THE HISTORY AND STATISTICAL DEVELOPMENT OF THE NEW ASME-SAE-AIAA-ISO MEASUREMENT UNCERTAINTY METHODOLOGY

Dr. Robert B. Abernethy*
Barbara Ringhiser
Pratt & Whitney Engineering Division
West Palm Beach, Florida

Abstract

A standard measurement uncertainty methodology has been adopted by the SAE,[1,2,3], ASME,[4,5,6] and AIAA[7,8,9]. The same standard has been recognized and adopted by ISA[10,11], ISO[12], USAF[13] and JANNAF[14]. This is remarkable after decades of striving to reach agreement. The objective of this paper is to document some of the history, statistical evaluation and significant contributions that led to this national and international consensus. The statistical Monte Carlo simulations that led to selecting this methodology will be described. It is hoped that the novice measurement uncertainty analyst will gain some understanding of the reasoning that supports the standard.

Standard Methodology

It is not the purpose of the paper to detail the methodology as that is available in the above references, particularly[1,4,5,12]. However, a brief summary may be helpful and is provided in Appendix A.

History

One organization, the National Bureau of Standards, has provided leadership in the field dating all the way back to Mayo Hershey[15] in 1911. Dr. Harry Ku, Dr. Joan Rosenblatt, Mr. J. Cameron, Churchill Eisenhart, Mary Natrella, Dr. Clifford Spiegelman and others have provided a wealth of noteworthy papers and reports.[16,17,18,19] In addition, Drs. Ku, Rosenblatt and Mr. Cameron provided invaluable guidance and consultation during the two decades of development of this methodology.

The hallmark paper by Professor Kline and McClintock[20] published in 1953 was used by many engineers within ASME for guidance. Starting in the fifties A.S.M.E.'s Performance Test Codes Committees made attempts to write a standard measurement uncertainty document which finally reached success in 1985[6]. In 1965, JANNAF, then called the ICRPG (Interagency Chemical Rocket Propulsion Group) organized a Performance Standardization Committee to develop standards for the rocket engine industry. The first meeting produced arguments about measurement error. A survey of the industry showed that there were many different methods in use and a contract was awarded to write a standard handbook. The effort took two years and produced many heated disagreements. The ICRPG Committee under Don Bartz of JPL (Chairman) rejected the handbook and awarded a second contract that produced an acceptable document[14] in 1968.

The success of the second effort was due to the strong support from NBS and to Monte Carlo simulation of alternatives that provided objective comparisons. These comparisons led to recommendations proposed to the NBS team and the ICRPG Committee. Four of the five final solutions employed in the standard methodology were developed at this time. These five problems will be discussed later in this paper.

The ICRPG Handbook was presented at the 1969 SAE-ASME-AIAA Propulsion Conference[8] and received widespread acceptance. The propulsion community successfully applied the rocket methodology to gas turbines which inspired a contract from the USAF Applied Propulsion Laboratory to produce a similar handbook for jet engines. This document[13] was given worldwide distribution by the USAF after introduction at the 1972 SAE-AIAA-ASME Propulsion Conference[9]. It was widely used within the aerospace industry.[21,26,43] In 1980 the Instrument Society of America reprinted it as an ISA Handbook[10]. However, the fifth problem, the dispute over how to combine random and systematic errors to obtain an uncertainty continued to rage through the seventies and even the authors were pessimistic about the prospects of settling the issue. The "great compromise" provided the solution.

The five problems and solutions follow:

1. Random Error Uncertainty

Precision error, repeatability, sampling error, and random error are all synonyms for the scatter we see in repeated measurements of the same thing. We do not expect repeated measurements to agree exactly. Fortunately, we have good statistical methods to estimate the random error uncertainty. There seemed to be general agreement that some multiple of the standard deviation should be used[17,20,21]. Two or three were the popular choices, or for small samples, Student's t_{95} or $t_{99.7}$. Monte Carlo simulation showed that three standard deviations when combined with bias limits produced large, very conservative uncertainty intervals that were unacceptable to instrumentation engineers. Eventually, two (or t_{95} for small samples) standard deviations became the consensus. (All Monte Carlo simulations discussed herein were programmed by the authors.)

However, a problem appeared when several measurements were combined to calculate a final test result. How should the estimates of measurement precision errors be propagated or combined to obtain the precision error estimate of the test result? For example, let us assume we have a compressor on test to measure thermodynamic efficiency. If we calculate the sample standard deviation(s) for each temperature and pressure in the inlet and exhaust, how should they be combined or propagated to obtain the corresponding standard deviation of efficiency, the test result? Many engineers opted for root-sum-squaring the products.[20,21,22]

$$(t_{95}S)_{Result} = \sqrt{\sum_{\substack{\text{all}\\ \text{measurements}}} (t_{95}S)^2} \qquad (1)$$

* Manager, RMS Engineering
Associate fellow AIAA

They reasoned, correctly, that for very large samples:

$$(2S)_{Result} = \sqrt{\sum_{\substack{all \\ measurements}} (2S)^2} \qquad (2)$$

However, this logic falls apart for small samples because the characteristic of a confidence interval is not invariant under transformation. If 95% small sample confidence intervals are root sum squared, the resultant interval will not be a 95% interval. This was demonstrated with Monte Carlo simulation in 1967 for the ICRPG Committee. (Figure 1).

