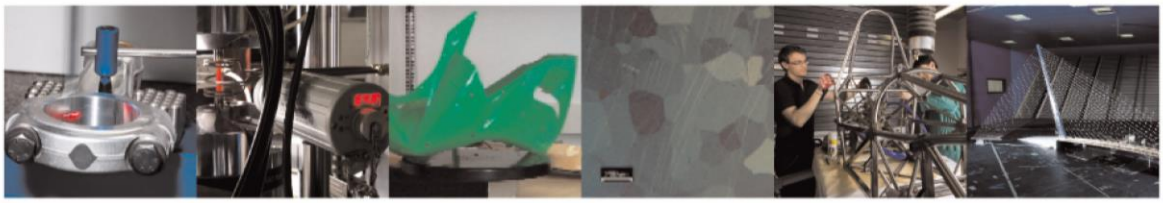

Statistics-based decision rules for the ISO 10360 series of standard tests

Stefano Petrò, Giovanni Moroni

This is a post-peer-review, pre-copyedit version of an article published in CIRP Annals. The final authenticated version is available online at: http://dx.doi.org/10.1016/j.cirp.2021.04.076

DIPARTIMENTO DI MECCANICA ■ POLITECNICO DI MILANO
via G. La Masa, 1 ■ 20156 Milano ■ EMAIL (PEC): pecmecc@cert.polimi.it
http://www.mecc.polimi.it
Rev. 0

Statistics-based decision rules for the ISO 10360 series of standard tests

Stefano Petrò, Giovanni Moroni (2)

Mechanical Engineering Department, Politecnico di Milano, Via La Masa 1, 20156, Milano, Italy

Verification tests defined in the ISO 10360 series of standards guarantee that coordinate measuring systems (CMS) have consistent performance. The development of these verification tests is focused on practical industrial applicability. All proposed tests are based on multiple measurements of probing points. This gives the tests a statistical nature. As each measurement is currently treated separately and must conform to the maximum permissible error, a considerable risk of false acceptance/rejection is present. An approach based on a statistical model of the current test is proposed instead. The approach can effectively manage customer and producer risks in a way that is consistent with ISO/IEC GUIDE 98-4.

Coordinate measuring machine (CMM), Quality assurance, Decision-making

1. Introduction

Performance verification guarantees the best functionality of coordinate measuring systems (CMSs). Acceptance tests verify conformance with the producer's specifications when a new system is acquired by a company. Reverification tests periodically check whether the system no longer conforms to the customer's specifications. As the context is customer/producer relationships, recognized standard tests must be considered. The ISO 10360 series of standards defined these tests. The first edition of the standard [1] defined two performance indicators related to pure probing and the measurement of length. These indicators have changed names several times and are currently known as the maximum permissible error of length measurement $E_{\mathrm{L,MPE}}$ [2] and maximum permissible single-stylus form error $P_{\mathrm{Form.Sph.1x25:SS:Tact,MPE}}$ [3] (for the sake of compactness, this paper adopts the previous symbol $P_{\mathrm{FTU,MPE}}$ [4]). The specific tests used to verify these indicators were defined as well. Tests were designed to guarantee the complete check of system performance in a reasonable amount of time (typically a workday). Even if the most recent revisions have introduced more parameters and have added parts of the series (e.g., the maximum permissible limit of the repeatability range [2] or the maximum permissible location error [3]), the original two parameters are still the most frequently considered because they are easy to understand and generally applicable, so they will be considered for the rest of this discussion.

All tests have a similar structure. Multiple measurements are performed on some reference artefacts, and all measurement results must fall within a specified interval. Although this approach is very simple, it neglects the statistical structure of a multiple-results experiment. This can lead to incorrect statements on the system's performance. In a recent paper [5], it was shown that test results are inconsistent with the concept of the coverage factor for uncertainty. In addition, the most recent revision of ISO 14253-1:2017 [6] introduced the need to consider the statistics behind the statement in conformance and nonconformance verification. This coheres with the Guide to the Expression of Uncertainty in Measurement (GUM) [7].

