Principles and Applications of Measurement Uncertainty Analysis in Research and Calibration

Chester V. Wells

Presented at the
*Third Annual Infrared Radiometric
Sensor Calibration Symposium*
September 14–17, 1992
Utah State University
Logan, Utah

◆◇◆ NR≡L

NOTICE

PRINCIPLES AND APPLICATIONS OF MEASUREMENT UNCERTAINTY ANALYSIS IN RESEARCH AND CALIBRATION

Chester V. Wells
National Renewable Energy Laboratory
1617 Cole Boulevard
Golden, Colorado 80401-3393

INTRODUCTION

Interest in Measurement Uncertainty Analysis has grown in the past several years as it has spread to new fields of application, and research and development of uncertainty methodologies have continued. This paper discusses the subject from the perspectives of both research and calibration environments. It presents a history of the development and an overview of the principles of uncertainty analysis embodied in the United States National Standard, ANSI/ASME PTC 19.1-1985, *Measurement Uncertainty*. Examples are presented in which uncertainty analysis was utilized or is needed to gain further knowledge of a particular measurement process and to characterize final results. *Measurement uncertainty analysis provides a quantitative estimate of the interval about a measured value or an experiment result within which the true value of that quantity is expected to lie.*

Years ago, Harry Ku of the United States National Bureau of Standards (now called the National Institute of Standards and Technology/NIST) stated that "The informational content of the statement of uncertainty determines, to a large extent, the worth of the calibrated value." [1] Today, that statement is just as true about calibration or research results as it was in 1968 when he wrote it.

Why is that true? What kind of information should we include in a statement of uncertainty accompanying a calibrated value? How and where do we get the information to include in an uncertainty statement? How should we interpret and use measurement uncertainty information? Where can we get more information on the subject?

This discussion will provide answers to these and other questions about uncertainty in research and in calibration. The methodology to be described has been developed by national and international groups over the past nearly thirty years, and individuals were publishing information even earlier. Yet the work is largely unknown in many science and engineering arenas. I will illustrate various aspects of uncertainty analysis with some examples drawn from the radiometry measurement and calibration discipline from research activities at the National Renewable Energy Laboratory, in Golden, Colorado (formerly called the Solar Energy Research Institute).

Here are some of the benefits of performing uncertainty analysis:

- Increases credibility of research or calibration results

- Increases validity of comparisons of research or calibration results from different labs or processes

- Improves design of experiments and measurement or calibration schemes (as a result of the pre-test analysis)

- Rigorously validates the final results (through the post-test analysis)

- Formalizes the method for incorporating uncertainty information from one measurement or experiment into succeeding results, for formal measurement and calibration traceability

- Helps meet documentation requirements showing compliance with laws and regulations, such as meeting quality requirements for laboratory test or measurement results

- Provides a reliable tool for quality assurance of measurements and research results in the R&D environment.

We encounter data of all types and quality. Data is truly *Valid Data* if it has known and documented paths of origin, including associated theory; measurements and their traceability to measurement reference standards; computations; and, uncertainty analysis of the results. Data with less than this may have questionable validity.

HISTORICAL BACKGROUND

The concept of measurement error and error analysis has long been employed by individuals making measurements in many fields. However, measurement <u>uncertainty</u>, as a quantifiable attribute of a measurement process or result is a relatively new concept.

Tracing the history of the development of one standard will help us better understand the process of measurement uncertainty analysis and the standards. I have drawn heavily on the paper by R.B. Abernethy and B. Ringhiser [2] and courses taught by, and discussions with, Robert Abernethy.

S.J. Kline and F.A. McClintock's hallmark paper, "Describing Uncertainties in Single-Sample Experiments" [3], was published in 1953 and has been used by mechanical engineers. In the 1950s, some committees of the American Society of Mechanical Engineers (ASME) began trying to write a standard for measurement uncertainty analysis. In 1961, the National Aeronautics and Space Administration and the U.S. Air Force asked for quarterly reports on measurement errors in rocket-engine testing [4]. In 1965, the Interagency Chemical Rocket Propulsion Group (ICRPG) organized a committee to develop a measurement uncertainty standard for the rocket

engine industry. A survey taken in preparation for developing this standard showed that, among 24 laboratories and contractors involved in rocket propulsion, there were 39 different methods in use [4]. It took two years to prepare a draft for the first uncertainty standard handbook, and that handbook was rejected by the ICRPG Committee. A second contract was awarded, and that attempt produced a document that was accepted [5]. This second effort was largely successful because of the involvement and strong support from staff at the National Bureau of Standards (NBS), and because of the use of Monte Carlo simulations of alternatives, which provided objective comparisons of competing methodologies [2].

These comparisons led to proposed recommendations sent to the NBS team and the ICRPG Committee, which resulted in adoption of solutions to the first four of the five problems faced in developing the standard methodology. The solution to the fifth problem, agreement on reporting the uncertainty interval, didn't come until much later [2]. This "great compromise" will be discussed later in some detail.

The five problems addressed were

- evaluation of random error uncertainty

- evaluation of systematic error uncertainty

- use of signed bias limits

- the defined measurement process

- uncertainty intervals and models (which needed the "great compromise").

The *ICRPG Handbook for Estimating the Uncertainty in Measurements Made with Liquid Propellant Rocket Engine Systems* [5] was widely accepted, and the rocket uncertainty methodology was then applied successfully to gas turbines (jet engines), resulting in a contract to produce a similar handbook for testing jet engines. That 1973 publication, *Handbook, Uncertainty in Gas Turbine Measurements*, [6], was reproduced by the Instrument Society of America in 1980 with slight revisions as the ISA *Measurement Uncertainty Handbook* [7] and was widely distributed and used for ISA Short Courses by Bob Abernethy and Ron Dieck.