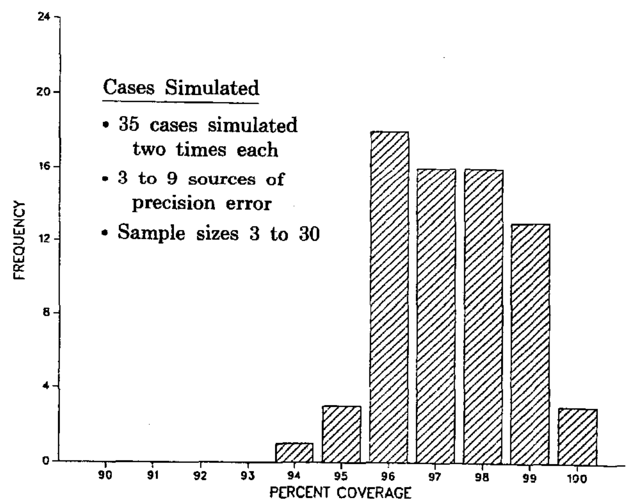

Fig. 1 Root-sum-square of ($t_{95} \times S$) does not produce 95% coverage.

Statistically, a better solution requires two steps:

$$(1) \quad S_{Result} = \sqrt{\sum_{\substack{all \\ measurements}} (S^2)} \qquad (3)$$

$$(2) \quad \text{Random Uncertainty} = \frac{t_{95}\, S_{Result}}{\sqrt{N}} \qquad (4)$$

If the result is an average the random uncertainty is reduced by the $\sqrt{N}$ when N points are averaged. Of course, if the result is a single value, $\sqrt{N} = 1$ This situation can arise if the test result is a single value and the standard deviation is based on prior test data.

However, the second step contains a formidable problem. Student's t_{95} is tabled as a function of degrees of freedom (ν). The usual sample standard deviation is distributed as the Chi-Square distribution, but the root-sum-square of a number of sample standard deviations tends toward a normal distribution because of the Central Limit Theorem. This was verified by Monte Carlo Simulation for SAE Committee E33 in 1982. As a result, there is no known exact solution to the problem of combining or propagating the degrees of freedom of the sample standard deviations to the degrees of freedom of the end test result, but there are approximations. The Welch-Satterthwaite approximation was selected as best based on Monte Carlo simulation[23]. Unfortunately, it is a rather complex solution:

$$\nu_{Result} = \frac{[\Sigma S_i^2]^2}{\Sigma \frac{S_i^4}{\nu_i}} \qquad (5)$$

Using the Welch-Satterthwaite equation[24,25] produces random uncertainty confidence intervals that are well behaved as shown by Monte Carlo simulation (Figure 2).

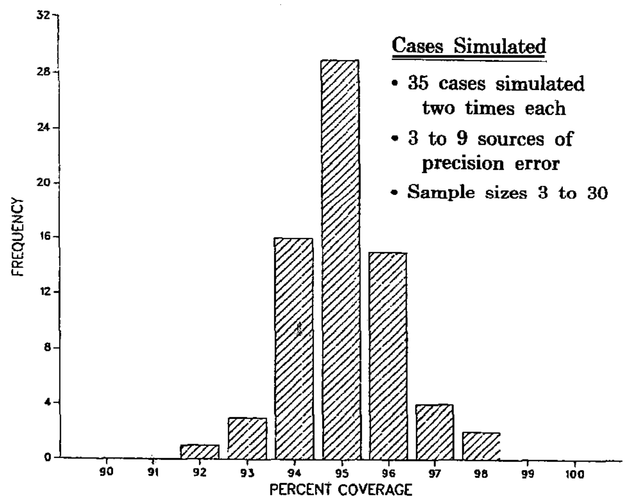

Fig. 2 Welch-Satterthwaite $t_{95} \times \sqrt{\Sigma S^2}$ produces 95% coverage.

This solution was quite acceptable, although complex. In an effort to simplify, Abernethy suggested that Student's t and Welch-Satterthwaite are only required for small samples. He recommended that above sample sizes of 30, t_{95} be approximated by 2.0 and degrees of freedom may be ignored. Examination of a Student's t table shows that for a sample of 30, degrees of freedom of 29, $t_{95} = 2.045$ and for an infinite sample, $t_{95} = 1.96$, so 2.0 might be an acceptable approximation. This suggestion was accepted by all of the committees as it produces a useful simplification. In most cases, small sample methods are not needed. This simplification has been described as "two for t for simplicity."

