AIAA JOURNAL
Vol. 31, No. 10, October 1993

Use of Previous Experience to Estimate Precision Uncertainty of Small Sample Experiments

W. G. Steele,* R. P. Taylor,† and R. E. Burrell‡
Mississippi State University, Mississippi State, Mississippi 39762
and
H. W. Coleman§
University of Alabama in Huntsville, Huntsville, Alabama 35899

The determination of precision uncertainty is described for experiments where the test measurements are obtained from single readings or from a limited number of independent readings. For these cases, the limits of precision uncertainty are best determined from previous experience based on larger sample sizes over the full range of variations that occur for a specific test. This paper describes the appropriate procedures to determine the limits of precision uncertainty based on previous experience for small sample experiments. A Monte Carlo simulation technique is used to model the experiments to demonstrate the effectiveness of using the precision uncertainty to determine the range that covers the true test result. Coverages of 95% or greater are found for the techniques investigated.

Nomenclature

B_r = bias limit
D = pipe diameter
f = friction factor
J = number of test variables
N_i = number present readings
n_i = number previous readings
P_r = precision limit
P_r AVG = average precision limit
$P_{\bar{X}_i}$ = precision limit of the mean
Q = flow rate
R_{95} = 95% limits of results
r = experimental result
r_{true} = true result
S_r = sample standard deviation
s_{x_i} = sample standard deviation for previous readings
$S_{\bar{X}_i}$ = sample standard deviation of the mean
t = 95% value of Student's t distribution
U_r = uncertainty
X_i = present readings
$\bar{X}_i$ = mean of the present readings
x_i = previous readings
$\bar{x}_i$ = mean of the previous readings
β = true bias error
ΔP = pressure drop
Δt = time interval
Δx = length
θ_i = sensitivity, Eq. (6)
μ = population mean
ν_i = degrees of freedom for single variable
ν_r = degrees of freedom for combination, Eq. (8)
ρ = density

Presented as Paper 92-3954 at the AIAA 17th Aerospace Ground Testing Conference, Nashville, TN, July 6–8, 1992; received Aug. 8, 1992; revision received March 16, 1993; accepted for publication March 16, 1993.

*Professor and Head, Mechanical and Nuclear Engineering Department. Member AIAA.
†Professor, Thermal and Fluid Dynamics Laboratory, Mechanical and Nuclear Engineering Department. Member AIAA.
‡Graduate Assistant, Thermal and Fluid Dynamics Laboratory, Mechanical and Nuclear Engineering Department.
§Eminent Scholar in Propulsion and Professor, Propulsion Research Center, Mechanical Engineering Department. Member AIAA.

Introduction

IN large-scale testing the determination of the precision (or random) uncertainty of the test measurements and of the calculated results from the test is not always a straightforward task. In many cases, only a single reading or a limited number of independent readings are available for each measurement. Because of this limitation on the available data from a particular test, the direct determination of the precision uncertainty can be difficult for small samples and impossible for single readings.

To define what is meant by precision uncertainty, consider the definition of the basic terms in uncertainty analysis. The goal of an experiment is to answer a question. If the true result which answers the question is called r_{true}, then our answer would be our best estimate of r_{true}, which we will call r, and the interval around r in which we are 95% confident that r_{true} lies. This relationship is shown in Fig. 1. The possible results of the test are scattered in a distribution (here assumed Gaussian) around the parent population mean of the results μ. This parent population mean is different from r_{true} by an amount called the bias β.

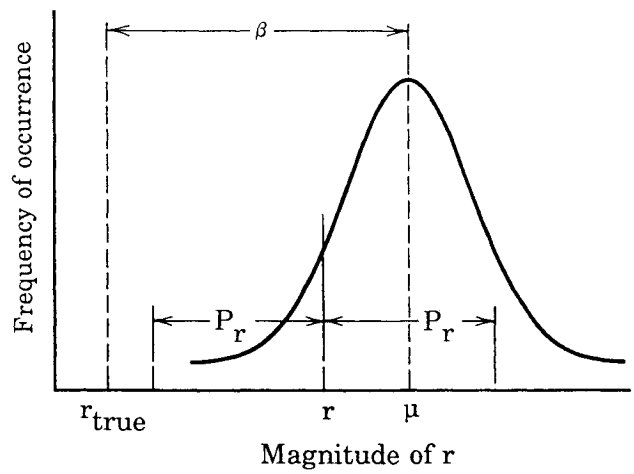

Fig. 1 95% confidence precision limit interval around an experimental result.