Abernethy and Ringhiser [2] also identified three areas in which future research work was needed and still continues: (1) curve fitting and the effects on uncertainty arising from errors in both the independent and dependent variables; (2) weighting competitive answers on the basis of calculated uncertainties; and (3) outlier detection and rejection methods.

The uncertainty analysis methods documented in the above resources have served the engineering field well for nearly two decades, and still do today. The utilization of this methodology by nine organizations was documented by Abernethy and R.P. Benedict [8]. We have used this methodology in a variety of calibration and research activities at NREL. In 1985, the ASME

3

Measurement Uncertainty Standard, based on this methodology (including the "great compromise"), was approved and adopted by the American National Standards Institute as a formal American National Standard [9]. It is also the basis for the International Organization for Standardization draft international standard, ISO/DIS 5168 *Fluid Flow Measurement Uncertainty* [10]. The 1987 draft has been approved by the nations voting on it and will be issued with added comments and information resulting from the balloting. The final translations into French and Russian are being prepared now. As soon as those translations are completed, all versions will be published [11].

The above activities and applications of uncertainty analysis are best characterized as being in the engineering and test environment, as distinguished from calibration and metrology. Yet they are all various forms of "physics experiments" involving physical measurements and calculation of final results, and users want to know how good those results really are.

THE TRUE VALUE

The "true value" is an important concept in experimental work. As R.J. Moffat points out so well, "...almost all situations where uncertainty analysis seems to fail arise because too little attention has been paid to identifying the intended true value of the measurement. Such an identification is always the first step in an uncertainty analysis, and lack of precision in that identification almost always causes serious trouble." [12] Deriving the statement of the true value requires careful painstaking work; it is often not easy to define, but it should be done in writing.

By *true value* we mean the actual value sought for the intended quality that existed at the intended time of the measurements. The measured sought may be a directly observable phenomena (e.g., the temperature of the liquid at a particular location in a storage tank) or one which exists only as a concept (e.g., the mean bulk liquid temperature in the storage tank).

The exact true valuable is unknowable—we won't even know when our result is exactly on the true value! It is unknowable because of the limitations (errors) of our measurement systems, the perturbing of the quantities to be measured by the presence of measurement sensors, an inadequate understanding of the physics involved, and often by a disparity between the definition of the true value sought and the end-use equations and design of the experiment implemented.

A common source of problems in experiments and uncertainty analysis is that the end-use equations are written assuming one definition of the true value, but the data are taken assuming or representing a different definition. If the physics of the measurement situation is adequately addressed, then various sources can be identified with their corrections and uncertainities, so that the true value sought can be approached more closely and the uncertainties identified more completely. A careful pre-test uncertainty analysis will help reduce these sources of uncertainty.

THE UNCERTAINTY INTERVAL AND THE TRUE VALUE

The goal of uncertainty analysis is to achieve a good quantitative <u>estimate</u> of the interval on either side of a measured value or experimental result within which the true value is expected (with a given confidence) to lie. The uncertainty interval is determined through careful statistical **and** engineering analysis work. But the uncertainty interval derived through this analysis method applies only if the statement of the true value is appropriate to the measurement and experimental processes performed and analyzed.

THE NATURE OF EXPERIMENTAL ERRORS

Whether the measurement process is the simple measurement of a single quantity or the final result of a complex experiment, the errors associated with a measurement process are commonly divided into three categories: random, systematic, and gross errors. Error is sometimes defined as the difference between the value of a measurement result and a given physical standard. But in the more general case, errors should be defined as the measurement or experiment result minus the true value sought. Since we cannot know the true value with complete certainty, our ability to determine errors is limited; we cannot know the error exactly, so there is uncertainty associated with knowing it, and an *estimate* of the actual error must be made using uncertainty analysis instead.

Even though we all may be familiar with these three categories of errors, I want to review them briefly and make some comments about them that are vital to our discussion. These comments are also important for our understanding of uncertainty. This will help us all to use the same nomenclature.

Random or precision errors are the ones most often addressed in reports of measurement results. This is likely because random errors are the ones that produce "scatter" in the final result, so we have to say something about them! Also, we can deal more rigorously with random errors because of the great body of knowledge about the statistical techniques and the ease (and enjoyment?!) with which those techniques can be applied using today's computers or even our pocket scientific calculators. In speaking of random errors producing scatter in the final result, we must make it clear that we are not referring to varying measurement results caused by variations in the quantity being measured. There is a method of dealing with that situation also [8, p. 5].

It is important to recognize that, for most of us, this was where our school lab experiences began, and we may have been led to believe that all there was to error analysis was average and standard deviation. Our statistical techniques deal most easily with random errors that are truly random, i.e., have a Gaussian or "normal" distribution. We often assume that these random or precision errors are normally distributed, yet we never confirm this, so we go on blindly applying statistical techniques that are valid only for normal distributions.

All of us are probably familiar with the concept that, for a truly normal distribution of randomly distributed measurements, approximately 68% of all the measurements fall within ±1σ (±1 standard deviation) of the mean, 95% are within ±2σ of the mean, and 99.7% of the measurements are within ±3σ of the mean of the individual measurements.

