The Validation of CMM Task Specific Measurement Uncertainty Software

S. D. Phillips, B. Borchardt, A. J. Abackerli,[†] C. Shakarji, and D. Sawyer
National Institute of Standards and Technology, Gaithersburg, MD 20899

P. Murray and B. Rasnick
BWXT Y-12, L.L.C., Oak Ridge, TN 37831

K. D. Summerhays, J. M. Baldwin, R. P. Henke, and M. P. Henke
MetroSage L.L.C., Volcano, CA 95689

1. Introduction

Historically metrologists have attempted to purchase accurate coordinate measuring machines (CMMs) by specifying small values of standardized performance test results such as those defined by the ASME B89.4.1 or ISO 10360 standards. These standardized tests typically involve very simple measurands, such as point-to-point length, measured on highly idealized and geometrically simple artifacts such as gauge blocks, step gauges, and ball bars. While such tests are useful from a commercial transaction viewpoint, i.e., all manufacturers specify their CMM performance to the same tests, which allows easy comparisons between brands, the test results are usually only vaguely indicative of accuracy for the complex tasks that the metrologist performs on CMMs. Consequently, the evaluation of task specific CMM measurement uncertainty was either an outright guess, or a time intensive investigation involving measurements of calibrated artifacts geometrically similar to, and measured in a manner specific to, the task under consideration.

In the last several years, computer (Monte Carlo) techniques have arisen to assist in CMM task specific uncertainty evaluation. These simulation methods offer the potential advantage of highly detailed and advanced mathematical models of CMM error propagation that can be easily configured to the specific measurement task under consideration. Hence metrologists have a new and powerful tool that can answer such questions as:

- "Will the CMM I intend to purchase meet the accuracy requirements for my unique workpieces?"
- "What is the uncertainty of my current measurement results?"
- "If I change some aspect of my measurement (e.g. the thermal environment, the stylus configuration, or the point sampling strategy) what will be the consequence for the measurement uncertainty?"

The ability to simulate the details of the measurement task also implies significant complexity as each new task option multiplies the number of possible measurement scenarios. Consequently such simulation programs are often comprised of many megabytes of compiled computer code. Unfortunately, like all complex software, validation of the code is a nontrivial task. Indeed, even the (usually less complex) CMM software that actually performs and analyzes the measurements is only validated at a very elementary level. (Typically this involves testing the geometry fitting algorithms for mathematical correctness, but omits all the relational aspects (e.g. concentricity), datum reference frames, and material conditions (e.g. Maximum Material Condition)).

In this paper we discuss several methods of identifying weaknesses in CMM simulation software. In particular, we have the benefit that some of the authors are actually developing CMM task specific simulation software and we have listed some of the difficulties addressed during the development of that software project.

In principle, developing software testing procedures is much more straightforward if the inner workings of the simulation software are well understood. Some Monte Carlo methods use sophisticated first principle mathematical models to propagate CMM error sources. Two such methods are the "virtual CMM" [1] and "PUNDIT/CMM" [‡] [2] both of which use, for example, full rigid body parametric error models and kinematic equations to propagate these error sources forward into point coordinate errors and ultimately into corresponding errors of the parameters, e.g. diameter, of the fitting algorithms. Other systems use only random perturbations of the measurement points (which contain no information about correlated errors, e.g. CMM axis squareness) and calculate the corresponding errors of

---

[†] Permanent address: Universidade Metodista de Pitracicaba – UNIMEP, Piracicaba-SP, Brasil, 123400-911
[‡] Disclaimer: The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology or by BWXT Y12.

the parameters of the fitting algorithms. Still other systems use empirical models. Unfortunately, since the source code is unavailable, any of these systems can be examined only as a "black box" where a specified set of input values results in an observed output value, a calculated expanded uncertainty.

2. Factors for Consideration in Testing

Blunders: It has often been said that the largest source of measurement error is misinterpretation of a Y14.5 drawing. While such problems clearly exist, mistakes of this type are considered by the GUM (ISO, "Guide to the Expression of Uncertainty in Measurement," Switzerland, 1993, corrected and reprinted 1995) as "blunders" and are not to be included in the measurement uncertainty statement (GUM 3.4.7). To the list of blunders we include improperly setting up an uncertainty simulation run and hence getting inappropriate uncertainties.