In summary, the recommended random uncertainty ($t_{95} \times S$) provides a rigorous statistical confidence interval (if the systematic error is zero or negligible) for the unknown true mean value. This is precisely what Gossett (pen name "Student") intended when he derived his t statistic at the Guinness Brewery in Dublin in 1905. Incidentally, he also used Monte Carlo simulation to verify his work, even though the computer had not yet been invented.

2. Systematic Error Uncertainty

Uncertainty analysis assumes a carefully controlled measurement process within which every calibration constant and correction has been applied. Systematic error is a synonym for fixed or bias errors. The systematic uncertainty is an estimate of the upper limit of the systematic error. In this country it is popularly called the bias limit, B. As all corrections have been made there is usually no data to calculate B. Therefore it is usually estimated from judgement and experience. Statistical methods do not apply.

Further, the distribution of bias limits is both unknown and probably unknowable, and therefore, how they should be combined is not obvious.

The Central Limit Theorem of Statistics leads us to believe that the sum of the true unknown bias errors is normally distributed assuming there are many, say more than ten. Some experts believe the test result bias limit should also be normally distributed although this is argumentative. If so, it would be logical to root-sum-square the bias limits and even if normality is not assumed, there are arguments to RSS, i.e. it is analogous to calculating the standard deviation.

Some Soviet uncertainty analysts believe the bias limits are randomly distributed (rectangular distribution). If so, the Central Limit theorem would lead us to believe that the test result bias limit would be normally distributed and the bias limits should be combined by RSS or quadrature. At the request of the Soviet delegation, ISO TC30 SC9 Revised D.P. 5168[12] has an appendix illustrating this approach.

Two arguments against the root-sum-square can be found in literature. If there were only three of four bias limits, the probability that they would have the same sign is significant, say one fourth or one eighth. This has inspired some to argue that the bias limits should be added together. However, Professor S. Kline argues and we agree, that this requires that they all be at their maximum value in addition to having the same sign. The probability of both these requirements being met is very small. For example, the probability of four bias limits all being within 1% of their maximum positive value would be $(0.01)^4$ or 10^{-8}, assuming a random (rectangular) distribution.

Also, if a measurement uncertainty analysis identified four or fewer sources of bias, there should be concern that some sources have gone unrecognized.* The analysis should be redone and expert help should be recruited to examine the calibration hierarchy, the data acquisition process, and the data reduction procedure for additional sources. This comment originated with Mr. Cameron of NBS.

The second argument is that one or two bias limits may be an order of magnitude or more larger than all the others. If so, it would be prudent to add the large ones to the RSS of the small ones. The counter argument is that such a situation would lead to corrective action to reduce the enormous source(s) of error and would contradict the "controlled measurement process" that we have assumed.

Therefore, for all the above reasons the root-sum-square** combination of bias limits is widely used and recommended.

* A.T.J. Hayward[26] indicates: "A full breakdown would probably reveal several dozen primary sources of uncertainty in the measurement..." p. 10.

** Hayward[26] further indicates: "The real justification for adding uncertainty components in quadrature is that it seems to work. Experience has shown that arithmetic addition of components often leads to a large overestimate of total uncertainty." p.19.

3. Signed Bias Limits

Some types of bias limits may have a know sign. Leakage errors in the process of measuring a high pressure or radiation errors from a thermocouple in a hot combustor would both be negative if they exist. The bias limits which result are non-symmetrical, i.e., not of the form ± B. They are of the form zero to minus B (or plus B). Table 1 lists several non-symmetrical bias limits for illustration.

Table 1 Non-symmetrical Bias Limits

Bias Limits	Explanation
0, +10 deg	The bias will range from zero to plus 10 deg
0, +7 psia	The bias will range from zero to plus 7 psia
−8, 0 deg	The bias will range from minus 8 to zero

Therefore, the combination of biases in this case cannot be handled by simply root-sum-squaring them — the upper and lower ranges must be combined separately.

For example, assume that B_2 and B_7 are non-symmetrical, and designated by B_2^+, B_2^-, B_7^+, and B_7^-. The combination with the symmetrical bias limits B_1, B_3, B_4, B_5, and B_6 is handled as follows:

$$B^+ = \left[B_1^2 + (B_2^+)^2 + B_3^2 + B_4^2 + B_5^2 + B_6^2 + (B_7^+)^2 \right]^{1/2} \qquad (6)$$

and

$$B^- = \left[B_1^2 + (B_2^-)^2 + B_3^2 + B_4^2 + B_5^2 + B_6^2 + (B_7)^2 \right]^{1/2} \qquad (7)$$

4. Defined Measurement Process

Review of the many methods used in the sixties showed that one reason for the variation was that different measurement processes were involved. For example, Kline assumed a single sample, (no averaging) while many writers were averaging to reduce the random uncertainty. Abernethy, Ku, Eisenhart, Rosenblatt and others recognized that the test result uncertainty depends on the defined measurement process. The standard method must be applicable to any and all processes.

