

Concept for an Algorithm Testing and Evaluation Program at NIST

Cathleen Diaz
Factory Automation Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

1 Introduction

A coordinate measuring system (CMS) is "any piece of equipment which collects three dimensional coordinates (points), calculates and displays additional information using the measured points." [1]. Figure 1 displays a high level view of how a CMS works. A CMS—such as a coordinate measurement machine, vision system, or theodolite—works by measuring points on the part surface and then analyzing the points to evaluate the part. The coordinate data are evaluated by data analysis software embedded in the CMS. The data analysis software is what produces the dimensional measurements—surface fits.

CMSs are widely used by the dimensional metrology community during the inspection of manufactured parts. The data analysis software within CMSs is becoming increasingly important. During the last few years research [2] shows that data analysis software contributes significantly to the total measurement error (measurement uncertainty) of a CMS. Hence, the overall performance of CMSs is affected.

The performance of data analysis software is affected by the choice of analysis method, quality of the software, and characteristics of the specific measurement tasks. There are no standards for evaluating these and other effects of software on the overall uncertainty of measurement. Through standard efforts by the dimensional metrology community, industry has identified the need for a formal mechanism to test and evaluate data analysis software within CMSs.

This report presents a concept for a formal mechanism, the NIST Algorithm Testing and Evaluation Program (ATEP). Section 2 describes the current problem with CMS data analysis software. Section 3 presents a proposed solution to the problem. Section 4 explains the technical approach to the proposed solution. Section 5 presents a summary.

2 Problem Statement

Concerns in the confidence of results reported by CMSs have been growing since the 1980's. European efforts of the Physikalisch-Technische Bundesanstalt (PTB) and the National Physical Laboratory (NPL) in addressing CMS data analysis software began in the early '80s [3].

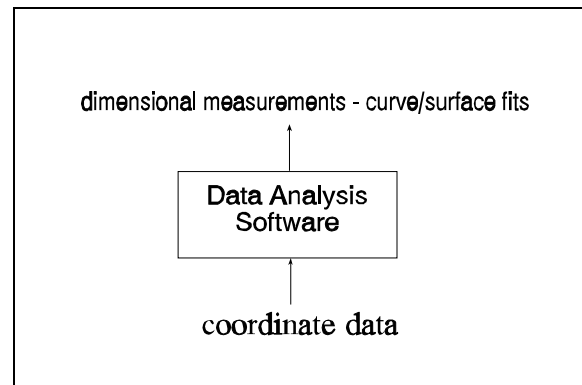


Figure 1 Coordinate measuring system—data analysis software analyzes coordinate data, producing surface fits.

In the mid '80s the American Society of Mechanical Engineers (ASME) established a working group on Coordinate Measurement Machine software.

The behavior, characteristics, and limitations of CMSs are not fully understood. This became evident at the 1988 ASME/NSF Workshop on Mechanical Tolerancing [4] and through the 1988 Government-Industry Data Exchange Program (GIDEP) Alert X1-A-88-01 on CMM software [5]. Results from the workshop stated that algorithm and software error effects on measurement uncertainty were often perceived as negligible. These assumptions led to overconfidence in results reported by CMS software. The GIDEP alert documented problems with tolerance computation software from several different vendors.

There are no guidelines or formal mechanisms for testing and evaluating CMS data analysis software in the United States. Europeans are currently addressing the need for formal mechanisms for testing data analysis software. NPL has prepared a document entitled *Proposed Draft ISO Standard: Method for Testing Software for Computing Gaussian Substitute Elements in Co-ordinate Metrology*. The U.S. and the Europeans are coordinating efforts in the development of such a standard.