Uncertainty of the Measurand: The GUM identifies that a poorly specified measurand is a legitimate source of measurement uncertainty as it gives rise to multiple "true values" all of which can be assigned as the "value of the measurand". It is the metrologist's obligation to consider this issue and quantify it; hence we do not consider this source of uncertainty (which is also frequently overlooked in traditional uncertainty analysis) in our investigation.

Simulation Fidelity: We define the "simulation fidelity" as the ability of the simulation software to faithfully represent the details of a real measurement in the simulation; how accurately the software models these details is a separate issue addressed elsewhere. All real measurements have an almost infinite list of influence quantities that affect the measurement result and all simulation software has a finite list of uncertainty sources that can be included in the calculated uncertainty. Poor fidelity does not allow the details of the measurement to be specified in the simulation software, and hence the simulation provides an uncertainty statement for a measurement scenario different from the real measurement. Good fidelity simply means that details of the measurement can be represented in the simulation; whether the software considers these details correctly is a separate issue. Uncertainty contributors can be considered either as intrinsic to the measurement system (i.e. sold with the CMM) or extrinsic. A few examples of influence quantities include:

> Extrinsic factors: CMM operator effects, especially workpiece fixturing variation; workpiece form error and how it interacts with the probing point sampling strategy; thermal properties and conditions of the workpiece; workpiece contamination (dirt or coolant…)

> Intrinsic factors: Multiple styli (either fixed star probe or an articulated stylus); scanning probes; rotary tables; CMM dynamic effects (changes in acceleration and velocity values)

In real measurements there will always be some factors that are not represented in the simulation, i.e. the simulation fidelity will not be perfect. This has two consequences: (1) in the evaluation of measurement uncertainty the CMM user is obligated to account for uncertainty sources that are not included in the simulation software; (2) in testing simulation software great care must be given to minimizing the influence quantities that are not accounted for in the simulation software; hence testing measurements must be constructed such that they can be simulated with good fidelity and the calculated uncertainty can be compared to the observed measurement errors.

Simplified, Incomplete, or Incorrect Mathematical Modeling: Assuming that the simulation software allows the inclusion of a particular influence quantity, e.g. describing the form error on a workpiece, or allowing the use of multiple styli, some sort of mathematical model must be invoked to calculate the errors due to this influence. These models can range from complex first principle derived models, to very simple approximations, to just plain wrong models that do not describe the behavior of the influence quantity.

Input Parameters: All models require some form of input to quantify the influence quantities. These input values may be extracted from actual measurement data collected on the CMM under simulation or may be user supplied. These input values describe such things as the type of probe, the magnitude of the probing error, the type of CMM, the magnitude of the CMM structural errors, the type of workpiece form error and its magnitude, etc. Even a very detailed and accurate model will yield nonsense if the input parameters are incorrect.

Coding Errors: In all large complex software systems, coding errors will be present. These can range from a simple incorrect logic statement to more subtle effects such as failing to clear registers or unanticipated interactions when function or object calls occur in different sequential order.

3. Methods of Testing

The procedures for testing CMM measurement uncertainty simulation software are still in an embryonic stage. Indeed, even testing the reliability of uncertainty statements produced by conventional means (e.g. using sensitivity coefficients as described in the GUM) is still controversial and under development. The GUM clearly recognizes the distinction between error and uncertainty, and in particular, points out that a measurement result might (unknowingly) have little error yet be assigned a relatively large uncertainty. This would correctly describe the metrologist's uncertainty about the result. Consequently an uncertainty statement cannot be invalidated because it is large relative to the measurement errors. However, an uncertainty statement can be invalidated if a significant fraction of errors lie outside the uncertainty interval, since the GUM interpretation of expanded uncertainty is that roughly 95 % of the errors (unless otherwise stated by the metrologist) will lie within the uncertainty interval. It is this criterion that we use to test simulation software. We believe that vast overestimation of measurement uncertainty will be eliminated in the commercial marketplace as uncertainty statements are used either to indicate the quality of a measurement result (e.g. calibration laboratories) or to establish gauging guardbands on workpiece tolerances to insure product reliability (acceptance / rejection decisions); in either case overestimation of measurement uncertainty will have large negative economic consequences.