But, all this is true only for large amounts of data that are truly normally distributed. If we compute the standard deviation from a more limited sample of measurements (fewer than 30) out of a much larger population, then ±2 standard deviations is not likely to correctly predict the boundaries within which 95% of the measurements in the larger or total population would lie. To adequately infer the 95% area from a small sample, we must use a multiplier larger than 2 times the standard deviation. This multiplier is given by the two-tailed "student's *t*" distribution for the 95% confidence level. Refer to the discussions and tables in references 6, 7, or 9 for details, because most statistics texts provide little or no relevant information on "student's *t*" for this application.

Systematic or bias errors produce no scatter in the final result, if they are truly fixed, unchanging, and systematic. Systematic errors are often overlooked or ignored, I believe, for several important reasons. First, because they are unchanging, they escape our attention. Second, our teachers may never have discussed them, or else they attributed them to mistakes, et cetera, or stated simply that they "are calibrated out." In his fairly well-known book, Professor P.R. Bevington is possibly typical of many teachers and writers on the subject: *"Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique."* "Throughout this book we will ignore the problem of systematic errors and concentrate on the random errors resulting from uncertainties in our results." [13] These quotations call to our attention several problems.

A portion of our educational system has impressed upon several generations of students that systematic errors are due to mistakes or failed (or uncalibrated) equipment; therefore, repair and/or calibration of the equipment will eliminate entirely the systematic errors. Today, in our technical society, we have to address this issue of misinformed graduates and show them the truth in the matter—systematic errors are frequently larger, even much larger, than the random errors—they cannot be ignored. They are not necessarily due to faulty equipment or calibrations or techniques, though indeed they can be.

At even the highest national levels of measurement standards maintained by NIST, systematic errors are present (so far as I have seen) in every measurement standard and in every calibration they perform. In a path of measurement traceability from NIST, the systematic error component of uncertainty grows at least some at every step in the measurement chain. This is not because of faulty equipment or calibrations or techniques; it is simply because of the limitation of our knowledge, instrumentation, and technique, and should not be "credited" to someone as a mistake. Replacing a 5-½ digit digital multimeter in the system with an 8-½ digit DMM may reduce but it won't entirely eliminate the systematic error. And, there are many situations in which random errors at one step in the measurement chain become fixed ("fossilized") and add to the systematic uncertainty [2]. This occurs frequently in the calibration process.

Systematic errors produce no scatter in the final result; therefore, they are not detectable through statistical techniques, such as determining the mean and standard deviation. Since the systematic error is truly constant, no change in the mean or standard deviation will be detected, so we might be led to presume that no systematic errors exist. Most practitioners today believe that the usual statistical analytical techniques are not appropriate for handling systematic errors. To adequately estimate these errors requires real engineering knowledge and judgement based on experience. It becomes obvious, then, that it is more difficult to teach inexperienced students to recognize and estimate systematic errors than to teach them to recognize and characterize random errors. Yet, we must find effective means of teaching our students about systematic errors.

The term "Bias Limit" (**B**) is used to indicate the estimate of the maximum possible bias or systematic error that may exist in a measurement process or experiment result. If properly evaluated, the bias error, β, will lie between the two bias limits:

$$-\mathbf{B} \leq \beta \leq +\mathbf{B}$$

Arriving at proper bias limits is where practical experience and engineering judgment are absolutely necessary.

Table 1 identifies five types of systematic or bias errors. There are those that can and should be "calibrated out"—type (1). That is the purpose of calibration, to remove or correct for systematic differences that we can identify, measure, and remove. But we should expect to encounter some random error component in that calibration process, as well as have some small residual systematic error of unknown magnitude and unknown (4) or known sign (5) that will contribute to the bias limit. Large systematic errors of unknown magnitude—type (3)—are assumed to have been eliminated because they are the gross errors, which is the last category to be discussed.

The small systematic errors (types 2, 4, 5) are to be identified and accounted for in the uncertainty analysis, as appropriate. If they are truly negligible, they should still be listed with the elemental errors, but with their uncertainty values given as "negligible."

Gross or spurious errors are large errors that are <u>assumed to be zero for this analysis</u>. That is, it is assumed they have all been found and eliminated. Such errors cannot be incorporated into any statistical or engineering analysis. If gross errors do exist, the measurements will have to be discarded, for the results would be in error by an undetermined amount, and these analysis techniques would not produce a correct estimate of the uncertainty interval. Some common sources of gross errors include human error (transposing digits while copying data); operator error (placing an instrument on a DC range rather than AC); equipment failure (internal reference power supply has changed output to some value outside its normal specification, yet the instrument may appear to be operating normally); outside interferences (electromagnetic interference, air pockets in tubing for a water line to a manometer); and, transducer installation errors (axis of linear position transducer is placed at some angle, Θ, to the direction of motion of the object for which the transducer is to indicate the position, leading to unknown and possibly

large biases). Peter Stein addresses this issue very deliberately in his Measurement Systems Engineering courses, listed in the Resources section at the end of this paper.