The terms in Fig. 1 are defined in the following along with the terms necessary to determine the 95% confidence interval covering the true result.

Precision limit P_r: The $\pm P_r$ interval about a result r (single or averaged) is the experimenter's 95% confidence estimate of the band within which the mean μ of many such results would fall if the experiment were repeated many times under the same conditions and using the same equipment. The precision limit is thus an estimate of the scatter or lack of repeatability caused by random errors and unsteadiness.

Bias limit B_r: The bias limit is an estimate of the magnitude of the fixed, constant errors. When the true bias error in a result is defined as β, the quantity B_r is the experimenter's 95% confidence estimate such that $|\beta| \leq B_r$.

Uncertainty U_r: The $\pm U_r$ interval about the result is the band within which the experimenter is 95% confident the true value of the result lies. The 95% confidence uncertainty is calculated as[1]

$$U_r = [B_r^2 + P_r^2]^{1/2} \tag{1}$$

In this study, we are only considering the precision uncertainty. In the vast majority of practical situations, bias uncertainty is uncorrelated with respect to the precision uncertainty. In that case, the precision can be uncoupled from the bias and considered independently. The existence of bias error in actual experiments would not change the conclusions of this study.

The determination of the precision limits of experimental measurements and of test results determined from these measurements will be examined using the techniques given in Refs. 1 and 2. Before we discuss these techniques, let us consider some typical test measurement situations.

Consider an experiment where some of the test variables have time scales such as that shown in Fig. 2. If the question in the experiment is "what is the result for time interval Δt?" then measurement readings taken over that time frame can be averaged, and the statistical techniques presented in Refs. 1 and 2 can be used to determine directly the precision limit of the result.

A more typical engineering situation is where the test variables X_i are held at nominally steady conditions with the actual variation of X_i with time similar to that shown in the entire span of Fig. 2. In this case, we typically desire the result to be indicative of the value over the time frame which includes all of the important time scales. However, it is often not practical to take readings over that entire time frame, as some of the variations may have time scales of hours or even days.

In a wind-tunnel test, for example, measurements are taken over a short period of time with the full understanding that the time scale of some of the variations is much longer than the test measurement time scale. In such a case, a value of X_i determined over some relatively short time period Δt should be considered as a single reading even if X_i is the average of 1000 readings or more during Δt. If the test measurements were repeated during the same steady test conditions over a different Δt time, different values would be obtained. This apparent variation in the test measurement values with time is representative of the precision uncertainty which should be assigned to these single readings.

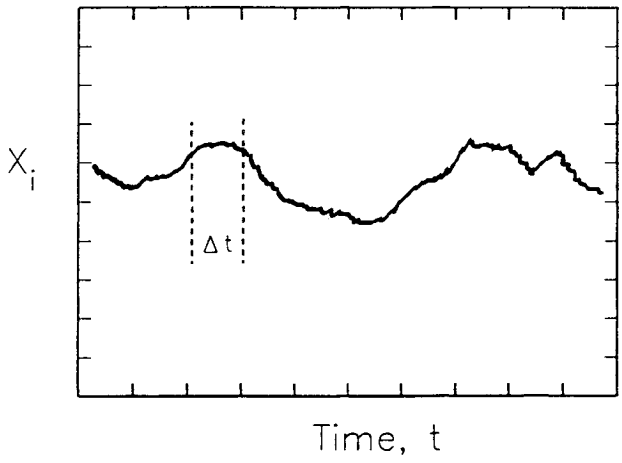

Fig. 2 Variation of variable X_i with time for a "constant" experiment condition.

