Surface Topography: Metrology and Properties

Modeling the systematic behavior at the micro and nano length scales

Danilo Quagliotti
Technical University of Denmark, Department of Mechanical Engineering, Productionstorvet 427A, Kgs. Lyngby, Denmark
E-mail: danqua@mek.dtu.dk

Keywords: systematic behavior, frequentist statistics, uncertainty modeling, GUM, surface metrology, optical microscopy, micro/nano metrology

Abstract
The assessment of the systematic behavior based on frequentist statistics was analyzed in the context of micro/nano metrology. The proposed method is in agreement with the well-known GUM recommendations. The investigation assessed three different case studies with definition of model equations and establishment of the traceability. The systematic behavior was modeled in $S_q$ roughness parameters and step height measurements obtained from different types of optical microscopes, and in comparison with a calibrated contact instrument. The sequence of case studies demonstrated the applicability of the method to micrographs when their elements are averaged. Moreover, a number of influence factors, which are typical causes of inaccuracy at the micro and nano length scales, were analyzed in relation to the correction of the systematic behavior, viz. the amount of repeated measurements, the time sequence of the acquired micrographs and the instrument-operator chain. The possibility of applying the method individually to the elements of the micrographs was instead proven not convenient and too onerous for the industry. Eventually, the method was also examined against the framework of the metrological characteristics defined in ISO 25 178-600 with hints on possible future developments.

1. Introduction
In recent years, the need for shared rules for the uncertainty evaluation in manufacturing metrology at the micro- and nano-dimensional scales has been discussed largely in an international comparison of optical instruments, measuring the surface topography by areal parameters [1]. Among the important issues emphasized by the comparison, it emerged clearly the lack of specific guidelines for the uncertainty evaluation. It resulted underestimated for the most, commonly based on the repeatability. However, in few cases the final uncertainty was stated with established traceability, but the use of non-suitable artifacts for the sub-micrometer length scale reversed the result towards an overestimated assessment.

The comparison also asserted the increasing interest for a three-dimensional (3D) characterization (areal method), which offered a more realistic representation of the surfaces, and a more complete understanding of their functions [2–4]. In fact, despite the profile method was well established [5–8] (at the time of writing a new framework for the profile method is under development in the new series ISO 21 920), it also became insufficient to determine the exact nature of 3D topographic features, and to cope with the constant reduction of critical dimensions, as it is inexorably occurring in micro and nano manufacturing [9–11]. Nonetheless, the impact of a non-proper evaluated measurement uncertainty, and the challenge in establishing the traceability were detrimental for the assessment of the manufacturing quality [12, 13].

Since then the technological rush has prompted into a digital manufacturing era with almost endless possibilities of ‘true 3D’, complex and multifunctional surfaces [14–17]. Furthermore, a new context for the instruments apt to measure the surface topography is available in the series ISO 25 178, based on the metrological characteristics [18, 19]. Even though different working principles exist, the metrological characteristics define sources of variation of the output in common to all the areal topography measuring instruments expressed as uncertainty contributors. Examples of earlier applications can be found elsewhere [20–23], while
A new ISO standard is currently under development, where the calibration of the metrological characteristics of areal topography measuring instruments is established through measurements of material measures [24]. An additional metrological characteristic has also been introduced recently to model the interaction between the surface under measurement and the instrument (topography fidelity—such quantity was commonly referred to in jargon as surface-instrument convolution). Eventually, the correction of the systematic effect is achieved as adjustment from the deviations with respect to the measured material measures. The state of the art of the metrological characteristics’ framework has been illustrated in a thorough review [25].

The use of metrological characteristics has many advantages. It is a unique framework with clear recommendations that could be adopted easily in any environment. Nonetheless, some ambiguities are still unsolved.

There are no indications regarding the topography fidelity assessment, and what kind of contributor it should be [24, 25]. The correction of the systematic behavior (systematics) is proposed as adjustment, while it should be sought in the instrument-surface interaction, which the topography fidelity aims to describe. In addition, the correction appears quite partial, achieved on the known systematic behavior in comparison with material measures. Most important, the whole framework relies on material measures, which are inescapably affected by manufacturing imperfections, and by quick aging and defects depending on the material used for the manufacture (nickel, steel, silicon substrates, etc.). They are also expensive, difficult to find in small dimensions and high aspect ratio above all at the nano scale, forcing to an extrapolation that enlarges even more the undetected variability. Material measures for use in manufacturing are classified in the geometrical product specification (GPS) system (see ISO25178–70 [26]), while typical examples and extensive considerations can be found elsewhere [27–31].

The general correction of the instruments’ systematic behavior subject matter of this paper aims to be in agreement with the international recommendations of the well-known ‘Guide to the expression of uncertainty in measurement’ (GUM—also GUM approach) [32], where the finally uncertainty statement is consequence of the correction of the systematics modeled by frequentist statistics.

The GUM approach in micro and nano metrology has already been considered in past works, mainly for the profile method [33–35].

Krystek [33] analyzed the case of roughness measurements, and investigated the propagation of the measurement uncertainty when filtering was applied. He also explained that filtered data are always correlated even if the unfiltered ones are not.

Morel [34] and Haitjema [35] pointed out the impossibility of applying the GUM approach to geometric and roughness measurements providing alternatives based on simulation. Haitjema also defined a type A model for the uncertainty assessment of the height values in areal topographical acquisitions, considered correlated. Nonetheless, the final equations resulted unmanageable because of the excessive amount of data. The correction of the systematics was not performed in any of the cases, but Morel and Haitjema considered the compensation of influence factors in line with the instruments’ metrological characteristics (amplification, linearity, residual flatness, noise, spatial resolution and probe size).

Eventually, it is worth to mention Moroni et al [36], who instead simulated the systematic behavior in 3D microscopy to be accounted in a task-specific uncertainty estimation.

A gap between the GUM recommendations and what is considerable achievable in manufacturing metrology nowadays appears evident. On the one hand, the GUM does not provide any indication on how to deal with measurements at the micro and nano length scales. These measurements are usually obtained as matrices of height values mapping portions of real surfaces (micrographs). This circumstance appears to be disregarded in such international recommendations. On the other, the latest developments in the context of micro and nano metrology—such as ISO 25178—do not take into account the correction of the unknown systematic behavior and the definition of a model equation, which are distinct recommendations in the GUM. Instead, the entire framework relies on calibrated artifacts that are inevitably affected by manufacturing imperfections. Therefore, the final uncertainty related to the sought non-ideal behavior of the instrument may result overestimated and inadequate for quality assurance.

Henceforth, this paper pursues clarifications about the correction of the unknown systematic behavior applied to micrographs. Thus, doubts and misconceptions related to the common practice are excluded one after the other, whereas the essential issues to be clarified remain in focus. Moreover, differences between the GUM and the recommendations in ISO 25178-600 and ISO 25178-700 are also emphasized throughout the paper whenever the opportunity arises (at the time of writing ISO 25178-700 is approved for registration as Final Draft International Standard).

1.1. Underlying hypotheses and structure of the paper

The GUM deals with quantities from any instrument, provided that a single measurand is a scalar [32]. Hence, micrographs produced by micro and nano scales instruments would not be suitable for the GUM approach because of the complexity of such measurands. Indeed, an ambiguity may arise from the fashion the measurands are defined. A micrograph is
essentially a matrix collecting a set of height values (related to pixels of an imaging sensor), which altogether are a three-dimensional representation of a portion of a surface [37]. To determine the metrological perspective, averages (arithmetic, integral, etc.) are extracted from the set of height values (matrix elements) according to the sought functional characteristics. Therefore, the work in the paper was carried out with the fundamental hypothesis that the analysis performed is valid on averages of matrix elements, and that such averages perpetuate the systematic behavior. The validity of such hypothesis is clarified in the discussion (see section 5).

In order to cope with the micrographs’ complexity the topic of correcting the systematic behavior at the micro and nano scales was carefully formulated and narrowed down through a sequence of three case studies, proven on surface topography measurements. In detail, the topography measurements were assessed by the areal parameter $S_q$ (quadratic integral average of micrographs’ elements) [38], and by step height measurements evaluated using the histogram technique (average distance between two distributions of elements in the same micrograph) [37]. Concerning the data processing, all the micrographs were assessed equally using the same commercial software [39]. The plane correction of the micrographs was achieved subtracting the least-square fit plane. Moreover, no filtering was applied. The application of filters to topography measurements is discouraged [37], especially concurrently with the application of the frequentist approach, because it would reshape the experimental distribution and, therefore, hinder the statistical analysis. An exception is the form correction, which is a merely non-destructive translation of pixels in the micrographs. Nonetheless, the form correction was not required in any of the considered cases. An outlier-removal method was occasionally applied to the extracted averaged measurands. The presence of outliers in the experimental distribution would counteract the efficacy of the normality tests. Barbato et al suggested diverse outlier-removal methods in a comprehensive work [40]. Who is writing deemed the Chauvenet’s method more efficient with surface topography areal averages. Chauvenet’s criterion considers a conventional 50% probability of having one or more rough errors in the whole data set. When applied, no more than two iterations of the criterion were fulfilled.

Finally, being the topic strictly related to metrology tasks, the manufacturing processes of the specimens used are only briefly introduced, and thus referred to dedicated sources for more details. Hence, the paper is structured as follows.

Section 2 describes the correction of the systematics carried out within the context of what is diffused routine in the field of metrology for manufacturing. Then, section 3 analyzes the effect on the systematics correction when the measurement conditions adhered to what was instead considered good practice in the field. The following section 4 summarizes the results of the previous two sections in a typical surface characterization task, and proves the reduction of the final measurement uncertainty in consequence of the systematics correction. Thus, the sequence of case studies is discussed thoroughly in section 5. Eventually, conclusions are drawn in section 6.

2. Dimensional and topographic metrology of micro structured tool inserts

Dimensions and micro topographies of proof-of-technology micro mold inserts (PoT) were case in point for an exploratory application of frequentist statistics to surface topography measurements. The PoT micro mold inserts are the two steel micro cavities shown in figure 1 as PoT-1 and PoT-2. They have been produced in steel by additive manufacturing and, successively, structured by Jet Electro-chemical Machining (Jet-ECM), in the context of the EU project Hi-Micro (European Commission’s 7th Framework Program) [41]. Details of the manufacturing process can be found in separate papers [42, 43], as well as the processing of the experimental data [44].

2.1. Analytical framework

The goal in this first instance was to establish a general method for analyzing and, eventually, correcting possible divergences among dimensional and surface topography measurements acquired using different optical instruments. Areal acquisitions from different optical technologies require considering specific critical aspects to perform a correct processing of the measurement data sets, and to allow for an effective comparison among the different instruments. Such matter has already been emphasized in past works. Mattsson et al [45] showed that agreement among surface roughness measurements can be limited by both inaccuracies in the definition of the acquisition areas, and the data set assessment, or post-processing. Moreover, Leach et al [46] stated that a comparison among heterogeneous measurements (i.e. measurements from different topography measuring instruments) is only possible when they have the same bandwidth. The bandwidth has been intended in the frequency domain, i.e. the one resulting from a Discrete Fourier Transform (DFT) applied to the set of height values in a micrograph. In the space domain it corresponds to the micrographs’ sampling interval. Since the sampling interval may be non-uniform when considering the $x$ and $y$ directions in the micrographs’ reference area, henceforth, it will be generically referred as quantization level. Thus, the same quantization level can be achieved practically when the areal acquisitions have (or have been tailored to) the same sampling length and sampling width or, equivalently, the same field of view and same number of height.
elements per direction. Yet, the purpose was to test the eligibility of the approach, inspecting if discrepancies, which are commonly due to the instrument-operator chain (magnification, quantization levels, etc.), could be corrected as systematic differences against a calibrated reference (CI). In fact, disregarding differences in heterogeneous micrographs is still a quite diffused habit when an exchange of topography measurements happens among laboratories with different capabilities for facing the same metrology task. Therefore, in this first implementation, the choice of the magnification (thus, of the quantization level) was left to the single operator.

