
Describing the Uncertainties in Experimental Results

Robert J. Moffat

*Professor of Mechanical Engineering,
Stanford University, Stanford, California*

■ It is no longer acceptable, in most circles, to present experimental results without describing the uncertainties involved. Besides its obvious role in publishing, uncertainty analysis provides the experimenter a rational way of evaluating the significance of the scatter on repeated trials. This can be a powerful tool in locating the source of trouble in a misbehaving experiment. To the user of the data, a statement (by the experimenter) of the range within which the results of the present experiment might have fallen by chance alone is of great help in deciding whether the present data agree with past results or differ from them. These benefits can be realized only if both the experimenter and the reader understand what an uncertainty analysis is, what it can do (and cannot do), and how to interpret its results.

This paper begins with a general description of the sources of errors in engineering measurements and the relationship between error and uncertainty. Then the path of an uncertainty analysis is traced from its first step, identifying the intended true value of a measurement, through the quantitative estimation of the individual errors, to the end objective—the interpretation and reporting of the results. The basic mathematics of both single-sample and multiple-sample analysis are presented, as well as a technique for numerically executing uncertainty analyses when computerized data interpretation is involved.

The material presented in this paper covers the method of describing the uncertainties in an engineering experiment and the necessary background material.

Keywords: *experimental uncertainty, error analysis, single-sample analysis, multiple-sample analysis, system errors*

INTRODUCTION

The error in a measurement is usually defined as the difference between its true value and the measured value. This definition is clear but not very helpful: the only real situations in which we even claim to know both the true value and the measured value are those in which we are calibrating or “qualifying” an experiment against baseline data or against one of the basic conservation laws of engineering. In most situations, we cannot talk very confidently about what the error in a measurement *is*, we can only talk about what it *might be*—about the limits that we feel bound the possible error.

The term “uncertainty” is used to refer to “a possible value that an error may have.” Kline and McClintock [1] attribute this definition to Airy [2], and it still seems an appropriate and valuable concept. The terms “uncertainty interval” and “uncertainty” are commonly used interchangeably, and they will be so used in this discussion, both referring to the interval around the measured value within which the true value is believed to lie. The term “uncertainty analysis” refers to the process of estimating how great an effect the uncertainties in the individual measurements have on the calculated result.

There is more to uncertainty analysis than just drawing error

bars on a plot of the data or calculating the root-mean-square deviation of data from a curve fit, or (worse yet) calculating the mean absolute value of that deviation. Those are simply techniques for providing some measure of the scatter in the results and provide no way of judging whether or not the observed scatter was “reasonable”; that is where uncertainty analysis comes in.

Uncertainty analysis began as the statistical interpretation of the errors in well-replicated experimental results. It became apparent during the early 1950s that many important engineering experiments could not be repeated enough times to provide useful statistical information, for reasons of economy or pressure of time. A rational way to use the framework of statistical inference to estimate the uncertainty in these single-sample experiments was described by Kline and McClintock [1] and still forms the basis for this branch of the art. Over the years, single-sample uncertainty analysis has been used more in research experiments than in production, and that association has led to its evolution as a diagnostic tool for the development of experiments. As a consequence, single-sample uncertainty analysis presents two uncertainty measures that are particularly useful during the planning and debugging stages of experiments, in addition to the usual “overall uncertainty.”

The distinction between single-sample and multiple sample

Address correspondence to Professor Robert J. Moffat, Mechanical Engineering Department, Stanford University, Stanford, CA 94305.

analysis hinges on whether or not a "large" or a "small" number of independent data points are taken at each test point and on how the data are handled. In this era of high-speed digital data acquisition, the issue of independence takes on more subtle overtones than it had in the early 1950s when the term was coined. Consider, for example, a measuring system capable of acquiring data at 100 kHz, a system readily available under current technology. Applied to an experiment whose output varied at approximately 10 mHz, this 100 kHz sample rate would produce about 100 readings in a 1 ms period. These can each be regarded as independent measures of the process, assuming the time between consecutive readings to be large compared with the autocorrelation time of the signal. This set of 100 readings would be, then, a multiple-sample set of observations. The same equipment, applied to a system whose output varied at 1 Hz, would produce a single-sample measure of the process, tainted, perhaps, by a multiple-sample contribution from the high-frequency random errors present in the measuring system.

Classification of a given experiment as single-sample or multiple-sample must take into account not only the number of observations made at each test point but also the data-sampling rate and the spectrum of significant frequencies in the process being studied and whether the data are averaged before processing or processed before averaging.

Single-sample uncertainty analysis has been described in the engineering literature by the works of Kline and McClintock [1] and Moffat [3, 4]. The techniques of multiple-sample analysis are described by Abernethy and Thompson [5] and summarized by Abernethy et al [6] and by ANSI/ASME PTC 19.1-1985 [7].

Both have the same final objective (to estimate the effect of the accumulated measurement uncertainties on the accuracy of the result), even though somewhat different procedures are required. Either the N th-order uncertainty interval of single-sample analysis or the $U_{0.95}$ interval of multiple-sample analysis answers the question: How close to this result does the true value probably lie, assuming that these are the right equations to use, that all the important variables have been included, and that the test situation is representative?

In addition to this final result, each method produces auxiliary information about the experiment, mainly useful as diagnostics, during the developmental phase of an experiment or in monitoring its "health" during a long series of runs. Single-sample uncertainty analysis generates two diagnostics: the zeroth-order uncertainty, which evaluates the contribution to the total uncertainty introduced by the measuring system, and the first-order uncertainty, which predicts the scatter that should be observed on repeated trials with the same equipment and the same instruments. Multiple-sample uncertainty analysis produces estimates of the total fixed error and the total random error in the result.

An uncertainty estimate of either type is only as good as the equation(s) it is based on. If those equations are incomplete and do not acknowledge all the significant factors that affect the result, or if falsely low values are used for the component uncertainties, then the analysis will underestimate the uncertainty in the result. On the other hand, if the component uncertainties are exaggerated, then the analysis will overestimate the uncertainty.

The most common, and most visible, use of uncertainty analysis is in reporting results to the technical community through publications, but it must be noted here that it has far broader uses. During the shakedown period of an experiment, it is a powerful diagnostic tool in seeking out the sources of residual error. In the early stages of an experiment, for example, when comparing the first results from a new test rig with those from an existing baseline set, the most frequent question is: Does the difference I

see mean that the new results are really different, or is this difference just consequence of the uncertainties in my measurements? Uncertainty analysis provides clear, unambiguous guidance: If the observed difference exceeds zero by more than the expected uncertainty interval for the difference, then the observed difference is probably significant. Even earlier in an experiment, uncertainty analysis can be used to help choose the most reliable technique for a given measurement or to identify the critically important instruments in a system (and thereby determine where expensive instruments are needed!).

THE BASIC MATHEMATICS

This section introduces the root-sum-square (RSS) combination, the basic form used for combining uncertainty contributions in both single-sample and multiple-sample analyses. In this section, the term δX_i refers to the uncertainty in X_i in a general and nonspecific way: whatever is being dealt with at the moment (for example, fixed errors, random errors, or uncertainties).

Describing One Variable

Consider a variable X_i , which has a known uncertainty δX_i . The form for representing this variable and its uncertainty is

$$X_i = X_i(\text{measured}) \pm \delta X_i \quad (20:1)$$

This statement should be interpreted to mean the following:

- The best estimate of X_i is X_i (measured)
- There is an uncertainty in X_i that may be as large as $\pm \delta X_i$.
- The odds are 20 to 1 against the uncertainty of X_i being larger than $\pm \delta X_i$.

The value of X_i (measured) represents the observation in a single-sample experiment or the mean of a set of N observations in a multiple-sample experiment.

The value of δX_i represents 2σ for a single-sample analysis, where σ is the standard deviation of the population of possible measurements from which the single sample X_i was taken. For multiple-sample experiments, δX_i can have three meanings. It may represent $tS_{(N)}/\sqrt{N}$ for random error components, where $S_{(N)}$ is the standard deviation of the set of N observations used to calculate the mean value \bar{X}_i and t is the Student's t statistic appropriate for the number of samples N and the confidence level desired. It may represent the bias limit for fixed errors (this interpretation implicitly requires that the bias limit be estimated at 20:1 odds). Finally, δX_i may represent $U_{0.95}$, the overall uncertainty in X_i .

The Student's t multiplier is a number, always larger than 2.0, that allows one to use $S_{(N)}$ rather than σ in estimating the uncertainty in the mean of a set. The Student's t statistic is tabulated in most statistical reference books under that name.

The result R of the experiment is assumed to be calculated from a set of measurements using a data interpretation program (by hand or by computer) represented by

$$R = R(X_1, X_2, X_3, \dots, X_N) \quad (2)$$

The objective is to express the uncertainty in the calculated result at the same odds as were used in estimating the uncertainties in the measurements. This issue was taken up by Kline and McClintock [1], who showed that the uncertainty in a computed result could be estimated with good accuracy using a root-sum-square combination of the effects of each of the individual inputs and that the RSS operation preserved the odds.

The effect of the uncertainty in a single measurement on the

calculated result, if only that one measurement were in error would be

$$\delta R_{X_i} = \frac{\partial R}{\partial X_i} \delta X_i \quad (3)$$

The partial derivative of R with respect to X_i is the *sensitivity coefficient* for the result R with respect to the measurement X_i .