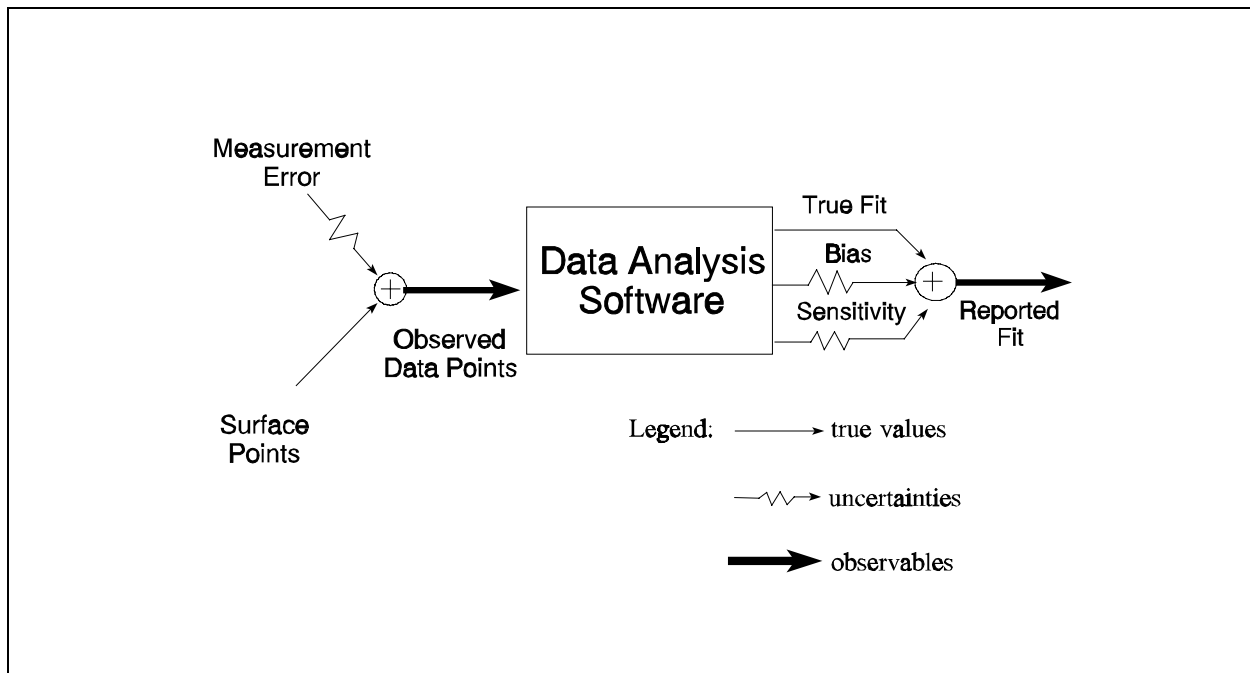


Figure 2 Data points are affected by measurement errors; fits are biased when tolerance theory and objective differ; software is sensitive to effect of measurement errors on reported fits.

Without a certification process, errors in the data analysis software go unidentified. Research demonstrates that software errors and measurement errors can propagate throughout the data analysis software, thus affecting the accuracy of quality control data for manufactured parts. Figure 2 shows an example of an error model demonstrating how measurement errors affect reported results from CMS software.

As a result of obtaining inaccurate reported fits, manufacturers can accept faulty parts or reject good ones. This problem worsens procurement methods, increases inspection costs, diminishes assurance of the quality of parts, increases measurement uncertainties, and lacks in traceability of measurement results.

3 Proposed Solution

NIST proposes the establishment of an Algorithm Testing and Evaluation Program (ATEP). The objective of ATEP is to provide the dimensional metrology community with a formal mechanism for testing and evaluating data analysis software found in CMSs. The approach focuses on three major goals:

- provide industry with a mechanism for evaluating CMS software,
- reduce measurement uncertainties associated with such software, and
- respond to industry’s need for an evaluation service that performs testing of CMS data analysis software.