2.2. Description of the experiment
Measurements of the micro cavities have been carried out inside the straight groove machined on PoT-1 according to the areas indicated in figure 1(a), namely the height of the steps identified as A11, A13 and A15, and the surface texture indicated as A12 and A14. While the sectioned surfaces on PoT-2 shown in figure 1(b) comprised the measurand areas A21–A24 (surface texture) and A25–A210 (height of step). The regions of areal acquisition were identified specifically by defining the on-specimen reference systems shown in figure 1. The height of the steps $H$ was extracted from the micrographs by the histogram technique, namely as distance between the two planes containing the maximum of sampled point occurrences at the corresponding height levels [37]. While the surface texture measurands $S_q$ were evaluated averaging the height values in the corresponding micrographs as root mean square value (RMS—$S_q$ areal parameter according to ISO25178-2 [38]).

Different measurement sessions were performed in three different laboratories using respectively focus-variations microscopy (FV), laser scanning confocal microscopy (LSC) and coherent scanning interferometer microscopy (CSI). In addition, contact measurements by a calibrated stylus instrument (CI) were used as reference for indirectly achieving the traceability. Thus, the measurement uncertainty was garnered from the correction of the systematic behavior, as the discrepancy between the optical instruments’ measurements and the calibrated contact ones (rough data are in appendix A).

2.3. Statistical modeling
The most favorable situation for the evaluation of the measurement uncertainty is a normal distribution of the experimental data. In such condition, the dispersion of the data is related to the combined influence of several small random causes, which is also the condition considered by the Central Limit theorem (CLT) [47]. Formal definition and different perspectives of the CLT can be found in subject-matter books [48–50]. An operational definition explains that independent random variables gathered from (almost) any distribution with mean $\mu$ approximate a normal distribution with the same mean, when their number tends to infinite. The only exception to the previous statement is regarding variables following the Cauchy distribution. In that case, in fact, the hypotheses of the CLT are not met [49]. Nonetheless, this statement can be considered valid in many situations dealing with experimental data. Specifically, a large number of replications of a measured quantity (i.e. random variables), which are independent in a sampling distribution (viz. experimental distribution), tends as a whole to a normal one. Thus, a large number of random variables, allegedly independent, altogether rejecting the normality reveals that a number of them might not be independent. A mathematical relationship can model the dependency among the random

![Figure 1. Overview of the proof-of-technology micro mold inserts by laser scanning confocal microscope. The regions of areal acquisition were identified over the specimens with respect to a local reference system. Specifically, acquisitions areas were in: (a) A machined straight groove (PoT-1). (b) Two sectioned surfaces at different height (PoT-2).](image)
variables inside an experimental distribution of measured quantities. Such relationship corresponds to the sought systematic behavior. Hence, the mathematical relationship subtracted from the experimental distribution allows to minimize the systematic behavior. The residual random variables are expected to be statistically quasi-independent, i.e. approaching a normal distribution [47, 51].

The large number of independent random variables (infinite in the mathematical formulation of the theorem) that allows for an approximated, nonetheless effective, application of the CLT can essentially be found comparing the $t$ distribution—used for small size samples—and the normal one at the same confidence level of 95%. The two distributions show the same trend with acceptable approximation when the $t$ distribution has degrees of freedom (DoF) from 10–15 and above. Hence, as a rule of thumb, such inferential statistics should be applied to experimental distributions with more than 10–15 repeated measurements per measurand.

Regarding the PoT, even though the measurement data were a substantial number when considering the areal acquisitions over the specimens, they were limited in terms of repeated measurements and unbalanced, i.e. discordant in terms of replicated measurements (see the raw data after the assessment of the data sets (post-processing) in appendix A, tables A1 and A2). To obtain a consistent number of data, and in agreement with the starting hypothesis of correcting the systematics for heterogeneous measurements, the results of the post-processing were normalized to their respective areal averages (subtraction of the corresponding averages in $A11$, $K$, $A210$, respectively). The normalized deviations are in appendix A, tables A3 and A4. They are also pictured in the box-plots in figure 2(a) and −b in terms of inter-quartile range (IQR) of the two groups of measurands.

The exclusion principle was applied next. In order to avoid the risk of excluding also suitable results, a full understanding of the experimental data was attained monitoring box-plots and histograms for, respectively, inspecting agreement among the experimental data and visualizing their distribution (see figure 2 and 3). The elimination was confirmed by the presence of disturbances in corresponding micrographs. The values excluded by two iterations of the Chauvenet’s criterion were four, and are indicated in parenthesis in the mentioned tables of appendix A. After the exclusion, the deviations were 41 representing the step

---

Figure 2. Box-plots of deviations with inter-quartile range (box), maximum (upper whisker), minimum (lower whisker) and median (column in the box) values. (a) Deviations $devH$ of step height before outliers removal. (b) Deviations $devSq$ of root mean square value before outliers removal. (c) Deviations $devH$ of step height after outliers removal (Chauvenet’s exclusion limits $x_{\text{min}} = -3.8 \mu m$, $x_{\text{max}} = 4.1 \mu m$). (d) Deviations $devSq$ of root mean square value after outliers removal (Chauvenet’s exclusion limits $x_{\text{min}} = -0.9 \mu m$, $x_{\text{max}} = 1.0 \mu m$).
heights and 38 the surface texture, and almost all the IQRs were in good agreement in each respective group, which means that the exclusion was effective. An exception was the area $A_{21}$. A closer look to the surface (see figure 4(b) in the next section 3) revealed some waviness only present in this measurand area. Nonetheless, no further value was evidenced by the exclusion principle. Moreover, no reasons were found to justify a third iteration of the Chauvenet’s criterion.

After the outliers removal, a departure of the experimental distributions from the comparable normal ones (theoretical normal distributions with same average and standard deviation of the corresponding experimental ones) can be deducted by inspecting the histograms in figure 3(b) and −d. In addition to the histograms, the normal probability plots (NPP) have trend clearly far from a straight line (see figure 5). Looking at the extremes of the curves, they show different trends of slope in both diagrams. Moreover, the slopes are not confirmed in the middle, where an arc is present. It is more pronounced and concave downwards for the deviations $devH$, which evidences a right asymmetric distribution (right-skewed). The arc is instead almost unnoticeable with opposite orientation for the deviations $devSq$, confirming a left-skewed distribution. Nonetheless, it is difficult to analyze completely the NPP curves because of their complexity, which suggests that different slopes piecewise can be

![Figure 3.](image)

**Figure 3.** Histograms of the experimental distribution, and comparable curves of the theoretical normal distribution for the deviations $devH$ of step height—before (a) and after (b) the outliers removal (Chauvenet’s exclusion limits $x_{L_{\min}} - 3.8 \mu m, x_{L_{\max}} = 4.1 \mu m$)—and for the deviations $devSq$ of root mean square value—before (c) and after (d) the outliers removal (Chauvenet’s exclusion limits $x_{L_{\min}} - 0.9 \mu m, x_{L_{\max}} = 1.0 \mu m$).

![Figure 4.](image)

**Figure 4.** Close-up of the measurand areas $A_{13}$ (a) and $A_{21}$ (b).
recognized (multimodal distributions), thus presuming that the systematic behavior was due to several factors.

An analysis of variance (ANOVA) was performed to relate the systematic behavior to the factors involved in the measurements. Being the data sets unbalanced, a general linear model was used to implement the ANOVA test, considering the three factors: instrument, magnification and measurand area. The test was performed using adjusted sum of squares, which means that the test was not considered dependent on any order of the factors into the model (the laboratories did not provide information regarding the order of acquisition). The ANOVA test is summarized in Table 1. Moreover, the experimental data were not enough for also considering the factors’ interactions. For the sake of completeness, the ANOVA test was also performed using sequential sum of square, in which case the instrument was also an influence factor for devSq (p-val \(\approx 0.02\)). Since no information was available about the sequence in which the data were acquired, the given order was an arbitrary one, nonetheless quite suitable to show a possible incompleteness of the test. In addition, the coefficients of determination \(R^2\) associated to the ANOVA models were not very robust, thus the models were not well fitted.

To solve this indeterminacy, the joint effect of random and systematic factors was described as a mixture of normal distributions \([52]\). When random and systematic factors are both present, in fact, the overall central tendency slightly split in sub-groups, each one dominated by a specific influence factor. Thus, the histogram envelope (kernel distribution—i.e. the probability density estimation of a measurand) was decomposed using theoretical normal distributions as individual kernels. The decomposition (kernel density estimation—KDE) was obtained performing a parametric mixture of individual kernels on the histogram data, and optimizing the Pearson’s \(\chi^2\) test. In detail, the individual kernels’ parameters in the mixture (average, standard deviation and percentage of incidence) were estimated minimizing the \(\chi^2\) statistics. They are given in the tables 2 and 3 for the measurands’ deviations. The number of individual kernels in the mixture was initially set as the number of systematic factors in the ANOVA test, and increased iteratively if required. Figure 6 shows the mixtures of three individual kernels for each measurand’s deviations, and the comparison between the estimated mixtures and the kernel distributions. As shown in the figure, the mixtures describe the kernel distributions’ shape with satisfactory fidelity. This suggested the presence of three dominant influence factors, even though the presence of other unknown systematic factors could not be excluded \([51, 48–50]\).

| Instrument | devH Influence (p-val < 0.01) | devSq No influence (p-val \(\approx 0.16\)) |
|------------|-------------------------------|---------------------------------------------|
| Magnification | Influence (p-val \(\approx 0.03\)) | Influence (p-val < 0.01) |
| Area | No influence (p-val \(\approx 0.69\)) | No influence (p-val \(\approx 0.25\)) |
| \(R^2\) | 61% | 84% |

Figure 5. Normal Probability Plot (NPP). (a) NPP of deviations of step height devH. (b) NPP of deviations of root mean square value devSq.

Table 1. General linear model ANOVA of deviations devH and devSq (adjusted sum of squares—confidence level 95%).

Table 2. Parameters of the individual kernels in the mixture of deviations devH/\(\mu m\) (the optimised \(\chi^2\) statistics is 0.53).

| Individual ker. | Average /\(\mu m\) | St. Dev./\(\mu m\) | Percent. Of incid./% |
|-----------------|------------------|-----------------|---------------------|
| \(\delta_1\) | 0.79 | 0.99 | 70 |
| \(\delta_2\) | −1.76 | 0.87 | 27 |
| \(\delta_3\) | 3.33 | 0.45 | 3 |
2.4. Modeling the systematic behavior

A measurement model can be selected among theoretical functional relationships, empirical ones or a mix of the two. Indirect measurements can be modeled by their functional relationships. More complex is the definition of a model when dealing with direct measurements. A complete discussion on the different types of modeling is beyond the purposes of this paper, and can be found elsewhere [53]. The author deemed more efficient to tackle the measurement modeling at the micro and nano scales by fitting statistical data.