In this paper, we propose a decision rule proving that CMSs conform to specifications that considers the statistical nature of the tests defined in the ISO 10360 series of standards. To do this, we will first establish statistical models for the tests and then calculate the probability that the test will be passed. Finally, we will introduce a decision rule based on this probability. The proposed method will be validated on experimental data.

2. Maximum permissible error of length measurement

$E_{\mathrm{L,MPE}}$ is probably the most diffused performance indicator for CMSs. This is particularly true for the form $E_{0,\mathrm{MPE}}$, in which the ram axis stylus tip offset is equal to 0 (stylus tip on the machine ram axis). It is a general system test based on the measurement of five length measurement standards. The standards are measured at seven different positions (orientations and locations) within the measuring volume. Each measurement was replicated three times, so 105 measurements were performed. The absolute difference between the calibrated and measured lengths E_0 shall not exceed $E_{0,\mathrm{MPE}}$. The test uncertainty U [8] should be considered in this verification, and the final condition of the verification is

$$|E_0| \leq E_{0,\mathrm{MPE}} \pm U. \tag{1}$$

The $\pm$ sign depends on whether a reverification or acceptance test is being performed. An acceptance test is performed when the system is first tested by the manufacturer to prove that it conforms to the specifications. In this case, the uncertainty is against the manufacturer and reduces the conformance zone. A reverification test takes place when the owner wishes to verify if the system still conforms to the specifications. In this case, the uncertainty is against the customer, and enlarges the conformance zone.

From a statistical point of view, the test procedure corresponds to a two-factor factorial design [9]. The two factors are the lengths of the length standards and the positions within the measuring volume. The number of replicas is equal to three. This kind of experiment is usually analysed by analysis of variance (ANOVA). ANOVA verifies which factors significantly affect the experimental (measured) result. Factors can be treated as fixed or random. The choice of either random or fixed factors defines the statistical model. In the authors' opinion, the $E_{0,\mathrm{MPE}}$ factors are fixed. Positions and standard lengths are suggested by the standard and chosen by the operator rather than randomly selected from a pool of available levels.

The effects model for a two-factor factorial design is

$$(E_0 =) y_{ijk} = \mu + \tau_i + \theta_j + (\tau\theta_{ij}) + \epsilon_{ijk} \tag{2}$$

in which y_{ijk} is the result of the experiment (measurement) in which the factor levels are set at i, j, and considering replica k, μ is the overall mean effect, τ_i is the effect of the ith level of the first factor (length measurement standard length), θ_j is the effect of the jth level of the second factor (position), $(\tau\theta_{ij})$ is the effect of the interaction between τ_i and θ_j, and ϵ_{ijk} is a random error component assumed to be distributed as the zero-mean normal independent random variable $N(0, \sigma^2)$. In practice, the measurement results are assumed to be independently distributed as random normal variables; $y_{ijk} \sim N(\mu + \tau_i + \theta_j + (\tau\theta_{ij}), \sigma^2)$. ANOVA allows us to simultaneously verify which factors are statistically significant (which terms among μ, τ_i, θ_j, and $(\tau\theta_{ij})$ significantly differ from zero) and estimate the various parameters (including σ^2).

ANOVA relies on three hypotheses of random error ϵ_{ijk}: normality, constant variance (homoscedasticity), and independence. These hypotheses can be verified a posteriori from the analysis of the residuals e_{ijk}:

$$e_{ijk} = y_{ijk} - \left(\hat{\mu} + \hat{\tau}_i + \hat{\theta}_j + \widehat{(\tau\theta)}_{ij}\right) \quad (3)$$

where $\hat{\mu}, \hat{\tau}_i, \hat{\theta}_j$, and $\widehat{(\tau\theta)}_{ij}$ are the estimates of the various terms of model (2) yielded by ANOVA. Some hypotheses are possibly not verified. A transformation of the data can usually solve violations of normality and homoscedasticity. Box-Cox [9] and Johnson's [10] transformations are usually applied. ANOVA is then repeated on the transformed data. Residual dependence cannot be easily solved and instead requires more complex models. If autocorrelation is present, the use of autoregressive integrated moving average (ARIMA) models can help solve the problem.