When several independent variables are used in the function R , the individual terms are combined by a root-sum-square method.

$$\delta R = \left\{ \sum_{i=1}^N \left(\frac{\partial R}{\partial X_i} \delta X_i \right)^2 \right\}^{1/2} \quad (4)$$

This is the basic equation of uncertainty analysis. Each term represents the contribution made by the uncertainty in one variable, δX_i , to the overall uncertainty in the result, δR . Each term has the same form: the partial derivative of R with respect to X_i multiplied by the uncertainty interval for that variable. The estimated uncertainty in the result has the same probability of encompassing the true value of the result as the uncertainties in the individual measurements have of encompassing their true values.

Equation (4) applies as long as [1]

1. Each of the measurements was independent
2. Repeated observations of each measurement, if made, would display Gaussian distributions
3. The uncertainty in each measurement was initially expressed at the same odds

The data interpretation program may be simple enough that all the partial derivatives can be evaluated analytically, or it may be too complex for that and require direct computer analysis; the procedures are the same in either case. A general technique for computerized uncertainty analysis is given in the next section.

The following comments might be helpful in "hand" analysis; they are mainly aimed at simplifying the task.

In most situations, the overall uncertainty in a given result is dominated by only a few of its terms. Terms in the uncertainty equation that are smaller than the largest term by a factor of 3 or more can usually be ignored. This is a natural consequence of the RSS combination: Small terms have very small effects. There are exceptions, of course, where there are many terms of approximately the same size, but in general that is not the case.

In many applications, the uncertainty estimate is wanted as a fraction of reading, rather than in engineering units. While this can always be calculated, starting from the results of the general form in Eq. (4), it is sometimes possible to do the calculation of relative uncertainty directly. In particular, whenever the equation describing the result is a pure "product form," such as Eq. (5), or can be put into that form, then the relative uncertainty can be found directly. That is, if

$$R = X_1^a X_2^b X_3^c \cdots X_M^m \quad (5)$$

then

$$\frac{\delta R}{R} = \left\{ \left(a \frac{\delta X_1}{X_1} \right)^2 + \left(b \frac{\delta X_2}{X_2} \right)^2 + \cdots + \left(m \frac{\delta X_m}{X_m} \right)^2 \right\}^{1/2} \quad (6)$$

This is a natural and convenient approach in situations where the uncertainties of the component measurements are described in terms of percent of reading and the result is needed in the same terms. The exponent of X_i becomes its sensitivity coefficient.

COMPUTERIZED UNCERTAINTY ANALYSIS

When R is calculated using a large-scale computer program or involves operations that are difficult to differentiate (e.g., table

look-ups or numerical integrations), the operations represented by Eqs. (3) and (4) either cannot or will not be done by hand. In most cases, it is not practical to write a separate computer program for the evaluation of uncertainties, both from the standpoint of complexity and because of the difficulty of ensuring that the uncertainty code is updated each time the main data interpretation code is revised. For these more complex experiments, the data interpretation program itself can be used to generate the uncertainty analysis, by sequentially perturbing the input values and accumulating the individual uncertainty contributions.

This direct computer-executed uncertainty analysis can be accomplished by sequentially perturbing the inputs according to the following procedure [4]:

1. Calculate the result R for the recorded data. Identify the value as R_0 and store it.
2. For $i = 1$ to N , where N is the number of variables in R : Increase the value of the i th variable, X_i , by its uncertainty interval, δX_i , and calculate the result, R_{i+} , using the augmented value of the i th variable with all other variables at their recorded (nominal) values. Find the difference $R_{i+} - R_0$ and store it as C_{i+} , the contribution to the uncertainty of R caused by the i th variable, assuming a positive excursion.
3. If the result R is likely to be a strongly nonlinear function of X_i , including consideration of the size of its uncertainty interval, then also calculate C_{i-} using $R_{i-} - R_0$. The present recommendation would be to use the average of the absolute values of C_{i+} and C_{i-} as the working value of C_i , but that recommendation must be regarded as tentative, since no definitive analysis has been done to investigate the issue. Next i .
3. The uncertainty in the result is the root-sum-square of the C_i .

A primary advantage of this method is that the actual working data interpretation program itself is used in the assessment of uncertainties. Thus, each time the program is modified, the modifications are automatically incorporated into the uncertainty calculations. The process of determining the uncertainty in the computed heat transfer coefficient from its basic measurements is illustrated in Fig. 1.

The method of sequential perturbations can also be applied to data interpretation programs of intermediate difficulty (too large to do by hand but not large enough to warrant a dedicated program) using spreadsheets on a personal computer, as pointed out by Catz [8].

THE SOURCES OF UNCERTAINTIES

The uncertainty attributed to a measurement is an estimate of the possible residual error in that measurement after all proposed corrections have been made. Although this discussion of the sources of uncertainties will begin with a discussion of errors, the distinction between the original error and the residual error must be clearly maintained. The uncertainty is determined by the residual error after correction, not the original error.

An error source is usually categorized as "fixed" or "random" depending on whether the error it introduces is steady or changes during the time of one complete experiment. The errors themselves are called *bias* or *precision errors*, and the precision error is presumed to behave randomly, with a zero mean. Both the bias and precision are presumed to represent stationary statistical properties of a Gaussian distributed data set.

Such a description tacitly assumes that each observation of the error is independent of its preceding observation—that the error source in question has no systematic variation. This is true only

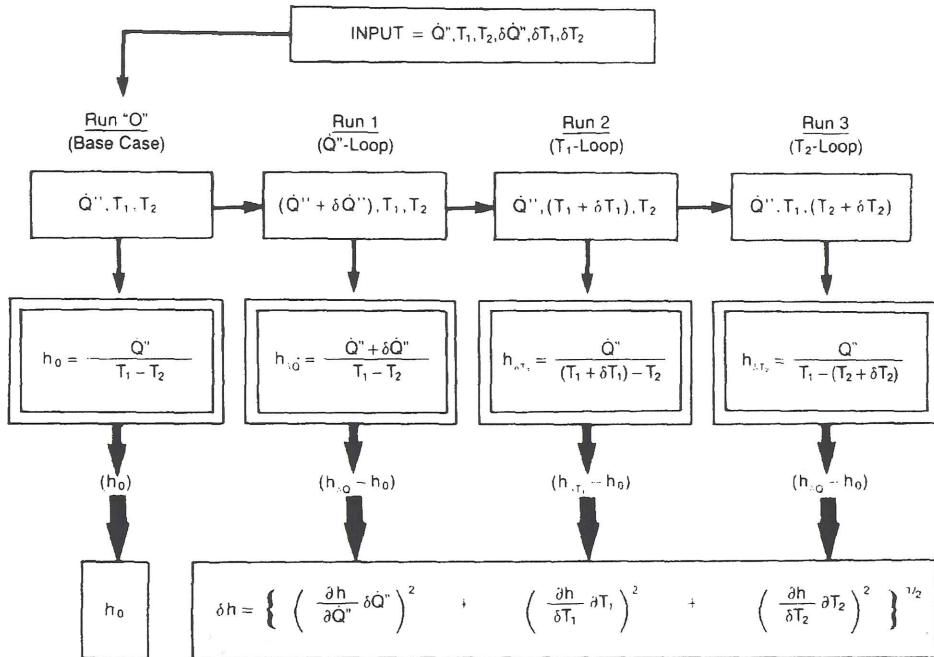


Figure 1. The method of sequential perturbation for calculating uncertainty intervals from the data reduction program.

when the interval between observations is longer than the autocorrelation time of the error source.

For most engineering experiments, a third category of error is necessary: a category for errors that change during an experiment, but not randomly. These will be described here as “variable but deterministic.” For example, the radiation error of a gas temperature sensor depends on the local wall temperature—if that changes during the test (for example, due to a change in test-cell ventilation rate), then the error will change. Such a change is by no means random.

In the present work, error sources will be classified as “fixed,” “random,” or “variable but deterministic.” The classification depends not only on how the source behaves with time, but also on the sampling frequency of the observations; any process will appear random if sampled slowly enough! The general term “variable error” will be used to include both the random and deterministic components of unsteady error. The residual error is always of the same type as the original error, so we expect to be dealing with fixed, random, and variable-but-deterministic uncertainty components.

The error in a measurement is defined as the difference between the observed value and the true value of the intended measurand. The “observed value” is easy to identify—it is the value returned by the measuring system after the appropriate corrections. It exists. The “true value,” however, is a not so easy to identify. To illustrate the options, consider a thermocouple probe installed downstream of a gas turbine combustor, in a duct with cold walls. There are (at least) four options that could be chosen for the definition of the “true” temperature. The classification and nomenclature follows that developed by Moffat [9]:

1. $T(1)$: the temperature of the thermocouple junction (the achieved value)
2. $T(2)$: the temperature of the gas at the junction location (the available value)
3. $T(3)$: the temperature the gas at the junction location would have had if the instrumentation system had not disturbed the distribution (the undisturbed value)

4. $T(4)$: the mass-flow-weighted average temperature the gas in the duct would have had, at the axial location of the thermocouple probe, had the instrumentation system not disturbed either the temperature or the flow distribution (the conceptual value)

The list of effects that must be counted as possible error sources depends on which of these options is elected as the definition of “true value.”

In the following paragraphs the various error sources are discussed in the order in which they implicitly appear in the hierarchy of possible true values listed above. This is a rather elaborate list. Measuring system errors are frequently the only ones dealt with in estimating measurement system uncertainty. This is unfortunate, especially in experiments involving heat transfer or temperature measurements, because the errors arising from system disturbance effects, system-sensor interactions, or conceptual difficulties are frequently larger than the measuring system errors.

Measuring System Errors

If $T(1)$ is elected as the true value, then only the measuring system errors need be considered. This category includes all the fixed and variable errors introduced by every component of the measuring system due to such effects as errors in gain (fixed error), ripple in power supplies (random), and drift due to temperature changes in the instrument. Keeping track of this collection of sources presents no small problem.