Figure 7 shows the deviations $devH$ and $devSq$ represented according to the sequence arbitrarily chosen for the current investigation. Their distributions clearly show tendencies that could be identified by regression models, and successively corrected (this case is widely discussed in the next section 3).

In the procedure proposed in this section, the correction was instead achieved against the CI reference measurements. This choice was intended for examining any detectable systematic behavior and, concurrently, establishing the traceability. Thus, a least squares regression was implemented onto the non-normalized values of deviations (de-normalized by their respective areal averages) to evaluate discrepancies between optical and contact reference measurements CI. The model equation found consistent with the experimental data of both step height $H$ and RMS $Sq$ was a straight line passing by the origin. In fact, if different instruments measure the same quantity, ideally, the results should equal (unitary angular coefficient). Hence, the best fit evaluated the mismatch between $x_{OPT}$, indicating any optical instrument measurements, and the reference $x_{CI}$ with respect to the unitary slope

$$x_{OPT} = q \cdot x_{CI}$$

Equation (1) was found consistent with the experimental data to mathematically describe the best-fit regression. Other factors influencing the evaluation were the reproducibility $\varepsilon_{Rep}$, i.e. the standard deviation of the regression results, and the resolution $\varepsilon_{Res}$. Thus, the metrological model equation used for the uncertainty evaluation was

$$y_{OPT} = q \cdot x_{CI} \pm \varepsilon_{Rep} \pm \varepsilon_{Res}$$

where reproducibility and resolution were considered random variables with null average. This is in line with GUM recommendations that require to consider all the relevant quantities for the evaluation of the uncertainty even if they are not immediately included in the mathematical model.

The instrument resolution is usually included in the reproducibility through repeated measurements (unless a coarse increment in basic instruments requires to be treated separately). Nevertheless, separate values for reproducibility and resolution may be considered when they refer to different contributors in the metrological chain, which is actually the case considered in this section. In fact, despite the stylus instrument resolution could be considered non-influential, the raw CI measurements were also analyzed by a post-processing software [39]. For this reason, the resolution was set equal to the numerical precision of the post-processing software. This was a conservative choice because, even though unlikely, the software significant figures might limit the instrument resolution.

Hence, the standard uncertainty was calculated propagating the uncertainty contributors to (2) by the usual method of combination of variances for uncorrelated quantities, where the contributors were as follows:

- The accuracy of CI ($U = 0.112 \, \mu m$, unfiltered) stated in the calibration certificate of the instrument for the range of interest (cf appendix C, too).
- The standard deviation of the coefficient of the model equation assessed in the best fit regression (see $\sigma_{q}$ in table 4).
- The reproducibility of the regression (i.e. the standard deviation of the residuals—see table 4).
- The post-processing software numerical precision (estimated to be 1 nm—conservative choice).

Finally, the expanded uncertainty was evaluated as the confidence interval corresponding to the conventional confidence level of 95%, and by a coverage factor calculated using the t(distribution with degrees of freedom given by the Welch-Satterthwaite formula [32].

2.5. Results of the correction

Equation (2) is the correction that was subtracted from the experimental data. The residuals after the correction are expected to exhibit random fluctuations, i.e. a quasi-horizontal trend line. If the tendency is not horizontal, the correction cannot be considered satisfactory. Such case is widely discussed in section 5 (see 7c). The results of the correction are summarized in the tables 5 and 6, with values of the reference CI, of the optical instruments and of the evaluated expanded uncertainties. The residuals of the correction are in figure 8. As revealed by the residuals’ trend, the correction was not very effective, achieving in both cases a partial compensation of the systematic behavior. Therefore, for highlighting the efficiency of each instrument-operator chain an uncertainty was also
evaluated considering a separate regression model for each optical instrument. Such uncertainties were evaluated similarly to the expanded uncertainty, nonetheless, they just had the intent to substantiate the performance of each single instrument-operator chain. It was found 

\[ ULSC \approx 3 \mu m, \quad UCSI \approx 4 \mu m \]

and 

\[ UFV \approx 6 \mu m \]

for step height measurements (maximum values). While it was 

\[ ULSC \approx 1.7 \mu m, \quad UCSI \approx 0.7 \mu m \]

and 

\[ UFV \approx 0.6 \mu m \]

for RMS measurements (maximum values). Thus, the closest instrument to the reference CI was LSC for step height measurements, and FV for RMS measurements.

3. Correction of the systematic behavior vs time sequence

In the previous instance, because of a shortfall of measurements, the correction of the systematic behavior and the establishment of the traceability were only possible against a reference (calibrated/calibratable instrument), matching the correspondent measurements. Nevertheless, if reference and instrument under test show incompatible trend the correction becomes inadequate (or impossible, as in the previous case).

Indeed, the development of a systematic behavior is to be sought in the non-ideal behavior and use of a measurement instrument, in actual environmental conditions, and independently from any reference measurements. Hence, such behavior is expected developing during the time of use, building up at each occurrence of the measurement events and, thus, is strictly related to their acquisition sequence.

To better inspect this matter, the measurand areas A13 (height of step—Figure 1(a)) and A21 (RMS value of texture—figure 1(b)) of the same specimens in the previous study were re-measured and corrected for the systematics according to the time sequence of acquisition (see also close-up of those measurand areas in figure 4). Several repeated measurements were performed in laboratory conditions by FV using different magnifications \((5 \times, 10 \times, 20 \times, 50 \times \text{and } 100 \times)\). Twenty-five repeated areal acquisitions were in the measurand area A13, and thirty in A21. Rough data for both measurand areas are in appendix B, tables B1 and B2, after the post-processing of the micrographs and value extraction. Details of the data processing can be found elsewhere [34].

Despite the Chauvenet’s criterion did not evidence any outlier, inspecting the box-plot representations both measurands show only reduced agreement, being it mostly for measurements from \(20 \times\) and \(50 \times\) objective lenses (see figure 9). Moreover, the histograms in figure 10(a) and -c demonstrated non-normal multi-modal trend of the experimental distributions. The issue was examined observing the following. Being the systematic behavior connected to the acquisition sequence, each time the instrument was reset (e.g. after being turned off) the sequence was interrupted. Thus,
the causes influencing systematically were ceased for that session. Since random and systematic factors act together, it is unlikely that a new acquisition sequence would have exactly the same trend of a previous one. In the specific case, the acquisitions were completed in six days, three per each specimen. The use of the instrument was continuous in each measurement day, and restarted the successive one. This particular condition was concisely termed day of acquisition (or simply day), and analyzed by a general linear model ANOVA considering the factors day of acquisition and magnification (sequential sum of squares). The test indicated that both influence factors were significant for both measurands (results in table 7).

A KDE successively confirmed the presence of two dominant influence factors. The associated individual kernels were identified by comparing their central values with the averages of the results corresponding to the sub-groups day and magnification. They are shown in figure 11, while the corresponding parameters are in the tables 8 and 9.

Unlike the previous case of section 2, in this instance the reference was not used in the regression, and the correction of the systematic behavior was performed with respect to the repeated measurements’ time sequence \( t_r \). Afterwards, the traceability was achieved as average distance of the experimental distribution from the average trend of the corresponding contact measurements \( x_{CI} \). Thus, the metrological model equation, giving the values corrected for the systematics, was

\[
\gamma_{OPT} = x_{CI} + x_{OPT} - x_{reg} \pm \epsilon_{Rep} \pm \epsilon_{res}. \quad (3)
\]

where \( x_{OPT} \) are the optical measurements, and \( x_{reg} \) is the mathematical model equation found consistent for fitting the experimental distributions. A first order polynomial was found for the measurand area A13:

\[ x_{reg\,H} = a + b \cdot t_r; \]

and a second order one for the measurand area A21:

\[ x_{reg\,s} = a + b \cdot t_r + c \cdot t_r^2 \]

The related parameters are in table 10.

For the sake of clarity, it should be noted that (2) is different from (3). The first one, if inverted, gives an estimate of a reference measurement, and the related uncertainty is consequence of the regression model. Equation (3), instead, achieves the correction of the systematic behavior as function of the time sequence \( t_r \) and the accuracy as distance from the reference (cf also with (4) in section 3.1, and (5) in section 4).

Similarly to section 2, the standard uncertainty was calculated propagating the uncertainty contributors to (3) by the usual method of combination of variances for uncorrelated quantities, where the contributors were as follows:

- The accuracy of CI, using the values \( H_{CI} = (162.30 \pm 0.11) \mu m \) and \( S_{CI} = (5.19 \pm 0.11) \mu m \) from the previous section 2.
- The standard deviation of the coefficient of the model equation assessed in the best fit regression (see table 10).
- The standard deviation of the residuals (see table 10).
- The post-processing software numerical precision (estimated to be 1 nm—conservative choice).

Finally, the expanded uncertainty was evaluated as the confidence interval corresponding to the conventional confidence level of 95%, and by a coverage factor calculated using the t(d)distribution with degrees of freedom given by the Welch-Satterthwaite formula [32].

The results are summarized in the tables 11 and 12 for the sub-groups of influence factors. Moreover, the residuals of the correction are in figure 12. The residuals’ trend indicated an effective correction (no tendency in the residuals’ distributions). Nonetheless, inspecting the histograms of the same distributions.
after correcting for the systematics (see figure 10(b) and (d) a partial compensation of the systematic behavior was observed in both cases. It was judged not acceptable for the measurand area $A_{2I}$.

3.1. Equalization of the micrographs’ quantization level

As a consequence of the results in measurand area $A_{2I}$ of section 3, the corresponding micrographs were all reviewed. The original measurements were matrices of several fields of view, sized to have approximately similar areas, but still heterogeneous. Taking care of the alignment, they were all re-sized to an area of $9320 \times 7200$ pixels, with pixel size of about $0.193 \, \mu m \times 0.181 \, \mu m$. Thus, the analysis was repeated on the equalized micrographs in the same fashion as in section 3. The corresponding raw data are in appendix B, table B3, after post-processing and value extraction.

The Chauvenet’s criterion indicated four outliers, two in the sub-group of $10 \times$ magnification and other two in the sub-group $100 \times$ magnification. The outliers were eliminated and replaced by the median of all the other values.

The histogram of the experimental distribution in figure 13 indicated a less uneven shape. Although, the results of the ANOVA test still confirmed the acquisition day and, surprisingly, the magnification as significant influence factors (see table 13—also, cf section 5).