Once the correct model has been identified, it is possible to estimate the probability that a single measurement meets condition (1) if the estimates $\hat{\mu}, \hat{\tau}_i, \hat{\theta}_j, \widehat{(\tau\theta)}_{ij}$, and $\hat{\sigma}^2$ are assumed in a normal model. As the measurement results are independent, determining the probability that all 105 measurements meet condition (1) is easy. For example, if all factors and interactions are significant, the probability is

$$P_{E_0} = \prod_{i=1}^{5}\prod_{j=1}^{7}\left[\Phi\left(\frac{E_{0,\mathrm{MPE}} \pm U - \hat{\mu}_{ij}}{\hat{\sigma}}\right) - \Phi\left(\frac{-(E_{0,\mathrm{MPE}} \pm U) - \hat{\mu}_{ij}}{\hat{\sigma}}\right)\right]^3 \quad (4)$$

where $\Phi(a)$ is the standard normal cumulative distribution function calculated at a, and $\hat{\mu}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\theta}_j + \widehat{(\tau\theta)}_{ij}$. If a transformation has been applied to verify the ANOVA hypotheses, the term $E_{0,\mathrm{MPE}} \pm U$ is transformed accordingly. P_{E_0} is an estimate of the conformance probability. The conformance probability is the probability that given the measurement result (the estimated parameters), the system conforms.

2.1. Proposed decision rule for the $E_{0,MPE}$ test

Suppose now that the CMS being tested is functional according to its specification. This means that system calibration would not improve the performance because, for instance, residual volumetric or scale errors cannot be further reduced due to technology limitations. However, a purely random measurement error could lead, according to the standard decision rule, to a nonconformance statement. This would be a 'false rejection'. The false rejection probability would be P_{E_0}. This is a producer's risk. The dual situation is also possible, in which the conformance of a CMS states that the system is not functional according to its specification (false acceptance). In this case, the false acceptance probability would be $1 - P_{E_0}$ and is the consumer's risk. ISO/IEC GUIDE 98-4:2012 asks that consumer and producer risks are considered in the decision rule. However, the current form of the decision rule does not consider this at all.

The proposed decision rules are as follows.

- In the case of an acceptance test, the aim is to prove that the system is functional. Therefore, the consumer's risk must be controlled. The system will be defined as conforming if the consumer's risk is lower than the agreed value β.
- In the case of a reverification test, the aim is to demonstrate that the system is not functional. Therefore, the producer's risk must be controlled. The system will be defined as nonconforming if the producer's risk is lower than the agreed value α.

Figure 1 schematically illustrates the overall procedure of the $E_{0,MPE}$ test.

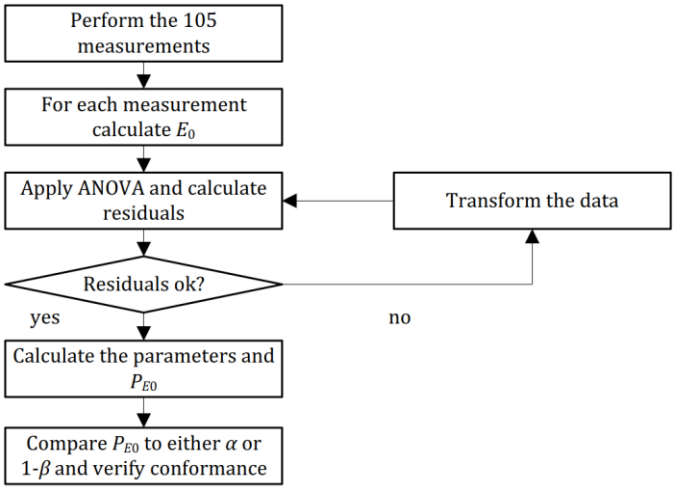

Figure 1. Proposed procedure for the $E_{0,MPE}$ test.