Errors introduced by the measurement system can be estimated by either experimental or analytical methods. Experimentally, one can do an “end-to-end” calibration, by providing a known and constant input to each channel of the measuring system and observing that channel’s output as a function of time; fixed errors are evidenced by an offset of the mean value, and variable errors by variations in the output. Typically, if the fixed errors exceed some acceptable level, they are “zeroed out” by pretest adjust-

ments to the system gain using a two-level check: zero input and "full scale for the test." It is the tolerance used to accept imperfect performance that establishes the residual fixed error in the measuring system—this includes the "fossilized" random error of the calibration, which is passed to the measuring system as a component of its fixed error. The total residual fixed error should be recorded for each channel and used as one of the inputs to the uncertainty analysis. Good practice requires both pre- and post-test checks to pick up evidence of low-frequency variable-but-deterministic error sources. When a change is found, the initial tolerance plus one-half of the change can be used as the estimate of fixed error.

In addition to the (fossilized) effects the random errors of the measuring system may have generated during the in situ calibration of the system, the random errors of the measuring system must also be accounted for in their effect on the acquired data. The variance of a set of repeated pretest trials can be used to measure the variable error that will be introduced by the measuring system, but the trials must cover a time interval representative of a true test, and the environmental conditions must be representative as well, or else the variable-but-deterministic components of variable error will not be picked up. A set of readings taken over a short time interval can detect only the variability that occurred during that interval. The variable error of the measuring system is usually not measured except specifically when a check on measuring system electrical performance is sought, since both single-sample

and multiple-sample uncertainty analyses pick up evidence of variable errors during their routine data acquisition.

The analytical approach to estimating measuring system errors requires an estimate of the fixed and random components of error from each component of the system, usually based on the manufacturer's specifications. Since few of those are described in terms of statistical properties of the population of instruments, this involves a great deal of interpretation. Taylor [10] presents a systematic approach to the identification and estimation of measuring system errors based on manufacturers' specifications. Figure 2 shows a typical worksheet used by Taylor in organizing the error analysis.

The manufacturer's specification description of an error must be converted into an estimate of the equivalent standard deviation σ , which would also describe the data. For example, if a manufacturer states that a component is accurate within 0.1% of full scale, this could be interpreted to mean that the odds are 20 to 1 against the error being larger than 0.1% of full scale. This interpretation is equivalent to claiming that σ for that error source was 0.05% of full scale. On the other hand, the original error statement could equally well have been interpreted to mean that 0.1% represented a 3σ excursion, in which case σ would have been judged to be 0.033% of full scale. The interpretation of manufacturer's specifications is a judgment call, based on experience. The σ estimates for the fixed and variable errors of each component must be entered on the worksheet in consistent

SENSOR	TEMPERATURE REFERENCE JUNCTION	DIFFERENTIAL AMPLIFIER	ACTIVE FILTER	MPLXR	PGA	S/H	ADC	
TYPE _____ RANGE _____	REF. TEMP. = _____ °. RANGE _____	CMV CMV GAIN = _____ °. ____ V CMV ____ V	δ_{CMV} GAIN = _____ °. ____ V		GAIN = _____ °. ____ V		F.S. = _____ V NO. BITS = _____	
MFG. _____ MODEL _____	MFG. _____ MODEL _____	MFG. _____ MODEL _____	MFG. _____ MODEL _____	MFG. _____ MODEL _____	MFG. _____ MODEL _____	MFG. _____ MODEL _____	MFG. _____ MODEL _____	
AMBIENT JUNCTION RTD MEAS. ERROR b7.1 = _____ S7.1 = _____ STABILITY S7.2 = _____	GAIN ACCURACY b6.1 = _____	GAIN ACCURACY b5.1 = _____	GAIN ACCURACY b4.1 = _____	GAIN ACCURACY b3.1 = _____	GAIN ACCURACY b2.1 = _____	GAIN ACCURACY b1.1 = _____	LINEARITY b1.2 = _____	
	LINEARITY b6.2 = _____	LINEARITY b5.2 = _____	LINEARITY b4.2 = _____	LINEARITY b3.2 = _____	LINEARITY b2.2 = _____	LINEARITY b1.2 = _____		
	OFFSET b6.3 = _____	OFFSET b5.3 = _____	OFFSET b4.3 = _____	OFFSET b3.3 = _____	OFFSET b2.3 = _____	OFFSET b1.3 = _____		
	JUNCTION ACCURACY b7.1 = _____	δ_{CMV} b6.4 = _____						
	TEMP. UNIFORMITY S7.1 = _____	NOISE S6.1 = _____	NOISE S5.1 = _____	NOISE S4.1 = _____	NOISE S3.1 = _____	NOISE S2.1 = _____		NOISE S1.1 = _____
	STABILITY S7.2 = _____	GAIN STABILITY S6.2 = _____	GAIN STABILITY S5.2 = _____	GAIN STABILITY S4.2 = _____	GAIN STABILITY S3.2 = _____	GAIN STABILITY S2.2 = _____		GAIN STABILITY S1.2 = _____
		ZERO STABILITY S6.3 = _____	ZERO STABILITY S5.3 = _____	ZERO STABILITY S4.3 = _____	ZERO STABILITY S3.3 = _____	ZERO STABILITY S2.3 = _____		ZERO STABILITY S1.3 = _____
		NORMAL MODE S6.4 = _____		CROSSTALK S4.4 = _____				QUANTIZATION S1.4 = _____
B8 = S8 =	B7 = S7 =	B6 = S6 =	B5 = S5 =	B4 = S4 =	B3 = S3 =	B2 = S2 =	B1 = S1 =	

Figure 2. A sample worksheet for estimating the errors in a measurement system. (From Taylor [10].)

terms, as well as that can be done, and the basis decision should be noted for the record: How was the specification interpreted? The σ estimates for the fixed (and variable) errors of the entire channel are then found as RSS combinations of the fixed (and variable) σ estimates of the components. To this point, the same procedure is followed for both single-sample and multiple-sample uncertainty analysis.

It is convenient to report all total error descriptors at the same confidence level; 95% is customary. For single-sample analyses, this is simply a matter of reporting the 2σ values, using the σ just estimated. For multiple-sample analysis the bias limit should also be reported as a 2σ value, but there is a complication with the precision index. In multiple-sample analysis the precision index must be related to the standard deviation of the mean of the sample, not that of the population. The standard deviation of the mean is $S_{(N)}/\sqrt{N}$, where N is the number of samples. In order to keep the present estimated values consistent with future measured values against which they may later be compared, it is important that the values reported here be appropriately scaled. The recommendation is to consider a hypothetical sample of 30: divide the estimated σ by 30 to estimate the standard deviation of the mean, and multiply by 2.0, the Student's t multiplier for $N \geq 30$. This has the desired effect of expressing the final value at 95% confidence. If, in some future test, the actual number of samples is less than 30, the processing of that data will involve the Student's t multiplier appropriate to the sample size and will produce a value comparable in meaning to the estimated value.

System-Sensor Interaction Errors

If the gas temperature $T(2)$ is used as the definition of the true value, instead of the thermocouple temperature, the system-sensor interaction errors (for example, the radiation, velocity, and conduction errors of a gas temperature measuring probe) must be added to the list of possible errors, since the gas temperature must be deduced from sensor temperature by correcting for these "errors" of measurement. These corrections are uncertain, due to the uncertainties in the heat transfer coefficient and the auxiliary information used in estimating the errors. This uncertainty remains as an uncertainty in the final measurement.

System-sensor interactions cause the achieved value of the measurand in the sensor to be different from the available value of the measurand in the system, at the location of the sensor, at the time the sensor is present. System-sensor interactions are not "errors," except to a measurements person; they are the natural consequences of the processes by which transducers respond to their environment.

The effects of these interactions are usually estimated analytically by deriving an equation relating the achieved value to the available value, considering the possible confounding mechanisms. As an example, consider the radiation error encountered when measuring the temperature of a hot gas in a duct with cold walls. The radiation error can be estimated by

$$T_{\text{gas}} - T_{\text{probe}} = \frac{\sigma\epsilon(T_{\text{probe}}^4 - T_{\text{wall}}^4)}{h} \quad (7)$$

where σ is the Stefan-Boltzmann constant, ϵ the emissivity of the sensor, T_{probe} the probe temperature (the achieved value), T_{gas} the gas temperature (the available value), T_{wall} the wall temperature, and h the heat transfer coefficient.

Equation (7) estimates the gas temperature (the available value) from the probe temperature (the achieved value), using estimates of the heat transfer coefficient, the emissivity of the wire, and the

wall temperatures, and can be used to correct the raw data. The uncertainty in the correction remains as an uncertainty in the final value of gas temperature and must be estimated.

There are four variables in Eq. (7), each of which is somewhat uncertain. The heat transfer coefficient h has the highest uncertainty (perhaps as much as $\pm 50\%$), but there is usually significant uncertainty in the "average" wall temperature as well as in the emissivity.

Equation (7) may be regarded as a small "data interpretation program" for calculating T_{gas} from T_{probe} and its uncertainty contribution calculated separately, or Eq. (7) could be imbedded in the main data interpretation program as a correction subroutine and the overall uncertainty in the final result calculated in one pass. From the developmental standpoint, it is better to keep it separate; then the contributions from different sources can be compared.

System Disturbance Errors

If the undisturbed value, $T(3)$, is used as the true value, then any system disturbance introduced by the measuring system must be corrected for, and the uncertainty in that correction remains as a residual uncertainty in the measurement.