The parameters of the regression for the correction are in tables 14, while the results for the sub-groups of influence factors are in table 15. Comparing figure 14

---

**Table 5.** Results of the step height measurements $H/\mu m$ for each measurand area, and related expanded uncertainties. The decimals in parentheses explicate the approximation on the last digit.

| CI  | LSC5× | LSC10× | CSI10× | CS120× | CS130× | FV20× | FV50× | U    |
|-----|------|-------|--------|--------|--------|-------|-------|------|
| A11 | 162.6| 162.1 | 162.1  | 160.0  | 160.2  | 158.9 | 4.3(0)|      |
| A13 | 162.3| 163.9 | 165.7  | 163.5  | 161.3  | 161.1 | 4.3(0)|      |
| A15 | 164.2| 167.3 | 169.0  | 167.6  | 165.2  | 163.7 | 4.3(0)|      |
| A25 | 97.3 | 96.2  | 95.9   | 96.7   | 91.6   | 91.6  | 4.2(0)|      |
| A26 | 84.8 | 83.9  | 83.7   | 84.8   | 85.2   | 85.2  | 4.2(0)|      |
| A27 | 177.2| 177.8 | 177.1  | 176.0  | 177.2  | 177.2 | 4.3(1)|      |
| A28 | 85.5 | 85.6  | 86.3   | 84.5   | 83.1   | 83.1  | 4.2(0)|      |
| A29 | 163.1| 163.9 | 161.5  | 162.6  | 158.1  | 158.1 | 4.3(0)|      |
| A210| 80.2 | 80.9  | 81.9   | 80.8   | 78.6   | 78.6  | 4.2(0)|      |

**Table 6.** Results of the RMS measurements $Sq/\mu m$ for each measurand area, and related expanded uncertainties.

| CI  | LSC5× | LSC10× | LSC100× | CSI10× | CS120× | FV20× | FV50× | U    |
|-----|------|-------|---------|--------|--------|-------|-------|------|
| A12 | 1.10 | 1.06  | 0.81    | 1.47   | 0.98   | 0.87  | 0.90  |      |
| A14 | 0.98 | 0.80  | 0.76    | 1.52   | 0.84   | 0.77  | 0.90  |      |
| A21 | 5.19 | 5.25  | 5.97    | 5.51   | 4.98   | 5.23  | 0.95  |      |
| A22 | 0.78 | 1.91  | 0.44    | 0.36   | 0.91   | 0.77  | 0.54  | 0.90  |
| A23 | 0.49 | 1.69  | 0.44    | 0.41   | 0.80   | 0.53  | 0.72  | 0.52  | 0.89  |
| A24 | 0.44 | 1.58  | 0.40    | 0.24   | 0.78   | 0.45  | 0.34  | 0.47  | 0.89  |
with figure 12(b), it is worth to note that the variability of the residuals (and, thus, the reproducibility) was reduced to about half range in consequence of the equalization.

3.2. Traceability by material measures
The use of a reference instrument for establishing the traceability may lead to the undesired possible presence of additional influence factors. A prevailing solution for the achievement of the traceability in manufacturing is through measurement standards that are material measures [55]. Measuring calibrated material measures with measurands’ dimensions very close to the quantities under investigation allows for the estimation of the instrument’s performance, thus, establishing the traceability. This was already noticed in the Introduction, concerning the metrological characteristics.
Figure 11. Comparison between mixture and estimation of the probability density of the experimental data (kernel). (a) Mixture of step height $H$. (b) Mixture of RMS values $Sq$. (c) Comparison between kernel and mixture of step height $H$. (d) Comparison between kernel and mixture of RMS values $Sq$.

Table 8. Parameters of the individual kernels in the mixture related to measurand area $A13$ (optimised $\chi^2$ statistics is 0.62).

| Dominant factor | Individual ker. | Average $/\mu m$ | St. Dev. $/\mu m$ | Percent. Of incid. $/%$ |
|-----------------|-----------------|------------------|-------------------|-------------------------|
| Day             | $\delta_1$      | 162.16           | 0.40              | 47                      |
| Magnification   | $\delta_2$      | 163.49           | 0.39              | 53                      |

Table 9. Parameters of the individual kernels in the mixture related to measurand area $A21$ (optimised $\chi^2$ statistics is 3.11).

| Dominant factor | Individual ker. | Average $/\mu m$ | St. Dev. $/\mu m$ | Percent. Of incid. $/%$ |
|-----------------|-----------------|------------------|-------------------|-------------------------|
| Day             | $\delta_1$      | 5.15             | 0.06              | 76.5                    |
| Magnification   | $\delta_2$      | 5.55             | 0.06(3)           | 23.5                    |

Table 10. Parameters of the least square regressions. Intercept $a$, slope $b$, quadrasic term $c$, standard deviation of the intercept $\sigma_a$, standard deviation of the slope $\sigma_b$, standard deviation of the quadrasic term $\sigma_c$, reproducibility $\epsilon_{rep}$ (standard deviation of the residuals), degrees of freedom and coefficient of determination $R^2$.

| $x_{reg}$ | $a/\mu m$ | $b/\mu m$ | $c/\mu m$ | $\sigma_a/\mu m$ | $\sigma_b/\mu m$ | $\sigma_c/\mu m$ | $\epsilon_{rep}/\mu m$ | DoF | $R^2/%$ |
|-----------|-----------|-----------|-----------|------------------|------------------|------------------|------------------------|-----|---------|
| $x_{reg} H$ | 162.83    | -0.058    | —         | 0.13             | 0.02             | —                | 0.63                   | 23  | 32      |
| $x_{reg} Sq$ | 5.03      | 0.01      | 0.002     | 0.03             | 0.002            | 0.003            | 0.11                   | 27  | 70      |

Table 11. Step height results for the measurand $A13$ after the correction. The values are for the sub-groups of influence factors, and related expanded uncertainties.

| $5\times$ | $10\times$ | $20\times$ | $50\times$ | $100\times$ | Day 1 | Day 2 | Day 3 |
|-----------|------------|------------|------------|------------|------|------|------|
| $H/\mu m$ | 162.63     | 162.89     | 161.58     | 162.25     | 162.14| 162.38| 162.32| 162.11|
| $U/\mu m$ | 1.41       | 1.38       | 1.37       | 1.38       | 1.41  | 1.41  | 1.38  | 1.41  |
Differences from the calibrated values can also be compensated in the model equation by the substitution method. For instance, the substitution method for uncertainty evaluation based on material measures, and concerning coordinate measuring machines, is described in the GPS standard ISO 15530-3 [56].

The use of calibrated material measures implies two sessions of measurements with the same instrument: measurements of the material measure $x_{\text{ref, gauge}}$ and measurements of the component under characterization $x_{\text{OPT}}$. Comparing $x_{\text{ref, gauge}}$ with the calibrated values in the calibration certificate $x_{\text{ref, cal}} \pm U_{\text{ref, cal}}$ the traceability for the characterization measurements can be achieved using a calibration factor given by the ratio $F_{\text{cal}} = x_{\text{ref, cal}} / x_{\text{ref, gauge}}$. Supposing to correct the systematics for both sets of measurements (same number of repeated measurements performed) with a first-order polynomial, namely $x_{\text{reg, ref, gauge}} = q_{g} \cdot t_{s} \pm \epsilon_{\text{Rep, gauge}}$ and $x_{\text{reg, OPT}} = p_{\text{OPT}} \cdot t_{s} \pm \epsilon_{\text{Rep, OPT}}$ the metrological model equation is

$$y_{\text{OPT}} = F_{\text{cal}} \cdot x_{\text{reg, OPT}} = \frac{x_{\text{ref, cal}}}{x_{\text{reg, ref, gauge}}} \cdot x_{\text{reg, OPT}} = x_{\text{ref, cal}} \cdot \frac{p_{\text{OPT}} \cdot t_{s} \pm \epsilon_{\text{Rep, OPT}}}{q_{g} \cdot t_{s} \pm \epsilon_{\text{Reg, gauge}}} \quad (4)$$

The uncertainty contributors to be referred to (4) are

- The standard deviations of the residuals of the best fit regressions ($\epsilon_{\text{Reg,OPT}} / \epsilon_{\text{Rep,OPT}}$).
- Uncertainty contributors of other possible influence factors not in the model equation (random variables with null average).

4. Correction of the systematic behavior in surface characterization

The results of the previous case showed a significant reduction of the measurement uncertainty as a consequence of the correction of the systematics, although the elimination of the influence factors was not always complete. Therefore, aim of the last instance was to test efficacy and generality of the measurement uncertainty possible reduction after the systematics correction.

The correction of the systematic behavior was implemented on the texture parameter $S_{q}$ of the topographic measurements of four different polished surfaces [38]. The approach was validated comparing the associated measurement uncertainties with the one of the same measurements and same uncertainty contributors assessed without correction.

The four surfaces were polished as follows [57]:

- T1: grade #15 diamond buff (nominal $Ra$ interval 51–76 nm).
- T2: 320 grit paper (nominal $Ra$ interval 229–254 nm).
- T3: 400 stone (nominal $Ra$ interval 254–305 nm).

|      | 5x | 10x | 20x | 50x | Day 1 | Day 2 | Day 3 |
|------|----|-----|-----|-----|-------|-------|-------|
| $S_{q} / \mu m$ | 5.23 | 5.10 | 5.21 | 5.07 | 5.40 | 5.18 | 5.22 | 5.16 |
| $U / \mu m$     | 0.45 | 0.44 | 0.44 | 0.45 | 0.46 | 0.44 | 0.44 | 0.45 |

Table 12. RMS results for the measurand $A_{21}$ after the correction. The values are for the sub-groups of influence factors, and related expanded uncertainties.

Figure 12. Residuals related to measurand area $A_{13}$ (a), and to measurand area $A_{21}$ (b), both represented as function of the acquisition sequence. The straight line indicates the horizontal tendency.
• T4: glass bead #11 dry blast (nominal Ra interval 635–711 nm).

At first, they were measured by a reference CI according to the on-sample reference system shown in figure 15(a) and −b, where the dominant texture was oriented orthogonally to the scanning direction (y(a)x) axis). Five acquisitions of a single evaluation area of 4 mm × 10 mm were performed on each surface, sampled at 8200 pixels along the x(a)xis, and 21 profiles along the y(a)x axis (nominal tip radius 2 μm). Considering the ranges of interest, the expanded uncertainties stated in the calibration certificate of the CI were 10 nm for Ra values below 229 nm, and 24 nm in the Ra interval 229–604 nm [57, 58] (rough data are in appendix C, table C1, while details of data processing can be found elsewhere [57]).

The surfaces were successively characterized by a LSC with 100× magnification (quantization level of 4096 × 4096 pixels, with pixel width of about 32 nm), and according to the local reference system and sampling scheme respectively in figure 15(c) and −d (laser scanning movement along the y(a)x axis and orthogonal to the dominant texture). Being the surfaces associated with four cylindrical steel tools (20 mm × 10 mm), the origin of the local reference system was set in the center of the cylinders as shown in figure 15(c). Thus, ten repeated measurements were performed, in the central position (0,0), and in the peripheral positions (x, y), (x, −y), (−x, y), (−x, −y), where |x| = |y| = 6.5 mm (rough data in appendix C, table C2).

The characterization of a surface normally implies the quantification of functional behaviors in terms of surface texture, which are to be proven on areas that are usually quite large with respect to the field of view of the common surface topography measuring instruments. Therefore, a prevailing choice is the acquisition of larger surface portions as matrices of fields of view (stitching). Such approach is useful when large features are to be characterized dimensionally. In this case, instead, the choice was to sample the surfaces at homogeneous spacing because of their expected uniformity. Thus, the investigation intended to average the assessment of the surface texture, so that possible variations due to dissimilarities or irregularities of the surface polishing among different areas could be included in the measurement uncertainty while compensating for the influence of the instrument. Hence, a larger area was chosen for the CI, not covering any of the areas chosen for the LSC to achieve a denser overall surface sampling.

### Table 13. General linear model ANOVA of measureand area A21 (sequential sum of squares—confidence level 95%).

| Day     | Influence (p-val ≈0.01) | Magnification | Influence (p-val < 0.01) | $R^2$ |
|---------|-------------------------|---------------|--------------------------|-------|
| Sq      |                         |               |                          | 69%   |

#### 4.1. Results

The data set of root mean square values Sq was examined for outliers by the Chauvenet’s criterion. Four outliers related to T3 were removed, one in each peripheral area (see appendix C—values in parenthesis in table C2). Examples of acquired micrographs for each polished surface are in figure 16 (3D views of micrographs).