2.2. Validation

ANOVA can be applied to real data with all hypotheses verified to validate the approach. If this requires some transformation, validation is still obtained.

Twelve instances of $E_{0,\mathrm{MPE}}$ were gathered on a 'Zeiss Prismo 5 VAST HTG' tactile CMM. The manufacturer states that for this machine, $E_{0,\mathrm{MPE}} = 2 + L/300$ µm, where L is the calibrated length of the length measurement standard in [mm].

ANOVA was used to analyse all twelve instances. In all cases, all factors (position within the measuring volume, length of the length measurement standard, and interaction between the two) were significant. This is qualitatively confirmed by the main effect and interaction plots in Figure 2 and Figure 3: the presence of residual scale and calibration errors explain the relevance of the length. Residual volumetric errors explain the relevance of the position and the interaction between position and length. As all factors are significant, the full model $\hat{\mu}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\theta}_j + \widehat{(\tau\theta)}_{ij}$ is applied to evaluate P_{E_0} by (4).

Concerning the verification of the hypotheses, the Anderson-Darling test verified the normality of the residuals e_{ijk}. Normality is verified in six of twelve datasets. However, taking a closer look at the residuals, in all cases, it is evident that non-normality depends on the presence of a few outliers. As the model is balanced, ANOVA is known to be robust and resistant to the presence of outliers, so the estimated model can be considered valid anyway. Levene's test verified homoscedasticity. The observation of the autocorrelation function (qualitatively) verified

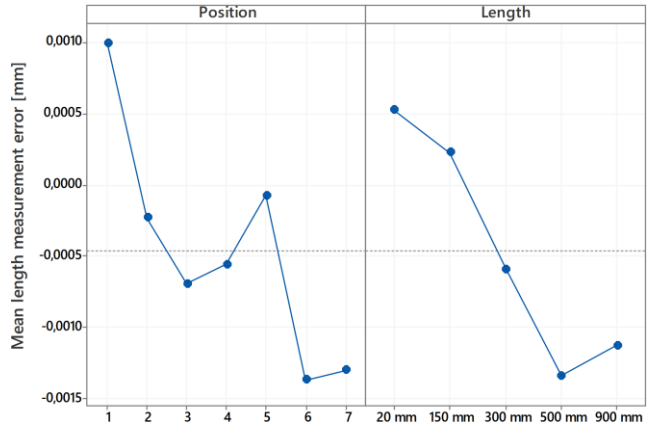

Figure 2. Example of a main effect plot for the position within the measuring volume and the length of the length measurement standard.

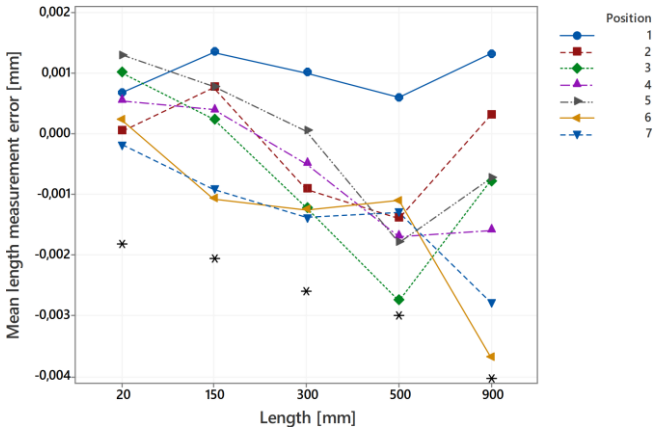

Figure 3. Example of an interaction plot for the position within the measuring volume and the length of the length measurement standard. Black asterisks indicate the acceptance test limit.