The decision on classifying elemental bias or precision errors depends on the defined measurement process. It should be noted that this may be an iterative process, redefining the measurement process until an acceptable pretest uncertainty is established. Moffat has championed this concept. Calibration errors are the most common initial error(s) which may be reclassified as a result of the defined measurement process. For example, if we are considering the measurement of air flow for gas turbine engines at a test facility over a long period such as many calibrations, the uncertainty of these measurements will contain errors due to variations in calibrations. The calibration process would contribute both bias and precision errors. However, if the same measurement was taken on a test involving only one calibration, then the precision in the calibration process would manifest itself as a bias. Classification of a calibration precision error(s) to a bias error(s) is marked with an asterisk (s*) to indicate that it is a fossilized or fixed error for this process (due to Ascough).

The uncertainty analysis for the above examples will be different from the uncertainty analysis for a comparative, back-to-back test to measure air flow on a single test stand for a single engine, which is a different measurement process. In a comparative test, the objective is to measure the change in the test result from the baseline test to the second test with the new design change. Calibration errors (all bias) may be ignored in comparative testing in that the same instrumentation and equipment must be used for both tests, and bias errors cancel out in the comparison or difference since calibration errors do not affect the comparison or difference between one test and another (the test objective being to determine if a design change is beneficial). In these three examples, the defined measurement process may have included the same engine, instrumentation, and test stand; however, there are differences in uncertainties due to the differences in test objectives and test duration.

The planned instrumentation, type and number, is also part of the definition of the measurement process. If the end measurement is an average of (1) a series of individual repeat points or (2) a number of simultaneous readings, or (3) a combination of both, this must also be specified, as the precision index depends on this information. Significant reductions in precision error can be obtained if averaging can be used under most conditions. (Averaging can be used (1) with repeated single measurements if the measured variable is constant, or (2) if redundant instruments can be recorded simultaneously).

In summary, the uncertainty analysis depends on a well defined measurement process. Before this was recognized, the classification of elemental errors was complex and frustrating. However, once it was recognized that the final classification depended on the defined measurement process, a simple rule for the initial classification could be adopted.

"The elemental error of a measurement should be put into one of two categories depending on how the error is derived. A random error is derived by a statistical analysis of repeated measurements while a systematic error usually must be estimated by nonstatistical methods[27]."

This approach makes a very complex situation manageable and keeps the statistical estimates and the judgment estimates error components separate until the appropriate time to combine them.

Note that this is consistent with the BIPM or CIPM recommendations[28].

A step-by-step measurement uncertainty analysis procedure is given in Appendix A.

5. Uncertainty Intervals

The question of how to combine the random and systematic uncertainties has been the major issue in the uncertainty debate during this century. Each side, addition versus root-sum-squaring (RSS), has attempted to build logical, objective cases without success. The intensity of the arguments was unbounded. Monte Carlo simulations originally produced by the authors in 1967 were later independently reproduced by John Ascough in England (1978), Robert Benedict (Westinghouse) and George Kelley (General Electric) in 1980. The simulation results provided quantitative comparisons and will be described later. However, the results did not settle the argument.

In general, the instrumentation and calibration people favored the optimistic, shorter RSS interval. The data analysts, the users, preferred the conservative, larger, addition interval proposed by NBS. Internationally, the British[21] leaned toward the RSS (so did Kline and McClintock[20]), even though Hayward[26] preferred addition.. The Soviets also preferred the addition.

By the late seventies, the authors concluded that the nation and world seemed equally divided. No standard would ever be possible. At that point, Dr. Joan Rosenblatt wrote to Abernethy[19] and proposed the "great compromise." She suggested that under certain constraints it would be acceptable to allow either addition or RSS or even some other combination. The constraints were that (1) the random and systematic components be separately propagated to the end test result (2) that the two components be reported separately and (3) that the choice of uncertainty formula be stated. Under these constraints, the uncertainty interval is the last calculation and may easily be redone if desired.

Abernethy easily agreed and proposed this compromise to SAE E33, ASME MFFCC, ASME PTC 19.1 and ISO TC30 SC9 Committees. All agreed (and breathed a sigh of relief). The arguments ceased. Formal NBS recognition of the compromise may be found in (NBS Special Publication 644). This led to standards[1,4,6,12]

The simulation comparisons of the addition and RSS methods are shown in Appendix B.

Areas for Future Research

There are at least three areas where more work is needed:

1. Calibration Curve Fitting
2. Weighting Competitive Answers
3. Outlier Methods

1) Curve fitting is usually based on least squares regression. The application to calibration produces some unusual problems. First, both the test meter and the master meter will have some precision error, which contradicts the least squares assumption that all the error is in the dependent variable. Secondly, if the test meter is regarded as the dependent (Y) variable because it usually has the larger error, the equation must be inverted to master meter (X) as a function of test meter (Y) to be used.

There have been many solutions proposed as to how to best fit calibration data considering these two problems. Simulation studies by the authors for the linear case show the Berkson[31] approach to be best, i.e. minimum error. Mandel[32] has also proposed a method which needs further study. However, for higher order there is no known best solution and the problem is very complex. Work is needed in this area.