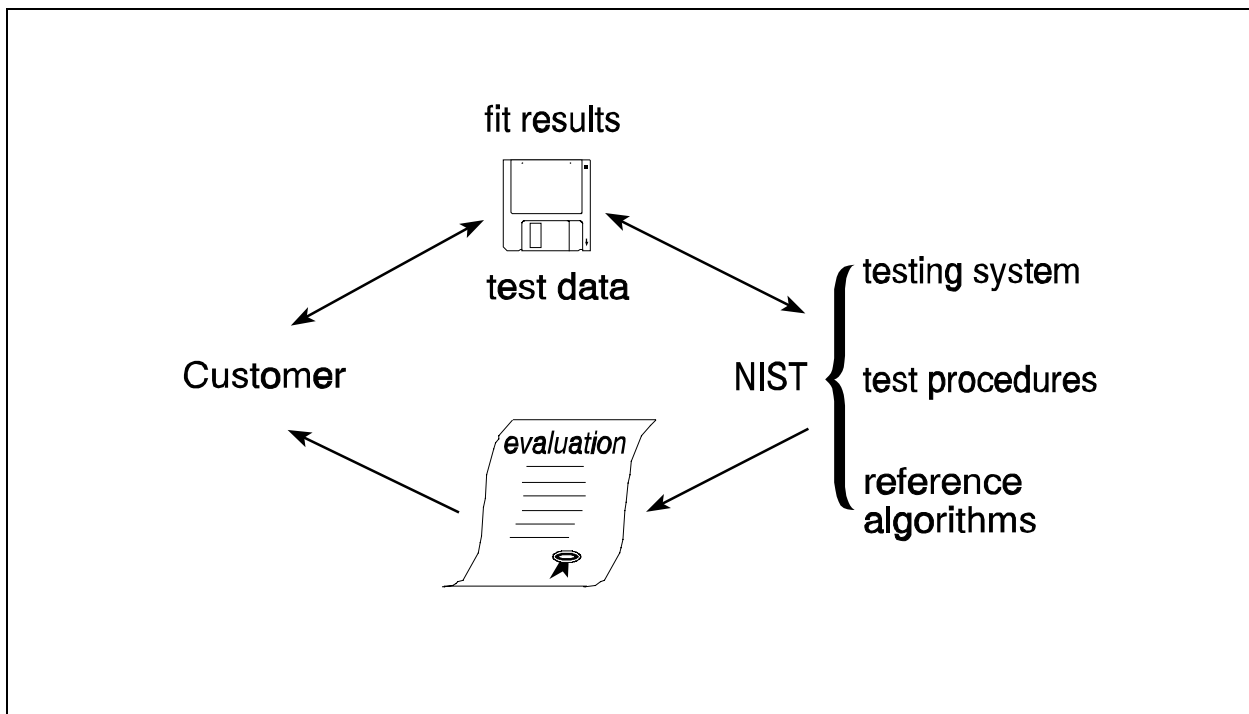


Figure 3 Based on the comparison of fit results, data analysis software is tested and evaluated by NIST for a given customer.

Figure 3 shows a high-level diagram of how ATEP will work. ATEP allows customers to submit a request to have their data analysis software tested and evaluated by NIST. NIST then provides the customer with NIST-generated data sets. The customers produce fit results from their data analysis software using the NIST-generated data sets. NIST generates fit results from

the same data sets using the ATS's reference algorithms. The two sets of fit results are then compared. In the comparison, NIST uses a "testing tool" (NIST's Algorithm Testing System (ATS)), geometric fitting reference algorithms, and test procedures developed in coordination with emerging ASME standards. NIST then provides the customer with a formal evaluation result.

In Figure 3 there are three components displayed under NIST's role in the concept for the Algorithm Testing and Evaluation Program. These components - a testing system, test procedures, and reference algorithms - are the key to the testing and evaluating process.

The testing system—NIST's ATS—is the tool used for generating test data, importing test data, generating fit results with the reference algorithms, and comparing generated fit results with the test fit results. The test procedures are currently being developed and are based on emerging standards from ASME Y14.5.1, B89.4.10, and B89.3.2 and from NIST's research efforts. The geometric fitting reference algorithms are incorporated within the ATS. These algorithms provide a baseline for performance comparison. Test fit results are compared to fit results generated by the reference algorithms in the ATS.