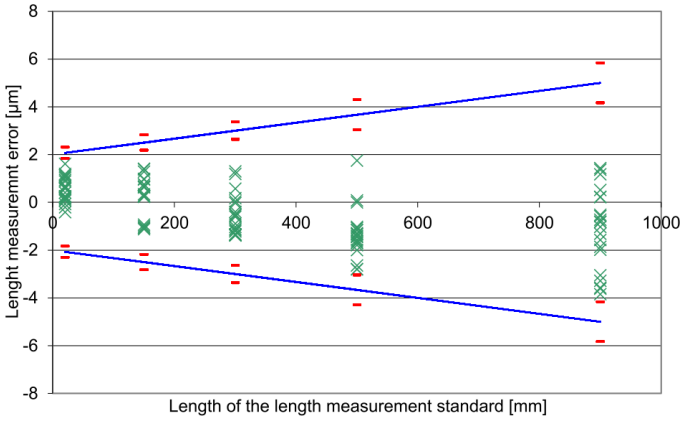

Figure 4. Example of results of the length measurement error test. Blue lines represent $E_{0,\mathrm{MPE}}$. Red lines indicate the acceptance and reverification limits. Green crosses indicate the measured length measurement errors.

independence. In both cases, no issue was found. Therefore, the proposed statistical model can be considered verified. In this case, no transformation was required.

Applying the conventional decision rule, all twelve datasets yield a conformance statement, both in verification and in reverification. If the proposed rule is applied instead, in the case of a reverification test ($\alpha = 0.05$), one dataset yields a nonconformance statement. In the case of the acceptance test ($\beta = 0.05$), all datasets yield a nonconformance statement. To understand this, consider how the average measurement error is distributed: the presence of residual volumetric, scale, and calibration errors bias the length measurement error until it is close to the limit of the conventional acceptance test. Figure 4 shows a series of measured length measurement errors very close to the acceptance limit, particularly for the 500 mm standard. This is confirmed by Figure 3, which shows that for the 500 mm length and position 3, the average length measurement error is very close to the limit. Even if the random error dispersion is small, this makes it so the probability that a single measurement error falls outside the acceptance test limit causes the test to fail.

Please note that the twelve datasets were collected over a period of approximately ten years as routine verification of a CMS. In this period, the corrective maintenance of the CMS was performed only once. The dataset that presents the most significant issues is the one collected just before maintenance, so we can conclude that the problem was related to the typical drift of a CMS.

3. Maximum permissible single-stylus form error

If the test for $E_{0,MPE}$ aims to verify the performance of the CMS in the whole measuring volume, the test for $P_{\mathrm{FTU,MPE}}$, also known as 'probing error', specifically verifies the performance of the probing system. In its original form, it considered discrete point probing and a single stylus tactile probing system. The same type of test is also considered for scanning, noncontact systems, and multiple stylus systems. Only the sample size and the probing pattern change. The single-stylus discrete point probing test is treated here. The other cases are derived by changing the sample size.

The $P_{\mathrm{FTU,MPE}}$ test is based on measuring a test sphere. The probing system probes the sphere in 25 locations. The ISO 10360-5:2020 [3] standard suggests a point pattern. An unconstrained least-squares sphere fits the point cloud. The 25 radii (distances) R_i of the single points from the centre of the sphere are calculated. The system conforms if the radius range (difference between the maximum, $R_{\max}$, and the minimum, $R_{\min}$, radius) P_{FTU} does not exceed $P_{\mathrm{FTU,MPE}}$. The test uncertainty U shall be considered in the test, so the form of the classic decision rule is

$$P_{\mathrm{FTU}} \leq P_{\mathrm{FTU,MPE}} \pm U \tag{5}$$

The $\pm$ sign depends on whether a reverification test or an acceptance test is being performed.