The ISO TC213/WG10 working group is developing a document (15530-4) addressing CMM measurement uncertainty via computer simulation. One aspect of that document is a disclosure of what measurements and influence quantities can be addressed by the simulation software, i.e. a description of the simulation fidelity. Also in the document, several general approaches are discussed: (1) comparison of errors found by the actual measurement of calibrated artifacts to the uncertainties calculated by simulation; (2) comparison of special case "reference values" uncertainties to the uncertainties calculated by simulation; and (3) comparison of "computer-aided verification and evaluation" (CVE) uncertainties to the uncertainties calculated by simulation.

Physical Measurements: The most direct method of examining the validity of an uncertainty statement produced by computer simulation is by physical measurements of calibrated artifacts. Ideally, the physical measurements will be performed in a manner that allows good simulation fidelity so that the calculated uncertainty statement can be directly interpreted as including at least 95 % of the measurement errors.

We have conducted several measurements of this type as shown in Figure 1. The figure shows both the calculated computer simulated uncertainty and the corresponding measurement error of a 150 mm diameter XX grade ring gauge positioned in multiple locations in the CMM workzone. The ring was sampled with 24 equally spaced points and the diameter and circularity were measured. The CMM was a moving bridge style with a workzone (X, Y, Z) of 460 mm, 460 mm, and 385 mm. The CMM was equipped with an articulated head and the probe used was a touch trigger type. We performed the measurements with the CMM error compensation (which corrects for geometrical errors in the CMM structure) both OFF and ON, giving in effect two different CMMs. Table 1 gives the performance parameters of the CMM for both cases. Maximum Permissible Error (MPE) is defined by ISO Standard 10360-5. Figure 1 (a) shows the diameter and circularity errors and simulated uncertainties for the ring gauge. The ring orientation follows the B axis of the articulation system to keep the ring plane perpendicular to the probe axis. Two measurements are shown at B= 0; for one the ring is in the XY plane and for the other the probe is pointed along the –Y axis of the CMM and the ring is in the XZ plane. The other orientations are taken every 45 degrees, with the ring and probe direction rotating about the CMM Z axis, e.g. B = 90 with the ring in the YZ plane. Figure 1 (b) displays the same type of data for the ring tilted 30 degrees with respect to the table and indexed every 45 degrees about the CMM Z axis; hence this includes positions with a diameter of the ring along the four CMM body diagonals. Figure 1 (c) and (d) show the corresponding information for the CMM with the compensation turned OFF.

Table 1

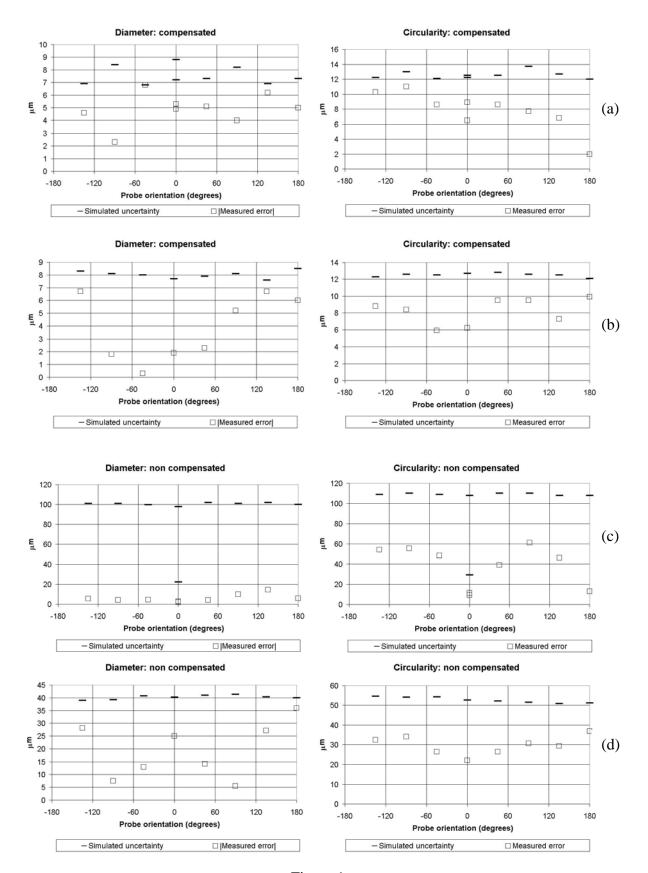| CMM Performance	Compensated	Non-compensated
X linear accuracy (B89.4.1)	5.5 µm	11 µm
Y linear accuracy (B89.4.1)	6.5 µm	12 µm
Z linear accuracy (B89.4.1)	3 µm	88 µm
Volumetric performance (B89.4.1)	13.5 µm	188 µm
Offset probe volumetric performance (B89.4.1)	50 µm/m	76 µm/m
Repeatability (B89.4.1)	0.8 µm	0.8 µm
Probe performance (25 points per ISO 10360-2)	7 µm	7 µm
MPE_{AF}, MPE_{AS}, MPE_{AL} (ISO 10360-5)	10 µm, 2 µm, 3.8 µm	16.6 µm, 3.1 µm, 5 µm