2) Another problem is that of how to best weight competitive solutions. For example, let us assume there are three different methods for determining in-flight propulsive thrust in a flight test program. If the bias errors were negligible, we can use straight forward least squares theory to determine the weighting factors as a function of the predicted precision errors. However, in the usual case these bias errors are not negligible. Adams[33] has suggested weighting by uncertainties and the ASME PTC19.1 committee adopted this suggestion[6]. The authors of this paper wrote the section in[6] based on Monte Carlo simulation studies. This approach needs to be validated, hopefully with more rigor. It would be a useful technique in many testing applications.

3) As to the third problem, uncertainty analysis assumes a carefully controlled measurement process such that there are no wild, spurious observations. The detection and rejection of these outliers is an intriguing statistical problem. Ideally, outlier detection leads to engineering analysis that provides a reason for the rejection of an outlier. However, sometimes the sheer size of the data precludes such an analysis and outlier rejection methods are needed.

Two methods have gained wide acceptance over the years — Grubbs[34] and Thompson's τ[35] and their validity has been shown by Monte Carlo simulation by the authors.

The first, Grubbs' method, has been adopted by the ISO[12], ASME[4], and others and is usually recommended for use in computerized outlier rejection because it rejects fewer points than Thompson's τ.

The second, Thompson's τ is recommended[4] for outlier detection where conservatism is not as important. Both methods assume normality of the data, which is a reasonable assumption. Non-parametric methods such as Tchebychev's inequality are needed in those rare cases where the data is non-normal.

If the data has some multivariate functional relationship such as a regression curve fit, outlier detection and rejection is more difficult. Several methods have been proposed[36,37]. The authors have investigated and recommend[38] for linear relationships but no formal acceptance has been given by the committees listed herein. For higher order, a simple approach is to use ± 1.9 standard errors of estimate (SEE) for all but the end points. A smaller multiple of SEE is more appropriate for the end points. This is definitely an area for further research.

Summary

A national and international consensus now supports a standard uncertainty methodology. The standard method is quite consistent with the CIPM or BIPM recommendations[28] if they are assumed to apply to the calibration portion of the error. The notation used in these standards is approximately the same as the International VIM[29] (even though the standards were written before VIM existed). The SAE E33 report represents the first international agreement to apply the method to aircraft performance. The ASME PTC19.1 standard is a readable presentation of the method which will be applied to all ASME Performance Test Codes. Internationally it has been presented in the Netherlands, Belgium, France, Great Britain, USSR and Israel by the authors.

Acknowledgements

This standard is the product of many people and many years of committee work. In addition to those mentioned herein the authors would like to recognize the contributions of the following:

Billy D. Powell (a), Ron Dieck (b,e) — Pratt & Whitney

Roger Dowdell (a) — University of Rhode Island

Dick Miller (a) — Foxboro

Robert B. Benedict (b) — Drexel University

George Kelley (b) — General Electric

Jim W. Thompson Jr. (c), S. Wehofer (c,d), and Dr. W. Kimsey (c) — Sverdrup

Gary Adams (c) USAF, WPAFB

John Ascough (c,d), RAE, Pyestock, England

Dr. A. Hayward, Dr. Colin Martin (d), Dr. Frank Kinghorn (d) H.M., Engineering Laboratory Scotland.

Dr. S. S. Kivilis (d) USSR

Dr. J. Mitchell (f), USAF AEDC

Rolf Schumacher (g), North American Rockwell

Committees
(a) ASME MFFCC and (b) PTC 19.1
(c) SAE E33
(d) ISO TC30 SC9
(e) ISA Measurement Uncertainty Committee
(f) NATO AGARD WG15 Uniform Engine Test Program
(g) ASQC (ANSI Z1) Writing Group on Calibration Assurance

Appendix A

Measurement Uncertainty Analysis Procedure

The procedure to follow in performing measurement uncertainty analyses is as follows.

(a) Analyze the formula or data reduction by which the final answer will be obtained to determine which values (measured or constant) must be investigated in the uncertainty analysis. Study the measurement system.

(b) For each measurement, make an exhaustive list of every possible elemental error, including calibration errors, data acquisition errors, and data reduction errors.

(c) Obtain an estimate of each elemental error and preliminarily classify according to the following:

"The elemental error of a measurement should be put into one of two categories depending on how the error is derived. A random error is derived by a statistical analysis of repeated measurements while a systematic error usually must be estimated by nonstatistical methods."[27]

(d) Make the final classification of the elemental errors into bias and precision based on consideration of the defined measurement process.

(e) Calculate the precision index S and estimate the bias limit B for each measurement.

(f) Propagate the precision index to the test result using the Taylor series expansion.

(g) Propagate the bias limit for the test result using the Taylor series expansion.

(h) Evaluate the degrees of freedom for the calculated parameter using the Welch-Satterthwaite formula if the precision sample standard deviation is based on small samples.