Unlike the test for $E_{0,\mathrm{MPE}}$, a single measurement result determines the conformance in the $P_{\mathrm{FTU,MPE}}$ test. However, the measured P_{FTU} depends on 25 probing points. Only the most deviating points are considered in the conformance statement. A single anomalous point would suffice for an incorrect statement. Therefore, the statistical study of the test is relevant.

Suppose that the radii are normally and independently distributed, $R_i \sim NIID(\mu, \sigma^2)$. Note that the independence is never verified, as the radius values depend on the fitting. However, as only three parameters are fitted, the correlation is expected to be small. The variance can be easily estimated from the R_i values. Hartley [11] demonstrated that the standardized range of a set of n $NIID$ random variables, i.e., $w = (R_{\max} - R_{\min})/\sigma$, is distributed according to the following cumulative distribution function:

$$P_{P_{\mathrm{FTU}}} = P(W) = \left(\int_{-\frac{1}{2}W}^{+\frac{1}{2}W} \phi(x)dx \right)^n + 2n \int_{\frac{1}{2}W}^{\infty} \phi(u) \left(\int_{u-W}^{u} \phi(x)dx \right)^{n-1} du \tag{6}$$

where $\phi(x)$ is the probability density of a standard normal random variable. The integrals can be solved numerically. If the radii are not normal, a suitable transformation can normalize them.

$P_{P_{\mathrm{FTU}}}$ is an estimate of the conformance probability. The conformance probability is the probability that given the measurement result (estimated variance), the system conforms.

3.1. Proposed decision rule for the $P_{FTU,MPE}$ test

Similar to the $E_{0,MPE}$ test, the $P_{\mathrm{FTU,MPE}}$ test is prone to false rejections, which create producer's risk $P_{P_{\mathrm{FTU}}}$, and false acceptances, creating consumer's risk 1-$P_{P_{\mathrm{FTU}}}$. Again, these risks should be considered in a decision rule, but the current form of the decision rule does not consider this at all.

The proposed decision rule indicated in §2.1 is newly proposed. Figure 5 illustrates the overall procedure used for the $P_{\mathrm{FTU,MPE}}$ test.

3.2. Validation

Demonstrating that the radii are independently distributed according to a normal distribution would validate the model. However, according to §3, the radii are never independent, but the correlation is small. Let us concentrate on normality.

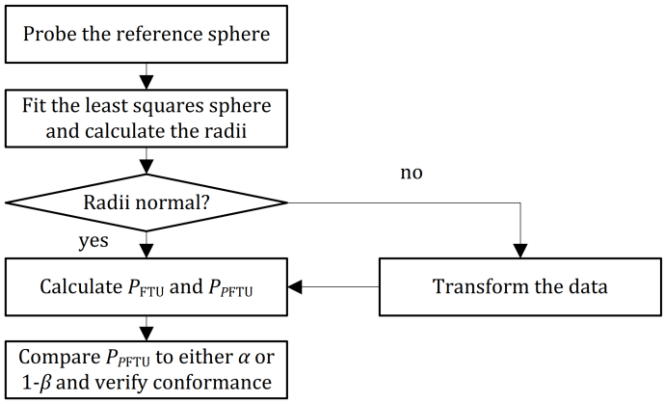

Figure 5. Proposed procedure for the $P_{\mathrm{FTU,MPE}}$ test.

One hundred instances of $P_{\mathrm{FTU,MPE}}$ were run on a 'Zeiss Prismo 5 VAST HTG' tactile CMM. The manufacturer states that for this machine, $P_{\mathrm{FTU,MPE}} = 2$ µm. All point clouds were fitted by unconstrained least-squares spheres, and the radii were calculated. The Anderson-Darling test for normality verified the normality of each set of radii. No set was found to be nonnormal. As there is no statistical evidence of nonnormality in any dataset, it is proven that in general, the radii probed on the reference sphere are normally distributed. Again, should the data be nonnormal, the method could be applied anyway after transformation.