It should be noted that depending on the defined test, the precision uncertainty could also include error sources other than random time variations, such as repeating test conditions, model positions, configuration variables, etc. For a fixed test configuration and a single operation of the facility, these errors would be fixed and therefore would be considered bias errors. However, if the test program of interest dealt with the operation of the test assembly multiple times and with the disassembly and reassembly of the test apparatus multiple times, then these errors would be part of the precision uncertainty.

In many test situations, the measurement for each of the test variables will essentially be a single reading because of the ideas just discussed. For some cases, we might be able to repeat the measurements during the test program to get a small number of repeated independent values for the test variables. With this information, we could calculate the precision limit of the test result. This situation will be discussed in the following section. It will be shown that when the precision limit is determined based only on these small data sets, considerable variations in P_r will occur from test to test.

A more appropriate method of determining the precision limit in these small data set cases (and especially for single data point cases) would be to use previous information about the variation of the readings over the full range of variations that could occur for a specific test. This preferred method is also discussed in the sections that follow.

Direct Determination of Precision Limit

Consider the case where the test result is determined once and the test variables X_i are measured N_i times where each N_i could be different. The result is then determined using the averages $\bar{X}_i$ as

$$r = r(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_J) \tag{2}$$

where

$$\bar{X}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} (X_i)_k \tag{3}$$

Note that we are assuming here that the N_i readings are taken over the appropriate time frame for variations in X_i.

According to Ref. 1 (standard method), the precision limit of the result is calculated as

$$P_r = tS_r \tag{4}$$

Here S_r is determined from the propagation equation as

$$S_r = \left[\sum_{i=1}^{J} (\theta_i S_{\bar{X}_i})^2 \right]^{1/2} \tag{5}$$

where

$$\theta_i = \frac{\partial r}{\partial \bar{X}_i} \tag{6}$$

and $S_{\bar{X}_i}$ is the sample standard deviation of the average value of the variable X_i for N_i repeated independent readings defined as

$$S_{\bar{X}_i} = \frac{1}{\sqrt{N_i}} \left[\frac{1}{N_i - 1} \sum_{k=1}^{N_i} [(X_i)_k - \bar{X}_i]^2 \right]^{1/2} \tag{7}$$

The value of t is the 95% confidence level value of the Student's t distribution for ν_r degrees of freedom where ν_r is estimated using

the Welch-Satterthwaite formula[1] as

$$\nu_r = \frac{\left[\sum_{i=1}^{J} (\theta_i S_{\bar{X}_i})^2\right]^2}{\sum_{i=1}^{J} [(\theta_i S_{\bar{X}_i})^4/\nu_i]} = \frac{S_r^4}{\sum_{i=1}^{J} [(\theta_i S_{\bar{X}_i})^4/\nu_i]} \quad (8)$$

with

$$\nu_i = N_i - 1 \quad (9)$$

An alternate method for determining P_r by directly propagating precision limits was proposed by Kline and McClintock[3] and is derived and illustrated by Coleman and Steele.[2] For this case P_r is determined from the propagation expression as

$$P_r = \left[\sum_{i=1}^{J} (\theta_i P_{\bar{X}_i})^2\right]^{1/2} \quad (10)$$

where $P_{\bar{X}_i}$ is the precision limit of the mean for the variable X_i defined as

$$P_{\bar{X}_i} = t_i S_{\bar{X}_i} \quad (11)$$

The t_i is the 95% confidence level value of the Student's t distribution for $N_i - 1$ degrees of freedom.

A comparison has been made of the coverage provided by the precision limits obtained from both the standard[1] method and the method in Ref. 2 (alternate method) for some common experiments. If the number of readings for each test variable is large ($N_i > 30$), then the two techniques will yield the same value for the precision limit. Also, Ref. 2 recommends that the standard[1] technique be used for determining P_r for single tests with averaged readings for the test variables when some of the measurements have sample sizes $N_i \leq 30$. However, the comparison of the two techniques for small sample experiments allows for a better understanding of the differences between them.