FIVE TYPES OF SYSTEMATIC ERRORS			
Sign and Magnitude	**Known Sign and Magnitude**	**Unknown Magnitude**	
Large	(1) Calibrated Out	(3) Assumed to be Eliminated	
Small	(2) Small or Negligible; Contributes to Bias Limit	(4) Unknown Sign	(5) Known Sign
		Contribute to Bias Limit	

Table 1.   Types of Systematic or Bias Errors [from references 6 or 7, p. 2]

Good measurement uncertainty analysis may include using "outlier" detection and rejection techniques to help identify and remove gross errors. Measurement results that appear to be outliers should not be discarded unless there is an independent technical reason to believe that they are truly spurious data. Data should be discarded only after careful consideration, for the outliers may be the best (or only!) indication we receive that something is truly wrong in the measurement or experimental processes [6 or 7, pp. 163-166; 9, pp. 23-25]. A fixed gross error will produce a grossly biased result rather than a wildly outlying changing result, so the common outlier identification techniques may fail to detect it.

THE MEASUREMENT UNCERTAINTY MODEL

The following sections will present only a brief overview of the methodology of uncertainty analysis that forms the basis for the ANSI/ASME PTC 19.1-185 standard [9]. For more detailed discussions of the basic methodology, refer especially to references 6 or 7, 9, and 10.

The uncertainty of a measurement is a function of the specific measurement process used to obtain the measurement result, whether it is a simple or a complex process. Measurement uncertainty analysis provides an estimate of the largest error that may reasonably be expected for that specific measurement process. If the measurement process is changed, then the uncertainty analysis must be re-examined, and changed as appropriate. Errors larger than the stated uncertainty should rarely occur in actual laboratory or field measurements, if the uncertainty analysis has been performed correctly. The steps involved in "correctly performing the uncertainty analysis" will be discussed in the next section.

8

As stated at the beginning of this paper, *measurement uncertainty analysis derives a quantitative estimate of the interval about a measured value or an experiment result within which the true value of that quantity is expected to lie.* That uncertainty interval is derived from information about the random and systematic errors associated with the specific measurement process. The two methods, or "models," used to compute the uncertainty interval have been designed for 99% and 95% "coverage." That is, if the measurement or experiment is performed many, many times, the true value sought for the result will be contained within the stated uncertainty interval about the result 99 times out of a 100 repetitions, or 95 times out of a 100 repetitions. Or, to say it another way, the true value will lie outside the uncertainty interval only 1 time out of 100 repetitions or 5 times out of 100 repetitions (that is, 1 time out of 20 repetitions), depending upon which uncertainty model is used. Those uncertainty models are labeled U_{99} and U_{95}, respectively. The differences between the two models lies only in how the total random and systematic errors are combined to derive the uncertainty interval.

In U_{99}—also called U_{ADD}—the two components (random and systematic) are added linearly

$$U_{99} = B + (2 \cdot S) \text{ or } U_{99} = B + (t_{95} \cdot S)$$

where B is the total bias limit, and S is the standard deviation in the final result, and t_{95} is "student's t."

In U_{95}—also called U_{RSS}—the two components are added in quadrature; that is, the square root of the sum of squares (RSS) of the two components is used to combine them.

$$U_{95} = \sqrt{B^2 + (2 \cdot S)^2} \text{ or } U_{95} = \sqrt{B^2 + (t_{95} \cdot S)^2}$$

U_{99} is a more conservative estimate of the uncertainty interval and will always be larger than U_{95}.

The individual systematic error components are added in quadrature ("RSSed") to obtain the systematic uncertainty component, the total bias limit, B. For the systematic errors, the uncertainty is

$$B = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial F(X_1, X_2, \ldots X_n)}{\partial X_i} \cdot B_i \right)^2}$$

where B_i is the Bias Limit for variable i, etc.

The random component of uncertainty for an individual measurement is taken as twice the standard deviation, $\pm 2S$ (or $\pm t_{95}S$ for small samples of data, 30 or fewer measurements). This corresponds to 95% of the area under the normal distribution of random measurements. S is the

standard deviation of the individual measurements, and t_{95} is the two-tailed "student's t" value for a 95% confidence level. The individual random error components are added in quadrature to develop the total random uncertainty component. For the random errors, the uncertainty is:

$$S = t_{95} \sqrt{\sum_{i=1}^{n} \left(\frac{\partial F(X_1, X_2, ..X_n)}{\partial X_i} \cdot S_i \right)^2} \; ; \; S_i = \frac{s_i}{\sqrt{N_i}}$$

where t_{95} is "student's t" for 95% confidence limits (use 2 if the number of measurements averaged is 30 or more); F is the function equation from which the final result result is computed; and the Xs are the independent variables. If the value of X_i is the group mean of M groups of N measurements, S_i is the standard deviation of the means of the M groups of measurements of the first variable X_i, and s_i is the standard deviation of the individual measurements that form the mean of a group of N measurements.

The degrees of freedom, df, in the final result arising from the various random errors sources is computed using the Welch-Satterthwaite equation [Ref 9, pp. 11-12]. If the degrees of freedom are more than 30, then use 2.0 instead of t_{95}. See the references.

FUNDAMENTALS OF HOW TO DO UNCERTAINTY ANALYSIS

We will now outline the steps for performing a measurement uncertainty analysis. For the details, consult references 6 (or 7), 8, 9, and 10. The ANSI/ASME PTC 19.1 [9] is the U.S. national standard we use at NREL for uncertainty analysis.

Step 1: Clearly define the "true value" sought in writing. It is well worth the time to do this in writing, for it will help clarify the measurement process and the experiment goal [12].

Step 2: Define the measurement process utilizing the statement of the true value and the research or calibration objectives. List all of the independent physical parameters to be measured and their nominal values or ranges. List all of the instruments and setups and their calibrations and characterizations that will be used to measure each parameter. Write the equations that define the functional relationships between the independent physical parameters (with their measurements) and the final result.

