

A General Quantitative Method to Validate Instrument Calibration Techniques

Steven D. Phillips

National Institute of Standards and Technologies, Gaithersburg MD 20899

As international standards proliferate and laboratory accreditation programs become more common, there is increasing interest in the calibration of measuring instruments. For sophisticated instruments such as coordinate measuring machines (CMMs) or laser trackers, historically there has been little agreement on what constitutes a calibration of these instruments. Calibration services have often designed their own instrument assessment methods, which often differ widely among practitioners. Hence, a well-defined quantitative determination of whether a proposed assessment *method* of an instrument rises to the level of a calibration is clearly needed. A typical example is a claim that completing the performance evaluation described in the ASME B89.4.1 standard constitutes a CMM calibration. The standard itself, which was designed to promote commerce, carefully avoids the topic of CMM calibration. We will discuss some of the issues involved in an instrument calibration and how such a calibration claim might be assessed.

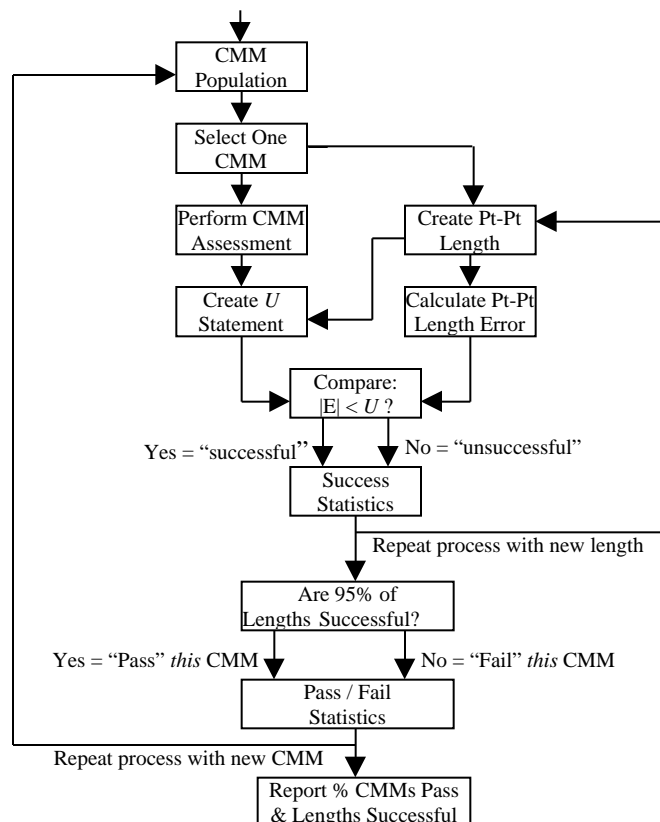
Recent thinking regarding the calibration concept might be summarized as stating that an instrument calibration specifies a well-defined measurand, a set of calibration conditions, and assesses all the reasonable error sources of the instrument that affect the accuracy of the measurement result. The accuracy is typically reported as an uncertainty statement that describes the dispersion of values about the measurement result that could reasonably be attributed to the measurand; sometimes the measurement bias is subsumed into the uncertainty statement and sometimes it is stated separately. Examining the previous “B89” CMM calibration claim, we see that the measurand is not identified. It is well known that the measurement of some features, e.g. partial arcs, are inherently less accurate than other features, hence the measurand must be identified and well specified. We will define the measurand of the proposed B89 calibration as the point-to-point length between any two points within the CMM workzone. Similarly, the conditions of the calibration are not defined; this would include specifying the probes, stylus configurations, environmental conditions and other factors that affect the CMM. For purposes of this example, we will adopt calibration conditions consisting of using a single probe and a simple straight stylus parallel to the ram axis, and ideal CMM environmental conditions (20 °C, negligible vibration etc.). More complex calibration conditions such as the use of multiple styli, stylus configurations that include a component perpendicular to the ram axis, or non-ideal environmental conditions would generally be included in an actual calibration but are omitted here for simplicity. Finally, we see that although a proposed error assessment method is described, i.e. as specified in the B89.4.1 standard, the construction of an uncertainty statement from the assessment results is absent. In this paper we will examine three different uncertainty statements and then explore how such statements might be validated.

While there are many different sources of CMM errors, in this paper we will only discuss those systematic errors associated with the CMM machine frame. For the calibration described above, the relevant B89 test procedures for these errors are the three linear displacement accuracy tests (LDAX, LDAY, LDAZ) and the volumetric performance test that yields the range of 20 ball bar measurements. We will construct different uncertainty statements using these test procedures and then attempt to validate the uncertainty statements based on computer simulation.