The amount by which a measurand changes because of the effect of the instrumentation on the system depends on both the process system and the measuring system. The usual advice is to keep the sensors as small as possible, to minimize the disturbance, and then estimate the remaining effect with a simple equation.

It is difficult to describe a representative system disturbance error for a gas temperature measurement, since the effect of the probe blockage on gas temperature depends on too many factors---any example would seem contrived. Instead, a simpler situation will be used to illustrate this effect: measurement of a surface temperature on a metallic specimen.

Consider a thermocouple attached to a hot metal part but exposed to a cooler gas flow. The thermocouple may act as a fin, transferring heat to the gas and locally cooling the metal part, just where the thermocouple junction is attached. This is a typical system disturbance effect.

For simplicity, assume no other sources of error; the thermocouple junction temperature (the achieved value) is assumed equal to the temperature of the metal where the junction is attached (the available value). The entire effect is due to the system disturbance caused by the heat extraction by the thermocouple. The effect is described by the following equation:

$$\frac{T_{\text{avail}} - T_{\text{undist}}}{T_{\text{gas}} - T_{\text{undist}}} = \frac{1 + \phi}{\phi} \quad (8)$$

where

$$\phi = \frac{1 + (hDk_w)^{1/2}/2k_s}{(hDk_w)^{1/2}/2k_s}$$

Equation (8) estimates the undisturbed value (metal temperature with no disturbance) from the available value (metal temperature with the disturbance), using estimates of the heat transfer coefficient, the wire diameter, and the conductivities of the materials involved. The thermocouple returns, at best, the available value. If the undisturbed temperature is needed, then a correction for the system disturbance must be made using Eq. (8) and the uncertainty in that correction must be estimated.

There are six variables in Eq. (8), each of which is somewhat uncertain, but most of the uncertainty will come from the heat

transfer coefficient h , which rarely can be trusted within $\pm 50\%$. If hand calculations are being used, a single-term uncertainty estimate would be justified.

Again, this correction can be done in a "correction" subroutine or inside the main program, but the "separate" approach is recommended for its diagnostic value.

Conceptual Errors

If the mixed mean fluid temperature, $T(4)$, is elected as the true value, then the effects of velocity and temperature maldistribution must be added to the list of sources of errors, and the uncertainties in the correction factors that account for the maldistribution must be considered. This requires estimates of the "pattern factors" that describe variations in velocity and temperature and an estimate of how the value at the probe locations compare with the mean values. In many situations the conceptual errors are the largest by far. There seems to be no limit to the mistakes one can make in assigning significance to what has been measured.

Identifying the True Value

Faced with the several options for the definition of the true value, how is the experimenter to decide which to use for a given measurement and thus identify the errors that must be considered?

The key question is: How will this value finally be used—what does it represent in the end-use equations? The end-use equations implicitly define the true value.

Consider the following examples:

$$Q = MC_p(T_1 - T_2) \quad (9)$$

$$Q = hA(T_s - T_g) \quad (10)$$

$$\text{BSFC} = \frac{W(t_1 - t_2)}{KTN} \quad (11)$$

Equation (9), an energy balance, requires that T_1 and T_2 be defined as the mass-flow-weighted average temperatures of the fluid flowing, while C_p must be the average value of its specific heat C_p over the temperature range T_1 to T_2 .

Equation (10), a heat transfer rate equation, requires that T_s be the average temperature of the surface over the area A , while T_g must be defined as required by the definition of h . The heat transfer coefficient is sometimes defined in terms of the mass-flow-weighted average gas temperature, sometimes in terms of the temperature of the gas far from the surface, and sometimes in terms of the adiabatic wall temperature of the surface. There is no way to specify the requirements for T_g without knowing which definition of h will later be used with it.

Equation (11), for calculating the brake specific fuel consumption (BSFC) of an engine, requires that W be the mass of fuel consumed by the engine over the time interval $t_1 - t_2$ and KTN (a calibration constant K times the torque times the speed) be the average power produced over that same interval. Someone must decide whether that "power" should be measured before or after allowance for the accessories (fan, air cleaner, generator, air conditioner, etc.). Uncertainties in those allowances constitute uncertainties in the corrected result. The torque, in this equation, must be the torque produced by the engine. Bearing and windage losses might make the torque different at the torque meter than at the output of the engine.

By looking at the end-use equations, the required definition of the "true" value can be identified for each measurement. Then

the possible error sources can be identified and the uncertainties in their corrections estimated. For any one measurement, there may be several error sources, each contributing a fixed error and a random error. These must all be dealt with according to the techniques appropriate for single-sample or multiple-sample uncertainty analysis, as discussed in the following sections.

The objective of the measurement must be thoughtfully and precisely identified by consideration of the end-use equations, as shown by the examples above, before one can list the "errors" in a particular "measurement." Those terms are in quotes to emphasize that they are dangerous words, subject to interpretation, and may mean different things to the talker and the listener, even though both may be nodding their heads in apparent agreement.

Overall Uncertainties

The overall uncertainties assigned to each measurement must include the measuring system uncertainties plus all those associated with system disturbances, system-sensor interactions, and the idealizations invoked in the data interpretation equations.

Subsequent calculations are simplified if the fixed and variable components of the overall uncertainty of each measurement are separately identified. Thus, for each measurement, there will be an overall "fixed error" estimate and an overall "variable error" estimate. The overall "fixed error" (or variable error) assigned to a measurement is calculated as the RSS combination of the estimated "fixed errors" (or variable errors) introduced by all of the possible sources.

In the following sections, whenever estimates of the fixed error or variable error are assigned to a measurement, those estimates will be the overall values. The reported value will, in every case, be treated as the best estimate of the true value of that measurement, as required in the end-use equations, and the fixed and variable error estimates for that measurement include the contributions from all the necessary corrections.

Note that failing to apply a correction for a recognized error whose magnitude can be estimated is equivalent to assigning a zero-valued correction along with an uncertainty large enough to include the required correction: the reported value is unchanged, but its uncertainty goes up.

SINGLE-SAMPLE UNCERTAINTY ANALYSIS

Single-sample experiments are those in which each test point is run only once or at most a very few times. Exploratory and research experiments in fluid mechanics and heat transfer are usually single-sample experiments, characterized by a broad parameter range with sparsely distributed data. Much effort is spent in establishing the credibility of the experiment during the "shakedown and debugging" and "baseline testing" phases. Such experiments are frequently configured like icebergs: Only about 10% of the effort shows as output data; the other 90%, which establishes the validity of the 10%, is unseen—relegated to the logbook. Single-sample uncertainty analysis has evolved with two objectives: first, to describe the uncertainty in the reported result and, second, to serve as a diagnostic tool during the long and tedious developmental phase of the experiments. Thus, in addition to estimating the overall uncertainty in the final result, single-sample uncertainty analysis generates two auxiliary descriptors for use during the developmental phases.

There are three questions an experimenter must deal with in developing and reporting an experimental result.

During the planning of the experiment:

Q(0). Is the proposed instrumentation system acceptable for the proposed measurements?

During the shakedown and debugging of the experiment:

Q(1). How much scatter in the results of repeated trials with this apparatus can be explained by what I already know about the instrumentation and the process?

When reporting the results in the literature:

Q(N). What is the overall uncertainty in the reported result?

To answer these questions, single-sample uncertainty analysis generates three descriptors for each result: the zeroth-, first-, and N th-order uncertainty estimates. Each of these is, in one sense, a "total uncertainty" for a particular kind of replication of the experiment—a measure of the scatter that would be encountered if the experiment were repeated subject to a particular set of constraints. The terminology arises from considering the role of replication in setting the levels of uncertainty in experiments. All of the techniques for estimating uncertainty either require actual replication (eg, multiple-sample uncertainty analysis) or require that one imagine that replications might be made (single-sample uncertainty analysis). From experience we know that when an experiment is run several times, the way in which the replications are conducted affects the scatter in the results. Thus it makes sense to assign different uncertainty intervals to different levels of replication. While any particular experiment could be "repeated" in hundreds of slightly different ways, experience has shown that three baseline levels of replication are useful to consider:

- Repeated trials with the process held absolutely constant
 - Zero variability in the process, but all fixed and variable errors of measurement accounted for
 - The zeroth-order replication
- Repeated trials with the process running, but no fixed errors
 - All the above variable errors plus one new source—process unsteadiness
 - The first-order replication
- Repeated trials with the process running and all fixed errors "on"
 - All the variable errors plus all the fixed errors
 - The N th-order replication

The uncertainty associated with each replication level describes the scatter that would be observed if the experiment were repeated under the constraints indicated above.

The zeroth-order uncertainty interval estimates the overall uncertainty (fixed errors and variable errors) arising from the instrumentation system itself. It is used to assess the suitability of a proposed instrumentation system. This is the smallest uncertainty interval that could be achieved with the proposed instrumentation. The only situation under which this could describe the overall uncertainty of the experiment would be one in which the experiment could be replicated with no variations in its process and no uncertainty in any error corrections.

The first-order uncertainty interval estimates the scatter in the results of repeated trials using the same equipment, procedures, and instrumentation each time but with the process running. It is used during the debugging phase of an experiment to assess the significance of scatter in the output. The first-order uncertainty interval accounts for all variable errors in the data, including the effects of process unsteadiness, but does not include any fixed

errors. It differs from the variable error component of the zeroth-order uncertainty because the first-order estimate includes the process variability.

The N th-order uncertainty interval estimates the overall uncertainty in the experiment as it was run, including the effects of process unsteadiness. It acknowledges the fixed and variable errors in the measuring system and in all the corrections applied to the observed value. This is the value that must be reported in publications. Note that the N th-order interval does not include any allowance for fixed errors that may have been introduced by the choice of apparatus or technique, unless those errors were picked up as system-sensor interaction errors, system disturbance errors, or conceptual errors. The replication envisioned under N th order allows for repeated trials with the present apparatus, with complete replication of all instruments and with variability in the conditions causing systematic errors.