For this comparison, a Monte Carlo simulation of each experiment was performed using a Gaussian random number generator. As an example, consider an experiment where a friction factor is determined from measurements of pipe diameter D, flow rate Q, test section length ΔX, pressure drop ΔP, and fluid density ρ as

$$f = \frac{\pi^2 D^5 \Delta P}{32 \rho Q^2 \Delta X} \quad (12)$$

First, hypothetical true values for the test variables were assigned along with standard deviations for the respective Gaussian distributions. The values used in this example are given in Table 1. For this simulation the bias error was assumed to be equal to zero; therefore, the true values of the test variables would correspond to the parent population means μ_i for the Gaussian distributions. The hypothetical true result for the friction factor was then calculated from Eq. (12) using these μ_i. By cycling through the random number generator N_i times for each test variable, the necessary information was obtained to determine P_r from both Eqs. (4) and (10).

Table 1 Hypothetical true values for the test variables and standard deviations for the respective Gaussian distributions for the friction factor example

Variable	True value	Standard deviation
Diameter D, m	0.05	1.875×10^{-4}
Flow rate Q, m^3/s	2.77×10^{-3}	1.700×10^{-5}
Test length ΔX, m	0.2	0
Density ρ, kg/m^3	1000	7.5
Pressure drop ΔP, Pa	80.06	0.6

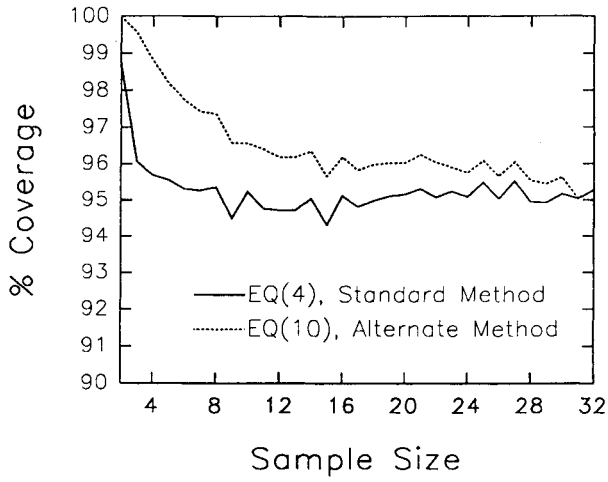

Fig. 3 Coverages for single, multisample experiments with averaged readings.

The experiments were run for sample sizes N_i ranging from 2 to 32. In each experiment, N_i was the same for each of the test variables. For each sample size, the numerical experiment was repeated 10,000 times. With $N_i = 2$, for example, 10,000 separate values of the P_r were determined along with 10,000 values of the result f. In each of the 10,000 experiments, a check was made for both precision limits to determine if the band $f \pm P_r$ contained the true value of f. If the band contained f_{true}, a counter was increased for that particular method of calculating P_r. This technique allowed the determination of a coverage fraction for each method for each sample size.

The coverages found for this example are shown in Fig. 3 where one curve is for the method in Ref. 2 [Eq. (10)] and the other curve is for the standard[1] method [Eq. (4)]. The standard method approaches 95% coverage after a sample size of about 7 whereas the alternate technique yields higher coverages. The alternate technique approaches a coverage of 96% after a sample size of 10.

Figures 4 and 5 show histograms of the precision limits determined for the 10,000 separate experiments for sample sizes of four ($N_i = 4$) using the standard method and the alternate method, respectively. These figures show that the distribution of these small sample precision limits is relatively broad. The standard method in Fig. 4 gives an average precision limit of 15×10^{-5} with values ranging from about 4×10^{-5} to 32×10^{-5}. The alternate method gives a similar distribution with an average precision limit of 19.2×10^{-5} and values ranging from about 7×10^{-5} to 36×10^{-5}. The alternate method will typically yield higher estimates of P_r. It should be recalled that each precision limit that is represented in these figures is the proper estimate for that particular four-sample experiment.