The correction of the systematics was carried out by a least square fit of the optical LSC measurements as function of the reference CI measurements, according to what is shown in section 2. The optical measurements corrected for accuracy and systematics are given by

![Figure 13. Histogram of the experimental distribution, and comparable curve of the normal distribution for the measureand area A21, after the equalization of the quantization level in the corresponding micrographs and outliers removal (Chauvenet’s exclusion limits $x_{L_{0.01}} = 5.1 \mu m$, $x_{R_{0.01}} = 5.4 \mu m$).](image-url)
where $x_{\text{OPT}}$ are the optical measurements with average $x_{\text{OPT,ave}}$, $x_{\text{CI}}$ is the mathematical model equation found consistent for best fitting the experimental data (straight line with null constant term). The parameters of the least square fit are in appendix C, table C3.

The measurement uncertainty evaluated according to (3) had the following contributors (the complete uncertainty budget is in table C4)

- $u_{\text{Repr,fit}}$: reproducibility values from the regression (standard deviation of the residuals).
- $u_{\text{slope}}$: standard deviation of the coefficient of the best fit regression.
- $u_{\text{Repe,opt}}$: standard deviation of the optical measurements (repeatability).
- $u_{\text{CI}}$: accuracy of the CI (see (C.1) in appendix C for more details).

When the correction of the systematic behavior was not performed, the traceability was achieved as the mismatch (distance) between optical and contact measurements, leading to the following equation

$$y_{\text{OPT}} = x_{\text{OPT}} + x_{\text{OPT,ave}} - x_{\text{regr}} \pm \epsilon_{\text{Rep}}$$

(5)

$$= x_{\text{OPT}} + x_{\text{OPT,ave}} - q \cdot x_{CI} \pm \epsilon_{\text{Rep}}$$

Table 14. Parameters of the least square regression. Intercept $a$, slope $b$, quadratic term $c$, standard deviation of the intercept $\sigma_{a}$, standard deviation of the slope $\sigma_{b}$, standard deviation of the quadratic term $\sigma_{c}$, reproducibility $\epsilon_{\text{Rep}}$ (standard deviation of the residuals), degrees of freedom and coefficient of determination $R^2$.

| $a$ $\mu$m | $b$ $\mu$m | $c$ $\mu$m | $\sigma_{a}$ $\mu$m | $\sigma_{b}$ $\mu$m | $\sigma_{c}$ $\mu$m | $\epsilon_{\text{Rep}}$ $\mu$m | DoF | $R^2$ % |
|-----------|-----------|-----------|----------------------|----------------------|----------------------|------------------|-----|-------|
| $x_{\text{opt, squ}}$ | 5.21 | -0.0001 | 0.001 | 0.02 | 0.0001 | 0.0002 | 0.07 | 27 | 24 |

Table 15. RMS results for the measurand $A_{21}$ after the correction of the equalized micrographs. The values are for the sub-groups of influence factors, and related expanded uncertainties. The decimals in parentheses explicate the approximation on the last digit.

| 5$x$ | 10$x$ | 20$x$ | 50$x$ | 100$x$ | Day 1 | Day 2 | Day 3 |
|------|------|------|------|------|------|------|------|
| $S_{q}$ $\mu$m | 5.22 | 5.17 | 5.18 | 5.31 | 5.10 | 5.18 | 5.20 | 5.20 |
| $U$ $\mu$m | 0.40(9) | 0.40(7) | 0.40(9) | 0.41(1) | 0.41(4) | 0.40(6) | 0.40(9) | 0.41(3) |

The measurement uncertainty evaluated according to (6) had the following contributors (the complete uncertainty budget is in table C5)

- $u_{\text{Repr}}$: maximum deviation of optical measurements (uniformly distributed).
- $u_{\text{Repe,opt}}$: standard deviation of optical measurements (repeatability).
- $u_{\text{Repe,CI}}$: standard deviation of reference measurements (CI) (reference repeatability).
- $u_{\text{CI}}$: accuracy of the CI (see (C.1) in appendix C for more details).
The contributors $u_{\text{opt,rep}}$ and $u_{\text{opt}}$ account respectively for the variability of the optical measurements $x_{\text{OPT}}$ (measurement process), and for the average of optical measurements $x_{\text{OPT,ave}}$ (traceability correction), both in (6).

Due to the complexity of the investigation and large amount of data, this third instance was summarized by the second-order interaction plot of a general linear model design of experiment (DOE—see figure 17). The DOE was implemented on the entire data set and, also, on partial results corresponding to the following factors:

- Measurements corrected for the systematic behavior: ‘Yes’, ‘No’.
- Evaluation in the central and peripheral areas: $(0, 0)$, $(x, y)$, $(-x, y)$, $(-x, -y)$.
- Surfaces with different polishing: ‘T1’, ‘T2’, ‘T3’, ‘T4’.

Referring to the numbers in figure 17, the results can be reviewed as follows:

(i) Systematics correction in different types of evaluation areas. The evaluated uncertainty was lower after the correction of the systematic effects, regardless of the evaluation area.

(ii) Systematics correction in different types of polishing. When the systematic effects were corrected, a lower uncertainty could be obtained for all the specimens except for T3 (stone polishing). This happened because some of the measurements related to T3 were affected by noise. In fact, they were in the only set of measurements in which an exclusion of outliers was necessary.

(iii) Efficacy of the systematics correction considering different evaluation areas (with respect to the specimen’s center) or different types of polishing. The correction produced lower uncertainties with different assessed values depending on the evaluation areas and on the surfaces. Such differences were small, and due to the sought unevenness in the different types of polishing.

(iv) Variations due to different sampling areas with respect to the specimen’s center. The largest spread of uncertainty values was on the evaluation areas of T3 (presence of noise).

(v) Variations due to different types of polishing. The lowest uncertainty was stated for T1 (diamond buff polishing), whilst the highest was for T4 (dry blast polishing). However, the final uncertainty for T4 was influenced by a higher corresponding traceability uncertainty in the calibration certificate of the CI.

5. Discussion

A gradual understanding of the differences between frequentist statistics limitations and the normal practice in the field of metrology for manufacturing was the endeavor of the first instance. The correction of the systematic behavior was a mere attempt, having the most remained unknown because of a general lack of complete information in the measurement data sets. The use of deviations, in fact, put together heterogeneous measurements, where the poorness of repetitions did not allow to discriminate among the three main influence factors highlighted by the KDE. The ANOVA test response identified as influence factors the instrument and the magnification used. Thus, a
third dominant influence factor evidenced by the mixture remained unknown. The instrument influence factor was possibly related to the instrument-operator chain—e.g. different light setting used for the same measurand. While the magnification influence factor clearly added differences in the micrographs like the number of pixels and their width. The regression models were also not adequate. The tendencies highlighted in figure 7 could not be corrected. The step height measurements would require a second order polynomial. Nevertheless, examining the \( x_{dev_Sq} \) distribution, only a correction by piecewise regressions (i.e. per instruments) would be effective. Different behaviors against one single reference instrument were inconsistent, possibly having also the reference a systematic behavior with different tendency.

Other dissimilarities were analyzed in the second instance, where the measurements were enough to inspect all the factors influencing the experimental distribution, and accounted in their natural sequence of acquisition. It was ascertained that measurements corresponding to different acquisition days were systematic outcome: the conditions responsible for the systematics were changed at the beginning of each measurement session. In fact, if the systematic behavior must be considered in its development (see section 3), it can be expected starting with the use of a specific instrument and ending when that instrument is turned off. Since optical measurements can be long-lasting, the influence factors must be considered consistently with a measurement session, expecting a possible different behavior of the same factors in other measurement sessions (i.e. implying an instrument turned off and on). Therefore, this was probably the unknown influence factor in the first case study.

Hence, for effectively applying the described method when there are long lasting acquisitions, the systematics should be related to, and corrected for, continuous measurement sessions planed in homogeneous groups. If so, a Gaussian mixture can be exploited for identifying the dominant influence factors.

The result after the equalization was another interesting outcome of the second case study. It revealed that the quantization level (or bandwidth [46]) is only one side of the problem when comparing heterogeneous micrographs. The other one is instead due to the surface mapping by finite resolution. The content of information was incomplete for 5× and 100× magnification. In the first case, the 5× objective lens was not capable of acquiring enough details on the measured surfaces because of its soft resolution. Therefore the micrographs resulted in an extremely low quality of the surface mapping. In the other case, the reduced field of view of the 100× objective lens could only analyze areas too small for the texture of the measured surfaces. Thus, particular features dominated the results because of the limited content of information. The surface mapping was, instead, equivalent for 20× and 50×, which accomplished the best trade-off quantization-information content for mapping the surface topography (at least for the non-smooth surfaces involved—cf figure 9). Moreover, the equalization granted more efficacy to the systematics correction, achieving a final uncertainty significantly reduced (cf table 12 with table 15).

The previous observations were conveyed to the third instance, which proved the correction of the systematics effective in reducing the final uncertainty. Moreover, sampling different areas on the surfaces and acquiring several repeated measurements for each area allowed for inherently compensating the influence of the instrument, and, concurrently, assessing a measurement uncertainty characterizing the surfaces’ unevenness. Specifically, multiple repeated measurements and multiple areal acquisitions together were exploited to separate the variability of the manufacturing process and the one of the instrument (an explicit example of use of both multiplicities can be found elsewhere [59]). Nonetheless, optical measurements are relatively long lasting. The validity of the approach is strictly related to the number of repeated measurements that is attainable in an acceptable time. Furthermore, the contact measurements were not adequate as reference. In fact, it emerged clearly that the CI uncertainties had a main impact on the final uncertainty budget, leading to a probable overestimation (cf uncertainty contributors in table C4). This was also the case of section 3.1, where the equalization of the quantization level reduced the spread of the residuals to about half of their variability, but the final uncertainty did not reduce consistently because the traceability uncertainty contributor was dominant.

Despite the caveat already introduced regarding the artifacts’ manufacture (see Introduction), a more adequate choice for establishing the traceability is still the use of material measures (at least, at the time of writing). They are not affected by the limitations inherent in the use of a stylus, viz. an additional systematic behavior, the finite tip radius and an expanded uncertainty inadequate for the present-day demand at the nano scale.

A controversy may arise from whether or not the third instance should instead be solved more efficiently by the assessment of the metrological characteristics. In such case, measurements of at least five material measures would be required in addition to the ones of the component under characterization. Each data set would add an uncertainty contributor for the traceability plus the detected non-ideal behavior of the instrument. Also, some of the metrological characteristics would be particularly affected by manufacturing imperfections (mainly linearity and \( x-y \) mapping deviations). Therefore, the systematics would only be accounted for, but not removed from the experimental data set. On the contrary, only one material measure is required for achieving the traceability according to the frequentist approach. Moreover, any
manufacturing imperfections can be compensated (or at least minimized) by the systematics correction.