4 Technical Approach

ATEP combines the ATS, test procedures based on the emerging national standards on specifying the performance of CMS software, and control over the ATS reference algorithms. NIST will use these components to test and evaluate the CMS data analysis software.

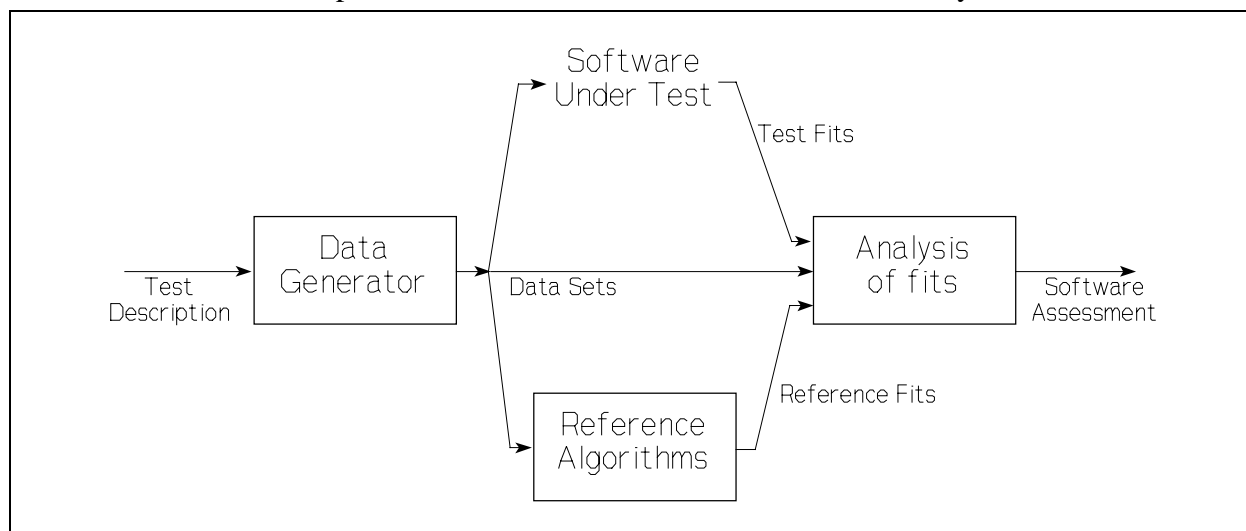


Figure 4 Architecture of the Algorithm Testing System—a tool for testing data analysis software.

The ATS was developed as part of NIST's contribution to the ongoing effort by industry standards groups to establish a standard for testing and evaluating data analysis software used in CMSs. The ATS is a software package for evaluating the performance of geometric fitting software (data analysis software). Figure 4 shows a diagram of the ATS architecture. The ATS consists of three components: a data generator, a set of reference algorithms, and analysis of fits.

The data generator produces data sets based on a test description. Four classes of information in the test description are used to generate the data sets: the nominal (ideal) geometry of the feature, the form errors of the feature being simulated, the sampling plan (distribution of points on the feature) for the data set, and the (random) measurement error distribution for the points. A test description consists of ranges of instance values for each of these information classes.

The data sets are processed by the software under test to generate test fits and by the reference algorithms to generate reference fits. The fits are analyzed in terms of geometric differences among fits. All fit geometries are bounded by the projection of the data onto the fit geometry. Test fits and reference fits are compared to evaluate the geometric differences between them. The data sets, test fits, and reference fits are used to assess performance measures for the software under test. For example, comparison of test fits to reference fits is used to detect code errors and to identify systematic bias. Comparison of fits among each other without a reference indicates how much they vary as a group. The geometric fitting reference algorithms provide a baseline for performance comparison.