Also shown in Figs. 4 and 5 is the number which corresponds to the limits $\pm R_{95}$ on the interval that enclosed 95% of the 10,000 friction factors determined for this four-sample case. Ideally, this limit would represent the minimum value that could be added to or subtracted from any single value of f, as shown in Fig. 1, to cover the true, unbiased result μ, 95% of the time. As can be seen, both of the methods for calculating the precision limit for this small sample size show significant variations around this minimum R_{95} value.

As shown in Fig. 3 for the four-sample case, these 10,000 separately calculated precision limits when combined with their respective results f yield a coverage of about 96% for the standard method and 99% for the alternate method; however, the use of any one of these individual precision limits to estimate the appropriate value of P_r for another experiment with a sample size of four could lead to either a significant overprediction or an underprediction of the range that contains the true value. One solution to this problem

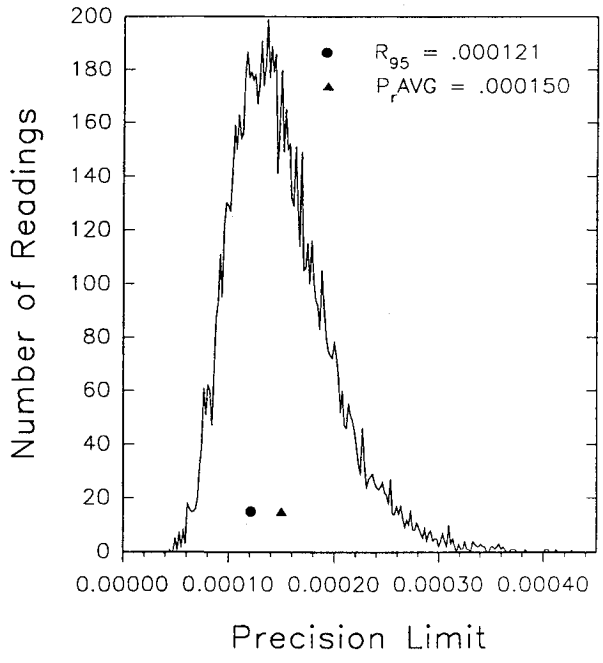

Fig. 4 Histogram of precision limits for four-sample experiments based on present information using the standard method,[1] Eq. (4).

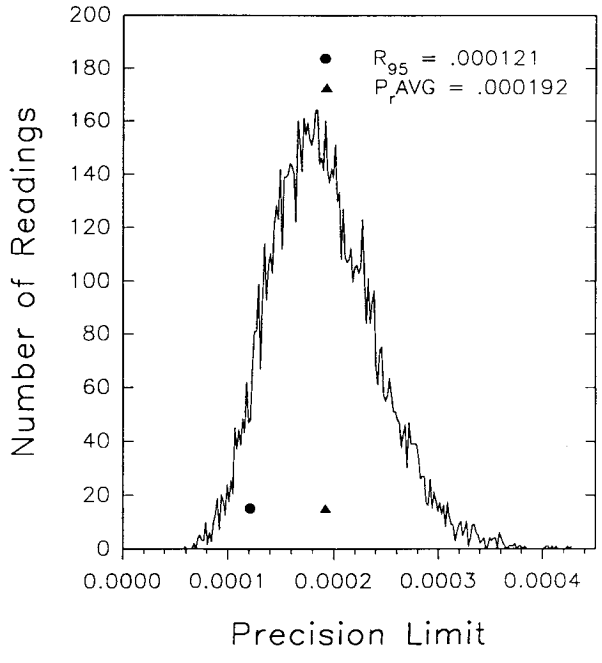

Fig. 5 Histogram of precision limits for four-sample experiments based on present information using the alternate method,[2] Eq. (10).

would be to increase the sample size. However, as discussed before, this is not practical for many cases. In the next section, we discuss the appropriate use of data from previous experience to determine better estimates of the precision limits for small sample experiments.