Conceptually, the validation procedure is based on a synthetic population of CMMs assumed to include any real CMM that a calibration service provider might encounter in actual practice. These CMMs are of different types, e.g. moving bridge or gantry, sizes, and most importantly, have different error characteristics. Some of these CMMs have large squareness errors, other have straightness errors, most have a mixture of many different parametric errors; some CMMs are in good working order and others have significant performance problems. The error behavior of each CMM is well defined, i.e. it has a specified set of parametric errors (in this paper we will consider only rigid body errors but this is easily

generalized), and the parametric error description of different CMMs is completely independent. Once a proposed calibration method is defined by specifying the measurand, calibration conditions, assessment method, and uncertainty statement construction, the proposed method can undergo validation. Consider selecting one CMM at random from the population of CMMs and performing the proposed assessment tests. Each aspect of the assessment test is computed just as it would occur in actual practice, e.g. for each LDA test, ten lengths are measured down the center of the workzone, parallel to the axis under consideration, and each of these 10 length errors is calculated. The results of the assessment tests are now available for the construction of the point-to-point length uncertainty statement (using a coverage factor of $k = 2$) applicable to future length measurements. To validate the uncertainty statement a length is generated between two randomly selected points in the CMM workzone. The details of this length (i.e. the start and end points) are available to the proposed uncertainty statement since the uncertainty may depend on details such as its the location, orientation and length. An uncertainty statement is produced and (since the complete error behavior of the CMM is known) the measurement error associated with this length can be calculated. If the error lies within the stated uncertainty then this is considered a success for *this particular length* measurement. Another length is randomly generated, an uncertainty statement produced, and the associated error is computed and compared with the uncertainty statement. This is repeated several hundred times so that a reasonable sample of all possible lengths in the workzone, i.e. the measurand, is produced. The percentage of uncertainty statements that are successful can now be computed. If at least 95 % of the cases are successful, i.e. their length errors lie within their associated uncertainty intervals, then the proposed calibration method is considered to have passed the validation test for *this particular CMM*. (The 95 % threshold is selected somewhat arbitrarily, see the discussion at the end of this paper.) Next, another CMM is drawn at random from the CMM population and the entire procedure repeated; note that this new CMM must undergo a new assessment as its error behavior is completely independent of the previous CMM. Once a few thousand CMMs have been selected from the CMM population and undergone the validation test, the percentage of CMMs that passed the test and the percentage of lengths that had their errors successfully contained within their uncertainty statements can be calculated. The validation test is outlined in Figure 1, and details of the verification test are briefly described in appendix I.

Figure 1:
An outline of the validation testing procedure.



Consider an uncertainty statement that proposes, for the calibration procedure previously defined, that any point-to-point length measurement within the workzone has an expanded uncertainty given by the root-sum-of-squares (RSS) of the three LDA values, i.e. $U = (\text{LDAX}^2 + \text{LDAY}^2 + \text{LDAZ}^2)^{1/2}$. Such an uncertainty statement seems inherently weak since the LDA tests are insensitive to many error sources, e.g. squareness. An interesting result of the simulation software is that this uncertainty statement passes 100 % of the (2,000) CMMs tested. It is relevant to note that the CMMs in the population have all the 21 parametric errors equally weighted, i.e. for a typical CMM all the parametric errors will be present at some level. This poses the question, “Does such an assessment method and associated uncertainty statement represent a valid calibration?” To address this issue we refer back to the definition of a calibration discussed earlier, and note that a calibration requires that all reasonable errors must be assessed by the procedure. Clearly the three squareness values are reasonable errors and subsequent simulation testing with the population of CMMs filtered to select those CMM with these parametric errors results in zero percentage of CMMs passing the validation test. Consequently, a proposed calibration procedure that involves an assessment involving only LDA tests fails to rise to the level of a calibration. This points out that the “reasonable errors” caveat requires that various sub-populations of CMMs, that display specific error behavior, be tested in addition to the general case of all errors being present. (A definitive statement about what constitutes “reasonable errors”, and hence should or should not be tested in a validation program, is best left to standards committees.) Subsequent investigation of the RSS LDA method show that the technique passes 100 % of the CMMs having only scale errors, and 90 % of the CMMs having only angular motion errors (pitch, roll, and yaw on each axis); see Table 1. Hence the explanation of passing 100 % of the CMMs that have all error types present is a result of the over estimation afforded by examining the *range* of the errors (as specified by an LDA test value), together with the over estimation from the RSS summation (roughly a factor of 1.7) of the three LDA test results. Apparently these two over-estimations together compensate for the insensitivity to squareness and straightness errors for CMMs with equally weighted parametric errors.