Regardless of the efficiency on the achievement of the traceability, the proven efficacy of the method on the final uncertainty reduction was also a confirmation of the validity of the starting hypothesis in section 1.1. In other words, average (height of step) and quadratic integral average (root mean square value) of measurands (sets of height values—as explained in section 1.1) met the hypotheses of CLT. In fact, even though the systematic behavior \( f(x_i) \) could not be recognized completely (\( \hat{f}(x_i) \) is an estimate in (7a) and (7b)), it was possible to correct the experimental distribution to a quasi-random one because the correction had zero expectation (7c), and the unrecognized systematic effects were randomized in the experimental distribution

\[
\begin{align*}
y_i &= x_i + f(x_i) - \hat{f}(x_i) \quad (7a) \\
f(x_i) - \hat{f}(x_i) &= 0 \quad (7b) \\
E\{f(x_i) - \hat{f}(x_i)\} &= 0. \quad (7c)
\end{align*}
\]

Conversely, failure in the validity of the starting hypotheses of CLT leads to a correction with a non-zero expected value of the randomized systematic components. Therefore, the non-zero expected value may become a bias increasing the overall evaluated uncertainty [30].

The possible use of averages for dimensional or topographic characterization is rigorously connected with the sought functionality of the component under test. Free-form surfaces are very common nowadays, where the metrology tasks urge more and more the adoption of distributed measurands (i.e. models of whole components with statistically characterized measurement uncertainty, and ready for on-demand calculations). Although the effective assessment of complex surfaces requires more advanced solutions than averaging the micrographs with roughness parameters, the instinctive belief of applying frequentist statistics pixel-wise would raise even more complications because of the nature of the micrographs.

Figure 18 shows portions of two normalized covariance matrices (Pearson’s correlation) as imaged in grey-scale palette, where the achievable maximum +1 and minimum −1 were associated respectively to white and black. The calculation was applied to micrographs of a lapped surface acquired by CSI using 20 × (Figure 18(a)) and 50 × (Figure 18(b)) magnification objectives. Being the micrographs matrices of 1000 × 1000 elements (one million elements each), only the central rows of the respective covariance matrices are visualized in the figure.

The correlation appears periodic and complex, with several inversions between positive and negative fluctuating values. The shown covariance also suggests

![Figure 16. Examples of acquired micrographs (3D views). (a) T1: diamond buff polishing (grade #15). (b) T2: 320 grit paper polishing. (c) T3: 400 stone polishing. (d) T4: dry blast polishing (glass bead #11).](image-url)
that the correlation between one specific pixel and the neighboring pixels fade away after a certain characteristic length (mainly depending on the nature of the imaging sensor, and on the degree of excitation of the constituent elements). It is clear that the application of the frequentist approach to inference was possible in this work because the correlation was resolved by averaging among height values, and the systematic behavior did not involve interactions among influence factors. Therefore, the correction of the systematics considering the measurand’s set of height values demands methods more sophisticated than the GUM approach, which are also required to be robust to noise. The noise in micrograph acquisitions is, in fact, the cause of different errors affecting the pixels’ measure. A typical example is the noise causing a sharp increase in the magnitude of some pixels’ measure (in jargon spikes). Spikes are outliers in the distribution of the surface mapped values. Nonetheless, the usual methods for the metrological outliers’ elimination would not be capable to resolve such presence, creating indeterminacy. In fact, metrological outliers are modeled statistically as rare events, thus, the removal methods are based on low probability of occurrence. On the contrary, diverse types of noise exist that have their own specific distributions (white noise, colored noise, flicker noise, etc.), which bring them closer to a systematic behavior rather than metrological outliers. Moreover, the presence of low-frequency noise (e.g. flicker noise from amplifiers) mixed with high-frequency noise hinders the possibility of elimination e.g. by filtering. A discussion on the effects of the measurement noise is beyond the scope of this work. While an innovative approach can be found elsewhere [60], the aim is to advise on the application of methods that are consolidated in many fields of metrology, although not verified and validated with micrographs acquisition due to the complexity of the operation of the instruments involved.

6. Conclusions

The sequence of the proposed case studies demonstrated that the frequentist approach can be applied to non-scalar measurands. The successful implementation of the method was achieved on averages and quadratic integral averages of matrix elements, which allowed the correction of the experimental distribution systematic behavior. The initial hypothesis of averaging micrographs’ elements, in fact, unfolded favorably, resolving the complex correlation among pixels. On the contrary, the correction of the systematics pixel-wise was recognized inadequate, requiring completely new approaches. Computers’ improved capability to store or promptly process information are opening opportunities across multi-disciplinary engineering fields. Therefore, new and manifold statistical solutions can certainly be sought in the boosted perspective of Artificial Intelligence, taking advantage of High Performance Computing.

The case studies also indicated that precautions are necessary. The first case analyzed bad habits in measurements and processing, still quite diffused in the field. The influence factors due to bad practice add to the systematic behavior of the instrument, counteracting the correction. Therefore, bad practice must be avoided to have an effective application of the method (but also when not applying it!). The second explanatory case complied with both good practice in the field of micro and nano manufacturing, and with the hypotheses of the CLT, showing how frequentist

![Figure 17. Second-order interaction plot comparing uncertainties related to different measurand areas, with and without correction of the systematic behavior (numbers are referred in the text).](image-url)
statistics can be useful in detecting and quantifying influence factors in the data set. The third case dealt successfully with a direct prove of a possible optimization of a typical surface characterization task, where the influence due to the variability of the instrument was minimized over the sought variability of the surfaces’ unevenness.

The minimization of the impact of the instrument on the measurement uncertainty is a strong point in favor of the method. When several manufacturing components are to be inspected at the micro and nano scales, the quality assessment can benefit significantly from a reduced instrument uncertainty contributor on the narrow tolerance verification intervals. The instrument uncertainty can be planned effectively by a KDE, identifying the dominant influence factors in the experimental distribution and, thus, minimizing their effect.

Drawbacks were also clearly indicated, mainly related to: (1) Difficulty of complying with CLT hypotheses when measurements are long lasting. (2) All the necessary cares needed for a correct implementation of the method. (3) Possible overestimation of the traceability uncertainty contributor when adequate reference measurements are not available, or even possible. Such inconveniences are not always avoidable, above all in an industrial environment. It is, in fact, unlikely that the burden of several repeated measurements for a large number of components to be verified could be considered cost-effective. As well as, it would not be acceptable planning long homogeneous measurement sessions.

Therefore, the main reason of dismissal of the analyzed method lies in the aforesaid disadvantages of formulating a stochastic model. Conversely, the deterministic framework of the metrological characteristics offers ease of application but is mostly affected by imperfections and inadequacies of the usable material measures, which in turn may lead to inaccuracies, and to an increase of the systematic behavior.

A possibility is foreseen in the use of self-calibration techniques [61], where rediscovered methods need to be integrated by new ones, which are robust to noise, correlation and imperfection of the material measures. Self-calibration techniques are especially needed for the inspection of the linearity and perpendicularity of the axes in the instruments’ reference system. For instance, the x and y mapping deviations are determined measuring the location of the center of mass of micro features in cross-grating material measures. Hence, they are heavily affected by any imperfection of the structured artifacts used, and depending on the result of the micrographs acquisitions.

In the end, a twofold change of perspective is needed in manufacturing metrology: To make easier the access to complex mathematics in the fashion of software tools for industrial users is a major objective. The free-form structuring of surfaces and volumes requires a real 3D metrology, where the traditional measurand must give way to the concept of distributed measurand.

Acknowledgments

The author would like to thank Giulio Barbato and Gianfranco Genta at Politecnico di Torino for fruitful discussions, and for sharing valuable advises concerning statistical analysis and uncertainty evaluation. Author’s gratitude also goes to Hans N. Hansen at Danmarks Tekniske Universitet—DTU for his continuous support, guidance and inspiration.
Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Dimensional and topographic metrology of micro structured tool inserts—supplemental data

Table A1. Height of the steps $H/\mu m$ in the indicated measurand areas. A double value in the same column stands for repeated measurement. The values in parenthesis were excluded by the Chauvenet’s criterion.

| Area | FV $5\times$ | FV $10\times$ | LSC $10\times$ | LSC $20\times$ | LSC $50\times$ | CSI $20\times$ | CI $^a$ |
|------|--------------|--------------|----------------|---------------|---------------|---------------|--------|
| A11  | 162.1        | 162.1        | 160.0          | 159.5, 161.0  | 158.9         | 162.6         |
| A13  | 165.9        | 165.7        | 163.5          | 161.1, 161.4  | 161.1         | 162.3         |
| A15  | 167.3        | 169.0        | 167.6          | (159.2), 165.2| 163.7         | 164.2         |
| A25  | 96.2         | 95.9         | 96.7           | 91.6          | 97.3          |
| A26  | 83.9         | 83.7         | 84.8           | 85.2          | 84.8          |
| A27  | 177.8        | 177.1        | 176.0          | 177.2         | 177.2         |
| A28  | 85.6         | 86.3         | 84.5           | 83.1          | 85.2          |
| A29  | 163.9        | 161.5        | 162.6          | 158.1         | 163.1         |
| A10  | 80.9         | 81.9         | 80.8           | 78.6          | 80.2          |

$^a$ Focus Variation microscope;

$^b$ Laser Scanning Confocal microscope;

$^c$ Coherent Scanning Interferometry microscope;

$^d$ Contact Instrument.

Table A2. Root-mean-square values $Sq/\mu m$ of the indicated measurand areas. A double value in the same column stands for repeated measurements. The values in parenthesis were excluded by the Chauvenet’s criterion.

| Area | FV $5\times$ | FV $50\times$ | FV $100\times$ | LSC $10\times$ | LSC $20\times$ | LSC $50\times$ | CSI $20\times$ | CSI $50\times$ | CI $^d$ |
|------|--------------|--------------|---------------|----------------|---------------|---------------|---------------|---------------|--------|
| A12  | 1.06         | 0.81         | 1.47          | 1.06, 0.89     | 0.87          | 1.10          |
| A14  | 0.80         | 0.76         | 1.52          | 0.80, 0.87     | 0.77          | 0.98          |
| A21  | 5.25         | (3.56)       | 5.97          | 4.98           | 5.23          | (6.32)        | 5.19          |
| A22  | 1.91         | 0.44         | 0.36          | 0.77           | 0.88          | 0.54          | 0.78          |
| A23  | 1.69         | 0.44         | 0.41          | 0.53           | 0.72          | 0.52          | 0.49          |
| A24  | 1.58         | 0.40         | 0.24          | 0.45           | 0.34          | 0.47          | 0.44          |

$^a$ Focus Variation microscope;

$^b$ Laser Scanning Confocal microscope;

$^c$ Coherent Scanning Interferometry microscope;

$^d$ Contact Instrument.