Determination of Precision Limit Using Previous Information

Consider again a test result that is determined once from J test variables where each variable X_i is measured N_i times. When n_i previous readings are known for each variable, the sample standard deviations for the variables can be calculated from these previous readings as

$$s_{x_i} = \left[\frac{1}{n_i - 1} \sum_{k=1}^{n_i} [(x_i)_k - \bar{x}_i]^2 \right] \tag{13}$$

where here

$$\bar{x}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} (x_i)_k \tag{14}$$

The appropriate sample standard deviation of the mean for N_i current readings of each variable is then

$$S_{\bar{X}_i} = s_{x_i} / \sqrt{N_i} \tag{15}$$

where N_i is the number of present readings averaged together to get $\bar{X}_i$, and n_i is the number of previous readings used to compute s_{x_i}. Here we are assuming that n_i is greater than N_i and is therefore a better data set for determining the precision limits for $\bar{X}_i$.

Using the standard method,[1] the standard deviation for the result is calculated from the propagation equation as

$$S_r = \left[\sum_{i=1}^{J} (\theta_i S_{\bar{X}_i})^2 \right]^{1/2} \tag{16}$$

The number of degrees of freedom for the result is determined as

$$\nu_r = \frac{\left[\sum_{i=1}^{J} (\theta_i S_{\bar{X}_i})^2 \right]^2}{\sum_{i=1}^{J} [(\theta_i S_{\bar{X}_i})^4 / \nu_i]} = \frac{S_r^4}{\sum_{i=1}^{J} [(\theta_i S_{\bar{X}_i})^4 / \nu_i]} \tag{17}$$

where

$$\nu_i = n_i - 1 \tag{18}$$

The derivatives in Eqs. (16) and (17) are evaluated at the current $\bar{X}_i$ values. Using the ν_r value from Eq. (17) to determine the t value from the Student's t distribution, the precision limit for the single result is calculated as

$$P_r = tS_r \tag{19}$$

The alternate approach[2] for determining the precision limit would be to calculate the precision limit for each reading as

$$P_{\bar{X}_i} = t_i S_{\bar{X}_i} \tag{20}$$

where $S_{\bar{X}_i}$ is from Eq. (15) and where t_i is the experimenter's t distribution value for $n_i - 1$ degrees of freedom. Then the precision limit of the result is calculated from

$$P_r = \left[\sum_{i=1}^{J} (\theta_i P_{\bar{X}_i})^2 \right]^{1/2} \tag{21}$$

The derivatives in Eq. (21) are evaluated at the current $\bar{X}_i$ values.

A Monte Carlo simulation was also performed for these two methods of determining the precision limits from previous information. For this case, n_i previous readings were made for each variable using the random number generator. Then N_i current readings were made for the variables, and the precision limits in Eqs. (19) and (21) were calculated. For a fixed number of previous readings and a fixed number of current readings the process was repeated 10,000 times. In each of the 10,000 tests, a check was made to determine if the result plus or minus the precision limit covered the true result. This technique allowed for the determination of a coverage fraction for each method for each combination of current and previous sample sizes.

Table 2 Comparison of precision coverages for the friction factor experiment when previous data are used

Current readings, N	Previous readings, n	Coverage for standard method, Eq. (19), %	Coverage for alternate method, Eq. (21), %
4	10	95.6	96.9
4	20	95.1	96.1
4	31	95.4	95.4

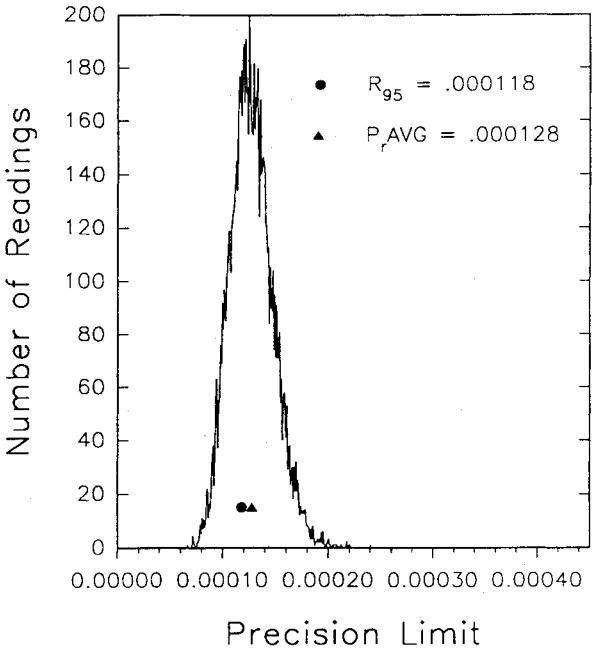

Fig. 6 Histogram of precision limits for four-sample experiments based on 10-sample previous information using the standard method,[1] Eq. (19).