Describing Fixed and Variable Errors in Measurements

The overall fixed error component of uncertainty is evaluated in the same way for single-sample and multiple-sample analysis: as the RSS combination of the component fixed errors.

Variable errors are handled differently in single-sample and multiple-sample analyses, in two ways. The first arises from the fact that each observation may be made only once in a single-sample experiment. Unlike a multiple-sample experiment, in which the variable error in a set of measurements can be determined from the variance of the set itself, single-sample experiments require an auxiliary experiment in order to estimate the variable component of uncertainty. This usually takes the form of a set of independent observations of the process at a representative test condition, over a representative interval of time, usually a set of 30 observations. The second difference arises from the fact that the uncertainty interval must be referred to an individual measurement, not the mean of a set. This means that the standard deviation of the population must be used, not the standard deviation of the mean of the set, and the data from the auxiliary experiment must be processed accordingly.

Consider a single observation of the parameter X_i , made with an instrument subject to normally distributed random errors but, for the moment, no fixed error. If one wishes to estimate where the true value lies with respect to that measurement, the standard deviation of the population of possible measurements from which that observation was taken must be used to estimate the uncertainty interval. The descriptive statement would be, for the usual 20:1 odds (95% confidence).

$$X_i = X_i(\text{measured}) \pm 2\sigma \quad \text{at (20:1) odds} \quad (12)$$

where X_i is the true value, X_i (measured) is the measured value, and σ is the standard deviation of the population from which the individual observation was taken.

Equation (12) follows from the assumption that if an infinite number of observations were made with the specified system, the mean value would be the true value (we have assumed no fixed error, for the moment) and the random errors would be normally distributed around the true value, with a standard deviation of σ . From the fact that 95% of all the elements in the population lie within $\pm 2\sigma$ of the mean, we can estimate, with 95% confidence, that the mean must lie within $\pm 2\sigma$ of the observation.

The principal difficulty here is in finding σ , the standard deviation of the population from a smaller-than-infinite set of observations. σ is different from the standard deviation of the set of observations made in the auxiliary experiment but can be estimated from it. The relationship between the standard deviation

Table 1. Estimating σ from S

n	σ/S (maximum)	σ/S (minimum)
1	31.62	0.45
2	6.26	0.52
5	2.45	0.63
10	1.75	0.70
20	1.44	0.76
30	1.34	0.80
40	1.28	0.82
50	1.24	0.84
60	1.22	0.85

of a sample and that of its parent population is described by the chi-squared statistic. This statistic must be used in estimating σ from the auxiliary data of a single-sample uncertainty analysis, as was pointed out by Taylor [11].

Table 1, adapted from Lindgren [12], lists the minimum and maximum values of σ/S , where S is the standard deviation of the sample of N measurements and σ is the standard deviation of the population from which the sample was taken. The parameter n is the number of degrees of freedom in the estimate of S , equal to $N - 1$, where N is the number of values in the set.

Given a set of 30 readings (strictly speaking, 31) from which a sample standard deviation S has been calculated, the most likely value for the standard deviation σ of the parent population is the value of S , but σ might lie anywhere between the maximum and minimum values from Table 1. Thus, the value that should be used for σ could be either S or $1.34S$ or $0.80S$, depending on whether one wished to use the most likely, the largest, or the smallest estimate of the random uncertainty component in the measurement.

The auxiliary experiment is usually conducted at the beginning of the main experiment, as part of the shakedown and debugging process. Using the value of σ from the auxiliary test to interpret the single-sample observations of the main experiment is akin to pooling the variance over the experiment. If conditions change markedly over the range of test conditions, more than one auxiliary data set may be needed.

Describing the Single-Sample Uncertainty in a Measurement

In single-sample uncertainty analysis, each measurement is assigned three uncertainty values: its zeroth-, first-, and N th-order uncertainties. The separate components of "fixed error" and "variable error" used in the calculations are no longer visible once the three uncertainties have been calculated.

The *zeroth-order uncertainty* of a measurement is the RSS combination of all the fixed and random uncertainty components introduced by the measuring system, as illustrated in Eq. (1).

$$\delta X_{i,0} = \{(\delta X_{i,\text{fixed}})^2 + (2\sigma_{i,0})^2\}^{1/2} \quad (13)$$

where $\delta X_{i,\text{fixed}}$ is the overall fixed error uncertainty of the measuring system, and $\sigma_{i,0}$ is the standard deviation of the population of individual measurements from the measuring system when its input is stationary.

The uncertainty $\delta X_{i,0}$ corresponds to the $U_{0.95}$ uncertainty of multiple-sample uncertainty for the special case where only measuring system errors are considered. It is not equivalent to the overall bias limit used in multiple-sample analysis, because it includes the random measurement errors.

The first-order uncertainty of a measurement describes the scatter that would be expected in a set of observations using the given apparatus and instrumentation system, while the observed process is running. The first-order uncertainty includes all effects of process unsteadiness as well as all variable error effects from the measuring system, both random and variable but deterministic, but does not include any fixed error effects from any source. It corresponds to a generalized precision index of the measurement, in the nomenclature of multiple-sample uncertainty analysis—generalized, because it includes the effect of process unsteadiness.

The first-order uncertainty interval for each measurement type must be measured in an auxiliary experiment, before the main data runs. It is found from the standard deviation of a set of 30 or more observations using the chi-squared statistic to estimate its maximum and minimum probable values. The set of data should be taken with the system running at a representative test condition and should be distributed over a representative interval of time, considering the process involved, with all test conditions representative of normal operation. This test is usually run during the shakedown and debugging phase of an experiment, and its result stored for later use in interpreting running results.

Given an auxiliary set of 30 readings (strictly speaking, 31) from which a standard deviation S has been calculated, the most likely value for σ_1 is the value of S , but σ_1 might lie anywhere between the maximum and minimum values from Table 1. The value that should be used for σ_1 could be either S or $1.34S$ or $0.80S$, depending on whether one wished to use the most likely, the largest, or the smallest estimate. That decision depends on the end use of the uncertainty analysis.

The *Nth-order uncertainty* of a result is a measure of its overall uncertainty, accounting for all sources of fixed and variable error in the experiment as it was run. This is the value that should be reported as the overall uncertainty when using single-sample uncertainty analysis.

The *Nth-order uncertainty* does not include any allowance for errors that might be embedded in the experiment because of the choice of technique or apparatus, except as those errors would have been picked up as system disturbances, system-sensor interactions, or conceptual errors. Those errors can only be identified by comparing the results from two different experiments and seeing whether or not they agree within the expected interval, considering their individual *N*th-order uncertainty intervals. If the two sets of results do not agree, that is evidence that the two experiments were different in some significant way.

The *Nth-order uncertainty* is calculated as the RSS combination of the first-order uncertainty $\delta X_{i,1}$ with the root-sum-square combination of fixed errors from every source.

$$\delta X_{i,N} = \{(\delta X_{i,1})^2 + (\text{RSS } \delta X_{i,\text{fixed}})^2\}^{1/2} \quad (14)$$

where $\delta X_{i,1}$ is the first-order uncertainty in X_i , and RSS $\delta X_{i,\text{fixed}}$ is the root sum square of all fixed error contributions from every level of error source.

The *Nth-order uncertainty interval* is equivalent in concept to the $U_{0.95}$ uncertainty of multiple-sample analysis, although the values would probably be different because the *Nth-order* measure includes the effects of the process unsteadiness, which may have been averaged out in the multiple-sample analysis.

These are the three fundamental possibilities, but intermediate replication levels can easily be described, and experimenters are free to define fractional replication levels as needed to clarify the discussion of uncertainties.

As an example, consider the problem of identifying the source of uncertainties in the electrical resistance of silicon solar cells. Taking the object of the measurement to be the cell resistance, the

test program might take the following form:

1. Repeated observations of the measuring instrument, left attached to a reference resistor. This samples the variable error of the measuring system—a component of the zeroth-order uncertainty in the measurement.
2. Repeated observations of the measuring instrument, left attached to a single working cell. This evaluates the first-order uncertainty δR_1 in the measurement—a measure of the total random error. The square root of the difference of the squares of the first-order uncertainties of these two data sets measures the uncertainty introduced by unsteadiness in the resistance of one specimen.
3. Repeatedly attaching and detaching the same measuring instrument to the same specimen, with one measurement for each attachment. This samples one more degree of freedom of the overall problem. The first-order uncertainty of this experiment might be called the 1.1th-order uncertainty. The square root of the difference of the squares of the first-order uncertainties of these last two data sets measures the uncertainty introduced by the act of attaching and detaching the instrument.
4. Another set of trials, denoted level 1.2, might involve repeated applications of the same measuring instrument to different sample cells. The square root of the difference of the squares of the first-order uncertainties of these last two experiments measures the increase in uncertainty associated with the use of different cells for each observation—this is a measure of the variance in the cell resistance.

When fractional replication level experiments are used, it is important to clearly describe the increase in complexity that accompanies each level and clearly define the operations involved. The purpose of fractional replication level experiments is usually to identify the contributions to scatter in the results of a complex overall experiment. In general, the same instrumentation is used throughout the series, and hence any fixed errors in their calibrations would affect each level of the trials in the same way. Thus, fixed errors may not enter the picture until the final (absolute) results are described.

