(s)}$ is represented by standard deviation, σ_x = 3% (see Table 4), it must be converted to standard deviation of the mean $\sigma_{\bar{x}} = \sigma_x / \sqrt{n} = 0.87$ % (n = 12). Then the required instrument standard uncertainty could be calculated as $u_{(i)} = [u_c^2 - u_{(s)}^2]^{1/2} = 1.22$ %. Based on this, the calibration laboratory might judge whether or not their calibration capability could be qualified for this calibration.

If the measurement uncertainty was specified for the stated mean Ra value without specifying how many readings were taken. This information was not sufficient for both calibration laboratories and users. Because the standard deviation σ_x could not be converted to the standard deviation of the mean $\sigma_{\bar{x}}$ without specifying the number of measurements n.

6. NIST Surface Calibration Reports

Most surface calibrations at NIST Surface and Microform Calibration Laboratory are performed using nine positions. However, when these calibrated specimens are used in the secondary laboratories, in many cases, only a single-trace calibration is performed. It is for this reason that the measurement uncertainty for NIST surface calibrations is reported for a single measurement except when the customer requests a report of measurement uncertainty for the mean Ra value of multiple measurements. Therefore, the standard deviation σ_x, rather than the standard deviation of the mean $\sigma_{\bar{x}}$, is used for representing the statistical variation in NIST calibration reports. As a result, a relatively conservative (large) measurement uncertainty would be introduced if customers use this specimen for multiple measurements and take the mean value as the result. However, in that case, the customer can follow the procedures illustrated in Table 3 by converting the standard deviation σ_x to the standard deviation of the mean $\sigma_{\bar{x}}$, and re-calculate the measurement uncertainty for the stated mean value of Ra.

References:

[1] Appendix A, "Measurement Conditions and Sources of Uncertainties for NIST Surface Roughness and Step Height Calibration Reports", last revised in July 2001, http://www.mel.nist.gov/div821/webdocs-13/surfcalib.htm.

[2] J. F. Song and T. V. Vorburger, Uncertainty Procedures for NIST Surface Finish and Microform Calibrations, Measurement Science Conference, CA, 1996.

[3] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, *Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1993.

[4] B.N. Taylor and C.E. Kuyatt, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297 (National Institute of Standards and Technology, Gaithersburg, MD, 1993).

[5] ISO 5436 – 1985, Calibration specimens – Stylus instruments – Types, calibration and use of specimens, ISO, Geneva, 1985.

[6] ISO/DIS 5436-1, *Geometrical Product Specifications (GPS) – Surface Texture: Measurement Standards Part 1: Material Measures*, 18 Dec, 1998.

[7] ASME B46.1-2002, *Surface Texture*, American Society of Mechanical Engineers, New York, 2002.