Figure 1

Figure 2 displays additional physical testing using a high accuracy CMM with an analog probe with performance values given in Table 2. In this case the artifact was a 300 mm diameter disc measured for diameter and circularity similarly to the measurements of Figure 1, but with 360 points. The styli used in these measurements were "L" shaped with each leg typically 80 mm. Due to some simulation fidelity issues with this version of the software we were forced to use styli of only approximate dimensions; however, we do not believe this significantly changed the calculated measurement uncertainty.

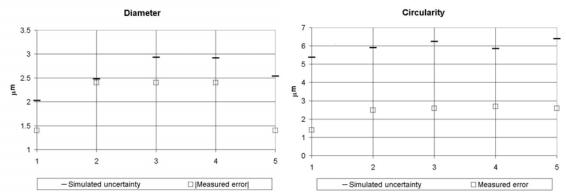| Table 2	
CMM Performance	**Values**
X linear accuracy (B89.4.1)	1.2 µm
Y linear accuracy (B89.4.1)	1.4 µm
Z linear accuracy (B89.4.1)	2.2 µm
Volumetric performance (B89.4.1)	3.9 µm
Offset probe performance (B89.4.1)	6.7 µm/m
Repeatability (B89.4.1)	0.34 µm
Probe ISO 10360-2 (25 points), 3 styli	0.82, 1.68, 1.82 µm

Figure 2

Reference Values: A powerful technique to detect problems in measurement uncertainty software involves the use of reference values. A reference value is a special case result that has a known value. This value can be obtained through a variety of means, including using other verified software known to produce a mathematically correct answer under specified conditions. Often the "reference value" can be obtained through some type of invariant quantity. Our use of reference values in this investigation is rather happenstance, however, we are in the process of a more systematic approach to this method. We describe below some anecdotal cases where reference values have proved useful.

Invariance: The simulation software under consideration can be configured in a variety of ways including deactivating entire classes of uncertainty sources. Thus it was possible to examine a geometrically perfect CMM and introduce as the sole uncertainty source only probing errors. Hence it can be expected that the uncertainty of some simulated measurement, e.g. the diameter of a ring gauge, should not depend on the particular location in the CMM workzone, provided the same number of probing points are measured with the same probe approach direction, etc. That is, while there may be some statistical variation dependent on the number of simulation cycles, in the limit of a large number of cycles this value should be invariant under changes in workpiece location. The simulation software under consideration has the fidelity to locate the workpiece at any location in the CMM workzone. It was observed that after a large number of workpiece relocations, the uncertainty of the simulated measurement was increasing. Eventually this was tracked down to the manner in which the relocations were calculated, as a series of transformation matrices from the previous position. Hence after dozens of workpiece relocations a very long sequence of matrices was being multiplied, allowing minor round off errors to get magnified. The solution was obvious, namely a workpiece relocation was always transformed from a reference position, not the previous position, hence involving only one transformation matrix for each location.