Describing the Single-Sample Uncertainty in a Result

In single-sample uncertainty analysis, each result is assigned three uncertainty values: its zeroth-, first-, and N th-order uncertainty. The separate components of fixed error and variable error are not described.

The form for calculating the single-sample uncertainty in a result is the same regardless of whether the zeroth-, first-, or N th-level interval is being calculated.

The result R of the experiment is assumed to be calculated from a set of measurements using a data interpretation program (by hand or by computer) represented by

$$R = R(X_1, X_2, X_3, \dots, X_N) \quad (15)$$

The effect of each measurement uncertainty on the calculated result if only that one measurement were in error would be

$$\delta R_{X_i} = \frac{\partial R}{\partial X_i} \delta X_i \quad (16)$$

The partial derivative of R with respect to X_i is the *sensitivity coefficient* for the result R with respect to the measurement X_i .

When several independent variables are used in the function R , the individual terms are combined by a root-sum-square method:

$$\delta R = \left\{ \sum_{i=1}^n \left(\frac{\partial R}{\partial X_i} \delta X_i \right)^2 \right\}^{1/2} \quad (17)$$

This is the basic equation of single-sample uncertainty analysis. Each term represents the contribution made by the uncertainty in one variable X_i to the overall uncertainty in the result R . Each term has the same form: the partial derivative of R with respect to X_i multiplied by the uncertainty interval for that variable. The estimated uncertainty in the result has the same probability of occurrence as have the uncertainties in the individual measurements.

Interpreting Single-Sample Uncertainty Statements

The *zeroth-order uncertainty* interval associated with a result estimates the contribution to its overall uncertainty arising from the instrumentation system itself. It is used in experiment planning to assess the suitability of a proposed instrumentation system for the projected task and in reporting to show the contribution of the instrumentation to the final uncertainty.

The *first-order uncertainty* interval associated with a result estimates the scatter that should be observed in the results of repeated trials using the same instrumentation each time, with the experiment running.

The first-order uncertainty interval should be used to assess the significance of scatter in the result of repeated trials with the same instruments and equipment—particularly helpful during the debugging phase of an experiment. First-order uncertainty intervals should not be used to assess the difference between data from two different rigs, because they do not acknowledge the possible fixed errors associated with the instruments or the corrections applied.

The *N th-order uncertainty* interval estimates the overall uncertainty in the result: it acknowledges all recognized sources of fixed and variable error in the experiment as it was conducted. It includes the fixed errors due to the measuring system and all corrections for system disturbances, system-sensor interactions, and conceptual errors, as well as the total variable error observed with the process running. It does not consider the uncertainty associated with the choices of apparatus or procedure, except as those are picked up in the above-mentioned error sources.

The *N th-order uncertainty* interval should be used to report the overall uncertainty in the measured result. The *N th-order uncertainty* intervals should not be used to assess the scatter on repeated trials with the same instruments and equipment, because it includes allowances for fixed errors that cannot contribute to the scatter on repeated trials.

MULTIPLE-SAMPLE UNCERTAINTY ANALYSIS

Multiple-sample tests or experiments are those in which enough data are taken at each test point to support a sound statistical interpretation of the random error characteristics of the set. Multiple-sample experiments are usually characterized by a compact test program calling for a sparse set of test points covered densely by data. Two typical multiple-sample experiments would be measuring a flow rate for custody transfer purposes or testing a product to demonstrate compliance with warranties. Such testing is common in industry and usually follows thoroughly standardized test procedures. Standardization of the tests is important to avoid arguments between buyer and seller as to the validity of the

test method. In addition, it is important to have a thoroughly standardized procedure for calculating measurement uncertainty in such cases, to avoid arguments between buyer and seller as to the meaning of "borderline" test results. Multiple-sample uncertainty analysis has responded to these pressures by leaning heavily on well-accepted statistical methods, in some instances forcing the structure of the experiment to conform to the requirements dictated by statistical theory.

Multiple-sample analysis concentrates on processing the means of sets of observations of each measurand. This reduces the visible effects of random errors due to both the instrumentation system and system unsteadiness, since those are "averaged out." As a consequence, multiple-sample uncertainty estimates are usually dominated by fixed errors, in contrast to single-sample experiments, which are usually dominated by variable errors. As a result of this viewpoint, multiple-sample uncertainty analysis discusses the calibration and traceability hierarchy in some detail and elaborates the statistical methods presented to allow the combination of calibration results with test results in a meaningful way. These features of multiple-sample analysis could be "borrowed" into single-sample analysis for situations where the calibration errors contribute significantly to the overall uncertainty.

As a consequence of the clear need for standardization in industrial acceptance testing, multiple-sample uncertainty analysis has moved much more rapidly toward this end than has single-sample analysis.

The present state of the art in specifications for multiple-sample uncertainty analysis is represented by ANSI/ASME PTC 19.1-1985 [7], Measurement Uncertainty. This document is a direct outgrowth of the early work of Abernethy and Thompson [5] and clearly reflects this view in its statement of purpose:

- [To identify] corrective action required to achieve test objective...
- [To provide] test validation...
- [To reduce] risk of making erroneous decisions...
- [To demonstrate] compliance with agreements

There are some differences of viewpoint and nomenclature that distinguish multiple-sample uncertainty analysis from single-sample analysis that should be pointed out at the outset.

In multiple-sample experiments, the reported value of the result is calculated from the mean of a set of observations at the test point, not from an individual observation. The uncertainty interval associated with the result surrounds that mean result.

The uncertainty in a result is found as a combination of the fixed error and the random error of the result and the term "uncertainty" is used only in connection with the final result. Individual measurements used in calculating the result are described as having "fixed" and "random" errors but are not said to have uncertainties per se, unless the measurement itself constitutes a result. For example, a measurement of "average gas temperature," as a result in its own right, would be said to have an uncertainty, but if this value were being used as one input to the calculation of a more complex parameter only its "fixed" and "random" errors would be described. All errors that are not "fixed" are presumed to be "random."

Perhaps because multiple-sample uncertainty analysis is so frequently concerned with standardized procedures, little emphasis has been given to the evolution of diagnostic techniques to help the experimenter in developing new experiments. The only outputs generated under multiple-sample uncertainty analysis, except for the overall uncertainty in the result, are the bias limit

and precision index of the result. This is not to say that a skillful practitioner cannot extract useful diagnostics from a multiple-sample uncertainty analysis—the information is certainly there—but simply to point out that single-sample and multiple-sample uncertainty analyses have somewhat different objectives and therefore somewhat different structures.

A Warning

The material in this section is taken from ASME PTC 19.1-1985 [7], except where otherwise noted. That document describes the statistical basis for multiple-sample uncertainty analysis and covers situations ranging from simple to complex.

The present discussion is limited to "simple" experiments, defined as follows. A "simple" experiment is one in which each sensor has been previously calibrated; a measuring system has been established by connecting appropriate sensors, amplifiers, and recorders; and N data sets have been taken, where each data set is the result of one sweep through all the measurement channels. A single data set consists of one observation of each variable, and the set of N observations constitutes one experiment. The mean value of each variable is calculated from the set of N observations and used to calculate some result R . The objective of the uncertainty analysis is to describe the uncertainty in the result.

The simplicity of this experiment arises from two sources: First, the calibration of the sensors is considered "done" before the present experiment starts and, second, each variable is measured the same number of times. Some experiments, particularly calibrations, involve simultaneous consideration of calibration, data acquisition, and data interpretation and involve different numbers of observations of different parts of the experiment. This complicates the statistical interpretation considerably. The average user should not try to interpret those situations, either from this document or from PTC 19.1-1985 [7], but should consult an expert in the statistical analysis of experimental error.

Describing Fixed and Random Errors in a Measurement

The residual fixed error and the random error of each measurement are described by its bias limit and its precision index, respectively, but other information must be available as well:

- The mean value of a set of N observations of the measurement, \bar{X}_i
- The precision index of the mean, $S_{\bar{X}_i}$, an estimate of the standard deviation of the mean of the set of N observations
- The number of degrees of freedom of the precision index of the mean, v
- The bias limit of the measurement, B_{X_i}

Any error that will not change during the conduct of the experiment is a "fixed error." Such errors arise from sources such as manufacturers' tolerances on instruments (eg, "accurate within ± 0.001 volts"), or tolerances on the setup of instrument channels (eg, "set the output to 1.000 volts ± 0.001 for a 1.00 mV input"). The fixed errors that remain embedded in a measurement, while "knowable," are not known; if they were known, a correction would have been applied and the "error" would no longer exist. These residuals must be estimated from manufacturers' specifications or by inspection of the setup tolerances on instruments.

The *bias limit* of a measurement is an estimate of the maximum probable value of its fixed error, usually estimated at 95%

confidence (ie, a value large enough that a prudent person would bet 20:1 against it being exceeded).

Consider a single variable X_i , which is being measured in a multiple-sample manner. Consider, first, the instrument or measurement system channel dedicated to that measurement—the instrument has been calibrated, and its calibration has an estimated fixed error B_{cal} . The calibration introduces no variable error, since an error in calibration cannot cause scatter on repeated readings. Other fixed errors may be introduced during the data acquisition and data reduction process, represented by B_{acq} and B_{red} . The overall bias limit of the measurement, B_{X_i} , is the root-sum-square combination of all these fixed error components.

$$B_{X_i} = \{(B_{\text{cal}})^2 + (B_{\text{acq}})^2 + (B_{\text{red}})^2\}^{1/2} \quad (18)$$

The first term on the right-hand side of Eq. (18), B_{cal} , can be interpreted in two ways, depending on the structure of the experiment, although this point is not discussed in PTC 19.1-1985 [7]. If the calibration is considered part of the present measurement act, then B_{cal} represents only the fixed errors introduced by the calibration, and its random errors will be picked up in the calculation of the precision index. If, as is more usually done, the calibration information has been “handed down” from a previous step, then the B_{cal} term on the right-hand side of Eq. (18) represents the “fossilized” overall uncertainty of that calibration (see the discussion of fossilization in the next section). In this latter case, the only random errors in the present experiment would be those apparent in the present data set. Shifting terms from precision to bias, through fossilization, does not alter $U_{0.95}$, since the RSS combination is unaffected, but it does alter the balance between fixed and random errors. If the precision index and the bias limit are being used separately as diagnostics, the effects of fossilization must be kept in mind.