Another possible uncertainty statement for point-to-point length measurements is given by the results of the volumetric performance test, i.e. the ball bar range (BBR) value. Since the volumetric performance test was designed to detect geometrical error in a CMM it is not surprising* that CMMs having only squareness errors or only angular errors pass the validation test at a level of 100 % and 99 % respectively. For CMMs having only scale errors, the validation test passes 93 % of the CMMs. The 7 % of the CMMs that failed, i.e. produced more than 5 % of the length errors outside of the uncertainty interval, typically have scale errors that are commonly biased, e.g. all positive; this common bias is undetectable using an uncalibrated ball bar. The results for the various sub-populations of CMMs examined are shown in Table 1, e.g. 98 % of those CMMs having only straightness parametric errors passed the validation test. This result involving straightness errors was an unexpectedly good showing considering that only half of the ball bar positions are sensitive to straightness errors and each position is only sensitive to the change in straightness between the two ball centers.

In light of the volumetric test’s relative weakness with regard to parametric scale errors, we consider an alternative uncertainty statement that includes aspects of the LDA test since it is particularly strong on detecting scale errors. Instead of just adding the LDA and BBR results together we consider each point-to-point length in the testing program as a vector with a magnitude equal to its length and having components (L_X , L_Y , L_Z). We linearly “pro-rate” a portion of the appropriate LDA test result, e.g. LDAX (L_X/X_M), where X_M is the length of the X axis; similarly for the Y and Z cases. Finally we RSS these three terms together with the BBR value to create the expanded uncertainty statement; note that this

*For a CMM having only squareness errors, and a length extending between two extreme corners of the workzone, the length error may be relatively large. Apparently the fact the BBR value is a *range* of all the ball bar lengths, and that such extreme lengths occur infrequently, results in less than 5 % of the errors associated with randomly selected lengths lying outside an interval given by $\pm U = \pm \text{BBR}$.

uncertainty statement depends on the length and orientation of the point-to-point length under consideration. As shown in Table 1, this significantly improves the passing percentage of CMMs having only scale errors.

Table 1. Three different Uncertainty Statements Tested Against Various CMM Populations

CMM Population Description by Parametric Error Type	RSS of LDAs		BBR		RSS of pro-rated LDAs & BBR	
	% Pass of CMMs,	% Success of all Lengths	% Pass of CMMs	% Success of all Lengths	% Pass of CMMs	% Success of all Lengths
Scale only	100	100	93	98.5	99	99.8
Straightness only	0	0	98	99.5	98	99.5
Angular only	90	98	99	99.5	99	99.7
Straightness & Angular	78	96	99	99.6	99	99.7
Squareness only	0	0	100	99.9	100	99.9
All Errors Present	100	99.9	98	99.5	99	99.8

(The % pass of CMMs, and the % success of all lengths are reproducible to 1 % and 0.5 % respectively.)

While we considered only three simple examples of uncertainty statements, all based on B89 test procedures, this validation methodology can be extended to any well-defined uncertainty statement and any assessment test procedure, including user concocted tests. Additionally the method can be applied to any system that can be mathematically modeled; we are currently developing this for laser tracker systems. In addition to being a useful tool in analyzing calibration procedures, this work raises several interesting questions. What are the “reasonable errors” of a measurement system? Since straightness errors and angular errors usually occur together and are kinematically related, should CMMs that only have straightness errors be considered “reasonable” errors? Similarly, on a mathematical level, what limitations should be placed on the functional forms of the parametric errors? Other interesting questions concern the application of uncertainty statements to the calibration process. For example, the standard interpretation of ($k = 2$) uncertainty statements implies that 5 % of the point-to-point length errors may lie outside their uncertainty intervals, however, due to the finite number of lengths tested per CMM, a 95 % success threshold may pass some “bad” CMMs and fail some “good” ones. Similarly, for any stated criteria for passing, what percentage of calibrated instruments are allowed to “fail” this criterion and still have the calibration *method* considered valid? For example, must each CMM population in Table 1 pass the validation test at the 100 % level? These and similar questions should keep standards committees debating for years.

Acknowledgments: This work was funded in part by the US Navy’s CCG and the US Air Force’s CCG program, and by NIST’s computational metrology program.

Appendix I: Details of the CMM Population

In this paper we only consider CMM states that can be described by rigid body kinematics. Each state is composed of 18 functions and three scalar values. The functional form of the parametric error functions is created using low order polynomials and sinusoidal functions with random frequencies; an upper limit, typically of several cycles per axis, is imposed to maintain a plausible error form. Each term is given an independent random amplitude drawn from a uniform distribution in the interval (-1,+1). All terms of a particular parametric error, e.g. all terms in the functional form of X axis pitch, are multiplied by a prefactor which allows each parametric error to be given a relative weighting, including zero. Additionally, all 21 parametric errors are multiplied by another prefactor that controls the overall magnitude of the CMM errors. In our examples where all parametric errors are equally weighted, this means the prefactor for each parametric error is identical. However, for any one CMM state created, the various parametric errors can differ widely in their relative magnitudes due to the random nature of the amplitudes of each term in a parametric error expression. Hence the phrase “equally weighted” should be taken in a statistical sense.