Step 3: List every possible elemental source of measurement error that you can think of, no matter what the source or how large or small the error may be thought to be. An excellent method for listing elemental error sources is illustrated in Table 5, with the associated discussion in reference 10.

Step 4: Group the errors into these three categories, by source: (1) calibration errors; (2) installation and data acquisition errors; and, (3) data reduction errors. Calibration errors are those associated with the calibration of each measuring instrument, sensor, transducer, etc. Installation

10

errors are those errors that arise from how and where the sensors are installed for the experiment. Be particularly alert here for systematic errors. Data acquisition errors are those associated with the performance of the data acquisition system (including sensors) in the environment in which they are used. Use manufacturer's specifications if you have no better data, and you have reason to believe you can trust those specifications. Data reduction errors are errors associated with the computer's algorithms and numerical handling routines, round-off errors, errors encountered in curve-fitting resulting experiment data, including calibration curves, errors arising through the limitations of statistical analysis software, etc. I prefer to assign calibration curve fitting errors to the calibration error category, not to the data reduction category. This permits me to know how large the total calibration error is, and to deal with it as necessary.

Step 5: Classify the errors into systematic and random errors. If data exist from which a standard deviation can be calculated or the random error estimated, consider it a random error. Random errors produce scatter in the final result. Otherwise, consider the errors to be systematic errors; systematic errors do not change for a given instrument and measurement process. A close evaluation of manufacturer's specifications can help in this step.

Step 6: Calculate the systematic and random errors for each physical parameter. Sometimes this information is obtained from previous tests, calibrations, or experiments.

Step 7: Separately propagate the two types of error to the final result. Use the Taylor series or small deltas ("dithering") to determine the sensitivity of the final result to each individual source of error. Simply adding the errors may lead to an uncertainty estimate that is too large or too small, depending on the sensitivity coefficients for each error. This is discussed in detail in the references.

An important caution: be careful not to mix percentage and absolute errors (percent added to watts/meter^2, for example).

Step 8: Calculate the Uncertainty Interval using either model (or both): $\mathbf{U_{99}}$ or $\mathbf{U_{95}}$.

Step 9: Use pre-test and post-test analyses. The use of both tests reduces the cost and risk of performing useless experiments, publishing invalid data, drawing wrong conclusions, or making wrong decisions.

Use the pre-test analysis to predict the uncertainty before an experiment or test is run. This can determine the appropriateness of measurement instruments and techniques before the investment is made to actually procure the equipment and run the proposed experiment. If the predicted uncertainty is not small enough to obtain the needed result, redesign the experiment—don't go on!

Use a post-test analysis to examine the final results for validity, problems, and the necessity of repeating the experiment. The uncertainty information for the final report is obtained in the post-test analysis.

<u>Step 10</u>: In the final report, show separately the final random error component of the uncertainty together with the associated degrees of freedom, the bias limit, and the uncertainty model used (either U_{99} or U_{95}, or both).

SOME EXAMPLES OF UNCERTAINTY ANALYSIS

To illustrate how uncertainty analysis has been or can be used, examples will be drawn from radiometry and optics-related measurement applications in the solar energy field, including photovoltaics (PV), solar irradiance measurements (both spectral and broadband) in the visible and infrared regions, and high-solar-flux measurements.

In a paper believed to be the first complete uncertainty analysis of PV reference cell efficiency calibration measurements [14], K.A. Emery, C.R. Osterwald, and I tried to include all sources of systematic and random errors associated with that measurement process. A total uncertainty of ±1.0% was shown to be possible *if all sources of systematic error were minimized.* A more typical value of ±7% was shown to be expected if some common sources of error were not minimized. In a typical case, the systematic errors were found to be five times as large as the random error. This demonstrates that using only the repeatability (standard deviation) to estimate the uncertainty in an efficiency measurement would greatly underestimate the actual error.

That paper also discussed a hypothetical example that demonstrates the value of knowing and minimizing the uncertainty of a measurement process. The example concerned measuring the improvements in PV device performance, during which the magnitudes of the improvements were assumed to decrease during the development of the device. After six improvement steps, a ±5% uncertainty measurement process is unable to confidently resolve smaller improvements, whereas the ±1% measurement process can resolve the magnitude of the progress through nine improvements.

This ability to resolve small changes is important, whether the goal is to measure research-device improvements, degradation, or stability. Uncertainty analysis gives us the insight to determine what size changes we should expect to confidently resolve and what changes "will be lost in the noise" of the measurement process. This is especially important information if we are measuring experimental devices that are degraded by the exposure to light during the measurement process.

Emery and others at SERI (now NREL) examined the uncertainties of a variety of methods used to calibrate PV reference cells [15]. Many sources of error in the various primary and secondary calibration methods were identified and discussed. The total uncertainty for the various methods was developed, allowing a comparison with one another and with AM0 calibrations. It was noted that systematic errors were larger, by 15% to 400%, than the random errors in all methods except one. Again, true systematic errors cannot be reduced by repeated measurements, and they are not related to the standard deviation of a calibration data set. Table 2 shows the summary table from that paper.