(i) Calculate the uncertainty of the calculated parameter using Eq. (8), i.e.,

$$U_{ADD} = \pm\left(B + t_{95}\frac{S}{\sqrt{N}}\right) \text{ and/or } U_{RSS} = \pm\sqrt{B^2 + \left(t_{95}\frac{S}{\sqrt{N}}\right)^2} \quad (8)$$

(t_{95} may be taken as 2.0 for large sample estimates of precision)

(j) Report
- The bias, precision, and total uncertainties of the calculated parameters.
- The equivalent degrees of freedom if the precision is based on small samples.
- The uncertainty formula employed.

Appendix B

Monte Carlo Comparison of Alternative Uncertainty Methods

If the bias and precision error estimates are propagated separately to the end test result and the equation used to combine them into uncertainty is stated, either U_{ADD} or U_{RSS} can be used. Monte Carlo simulation was used to compare the additive and root-sum-squared values.

Uncertainty Interval Coverage

A rigorous calculation of confidence level or coverage of the true value by the interval is not possible because the distributions of bias limits, based on judgment, cannot be determined. Monte Carlo simulation of the intervals can provide approximate coverage assuming various bias limit distributions.

For large samples (N > 30):

$$U_{ADD} = B + \frac{2S}{\sqrt{N}} \quad (B.1)$$

$$U_{RSS} = \sqrt{B^2 + \left(\frac{2S}{\sqrt{N}}\right)^2} \quad (B.2)$$

For small samples:

$$U_{ADD} = B + t_{95}\frac{S}{\sqrt{N}} \quad (B.3)$$

$$U_{RSS} = \sqrt{B^2 + \left(t_{95}\frac{S}{\sqrt{N}}\right)^2} \quad (B.4)$$

where t_{95} is determined from the degrees of freedom calculated from the Welch-Satterthwaite approximations. If the test result is an average of N points the $\sqrt{N}$ accounts for the improvement in precision.

As the actual bias error and bias limit distributions will probably never be known, the simulation studies were based on a range of assumptions.

Simulation Cases Considered (102 cases)

A number (102) of simulation cases were first considered. The results in Figures B.1 through B-4 are based on these cases. These included:

- From 3 to 5 error sources, both bias and precision
- Bias errors distributed both normally and rectangularly
- Precision distributed normally
- Bias limits at both 95% and 99.7% for both the normal and the rectangular
- Precision indexes based on sample sizes from 3 to 30
- Ratio of precision to bias errors at 1/2, 1.0 and 2.0

Another 102 cases were simulated with errors from up to 19 sources with similar results.

The result of the studies comparing the two intervals are:

- U_{ADD} averages 99.1% coverage while U_{RSS} provides 95% based on bias limits assumed to be 95% (2σ for normally distributed biases and 1.645σ for rectangularly distributed biases.) (Figure B.1).

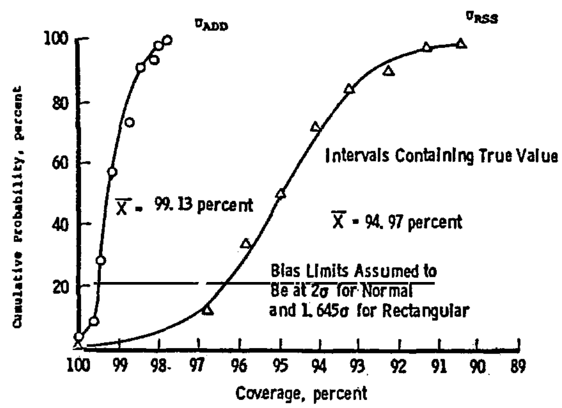

Fig. B-1 Coverage distribution for 95% bias limits.

- U_{ADD} averages approximately 99.7% coverage and U_{RSS} coverage is 97.5% if the bias limits are assumed 99.7% (3σ for normal and 1.732σ for rectangular bias limit.) (Figure B.2.).

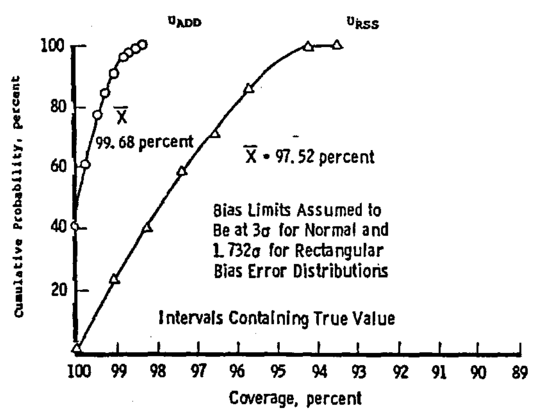

Fig. B-2 Coverage distribution for 99.7% bias limits.

- Because of these coverages, U_{ADD} is sometimes called U_{99} and U_{RSS} is called U_{95}.