Other geometric fitting algorithms could also serve as references. ATEP will address mechanisms for selecting and replacing reference algorithms with ones that provide an improved or expanded baseline for comparison. The selection process for reference algorithms will lead to identifying standard reference algorithms (SRA's) for different objective functions. A process for selecting standard reference algorithms for evaluating CMS software is described in a separate paper [6]. Briefly, that document presents a two-stage reference algorithm selection process: an alpha phase conducted by NIST and a beta phase of public review. Figure 5 shows an overview of the SRA selection process. The selection process will allow organizations to submit algorithms that could be useful to the dimensional metrology community. These algorithms would then be evaluated and possibly selected as a standard reference.

Test procedures within ATEP will be based on emerging national standards. NIST and the ASME B89.4.10 Working Group for Coordinate Measuring Systems Software Performance Evaluation are working together on a new U.S. standard, *Methods for Performance Evaluation of Coordinate Measuring System Software*. The purpose of this standard is to provide guidelines that benefit both CMS software suppliers and users in the development and application of inspection-related data analysis software. The ASME B89.4.10 standard will specify methods for characterizing and testing the performance of software used to reduce coordinate points to geometric parameters in metrology applications.

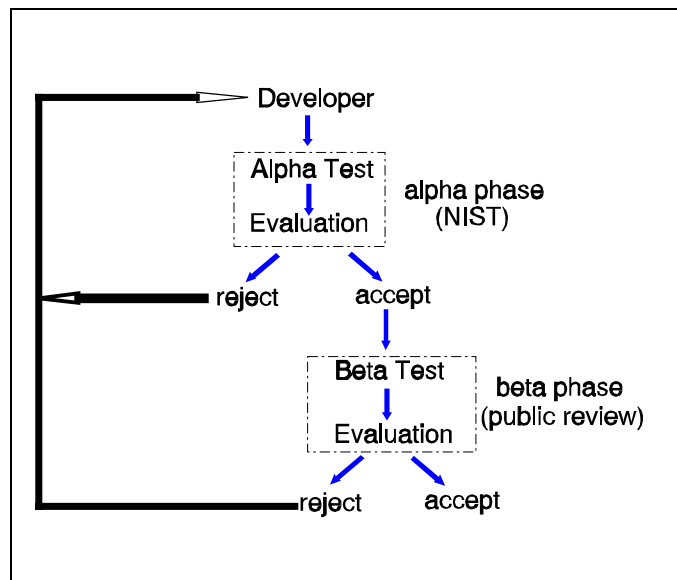


Figure 5 SRA selection process. The process proceeds in two phases—the alpha phase and the beta phase.

Other related emerging standards are the efforts of ASME Y14.5.1 draft standard *Mathematical Definition of Dimensioning and Tolerancing Principles*, and the ASME B89.3.2 draft standard *Dimensional Measurement Methods*. ASME Y14.5.1 establishes a mathematical definition of geometric dimensioning and tolerancing consistent with the practices and principles of ANSI Y14.5, enabling the determination of actual values. ASME B89.3.2 deals with the measurement of workpieces dimensioned and toleranced according to ANSI Y14.5.

In addition to ATEP's three components (the ATS, test procedures, and reference algorithms), it is being developed on a three-step process design philosophy [7]. The first step is to identify the performance factors of the fitting software—what conditions or influences affect the quantities of interest. The second step is to develop a performance model that relates the performance factors to levels of performance. The third step is to develop tests to evaluate the model parameters. The model can then be used to estimate measurement uncertainties of computed quantities.

The performance of fitting software is affected by two factors: the fitting objective chosen, and the implementation of that fitting objective in the software. Types of errors associated with the fitting objective are *bias* errors and *sensitivity* errors. A bias error is the deviation between the mathematical objective and the tolerance theory. A sensitivity error is the effect of point errors and sampling noise on reported fits. Factors that affect the implementation of the fitting objective in software include the optimization methods used to compute the fits, the computing environment, how extreme cases are handled by the software, and the correctness of the code.