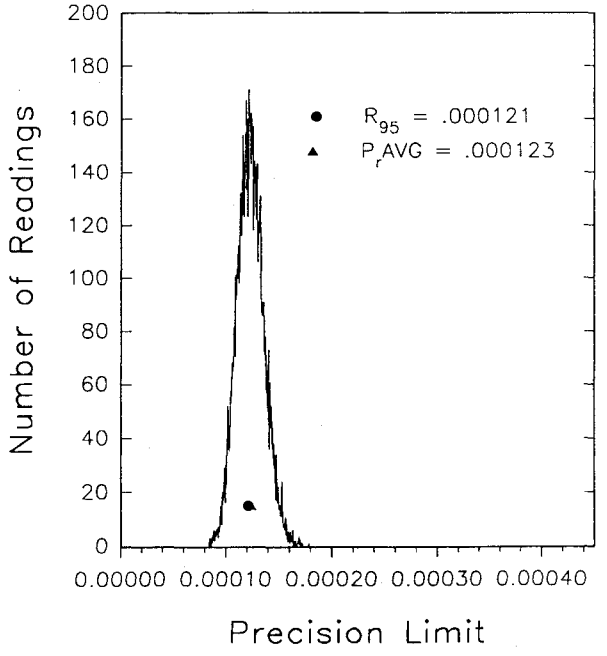

Fig. 7 Histogram of precision limits for four-sample experiments based on 20-sample previous information using the standard method,[1] Eq. (19).

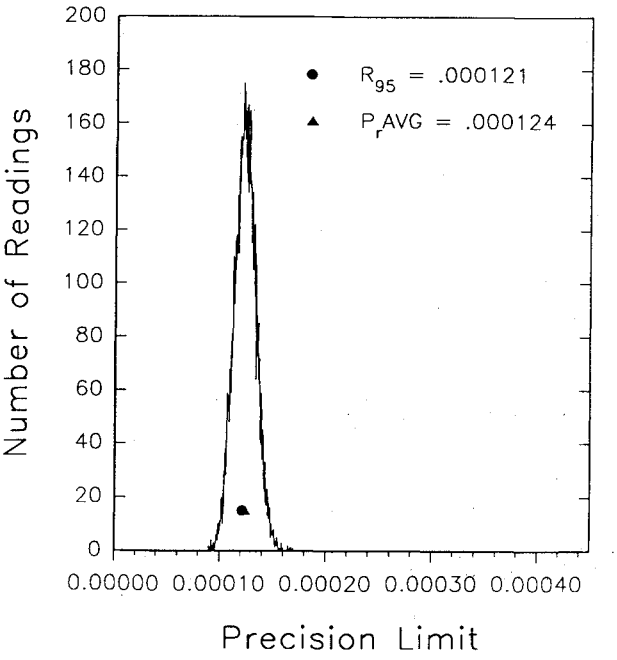

Fig. 8 Histogram of precision limits for four-sample experiments based on 31-sample previous information using the standard method,[1] Eq. (19).

Table 2 shows the results of these simulations for the pipe friction factor example of Eq. (12). The number of readings used to evaluate D, Q, ΔX, ρ, and ΔP is $N_i = 4$, and the number of previous readings used to compute the sample standard deviations was varied as n_i = 10, 20, and 31. The table shows that the standard method, Eq. (19), gives approximately 95% coverage for all of the cases whereas the alternate approach, Eq. (21), gives slightly higher coverages for all values of $n_i < 31$.