- If the bias error is negligible, both intervals provide 95% statistical confidence.

- If the precision error is negligible, both intervals provide 95% to 99.7% depending on the assumed bias limit size.

- When the interval coverages are compared, U_{ADD} provides a more precise estimate of the interval size (range of 98% to 100%) as opposed to 93% to 100% for U_{RSS} (Figures B.1 to B.4).

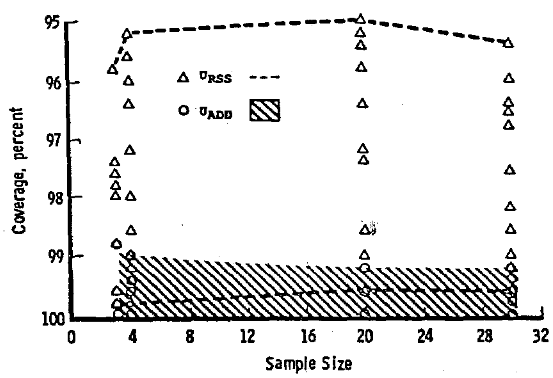

Fig. B-3 U_{ADD} and U_{RSS} sensitivity to variation in sample size.

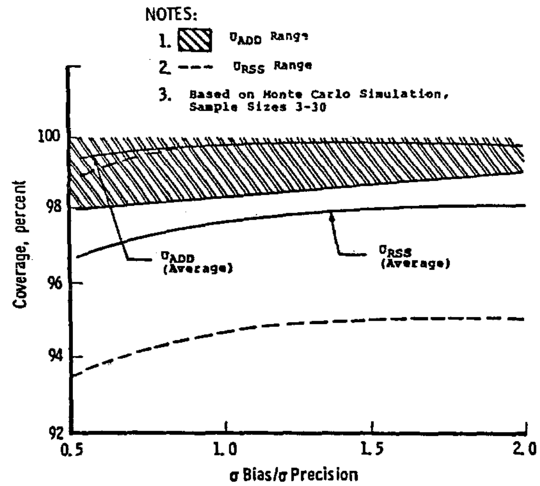

Fig. B-4 U_{ADD} and U_{RSS} sensitivity to bias / precision error ratio.

- U_{ADD} is larger than U_{RSS}. The average ratio of the U_{ADD} interval size to U_{RSS} interval size is 1.35:1.

The Simulation Procedure

1. Measured values, $X = b_1 + \ldots + b_N + \varepsilon_1 + \ldots + \varepsilon_N$ were generated. The b_i's are randomly generated bias errors generated from N distributions (rectangular or normal) and the ε_i's are randomly generated precision errors from the normal distribution. The means of the distributions were all zero. The precision indices were input as above.

2. N random samples of size $n_1, \ldots, n_N$ were drawn from normal distributions, $i = 1, \ldots, N$.

3. The standard deviations of the N distributions, the Welch-Satterthwaite degrees of freedom and the U_{ADD} and U_{RSS} uncertainty intervals were calculated.

4. It was determined whether

$$(X - U_{RSS}) \leq 0 \leq (X + U_{RSS})$$

$$(X - U_{ADD}) \geq 0 \geq (X + U_{ADD})$$

The percent of times the interval covered 0 was calculated for 500 cases. The results are shown on Figures B.1, 2, 3 and 4.

References

1. "Uncertainty of In-Flight Thrust Determination," SAE AIR 1678, 1985.*

2. R. B. Abernethy, "SAE In-Flight Propulsion Measurement Committee E33: Its Life and Work," SAE in Aerospace Engineering Volume 1, No. 1, July 1981.

3. R. B. Abernethy, et al, "Uncertainty Methodology for In-Flight Thrust Determination," SAE paper 831438, October 1983, see also SAE paper 831439 for application.

4. ANSI/ASME MFC-2M, "Measurement Uncertainty for Fluid Flow in Closed Conduits," 1983.

5. R. B. Abernethy, R. P. Benedict, and R. B. Dowdell, "ASME Measurement Uncertainty,", ASME paper 83-WA/FM-3.

6. ANSI/ASME PTC 19.1, "Measurement Uncertainty," Performance Test Codes Supplement, 1985*

7. AIAA Thrust — Drag Editorial Board, chaired by Dr. E. E. Covert, Director MIT, "Measure and Prediction of Thrust and Drag from Transports to Fighters," 1985.*

8. D. L. Colbert, B. D. Powell, and R. B. Abernethy, "The ICRPG Measurement Uncertainty Model," AIAA paper No. 69-734.

9. R. B. Abernethy, and J. W. Thompson, Jr., "Uncertainty in Gas Turbine Measurements," AIAA paper no. 73-1230.

10. R. B. Abernethy, et al, "Uncertainty in Gas Turbine Measurements," Revised 1980 Edition, ISA I-483-3.

11. R. B. Abernethy and R. P. Benedict, "Measurement Uncertainty: A Standard Methodology," ISA Paper, 1984, 0-87664-806-5/84.