12

PV REFERENCE CELL CALIBRATION UNCERTAINTIES				
Calibration Method	E_{tot}	Bias (±%)	Random (±%)	Total (±%)
Global fixed tilt	pyranometer	3.1	1.5	4.3
Global fixed tilt	direct + diffuse	2.3	1.2	3.2
Global-normal	pyranometer	2.5	1.5	3.7
Global-normal	direct + diffuse	0.8	1.2	2.5
Direct-normal	cavity radiometer	0.5	0.3	0.7
X25 simulator	±1% reference cell	1.0	0.2	1.1
SPI-SUN simulator	±1% reference cell	1.4	1.2	3.0
AM0 (JPL balloon calibration)†		0.5	0.2	0.7
AM0 (NASA airplane calibration)†		1.0	--	1.0
† A standard uncertainty analysis has not been performed for AM0 calibration methods; these are estimates based on published information.				

Table 2.   Summary of PV reference cell calibration uncertainties, from reference [15].

Table 2 clearly shows that the best method to use to calibrate PV reference cells from the ground is against an absolute cavity radiometer in the direct-normal mode, and that the best method to transfer that calibration to another reference cell is to use a model X25 solar simulator. The other methods do not rival these, except those made outside the atmosphere, for which a rigorous uncertainty analysis was not performed in the manner we are describing in this paper.

Daryl R. Myers examined the uncertainties encountered in acquiring field data included in the SERI solar spectral data base [16]. He listed and evaluated 21 major sources of uncertainty, identifying the systematic and random errors and total uncertainty for the calibration process and the field measurements with the spectroradiometers. The systematic and random uncertainties for the spectroradiometer itself, for the calibration process and for outdoor field measurements, were all tabulated and plotted as a function of wavelength. Many measurement processes have uncertainties that are a function of one or more variables, such as wavelength, amplitude, frequency, range, etc. It is valuable to shown such dependencies.

One of the key sources of measurement uncertainty we face in the field of solar energy is in the measurement of broadband (from about 300 to 3,000 nm) global solar irradiance (measurement

of irradiance over a 2π solid angle), either on a horizontal plane (called the global horizontal solar irradiance) or on a tilted plane, such as normal to the sun (called global normal). The instrument used for such global measurements is the pyranometer. Although simple in principle, all pryanometers have some performance characteristics that are difficult to measure and correct for. Myers examined the errors associated with calibrating these instruments and gave us a starting point in reporting uncertainties for the calibration process [17] in a manner similar to the method of the ANSI/ASME standard.

At NREL, our usual method for calibrating pyranometers is the component summation technique: the reference value of the global irradiance (whether on a horizontal or tilted surface) is measured as the sum of the direct beam from the sun and the diffuse radiation falling on the pyranometer from the sky and other sources scattering onto the pyranometer's sensor. In our Broadband Outdoor Radiometer CALibrations (BORCALs), we have identified these sources of error: the error in the direct-beam-irradiance measurement using an absolute cavity solar radiometer; the error in the diffuse-irradiance measurement using a solar tracking disk shading a pyranometer; departure of the response of the pyranometer from an ideal angular response (in both cosine and azimuth); the computation of the zenith angle; change in responsivity to temperature; generation of an output signal related to rate of change of temperature; linearity; leveling of the sensor; spectral response; time constants of the sensor response and the data acquisition system sampling rate; data acquisition system voltage measurement errors; thermal emfs (electromotive forces) and electromagnetic interference in the measuring circuit; and, the random error in measuring the calibration factor (CF) some 100 to 1,000 times during the outdoor calibration event. We include a formal discussion of the calibration process and error sources with the calibration report. The uncertainty in the calibration factor for each pryanometer is reported. The U_{95} uncertainty reported for the CF value for a typical pyranometer is most commonly in the range from ±3.2% to ±4%, with an occasional maximum near 8%.

In the 1978 to 1980 time frame, flat-plate solar collector efficiency tests performed at U.S. labs showed differences of about 3% in efficiency results among test laboratories. Similar differences were showing up among European laboratories and between the U.S. and European laboratories. The measurement process for an efficiency test included the measurement of the solar irradiance on the tilted surface and the measurement of the fluid-flow rate and the temperature rise (ΔT) in the fluid stream as it went through the collector. After examining the flow and temperature measurements, it was concluded that the cause must be due to the calibration and performance of the pyranometers. So, in March 1980, the International Energy Agency sponsored a comparison of 22 reference pyranometers from the various laboratories; it was held at the World Radiation Center (WRC), Davos, Switzerland. The result was that the sixteen "1% - WMO Class 1 Pyranometers" showed a 10.88% spread in measured irradiance, with the mean value offset by -6% from the value measured by the reference pyranometer from WRC [18]. Recalibration of that reference radiometer changed its CF by about 3%. Since then, a number of us have been working to better characterize and calibrate pyranometers, and we now approach ±3% uncertainty, as described above.

Uncertainties in solar infrared measurements are of growing interest. We are experiencing both optimism and pessimisism, for reasons you will see. Elsworth Dutton, of the Climate Monitoring and Diagnostics Laboratory (National Oceanic and Atmospheric Administration (NOAA) laboratories in Boulder, CO), performed two calibrations of each of 20 pyrgeometers and compared the results with factory calibrations of the same instruments. His results showed a net difference of 0.07% from the factory, with a standard deviation of 1.57% in the difference [19]: quite encouraging! In another case, a single pyrgeometer was calibrated 11 times, using a variety of techniques at several labs around the world. The result was a 5.5% standard deviation, with a range of 18.4% variation in the calibration factor for the one instrument. See Table 3 [20]. Such results encourage us to apply the method of ANSI/ASME PTC 19.1 to this area as well.