ATEP will make three basic testing assumptions for testing software performance. The first assumption is that all tests will conform to well-defined error models. The second assumption is that test results must relate directly to inspection tasks. Test results will quantify the uncertainties of geometric relationships computed by the software. The third assumption is that the software to be tested will be considered to be a "black box." As a result, the testing method is limited to supplying the software with fitting problems and analyzing the fit results [8].

5 Summary

The NIST Algorithm Testing and Evaluation Program will be made available through the Calibrations Program Special Test Service through NIST's Office of Measurement Services. Intended customers include CMS vendors as well as users. ATEP is projected to be placed in service during 1994.

Current work includes program documentation for the Special Test Service, completion of a methodology for software performance testing, completion of procedures for a standard reference algorithm selection process, and completion of administrative policies.

Also, the ASME working groups mentioned throughout this report are currently developing the related standards. Since the software testing methodologies are based on the emerging standards, ATEP is being coordinated with the acceptance of these methodologies into U.S. standards.

Additionally, other work in software testing is being carefully studied. For example, Germany offers a service to test CMS software by comparing results for test data sets to results obtained from reference software [9]. Additionally, Great Britain has proposed mechanisms for testing form assessment software [10].

For further information, questions, or suggestions, please contact the following people.

Cathleen Diaz
National Institute of Standards & Technology
Metrology Building, Room A127
Gaithersburg, MD 20899
telephone: (301) 975 - 2889
fax: (301) 258 - 9749
email: diaz@cme.nist.gov

Dr. Theodore H. Hopp
National Institute of Standards & Technology
Metrology Building, Room A127
Gaithersburg, MD 20899
telephone: (301) 975 - 3545
fax: (301) 258 - 9749
email: hopp@cme.nist.gov

References

1. ASME, 1993, ASME B89.4.10-199x, *Methods for Performance Evaluation of Coordinate Measuring System Software*, Draft, American Society of Mechanical Engineers, New York.
2. Estler, Tyler W., 1989, *Accuracy Analysis of the Space Shuttle Solid Rocket Motor Profile Measuring Device*, NISTIR 89-4171, National Institute of Standards and Technology, Gaithersburg, MD.
3. Cox, M.G., 1992, *Improving CMM Software Quality*, NPL Report DITC 194/92, National Physical Laboratory, Middlesex, U.K.
4. ASME, 1988, ASME CRTD-15, *Research Needs and Technological Opportunities in Mechanical Tolerancing*, American Society of Mechanical Engineers, New York.
5. GIDEP, 1988, GIDEP Alert X1-A-88-01, Walker, R., *CMM Form Tolerance Algorithm Testing*, Government-Industry Data Exchange Program, DOD, Washington, D.C.
6. Algeo, M.E.A., and Diaz, C., 1994, NISTIR 5374, *A Process for Selecting Standard Reference Algorithms for Evaluating Coordinate Measurement Software*, National Institute of Standards and Technology, Gaithersburg, MD.
7. Diaz, C., and Hopp, T.H., 1993, *Testing Coordinate Measuring Systems Software*, to appear in proceedings of ASQC Measurement Quality Conference, October 26-27, National Institute of Standards and Technology, Gaithersburg, MD.
8. Hopp, T.H., 1993, *Computational Metrology*, Manufacturing Review, Vol. 6, No. 4, pp. 295 - 304, American Society of Mechanical Engineers, New York.
9. Wäldele, F., Bittner, B., Busch, K., Drieschner, R., Elligsen, R., 1993, *Testing of Coordinate Measuring Machine Software*, Precision Engineering, Vol. 15, pp.121-123.
10. Cox, M.G., and Forbes, A.B., 1992, *Strategies for Testing Form Assessment Software*, NPL Report DITC 211/92, National Physical Laboratory, Middlesex, U.K.