12. ISO TC30 SC9, "Fluid Flow Measurement Uncertainty," Draft of January 1985. (Revision to ISO 5168)

13. R. B. Abernethy, and J. W. Thompson, Jr., "Handbook Uncertainty in Gas Turbine Measurements," AEDC-TR-73-5 February 1973.

14. R. B. Abernethy, et al, "ICRPG Handbook for Estimating the Uncertainty in Measurements Made With Liquid Propellant Rocket Engine Systems," JANNAF (formerly ICRPG) Performance Standardization Working Group Report CPIA No. 180 (AD 851 127), April 1969.

15. M. D. Hershey, "A Development of the Theory of Errors with Reference to Economy of Time," British Association for the Advancement of Science, 1913, Reprinted by Churchill Eisenhart, NBS, Journal of Research, 1965.

16. H. H. Ku, "Uncertainty is NBS's Business," presented to the NBA Advisory Panel, October 9, 1975.

17. H. H. Ku, "Precision Measurement and Calibration," NBS Special Publication 300, Volume 1, 1969.

18. J. Rosenblatt and C. H. Spiegelman, NBS, "Discussion," Technometrics Volume 23, No. 4, November 1981.

19. Letter Dr. Joan R. Rosenblatt, Director Center for Applied Mathematics, NBS to Dr. Robert B. Abernethy, December 8, 1980 & NBS Special Publication 644, "Expression of the Uncertainties of Final Measurement Results" Reprints, January 1983.

20. S. J. Kline and F. A. McClintock, "Describing Uncertainties in Single-Sample Experiments," Mechanical Engrg., Vol. 75, 1953.

21. AGARD Report AG-237, "Guide to In-Flight Thrust Measurement of Turbojets and Fan Engines," Chapter 4. MIDAP Study Group, Jan. 1979.

22. ISO 5168, "Measurement of Fluid Flow — Estimation of Uncertainty of a Flow-rate Measurement," 1976.

23. R. B. Abernethy, "An Improved Statistical Method for Measuring Reliability with Special Reference to Rocket Engines," PhD. Thesis, Univ. of London, July 1965.

24. K. A. Brownlee, Statistical Theory and Methodology in Science and Engineering, Second Edition, John Wiley, 1967.

25. B. L. Welch, "The Generalization of Students 's' Problem When Several Different Population Variances are Involved," Biometrika 34 (1947), 28-35.

26. A. T. J. Hayward, "Repeatability and Accuracy," Mechanical Engineering Publication, Ltd., 1977.

27. P. J. Campion, J. E. Burns, and A. Williams, "A Code of Practice for the Detailed Statement of Accuracy," NPL Rept. 1973, Section 5.

28. "Report of the BIPM Working Group on the Statement of Uncertainties," 1980.

29. VIM: "International Vocabulary of Basic and General Terms in Metrology," BIPM-ISO-CEI-OIML, VIM Working Group 1983.

30. R. W. Golden, "Semiscale Uncertainty Report: Methodology," prepared for U. S. Nuclear Regulatory Commission, NUREG/CR-2459, September 1982.

31. J. Berkson, "Are There Two Regressions?," Journal of the American Statistical Association, June 1950.

32. J. Mandel, "Fitting Straight Lines When Both Variables are Subject to Error," J.Quality Technology, Vol. 16, No. 1, January 1984.

33. G. R. Adams, "Improvement of Parameter Estimation Through Uncertainty Analysis," ASME Paper 80-GT-9 Presented at the Gas Turbine Conference and Products Show, New Orleans, La., March 10-13, 1980.

34. F. E. Grubbs, "Procedures for Detecting Outlying Observations in Samples," Technometrics, 11(1), 1-21, February 1969.

35. W. R. Thompson, "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of the Deviation to Sample Standard Deviation," Annals of Mathematical Statistics, Vol. 6, pp. 214-219, 1935.

36. G. L. Tietjen, R. H. Moore, and R. J. Beckman,. "Testing for a Single Outlier in Simple Linear Regression, Technometrics, 15, 717-721, (1973)

37. N. R. Draper, and J. A. John, "Influential Observations and Outliers in Regression," Technometrics 23, 21-26. (1981)

38. R. R. Lund, (1975). "Tables for an Approximate Test for Outliers in Linear Models," Technometrics 17,473-476.

39. A. T. J. Hayward, "The State of the Art of Accuracy Assessment: Total Confusion"' Control and Instrumentation, October 1977.

40. R. B. F. Schumacher, "Systematic Measurement Errors," ASQC Journal of Quality Technology, January 1981.

41. ASQC Writing Group Draft, "American National Standard for Calibration Systems," 1983.

42. R. E. Smith, Jr. and Sam Wehofer, "From Measurement Uncertainty to Measurement Communications, Credibility and Cost Control. . ." ASME paper 1983 WAM.

43. "Aircraft Gas Turbine Engine Monitoring System Guide," SAE, ARP 1587, 1981.

* To Be Published.