PYRGEOMETER CALIBRATION COMPARISON (PIR #26181F3)		
DATE	CALIBRATING ORGANIZATION	CALIBRATION CONSTANT ($\mu V/W/m^2$)
Spring 1986 Feb. 1987	Eppley (original purchase) Eppley	3.90 3.77
Aug. 1987	NARC (black body) NARC (sphere, cooling)	3.93 4.14
Nov. 1989	NARC (black body) NARC (sphere, cooling)	3.84 4.19
Mar. 1990	MOH (warm black body) MOH (ice block) MOH (warm black body + intermediate dome + ventilation)	4.46 4.16 4.19
Nov. 1991	BSRN (CSU BB tech at NOAA/Boulder) BSRN (ice dome, FIRE II, Coffeyville, KS)	3.72 4.02
Average CF = 4.029 (n = 11); Std. Dev. in CF = 0.222, which is 5.5% of mean CF; Range in CF values = 0.74, which is 18.4% of mean CF.		

Table 3.   Comparison of Calibration Factor (CF) for one pyrgeometer calibrated 11 times using 8 different techniques in 6 different lab settings [20].

Carl E. Bingham reported the uncertainties in making irradiance measurements at high solar fluxes in an award-winning paper [21]. He prepared an exhaustive list and examined, in great detail, the uncertainties for two different systems that measure the solar flux at levels of about 250 W/cm^2 and 2kW/cm^2, the equivalent of 2,500 to 20,000 suns. Bingham followed the ANSI/ASME standard closely.

SUMMARY:

Measurement uncertainty analysis yields a quantitative estimate of the interval about a measurement or final experimental result within which the true value of the measured quantity is expected to lie. The methodology has been carefully researched, and is documented in both a U.S. National Standard and an international standard. It is suitable for both simple measurement and complex experimental applications. The method utilizes our best knowledge in both statistics and engineering. By following the recommendations, uncertainty information provided with a final result can be carried forward with rigor from one user to the next.

The ANSI/ASME standard, PTC 19.1-1985, is a good starting point for performing uncertainty analysis in a variety of calibration and research applications, even though it is being revised and updated.

Even though we may have to begin an uncertainty analysis with less than complete information, these encouraging words from Charles Babbage are most applicable: "Errors using inadequate data are much less than those using no data."

RESOURCES CONCERNING UNCERTAINTY ANALYSIS:

There are several resources available from which to learn the details of measurement uncertainty analysis. Of course, there is no substitute for actual practice doing it. The books and reports cited in the References are a good starting place. The fastest way to learn the methodology embodied in the U.S. American National Standard, ANSI/ASME PTAC 19.1-1985, is to take one of the courses taught by Ron Dieck or Robert Abernethy. Additional resources are listed below, including references 22, 23, 24 and 25, which are not referenced in the body of this paper:

<u>Measurement Uncertainty and Related Courses:</u>

- "Measurement Uncertainty: Concepts and Applications" — Taught as a 2- or 3-day on-site course, and as a 2-day Instrument Society of America (ISA) course by Ronald H. Dieck; emphasizes systematic uncertainty and ANSI/ASME PTC 19.1. Contact: Ron Dieck Associates; 7 Dunbar Road, Palm Beach Gardens, FL 33418 (uses References 9 and 22, plus extensive notes and handouts).

- "Test Measurement Accuracy" — A measurement-uncertainty course with emphasis on systematic errors and ANSI/ASME PTC 19.1, taught by Robert Abernethy as a 2-day on-site course and for the American Society of Mechanical Engineers (ASME), ISA, and the University of Tennessee Space Institute. Contact: Robert B. Abernethy; 536 Oyster Road, North Palm Beach, FL 33408 (uses References 6, 9, and 10, plus extensive notes and handouts).

- "Measurement Uncertainty - Measurement Assurance" — A 5-day course taught by Rolf B. F. Schumacher; emphasizes random error and metrology applications; does not emphasize ANSI/ASME PTC 19.1-1985. Contact: Coast Quality Metrology Systems,

Inc.; 35 Vista del Ponto, San Clemente, CA 92672-3122 (uses extensive course notebook).

- "Measurement Uncertainty Training Course" — A introductory 4-day course taught by Al Catland; especially good for senior technicians; does not follow exactly ANSI/ASME PTC 19-1-1985. Contact: Measurement Technology Company; 12620 Avenida De Espuela, Poway, CA 92064-2535 (uses extensive course notebook).

- "Designing and Specifying Data Acquisition Systems" — A 2-day course, taught by James L. Taylor for ISA (Course #T330), that uses uncertainty goals as a design criteria. Contact: Instrument Society of America; 67 Alexander Drive, P.O. Box 12277, Research Triangle Park, NC 27709 (uses Reference 23).

- "Measurement Systems Engineering and Dynamics Courses" — Detailed engineering level on measurement systems, per se; highly insightful in means of avoiding gross errors in measurement systems; two consecutive 5-day courses organized by Peter Stein; generally taught in March each year; also given in on-site versions. Contact: Stein Engineering Services, Inc.; 5602 East Monte Rosa, Phoenix, AZ 85018-4646 (uses extensive course notebook and handouts. Reference 24 contains a portion of Stein's material).

Technical Paper Sessions:

The following meetings almost always have at least one technical session devoted to uncertainty analysis, plus possible tutorials or short courses on the subject.