Bias limits can be assessed either experimentally or analytically, following the procedures discussed earlier with respect to fixed errors. Experimentally, the acceptance tolerance for less-than-perfect response in a pretest system calibration check constitutes the bias limit for that channel. Analytically, Taylor's worksheet for cataloging the errors in a measurement, referred to in the earlier section, can be used. The symbols B and S in that worksheet stand for bias errors and precision errors, respectively. If, in a measurement channel, the sensor has been previously calibrated, its overall calibration uncertainty (the RSS combination of the fixed and random errors of that calibration) should be regarded as having been fossilized, and the overall uncertainty should be entered as the fixed error of calibration. No random error component would be entered for the sensor calibration. The other components of the channel, however, being involved in the present act of measurement, would contribute both fixed and random errors.

The bias limit is estimated at 95% confidence and therefore serves in a manner equivalent to a 2σ estimate of the fixed error. It is combined with the 95% confidence $tS_{\bar{X}}$ estimate of the random error in calculating the overall uncertainty. Thus, although the term has no statistical basis, it is used, as a practical matter, as though it did.

The *precision index* of a measurement is a measure of its random error and can be estimated from the test data alone without an external reference. The precision index of a data set is equal to its standard deviation $S_{\bar{X}_i}$. To be useful in estimating uncertainty, $S_{\bar{X}_i}$ must be accompanied by knowledge of the number of degrees of freedom it represents, so that an appropriate value of the Student's t multiplier can be used to describe the 95% confidence interval.

The precision index of a set of measurements is calculated from the data of the present act of measurement; it is not necessary to estimate it. Sometimes, however, one wishes to estimate what the precision index of a set of measurements would be, to evaluate the performance of a proposed measuring system. It is also helpful to understand how the contributions to random error arise in a system and interact. For these two reasons, it is worthwhile to note that, analytically, the overall precision index of a measurement is the root-sum-square combination of the precision indices of all the subordinate measurements considered to be part of the present measurement. A worst-case situation would involve calibration, acquisition, and data reduction, all being considered part of one operation. For that complex experiment,

$$S_{X_i} = \{(S_{\text{cal}})^2 + (S_{\text{acq}})^2 + (S_{\text{int}})^2\}^{1/2} \quad (19)$$

Each of the three terms on the right-hand side of Eq. (19)— S_{cal} , for example—may also be the root-sum-square combination of several random error components contributed by the measurements used in the process at that level, providing they are all considered to be part of the same act of measurement.

In a more usual situation, there is only one term to consider, since random errors can arise only with the present act of measurement—random errors arising in previous steps are “fossilized.” The issue of fossilization is not addressed in PTC 19.1 [7], except indirectly, but is important to understand. It arises most often in connection with calibrations. The purpose of a calibration is to evaluate the correction that must be added to the indicated value to obtain the true value. The correction is usually taken as the mean value of the error over the set of calibration trials and is therefore equal to the average error of the instrument in the calibration data set. There is some residual uncertainty in this calibration correction, since it was derived from a sample of finite size. A second calibration would quite likely yield a different value for C for the same nominal conditions. The variability of the mean of a set of observations, on repeated trials, is related to the standard deviation of the population of individual measurements and is closely related to the precision error of the calibration. By definition, the precision index of the calibration, $t(S_N)/\sqrt{N}$, measures the range within which the average of a set of N readings of the instrument might lie by chance alone. Thus the precision index of the calibration has two meanings: It is a direct measure of the uncertainty in the correction constant deduced by the calibration experiment as a consequence of random errors in the calibration process, and it is also a direct measure of the possible value of the residual fixed error that constant may have when the instrument is applied to a later experiment.

In the usual multiple-sample measurement situation, an instrument whose calibration uncertainty has been fossilized is being applied to a new task. The only sources of random error in this new act of measurement are those that arise from the new task. The precision index of the new set of measurements is determined directly from the new measurements themselves.

The precision index of an individual measurement, X_i , is determined directly from the data set, as follows:

$$S_{X_i} = \left\{ \sum_{i=1}^N \frac{(X_i - \bar{X}_i)^2}{N-1} \right\}^{1/2} \quad (20)$$

The precision index of the mean is what is needed, however, in the analysis of uncertainties, since the resulting uncertainty will be

associated with the mean value, not an individual value:

$$S_{\bar{X}_i} = \frac{S_{X_i}}{\sqrt{N}} \quad (21)$$

The number of degrees of freedom, ν , associated with the precision index of a simple experiment (one set of N observations) is $N - 1$. The same value applies to the precision index of the mean. This is the value to be used in what is described above as "the usual situation," where the random errors of previous steps have been fossilized and the only components of randomness are visible in the data themselves.

When complex measurements are considered (ie, when all the data from calibration, acquisition, and reduction are being processed as one measurement act), calculating the number of degrees of freedom is more complicated, and the full text of PTC 19.1-1985 [7] should be consulted.

Describing the Overall Uncertainty in a Single Measurement

The uncertainty is not usually calculated for a single measurement if it is used only as a component of a future calculation; instead, three pieces of information are provided:

- The bias limit, B_{X_i}
- The precision index of the mean, $S_{\bar{X}_i}$
- The number of degrees of freedom, ν

The uncertainty is calculated for a single measurement, which is considered to be an end in itself.

$$U_{0.95} = \{(B_{X_i})^2 + (tS_{\bar{X}_i})^2\}^{1/2} \quad (22)$$

Here t is the Student's t multiplier for 95% confidence and ν degrees of freedom. Note that the precision index $S_{\bar{X}_i}$ of the mean is used in calculating $U_{0.95}$.

The final statement describing X_i as having an uncertainty would be: The best estimate of X_i is the mean value \bar{X}_i plus or minus the uncertainty interval $U_{0.95}$.

Describing the Overall Uncertainty in a Result

The result R of an experiment is assumed to be calculated from a set of measurements using a data interpretation program (by hand or by computer) represented by

$$R = R(X_1, X_2, X_3, \dots, X_N) \quad (23)$$

Each measurement has a known bias limit and precision index, and the number of degrees of freedom for each precision index is known. The first step in estimating the overall uncertainty in the result is to calculate the bias limit and the precision index of the result.

The bias limit of each measurement affects the bias limit of the calculated result in proportion to its sensitivity coefficient (the partial derivative of R with respect to X_i is the *sensitivity coefficient* of the result R with respect to the measurement X_i).

If only one measurement has a nonzero bias limit, the bias limit of the result would be

$$B_{R,i} = \frac{\partial R}{\partial X_i} B_i \quad (24)$$

When several independent variables are used in the function R , the individual terms are combined by a root-sum-square method.

$$B_R = \left\{ \sum_{i=1}^N \left(\frac{\partial R}{\partial X_i} B_i \right)^2 \right\}^{1/2} \quad (25)$$

The precision index of the result, S_R , is affected by the precision index of each measurement in the same way, and the same form is used in calculating it:

$$S_R = \left\{ \sum_{i=1}^N \left(\frac{\partial R}{\partial X_i} S_i \right)^2 \right\}^{1/2} \quad (26)$$

The number of degrees of freedom ν_R is $N - 1$, where N is the number of observations in the data set, assuming each variable was measured the same number of times. For more complex situations, the full text of PTC 19.1-1985 [7] should be consulted.

The Student's t multiplier (using the two-tailed distribution) is next found, at the 95% confidence level, and the uncertainty in the result evaluated:

$$(U_R)_{0.95} = \{(B_R)^2 + (tS_R)^2\}^{1/2} \quad (27)$$

This is the interval around the mean value R within which the true value is believed to lie.

REPORTING THE MEASUREMENT UNCERTAINTY IN EXPERIMENTAL RESULTS

Authors are expected to report the measurement uncertainty in at least their principal results. The uncertainty can be estimated using either the single-sample or multiple-sample technique, depending on the nature of the experiment. Whichever method is used, the uncertainty statement should be accompanied by enough discussion, and enough supporting data, that the reader can properly interpret its significance.

The most useful uncertainty estimate, from a reader's viewpoint, is one that has the following meaning:

- If the present experiment were repeated using the same apparatus and techniques, and similar (but not the same) instruments, 19 of 20 repeated trials would produce results within $\pm XX$ (or $PP\%$) of the present value.

or

- The reported value is the best estimate for the result, and, with 95% confidence, the true value is believed to lie within $\pm XX$ of that value.

Such statements allow the reader to assess the significance of any differences between the newly reported work and other sources in the literature.

Authors not familiar with the benefits of formal uncertainty analysis sometimes arbitrarily inflate the uncertainty intervals assigned to their data according to what others seem to claim for the same type of work, even though they may, privately, believe that their work is more accurate. This may be done to avoid arguments, or perhaps from fear of being considered boastful or naive. In either case, this inflation is counterproductive. When results are compared between two laboratories, the two data sets will be judged to "agree" if their difference is less than the expected uncertainty in that difference. Thus, overestimation of the uncertainty intervals makes it difficult to spot real differences and allows bad data to stay unchallenged. Underestimation, on the other hand, provokes unnecessary arguments.