Table A3. Deviations $devH/\mu m$ representing the values in table A1 normalized (subtracted) to their respective areal averages, after applying the Chauvenet’s criterion (excluded values in parenthesis).

| Area | FV $5\times$ | FV $10\times$ | LSC $10\times$ | LSC $20\times$ | LSC $50\times$ | CSI $20\times$ |
|------|--------------|--------------|----------------|---------------|---------------|---------------|
| A11  | 1.47         | 1.53         | −0.57          | −1.08, 0.36   | −1.69         |
| A13  | 1.13         | 2.88         | 0.68           | −1.65, −1.33  | −1.71         |
| A15  | 1.98         | 3.71         | 2.22           | −(6.15), −0.10| −1.64         |
| A25  | 1.12         | 0.78         | 1.62           | −3.52         |
| A26  | −0.48        | −0.68        | 0.36           | 0.79          |
| A27  | 0.76         | 0.13         | −1.04          | 0.16          |
| A28  | 0.74         | 1.41         | −0.36          | −1.78         |
| A29  | 2.38         | 0.002        | 1.04           | −3.43         |
| A210 | 0.37         | 1.31         | 0.28           | −1.69         |

$^a$ Focus Variation microscope;

$^b$ Laser Scanning Confocal microscope;

$^c$ Coherent Scanning Interferometry microscope;

$^d$ Contact Instrument.
Appendix B. Correction of the systematic behavior vs time sequence—supplemental data

Table B1. Step height measurements $H/\mu m$ in the measurand area A13.

| Area | $FV^*\times$ 5 | $FV^*\times$ 10 | $LSC^*\times$ 10 | $CSI^*\times$ 10 |
|------|----------------|-----------------|------------------|------------------|
| A14  | 163.7          | 163.4           | 162.1            | 162.3            |
| A21  | 163.6          | 163.4           | 162.8            | 162.5            |
| A22  | 163.5          | 163.4           | 162.2            | 162.2            |
| A23  | 164.1          | 163.4           | 162.3            | 162.6            |
| A24  | 163.9          | 163.8           | 162.6            | 162.9            |

Table B2. Root-mean-square measurements $Sq/\mu m$ in the measurand area A21.

| Area | $FV^*\times$ 5 | $FV^*\times$ 10 | $LSC^*\times$ 10 | $CSI^*\times$ 10 |
|------|----------------|-----------------|------------------|------------------|
| A12  | 5.27           | 5.14            | 5.17             | 5.30             |
| A14  | 5.31           | 5.36            | 5.14             | 5.21             |
| A21  | 5.40           | 5.15            | 5.18             | 5.21             |
| A22  | 4.54           | 6.61            | 5.19             | 5.21             |
| A23  | 5.28           | 5.50            | 5.29             | 5.24             |
| A24  | 5.51           | 5.13            | 5.08             | 5.09             |

Table B3. Root-mean-square measurements $Sq/\mu m$ in the measurand area A21, re-sized at the same quantization level. The values in parentheses were excluded by the Chauvenet’s criterion.

| Area | $FV^*\times$ 5 | $FV^*\times$ 10 | $LSC^*\times$ 10 | $CSI^*\times$ 10 |
|------|----------------|-----------------|------------------|------------------|
| A12  | 5.27           | 5.14            | 5.30             | 5.23             |
| A14  | 5.31           | 5.36            | 5.21             | 5.24             |
| A21  | 5.40           | 5.15            | 5.18             | 5.21             |
| A22  | 4.54           | 6.61            | 5.19             | 5.21             |
| A23  | 5.28           | 5.50            | 5.29             | 5.24             |

Appendix C. Correction of the systematic behavior in surface characterization—supplemental data

The following considerations where made with regards to the contact instrument (CI) measurements:

- Measurements of $Ra$ in the calibration certificate led to evaluate the capability of the reference stylus instrument to measure average height variations on a surface within a confidence interval (expanded uncertainty). For this reason, the uncertainty of CI related to $Ra$ measurements was considered compatible with measurements of $S_a$.

- $Sa$ and $Sq$ are both amplitude parameters and strongly correlated (they measure the same quantity in different ways). The related uncertainties are normally in the same order of magnitude. Therefore, considering the same CI uncertainty for $Sq$ values is an estimation of the uncertainty based on previous knowledge and experience [32].

- The uncertainty contributor related to the calibration uncertainty is constant for all cases, related to both the correction of the systematic behavior and not. Hence, it is a common constant value and may affect the final evaluated uncertainty but not the investigation itself (correction of systematics). Nonetheless, a realistic final standard uncertainty for CI measurements was considered as $u_{CI} = \max\{u_{cal} \cdot u_{repeatability}, \sqrt{n_{input}}\}$, where $U_{cal} = k \cdot u_{cal}$ ($k = 2.2$ evaluated as inverse $t$ distribution on a confidence interval of 95% and degrees of freedom given by $n_{input} = 1$), and, thus, spread on the averaged input multiplicity. In other words, being the CI repeatability small, it was

\[
U_{CI} = u_{cal} \cdot \sqrt{n_{input}}
\]

with $n_{input} = 12$ repeated measurements stated in the calibration certificate.
Table C1. Cl measurements of the areal parameters Sa and Sq [38].

|       | T1    | T2    | T3    | T4    |
|-------|-------|-------|-------|-------|
| Sa/μm | 48.0  | 144.0 | 231.9 | 510.5 |
| Sq/μm | 47.6  | 133.3 | 232.6 | 510.5 |
| Uref/μm | 62.1  | 179.5 | 313.5 | 649.1 |

Table C2. LSC measurements of the areal parameter Sq. The values in parentheses were excluded by the Chauvenet’s criterion.

| (0,0) | (x,y) | (x,y) | (−x,y) | (−x,y) |
|-------|-------|-------|--------|--------|
| 49.2  | 48.8  | 50.4  | 45.9   | 56.0   |
| 48.9  | 49.5  | 49.9  | 45.0   | 46.3   |
| 47.9  | 44.0  | 49.6  | 43.5   | 55.1   |
| 48.2  | 48.5  | 49.9  | 41.1   | 54.2   |
| T1    | 49.8  | 49.0  | 50.1  | 50.8   | 55.8   |
| 49.8  | 49.4  | 44.5  | 44.3   | 56.3   |
| 48.8  | 49.7  | 48.9  | 43.9   | 45.8   |
| 49.2  | 44.0  | 42.8  | 39.4   | 53.2   |
| 49.5  | 48.7  | 50.4  | 48.7   | 55.0   |
| 48.5  | 43.1  | 43.6  | 42.9   | 55.2   |
| 157.7 | 175.3 | 192.2 | 153.8  | 139.3  |
| 157.2 | 176.2 | 195.1 | 152.0  | 138.4  |
| 157.2 | 174.6 | 192.0 | 156.1  | 139.3  |
| 159.1 | 175.6 | 191.6 | 153.6  | 135.9  |
| T2    | 155.2 | 177.2 | 195.3  | 151.3  | 141.7  |
| 166.5 | 179.4 | 193.8  | 158.7  | 142.5  |
| 160.2 | 179.8 | 190.9  | 153.9  | 141.0  |
| 163.1 | 174.7 | 195.7  | 152.5  | 142.3  |
| 166.3 | 174.7 | 194.5  | 154.1  | 134.5  |
| 157.9 | 174.6 | 193.1  | 153.0  | 141.0  |
| 387.0 | 407.0 | 305.9  | 218.1  | 220.0  |
| 392.9 | 404.0 | (326.2)| 223.7  | 225.0  |
| 388.9 | 409.0 | 299.2  | 223.1  | 218.4  |
| 387.1 | 410.8 | 305.3  | (262.1)| 224.0  |
| T3    | 391.5 | 408.7 | 303.8  | 223.6  | 220.8  |
| 385.9 | 409.1 | 308.6  | 221.1  | 221.2  |
| 384.2 | 400.2 | 302.6  | 225.4  | (258.2)|
| 385.3 | 411.1 | 308.9  | 221.8  | 218.9  |
| 385.0 | (343.4)| 302.6  | 221.5  | 219.2  |
| 388.3 | 402.8 | 318.9  | 228.0  | 225.9  |
| 577.2 | 497.4 | 572.8  | 525.1  | 594.9  |
| 574.0 | 494.2 | 570.3  | 522.8  | 597.6  |
| 574.8 | 490.0 | 571.1  | 526.3  | 584.4  |
| 579.0 | 498.7 | 573.3  | 524.4  | 593.9  |
| T4    | 582.3 | 486.7 | 566.3  | 523.6  | 592.3  |
| 577.4 | 484.0 | 568.4  | 523.0  | 595.3  |
| 575.7 | 490.3 | 566.0  | 521.3  | 596.7  |
| 573.2 | 483.8 | 573.6  | 528.0  | 604.0  |
| 571.8 | 486.8 | 565.0  | 528.4  | 591.3  |
| 576.2 | 499.7 | 567.9  | 524.6  | 592.0  |

* Laser Scanning Confocal microscope.

Table C3. Results of the analysis of all the average values in the surfaces’ sampling areas. Data corrected for the systematic behavior. \( q \) is the slope of the model equation. \( s_{\text{unc, fit}} \) is the standard deviation of the regression. \( s_{\text{unc, fit}} \) is the average of the root mean square values after correcting for the systematic behavior. \( U \) is the expanded uncertainty (see table C4 below).

|       | q/1  | \( s_{\text{unc, fit}} \)/nm | \( S_{\text{unc, fit}} \)/nm | U/nm |
|-------|------|-------------------------------|-----------------------------|------|
| T1    | 0.790 ± 0.009 | 1.8                          | 48                           | 28   |
| T2    | 0.922 ± 0.003 | 1.5                          | 165                          | 32   |
| T3    | 0.972 ± 0.002 | 1.8                          | 309                          | 35   |
| T4    | 0.849 ± 0.001 | 1.8                          | 551                          | 71   |

Table C4. Uncertainty budget of the measurements corrected for the systematic behavior, averaging in all areas. \(|c|\) is the absolute value of the sensitivity coefficient of the generic uncertainty contributor \( u(x_i)/u(y)\) (propagated contributors).

|       | \(|c|\) | \( u(x_i)/nm\) | \( u(y)/nm\) |
|-------|-------|----------------|-------------|
| T1    | 0.8  | 15.7           | 12.4        |
|       | 1.0  | 1.7            | 1.7         |
|       | 60.9 | 0.009          | 0.6         |
|       | 1.0  | 1.8            | 1.8         |
|       | 2.2  | 2.2            | 2.2         |
|       | 28   | 28             | 28          |
| T2    | 0.9  | 15.7           | 14.5        |
|       | 1.0  | 1.5            | 1.5         |
|       | 1.0  | 1.5            | 1.5         |
|       | 2.2  | 2.2            | 2.2         |
|       | 32   | 32             | 32          |
| T3    | 0.8  | 37.8           | 32.1        |
|       | 1.0  | 1.8            | 1.8         |
|       | 35   | 35             | 35          |
| T4    | 0.8  | 0.001          | 0.6         |
|       | 1.0  | 1.8            | 1.8         |
|       | 2.2  | 2.2            | 2.2         |
|       | 71   | 71             | 71          |

Table C5. Uncertainty budget of the non-corrected measurements, averaging in all areas (\(|c| = 1.0\) (\( u(x)/y \)) (propagated contributors).

|       | \( u_{CI}/nm\) | \( u_{opt, fit}/nm\) | \( u_{CI}/y\) |
|-------|----------------|-----------------------|---------------|
| T1    | 15.7           | 15.7                  | 15.7          |
|       | 1.5            | 1.5                   | 1.5           |
|       | 37.8           | 37.8                  | 37.8          |
| T2    | 0.4            | 1.2                   | 0.5           |
|       | 0.8            | 1.4                   | 1.7           |
|       | 2.2            | 2.2                   | 2.2           |
| T3    | 1.6            | 1.7                   | 1.5           |
|       | 2.2            | 2.2                   | 2.2           |
|       | 35             | 35                    | 35            |
| T4    | 35             | 35                    | 35            |
ORCID iDs