- ISA International Instrumentation Symposium (April or May each year)

- Measurement Science Conference (January or February each year)

- National Conference of Standards Laboratories Annual Workshop & Symposium (July or August each year)

- ISA Annual General Conference (October each year)

REFERENCES:

[1]   Ku, Harry H. "Expressions of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value." *Measurements & Data*; July-August 1968; Reprinted with corrections, November 1968, p. 72. [Reprinted in *Precision Measurement and Calibration,* NBS Special Publication 300, Vol. 1: "Statistical Concepts and Procedures;" Harry H. Ku, Editor. National Bureau of Standards, February 1969; p. 73].

[2]   Abernethy, R.B. and Ringhiser, B. "The History and Statistical Development of the New ASME-SAE-AIAA-ISO Measurement Uncertainty Methodology." Paper No. AIAA-85-1403. Presented at the AIAA/SAE/ASME/ASEE 21st Joint Propulsion Conference, Monterey, CA, 8-10 July 1985.

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

[4]   Abernethy, R.B.   Personal communication to Short Course, "Test Measurement Accuracy."   29th International Instrumentation Symposium.   Instrument Society of America.   Albuquerque, NM.   May 1983.   Verified 10 September 1992 by Robert Abernethy.

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

[6]   Abernethy, R.B. and Thompson, J.W, Jr.   *Handbook, Uncertainty in Gas Turbine Measurements.*   AEDC-TR-73-5.   Arnold Air Force Station, Tennessee: Arnold Engineering Development Center February 1973.   Available from the National Technical Information Service, Springfield, VA 22161.

[7]   Abernethy, R.B. and Thompson, J.W, Jr. *Measurement Uncertainty Handbook.* Revised 1980.   This is a reprint with minor revisions of AEDC-TR-73-5, reference [6] above, but is now out of print.   Instrument Society of America.   Research Triangle Park, NC.   1980.

[8]   Abernethy, R.B. and Benedict, R.P.   "Measurement Uncertainty: A Standard Methology." *ISA Transactions*; Vol. 24, No. 1 (1985), pp. 75-79.   Presented at the 30th International Instrumentation Symposium, Instrument Society of America, Denver, CO, May 1984.

[9]   ANSI/ASME. *Measurement Uncertainty.* Supplement to ASME Performance Test Codes, PTC 19.1-1985.   The American Society of Mechanical Engineers, New York.   1985.

[10]   ISO.   *Fluid Flow Measurement Uncertainty.*   International Organization for Standardization, Technical Committee TC30 SC9.   Draft Revision of ISO/DIS 5168.   May 1987.

[11]   Abernethy, R.B.   Personal communication to Chester Wells, 10 September 1992.

[12]   Moffat, Robert J.   "Identifying the True Value—The First Step in Uncertainty Analysis." *Proceedings of the 34th International Instrumentation Symposium.*   Paper No. 88-0729. Instrument Society of America, May 1988.

[13]   Bevington, Philip R.   *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill Book Company, New York, 1969; pp. 8, 6.

[14]   Emery, K.A., Osterwald, C.R., and Wells, C.V. "Uncertainty Analysis of Photovoltaic Efficiency Measurements." *Proceedings of the Nineteenth IEEE Photovoltaic Specialists Conference; May 4-8, 1987, New Orleans, LA.*

[15]   Emery, K.A., Osterwald, C.R., Rummel, S., Myers, D.R., Stoffel, T.L., and Waddington, D. "A Comparison of Photovoltaic Calibration Methods." *Proceedings of the 9th European Photovoltaic Solar Energy Conference; September 25-29, 1989, Freiburg, W. Germany.*

[16]   Myers, D.R. "Estimates of Uncertainty for Measured Spectra in the SERI Spectral Solar Radiation Data Base." *Solar Energy* Vol. 43, No. 6, 1989. pp. 347-353.

[17]   Myers, D.R. *Uncertainty Analysis for Thermopile Pyranometer and Pyrheliometer Calibrations Performed by SERI*, SERI/TR-215-3294. Solar Energy Research Institute (now National Renewable Energy Laboratory). April 1988.

[18]   Talarek, H.D. *Results of a Pyranometer Comparison; Davos, March 5 and 6, 1980.* International Energy Agency Program to Develop Solar Heating and Cooling Systems, Task III: Performance Testing of Solar Collectors. World Radiation Center, Davos, Switzerland. June 1980

[19]   Dutton, Ellsworth G. *An Extended Comparison Between LOWTRAN7-Computed and Observed Broadband Thermal Irradiances: Global Extreme and Intermediate Surface Conditions.* Accepted for the Journal of Atmospheric and Oceanic Technology. Expected to be published June 1993.

[20]   DeLuisi, John. February, 1992.

[21]   Bingham, Carl E. "Uncertainty of Calorimeter Measurements At NREL's High Flux Solar Furnace." Prepared for the ASME International Solar Energy Conference, Maui, HI, April 4-8, 1992. NREL/TP-254-4593, December 1991.

[22]   Dieck, Ronald H. *Measurement Uncertainty: Concepts and Applications.* Research Triangle Park, NC: Instrument Society of America, (to be published Oct. 1992).

[23]   Taylor, James L. *Computer-Based Data Acquisition Systems - Design Techniques.* 2nd Edition. Research Triangle Park, NC: Instrument Society of America, 1990.

[24]   Benedict, Robert P. *Fundamentals of Temperature, Pressure, and Flow Measurements.* 3rd Edition. New York: John Wiley & Sons, 1984. Chapter 10, "Uncertainties and Statistics."

[25]   Stein, Peter K. *The Unified Approach to the Engineering of Measurement System; Part I - Basic Concepts.* Stein Engineering Services, Inc. Phoenix, AZ. April 1992.