SUMMARY AND CLOSURE

The process of uncertainty analysis begins with identification of the desired true value of the measurement in question. The basis for this identification must be found in the end-use equations, where that measurement is used in a conceptual sense. The key question is: What does that term stand for in the equation? Five options for the true value of any measurement can be identified: the observed value, the value reported from the measuring system; the achieved value, the value of the measurand in the sensor; the available value, what the achieved value would have been if the sensor had no installation error; the undisturbed value, what the available value would have been if the instrumentation had not disturbed the system; and the conceptual value, the value the experiment planner had in mind requesting the measurement.

Once the true value has been identified, the various error mechanisms (measurement system errors, system-sensor interactions, system disturbances, and conceptual errors) can be examined systematically to see how much error each contributes. With the error sources identified, corrections can be applied for the fixed errors each contributes—but not exactly. The uncertainty in these corrections contributes to the fixed error component of the overall uncertainty in the final result. All of the random error in a measurement is contributed by the present experiment—partly from the instrumentation and partly by the process instability. Once the fixed and random errors are known, the uncertainty descriptors can be calculated by either single-sample or multiple-sample methods.

Single-sample uncertainty analysis creates three descriptors for each result: the zeroth-order-, first-order-, and N th-order uncertainty estimates. Zeroth-order uncertainty estimates the possible error arising from the measuring system alone, including all fixed and random errors. It is used to evaluate the suitability of the instrumentation for the proposed task. First-order uncertainty estimates the scatter that should be expected on repeated trials with the same apparatus and instrumentation. It is used as a benchmark to judge whether the scatter actually observed is “reasonable” or is evidence of a problem with the experiment. N th-order uncertainty includes all recognized fixed and random errors and estimates the range within which all similar results should lie if obtained from a similar apparatus by means of a similar technique.

Multiple-sample uncertainty analysis also creates three descriptors for each result: the overall fixed error, the overall random error, and the overall uncertainty. The overall fixed error is the root sum square of all residual fixed errors in the present experiment. The overall random error is deduced from the observed multiple-sample data set. The overall uncertainty is calculated as the root sum square of the fixed error and random error terms, taking proper account of the number of degrees of freedom represented by the data set. The detailed mathematics of single-sample and multiple-sample analyses differ only slightly, and that difference is due mainly to the fact that multiple-sample analysis deals with the error in the mean of a set of data, while single-sample analysis necessarily deals with the error in a single observation.

When experimental results are reported in the literature, the overall uncertainty must be reported. The uncertainty could be calculated using either the single-sample or multiple-sample methodology, but the method and the basis data should be described. Conventionally, the engineering community accepts uncertainty estimates at 95% confidence (20:1 odds), but some

situations may require more stringent standards. The confidence level should always be explicitly noted.

Author's Note

All of the above, with its emphasis on statistics and equations, may sound as though uncertainty analysis is a pretty routine operation. Like most engineering subjects, it may seem cut and dried to the uninitiated—my colleagues in the “soft” sciences tell me how lucky I am to be in a field where opinion plays such a small role. I agree that once the equations and values are chosen, the solution follows—no question about that—but they might be surprised to learn about the arguments engineers have over *which* equations to use, and *what values* to install in them! The art of engineering lies not in solving the equations but in picking the “right” equations and installing the “right” values so the result of our mathematics answers the question we had in mind at the beginning. This is particularly true in uncertainty analysis. The equations are very simple, as are the statistical ideas, yet applying them to any real experiment raises a host of questions: How should this item be treated? Is this a fixed or a variable term? What does this result mean? Should this effect be considered an error? In the light of the expected uncertainty, can we say that these two results are really different? Which design should we go ahead with?

One of the key points in the present paper is the notion of starting the uncertainty analysis at the “bottom line”: identifying the implicit meaning of the result in its end-use equations. Starting from what the result must mean in the equations that use it and working through the different ways in which errors intrude into measurements leads, rather certainly, to a list of all the assumptions and errors that might stand between the concept and the measurement. That approach represents an attempt at systematizing the process of “picking the right equations.” Without agreement as to the real objective of a measurement, there can be endless arguments about what errors to include in the analysis. I would be interested in comments about this approach and examples of its application.

There are many areas in which this field is still developing. The term “fossilization,” for example, is less than 10 years old (as of this writing), has not been discussed much, and is still worth a good deal of thought. A number of issues about fossilization remain unclear: How and where should the idea of fossilization be applied in complex situations? How should one describe “fossilization” so that most users will come to the same conclusions, given the same situations? We need more discussion and more papers in this area, for ventilating the different opinions, identifying the unsolved problems, and reducing this idea to practice.

I am indebted to my friends for their patience and good will in answering my phone calls at odd hours and for their willingness to discuss the points I found hard to understand. In particular, I would like to thank Steve Kline, my colleague at Stanford; Ron Dieck, at Pratt & Whitney; Mike Englund, at Garrett Turbines; Jerry Catz, at the University of Miami; and Bob Abernethy, now retired from Pratt & Whitney. I have appreciated their help, their opinions, and their arguments and have tried to reflect their comments in the text as well as I could.

NOMENCLATURE

B	bias limit, generic
B_{cal}	bias limit due to calibration; see Eq. (18)

B_{acq}	bias limit due to acquisition; see Eq. (18)	δR_N	Nth-order uncertainty in the result
B_{int}	bias limit due to interpretation; see Eq. (18)	δR_{X_i}	contribution to the uncertainty in the result from variable X_i , generic
B_R	bias limit of the result; see Eq. (25)	$\delta R_{X_{i,0}}$	contribution to the zeroth-order uncertainty in the result from variable X_i , generic
$B_{R,i}$	contribution to the bias limit of the result from X_i ; see Eq. (24)	$\delta R_{X_{i,1}}$	contribution to the first-order uncertainty in the result from variable X_i , generic
C_i	contribution to the uncertainty of the results from the i th variable	$\delta R_{X_{i,N}}$	contribution to the N th-order uncertainty in the result from variable X_i , generic
C_P	specific heat of the fluid flowing; see Eq. (9), $J/(kg \ ^\circ C)$	ϵ	emissivity for thermal radiation; see Eq. (7)
D	thermocouple diameter; see Eq. (8), m	σ	standard deviation of a population, generic
h	heat transfer coefficient; see Eqs. (7) and (8), $W/(m^2 K)$	σ_0	standard deviation of a population of zeroth-order measurements
k_w	thermal conductivity of the thermocouple wire; see Eq. (8), $W/(m K)$	σ_1	standard deviation of a population of first-order measurements
k_s	thermal conductivity of the substrate; see Eq. (8), $W/(m K)$	σ_N	standard deviation of a population of N th-order measurements
M	mass flow rate of fluid; see Eq. (9), kg/s	ν	number of degrees of freedom
N	number of elements in a sample		
R	result of a calculation based on one or more measurements		
R_0	result calculated with all nominal values of the inputs; see the method of sequential perturbations		
R_{i+}	result calculated with the i th variable set high, in the method of sequential perturbations		
R_{i-}	results calculated with the i th variable set low, in the method of sequential perturbations		
S	standard deviation of a finite set, generic		
S_{cal}	precision index of the calibration experiment		
S_{X_i}	standard deviation of the measurements of the i th variable		
$S_{\bar{X}_i}$	precision index of the mean of the measurements of the i th variable		
T	temperature, used with various subscripts; see Eqs. (7) and (8), K		
$U_{0.95}$	overall uncertainty, as calculated in multiple-sample uncertainty analysis using the root-sum-square method		
X_i	the i th variable		
\bar{X}_i	the average value of the i th variable		
Greek Symbols			
δR	uncertainty in the result, generic		
δR_0	zeroth-order uncertainty in the result		
δR_1	first-order uncertainty in the result		

REFERENCES

1. Kline, S. J., and McClintock, F. A., Describing Uncertainties in Single Sample Experiments, *Mech. Eng.*, 3–8, Jan. 1953.
2. Airy, Sir George Biddle, *Theory of Errors of Observation*, Macmillan, London, 1879.
3. Moffat, R. J., Contributions to the Theory of Single Sample Uncertainty Analysis, *Trans. ASME, J. Fluids Eng.*, **104**, 250–260, 1982.
4. Moffat, R. J., Using Uncertainty Analysis in the Planning of an Experiment, *Trans. ASME, J. Fluids Eng.*, **107**, 173–178, 1985.
5. Abernethy, R. B., and Thompson, J. W., Jr., Handbook Uncertainty in Gas Turbine Measurements, AEDC-TR-73-5, AD 755356, Feb. 1973. Also available as *Measurement Uncertainty Handbook*, Revised, 1980, Instrument Society of America.
6. Abernethy, R. B., Benedict, R. P., and Dowdell, R. B., ASME Measurement Uncertainty, *Trans. ASME, J. Fluids Eng.*, **107**, 161–164, 1985.
7. ANSI/ASME PTC 19.1-1985, Supplement to Performance Test Coded, Instruments and Apparatus. Part 1—Measurement Uncertainty, The American Society of Mechanical Engineers, New York.
8. Catz, J., personal communication, Dept. of Mechanical Engineering, University of Miami, Coral Gables, Florida, 1987.
9. Moffat, R. J., The Measurement Chain and Validation of Experimental Measurements, Paper B-103, ACTA IMEKO, p. 45, 1973.
10. Taylor, J. L., *Computer-Based Data Acquisition Systems: Design Techniques*, Instrument Society of America, 1986.
11. Taylor, J. L., personal communication, May 1987.
12. Lindgren, B. W., *Statistical Theory*, 3rd ed., p. 575.