Danilo Quagliotti @ https://orcid.org/0000-0002-8347-2884

References

[1] Tosello G, Haitjema H, Leach R K, Quagliotti D, Gasparin S and Hansen H N 2016 An international comparison of surface texture parameters on polymer artefacts using optical instruments CIRP Ann. 65 529–32
[2] Hansen H N, Carneiro K, Haitjema H and De Chiffre L 2006 Dimensional micro and nano metrology CIRP Ann. 55 721–43
[3] Hansen H N, Hocken R and Tosello G 2011 Replication of micro/nano surface geometries CIRP Ann. 60 695–714
[4] Leach R 2013 Introduction to surface topography Characterisation of Areal Surface Texture ed R Leach (Heidelberg, Germany: Springer Berlin) 1–13
[5] ISO 3274:1996 Geometrical Product Specifications (GPS)—Surface texture: Profile method—Nominal characteristics of contact (stylus) instruments. ISO 3274 (Geneva: International Organization for Standardization).
[6] ISO 4287:1997 Geometrical Product Specifications (GPS)—Surface texture: Profile method—Terms, definitions and surface texture parameters. ISO 4287 (Geneva: International Organization for Standardization).
[7] ISO 4288:1996 Geometrical Product Specifications (GPS)—Surface texture: Profile method—Rules and procedures for the assessment of surface texture. ISO 4288 (Geneva: International Organization for Standardization).
[8] ISO 12 179:2000 Geometrical Product Specifications (GPS)—Surface texture: Profile method—Calibration of contact (stylus) instruments. ISO 12 179 (Geneva: International Organization for Standardization).
[9] Tosello G, Hansen H N, Marinello F and Gasparin S 2010 Replication and dimensional quality control of industrial nanoscale surfaces using calibrated APM measurements and SEM image processing CIRP Ann. 59 563–8
[10] Tosello G, Marinello F and Hansen H N 2012 Characterisation and analysis of microchannels and submicrometre surface roughness of injection moulded microfluidic systems using optical metrology Plast. Rubber Compos.-Macromol. Eng. 41 29–39
[11] Tosello G, Hansen H N, Calason M and Gasparin S 2014 Challenges in high accuracy surface replication for micro optics and micro fluids manufacture Int. J. Precision Technology 4 412–44
[12] Weckemann A and Rinnagl M 2000 Acceptance of processes: do we need decision rules? Praxis. Eng. 24 264–9
[13] Kunzmann H, Pfeifer T, Schmitz H and Weckemann A 2005 Productive metrology—adding value to manufacture CIRP Ann. 54 155–68
[14] Jiang X, Senin N, Scott P J and Blatayron F 2021 Feature-based characterisation of surface topography and its application CIRP Ann. 70 681–702
[15] Brinksmeier E, Karpuschewski B, Yan J and Schönenm L 2020 Manufacturing of multiscale structured surfaces CIRP Ann. 69 717–39
[16] Mendez Ribó M 2020 Vat photopolymerization process chain Ph.D Thesis Technical University of Denmark
[17] Leach R K, Bourell D, Carmignato S, Dommez A, Senin N and Deswalt W 2019 Geometrical metrology for metal additive manufacturing CIRP Ann. 68 677–700
[18] Leach R K, Giuss C L, Haitjema H, Evans C and Jiang X 2015 Calibration and verification of areal surface texture measuring instruments CIRP Ann. 64 797–813
[19] ISO 25 178–600:2019 Geometrical product specifications (GPS)—Surface texture: Areal—Part 600: Metrological characteristics for areal topography measuring instruments. ISO 25 178–600 (Geneva: International Organization for Standardization).
[20] Giuss C L, Leach R K, Helary F, Gutasukas T and Nimishakavi L 2012 Calibration of the scales of areal surface topography–measuring instruments: I. Measurement noise and residual flatness Meas. Sci. Technol. 23 035008
[21] Giuss C L, Leach R K and Helary F 2012 Calibration of the scales of areal surface topography measuring instruments: II. Amplification, linearity and squareness Meas. Sci. Technol. 23 065005
[22] Giuss C L and Leach R K 2013 Calibration of the scales of areal surface topography measuring instruments: III. Resolution Meas. Sci. Technol. 24 105010
[23] MacAulay G D and Giuss C L 2016 Assessment of uncertainty in structured surfaces using metrological characteristics CIRP Ann. 65 533–6
[24] ISO/DIS 25 178–700 Geometrical product specifications (GPS)—Surface texture: Areal—Part 700: Calibration, adjustment and verification of areal topography measuring instruments. ISO 25 178–700 (Geneva: International Organization for Standardization).
[25] Leach R, Haitjema H, Su R and Thompson A 2021 Metrological characteristics for the calibration of surface topography measuring instruments: a review Meas. Sci. Technol. 32 032001
[26] ISO 25 178–70:2014 Geometrical product specification (GPS)—Surface texture: Areal—Part 70: Material measures. ISO 25 178–70 (Geneva: International Organization for Standardization).
[27] De Chiffre L, Carli L and Eriksen R S 2011 Multiple height calibration artefact for 3D microscopy CIRP Ann. 60 355–8
[28] Leach R, Giuss C, Guttmann M, Jakobs P J and Rubert P 2015 Development of low-cost material measures for calibration of the metrological characteristics of areal surface texture instruments CIRP Ann. 64 545–8
[29] Fortmeier I, Schulz M and Mee R 2019 Traceability of form measurements of freeform surfaces: metrological reference surfaces Opt. Eng. 58 092602
[30] Eßler M, Herig J, von Freymann G and Seewig J 2018 Calibration sample for arbitrary metrological characteristics of optical topography measuring instruments Opt. Express 26 1669–23
[31] Carmignato S, De Chiffre L, Bosse H, Leach R K, Balsamo A and Estler W T 2020 Dimensional artefacts to achieve metrological traceability in advanced manufacturing CIRP Ann. 69 693–716
[32] JCGM 100:2008 Evaluation of measurement data—Guide to the expression of uncertainty in measurement Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures (BIPM), Sèvres, France
[33] Krystek M 2001 Measurement uncertainty propagation in the case of filtering in roughness measurement Meas. Sci. Technol. 12 63–7
[34] Morel M A A 2006 Uncertainty estimation of shape and roughness measurement Ph.D Thesis Technische Universität Eindhoven
[35] Haitjema H 2015 Uncertainty in measurement of surface topology Surf. Topogr., Metrol. Prop. 3 33004
[36] Moroni G, Syam W P and Petrò S 2018 A simulation method to estimate task-specific uncertainty in 3D microscopy Measurement 122 402–16
[37] Quagliotti D, Calason M and Tosello G 2018 Metrological quality assurance in micro injection molding Micro Injection Molding ed G Tosello (München: Carl Hanser Verlag GmbH & Co. pp) 241–88
[38] ISO 25 178–2:2012 Geometrical product specification (GPS)—Surface texture: Areal—II: Terms, definitions and surface texture parameters. ISO 25 178–2 (Geneva: International Organization for Standardization).
[39] Software used: SPIPSTM v. 6.2.8, Image Metrology A
[40] Barbaty G, Barini E M, Genta G and Levi R 2011 Features and performance of some outlier detection methods J. Appl. Stat. 38 2133–49
[41] Internet site: www.hi-micro.eu
[42] Hackert-Oschätzchen M, Meichsner G, Zeidler H, Zinecker M and Schubert A 2011 Micro machining of different steels with closed electrolytic free jet AIP Conf. Proc. 1337 1337–43
[43] Hackert-Oschätzchen M, Meichsner G, Zinecker M, Martin A and Schubert A 2012 Micro machining with continuous electrolytic free jet Precis. Eng. 36 612–9
[44] Quagliotti D, Tosello G, Islam A M, Hansen H N, Zeidler H, Martin A, Schubert A, Brandao C and Riemer O 2015 A method for dimensional and surface optical measurements uncertainty assessment on micro structured surfaces manufactured by jet-ECM Proc. of the Int. Conf. on Micro-Manufacturing–4M, Milan, Italy, 31 March–2 April
[45] Mattsson L et al 2008 How reliable are surface roughness measurements of micro-features? — Experiences of a Round Robin test within nine 4M laboratories Proc. of the Int. Conf. on Micro-Manufacturing—4M, 9-11 September 2008, Cardiff, UK
[46] Leach R and Haitjema H 2010 Bandwidth characteristics and comparisons of surface texture measuring instruments Meas. Sci. Technol. 21 1–9
[47] Barbato G, Genta G and Germak A 2013 Measurements for Decision Making (Société Editrice Esculapio)
[48] Janke S J and Tinsley F 2005 Introduction to Linear Models and Statistical Inference (Hoboken, NJ: Wiley)
[49] Sivia D and Skilling J 2006 Data Analytic: A Bayesian Tutorial II revised edn (Oxford, UK: Oxford University Press)
[50] Pavese F and Forbes A 2009 Data Modeling for Metrology and Testing in Measurement Science ed F Pavese and A B Forbes (Boston, MA: Birkhäuser Boston)
[51] Bailey David G 2017 Not Normal: the uncertainties of scientific measurements R. Soc. Open Sci. 4 160600
[52] Aggogeri F, Barbato G, Genta G and Levi R 2014 Statistical modelling of industrial process parameters Procedia CIRP IX CIRP Conference on Intelligent Computation in Manufacturing Engineering 23–25 July 2014, Capri, Italy 2014
[53] JCGM GUM-6:2020 Guide to the expression of uncertainty in measurement—VI: Developing and using measurement models (GUM-6) Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures (BIPM), Sèvres, France
[54] Quagliotti D, Tosello G, Islam A M, Hansen H N, Zeidler H, Martin A, Schubert A, Brandao C and Riemer O 2015 Optical micro-metrology of structured surfaces micro-machined by jet-ECM Proc. of the XV international conference of the european society for precision engineering and nanotechnology (euspen) (Leuven: Richard Leach) 167–168 Leuven, Belgium, 1-5 June
[55] JCGM 200:2012 International vocabulary of metrology—Basic and general concepts and associated terms (VIM) Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures (BIPM), Sèvres, France
[56] ISO 15 530-3:2011 Geometrical product specification (GPS)— Coordinate measuring machines (CMM): Technique for determining the uncertainty of measurement—Part III: Use of calibrated workpieces or measurement standards. ISO 15 530-3 (Geneva: International Organization for Standardization).
[57] Quagliotti D, Baruffi F, Tosello G, Gasparin S, Annoni M, Parenti P, Sobiecki R and Hansen H N 2016 Correction of systematic behaviour in topographical surface analysis Proc. of the 4M/IWMI2016 Conference (Copenhagen: Tosello G, Hansen H N, Kornel E and Dimov S) 277-80 Copenhagen, DK, 13-15 September
[58] CGM (Center for Geometrical Metrology) DANAK Calibration certificate n. RUM1101
[59] Quagliotti D, Hansen H N and Tosello G 2018 Multiscale dimensional tolerance specifications established on shrinkage assessment in ceramic micro injection molding Procedia CIRP —Proc. of the XV CIRP CAT Conference Milano, Italy, 11–13 June 2018 75 145–8
[60] Maculotti G, Genta G, Quagliotti D, Galetto M and Hansen H N 2021 Gaussian process regression-based detection and correction of disturbances in surface topography measurements Qual. Reliab. Eng. Int. 00 1–18
[61] Evans C J, Hocken R J and Estler W T 1996 Self-Calibration: Reversal, redundancy, error separation, and absolute testing CIRP Ann. 45 617–34