Applying the proposed decision rule to the one hundred test runs, the CMS is always shown to be conforming both in acceptance and reverification tests ($\alpha = \beta = 0.05$). Applying the conventional rule, the CMS always conforms to reverification tests. Considering acceptance, it would be shown to be conforming in 91/100 tests. This discrepancy indicates that the nine tests in which the CMS is shown to be conforming in reverification and nonconforming in acceptance are due to slightly anomalous probing points strongly influencing the test result. The proposed approach instead, considering all points, is not prone to this.

4. Conclusions

The presence of multiple measurements in the tests indicated in the ISO 10360 series of standards can alter the conformance statement. Performing de facto multiple tests (one per measurement) significantly reduces the probability that a CMS is defined as conforming.

In this paper, decision rules that correctly handle multiple measurements were proposed. The decision rules are based on simple statistical models of the measurement results. The test uncertainty is considered in the decision rule. The producer's and consumer's risks are assessed and considered.

One could argue that compared to the standard rule, the proposed rule is more complex and yields results that are not directly interpretable. However, performance tests are routinely performed by expert operators from system manufacturers. These operators can be trained for the new rule, and the analysis required could even lead to an increased knowledge of the system.

This research will be further developed to include performance indicators that have not been considered so far (e.g., the maximum permissible limit of the repeatability range or the maximum permissible location error). An international round robin could be conducted to consider different CMSs at different stages of the lifecycle.

Acknowledgements

We gratefully acknowledge the Italian Ministry of Education, University and Research for the support provided throughout the Project "Department of Excellence LIS4.0 - Lightweight and Smart Structures for Industry 4.0" (CUP: D56C18000400006).

References

[1] International Organization for Standardization., 1994, ISO 10360-2: Coordinate metrology - Part 2: Performance assessment of coordinate measuring machines.

[2] International Organization for Standardization., 2009, ISO 10360-2: Geometrical Product Specifications (GPS) - Acceptance and reverification tests for coordinate measuring machines (CMM) - Part 2: CMMs used for measuring linear dimensions.

[3] International Organization for Standardization., 2020, ISO 10360-5: Geometrical product specifications (GPS) — Acceptance and reverification tests for coordinate measuring systems (CMS) - Part 5: Coordinate measuring machines (CMMs) using single and multiple stylus contacting probing systems using.

[4] International Organization for Standardization., 2010, ISO 10360-5: Geometrical Product Specifications (GPS) - Acceptance and reverification tests for coordinate measuring machines (CMM) -- Part 5: CMMs using single and multiple stylus contacting probing systems.

[5] Petrò, S., Moroni, G., 2021, A statistical point of view on the ISO 10360 series of standards for coordinate measuring systems verification, Measurement, 172:108937, Doi: 10.1016/j.measurement.2020.108937.

[6] International Organization for Standardization., 2017, ISO 14253-1: Geometrical product specifications (GPS) - Inspection by measurement of workpieces and measuring equipment - Part 1: Decision rules for proving conformance or nonconformance with specifications. 3rd ed., Geneva, Switzerland.

[7] Joint Committee for Guides in Metrology., 2012, ISO/IEC GUIDE 98-4: Uncertainty of measurement - Part 4: Role of measurement uncertainty in conformity assessment. 1st ed., Geneva, Switzerland.

[8] International Organization for Standardization., 2006, ISO/TS 23165: Geometrical product specifications (GPS) - Guidelines for the evaluation of coordinate measuring machine (CMM) test uncertainty.

[9] Montgomery, D.C., 2013, Design and analysis of experiments. 8. ed, Singapore: Wiley.

[10] Yeo, I.N.K., Johnson, R.A., 2000, A new family of power transformations to improve normality or symmetry, Biometrika, 87/4:954–9, Doi: 10.1093/biomet/87.4.954.

[11] Hartley, H.O., 1942, The Range in Random Samples, Biometrika, 32/3/4:334–48.