As seen earlier in Figs. 3–5, the precision limits determined directly from small samples were on average significantly different from the 95% limits for the friction factor population. Figures 6–8 show the histograms of the precision limits calculated from previous information as discussed in the preceding paragraphs.

The results obtained when the standard method is used to compute the precision limit for the mean of four readings are shown in Figs. 6–8. It is seen that as the sample size of the previous data increases, the average precision limit approaches the band that encloses 95% of the population of results R_{95}. Also, the dispersion of the precision limits decreases. As noted earlier in Fig. 4, when the four-sample readings were used alone to compute the precision limits, these limits ranged from about 4×10^{-5} to about 32×10^{-5} with an average precision limit of 15×10^{-5} compared to an R_{95} value of 12.1×10^{-5}. In Fig. 6, when 10-sample previous sets were used to compute the precision limits of the four-sample means, the values ranged from 7×10^{-5} to 20×10^{-5} with an average precision limit of 12.8×10^{-5} compared with an R_{95} value of 11.8×10^{-5}. For previous sample sizes of 20 and 31, the average precision limits and R_{95} values almost exactly coincide, and the dispersion becomes smaller and smaller. Note, however, that a sample size of 31 still gives some dispersion. The values of R_{95} vary slightly in the figures because 10,000 different experiments were considered for each case.

Similar histogram comparisons were made for the results using the alternate scheme of Eq. (21). For previous sample sizes of 10 and 20, the histograms were found to be shifted to the right with the average precision limits being somewhat larger than those computed with the standard scheme using Eq. (19) (0.000139 vs 0.000128 for 10 previous readings and 0.000129 vs 0.000123 for 20 previous readings). The two schemes coincide for previous sample sizes of 31 and larger.

Typically, in the early phases of an experimental program such as the debugging phase, large data sets are taken to qualify the

instrumentation and the test procedures. These large data sets can provide insight into the appropriate precision information to be used later in the actual execution phase of the experiment. However, as noted earlier, even when a previous sample size of 31 readings is used to determine the precision limit, a variation in the precision limits of about ±12% around the R_{95} value is observed. The possibility exists that the precision information determined from one set of 31 previous readings would be on the low side and would not provide 95% coverage of the truth when used with future readings to find a precision limit.

The key point here is that multiple data sets should be taken during the debugging phase of an experiment, and the sample standard deviations from Eq. (13) for these data sets should be compared. After the experiment is fully debugged, these standard deviations can then be used to determine the appropriate precision information for the calculation of the precision limits of the results determined later from smaller samples.

If large data sets are used for the previous readings for all of the test measurements ($n_i > 30$), then both the standard method and the alternate method yield the same value for the precision limit of the result. In this case it is convenient and informative to go ahead and calculate the precision limit for each variable using Eq. (20). These precision limits provide an approximation of the scatter expected in the measurement of each variable. If during the execution phase of the experiment a scatter in the readings of a variable is observed which is greater than this previously determined precision limit, then a check should be made to determine if something has changed or if testing is being done in a range that was not anticipated.

Summary

The determination of the appropriate precision uncertainty for small sample experiments is not always a straightforward task. When precision limits are determined directly from current readings, care must be taken to ensure that the readings are taken over the full range of variations that occur during the test. In some cases, this direct determination of precision limits can lead to a very lengthy test program. In other tests, it may not be possible or desirable to take readings over the entire range of variations, and instead only single readings or small samples of readings are taken. In these cases it is necessary to use previous information to estimate the appropriate precision limits for the test variables.

References

[1]American National Standards Institute/American Society of Mechanical Engineers, *Measurement Uncertainty*, PTC19.1-1985, American Society of Mechanical Engineers, New York, 1986, Secs. 2–3.

[2]Coleman, H. W., and Steele, W. G., *Experimentation and Uncertainty Analysis for Engineers*, Wiley, New York, 1989, Chaps. 2,4.

[3]Kline, S. J., and McClintock, F. A., "Describing Uncertainties in Single-Sample Experiments," *Mechanical Engineering*, Vol. 75, Jan. 1953, pp. 3–8.