Known values: Zero is a powerful reference value. When all inputs are set to zero, a perfect measurement is being considered, and so the simulation should yield zero measurement uncertainty. While most of the various simulated uncertainty sources satisfied this result, we have found one probe model that yields small (submicrometer) expanded uncertainties indicating a problem that is currently under investigation.

Other known results can be used to test the simulation software. For example, a workpiece temperature can be set to 21 °C with zero uncertainty and a thermal expansion coefficient set to 10^{-5} / °C with zero uncertainty, hence the

result for a one meter feature of size measured without thermal compensation should yield a systematic error of +10 μm and the simulation software should produce this result, and for our test software it did so.

If well-tested reference software is available it can be used to produce known values. For example, NIST has an algorithm testing software package used to check least squares fitting algorithm for the mathematically correct result. We used this software to produce the diameter of a circle that was perturbed by a systematic form error (e.g. three lobes) of specified amplitude. Using the reference software, a large set of results was produced for different phase combinations of the sampling points and the form error. Twice the standard deviation of these results can be compared to the CMM simulation of a similar form error and sampling strategy. We noticed that the reference software and the CMM simulation software were converging to different values for the uncertainty in the circle diameter. Upon further investigation the problem was discovered to be a coding error in the simulation software that caused N equally spaced points on a circle to have a point both at 0 degrees and 360 degrees, i.e. one location on the circle had two superimposed points. A revised version of the simulation software corrected the problem leading to equivalence between the reference and CMM simulation uncertainties.

Sometimes specific uncertainty results can be determined analytically. For example, the measurement of a circle with three points each separated by an angle θ is known to produce an uncertainty in the radius given by the equation below, where σ is the standard deviation of a Gaussian distribution of radial perturbations [3]. In these analytic cases it is straightforward to check the simulation results.

$$u^2(radius) = \frac{1 + 2\cos^2\theta}{2(1 - \cos\theta)^2}\,\sigma^2$$

Computer-aided Verification and Evaluation (CVE)
This powerful testing method allows for great control in determining the ability of the software under test to reflect certain isolated influence factors in its reported uncertainty. The details of the method are given in [4], but we outline the basic concepts here. In physical testing, several measurements are made and compared with calibrated values. The CVE approach is similar, but measurements are made by a CMM that exists—not physically—but rather as a model within a computer program. Given a CMM model, a program can easily simulate thousands of measurement tasks with exactly known measurement errors. The same CMM model can be used to obtain the needed input quantities to the uncertainty evaluating software. The known measurement errors can be compared with the uncertainty reported by the software under test to find the percentage that lie within the stated uncertainty. One can then repeat this test over many modeled CMMs—a prohibitive task if relying solely on physical testing. This CVE technique works with other influence factors (besides the CMM itself) provided realistic models are known. While we did not investigate CVE techniques in this paper, they represent a generalization of the reference value method and will be presented in a future publication.

3 Conclusions
Our preliminary investigation of validating CMM measurement uncertainty statement produced by computer simulation produced a variety of results. All of the measurement errors found in the physical measurements were inside their corresponding uncertainty intervals. The physical measurements are only able to test error sources that are actually present during the measurement. For example, if our CMM had no XZ or YZ axis squareness error then we would be unable to determine if the simulation software would correctly handle this class of CMM errors. Since in general the details of the CMM errors are unknown, physical testing on several CMMs is warranted in order to get a good representation of all potential CMM error sources and exercise the simulation software over this population of error sources. To the extent that the uncertainty appears to be overestimated in some cases of the physical measurements (Figures 1 and 2), this may indicate that further refinements in the simulation software may be useful.

The physical measurement results were unable to detect the problems in the simulation software involving the part placement transformation matrices, the switching probe error, or the redundant point sampling strategy error, as these effects were typically submicrometer and not discernable in the results. The reference value testing can catch these errors relatively quickly in cases where one knows what reference value to employ and corresponding measurement to simulate. This suggests that a well-documented list of reference value tests might be a useful tool to employ before starting the more expensive aspect of real physical measurements of calibrated parts.

1. http://www.ptb.de/en/org/5/53/532/vcmm.htm
2. www.metrosage.com
3. S.D. Phillips, et. al., Precision Engineering 22: 87-97, 1998
4. S.D. Phillips, Proceedings ASPE 1999 Annual Meeting 525-528