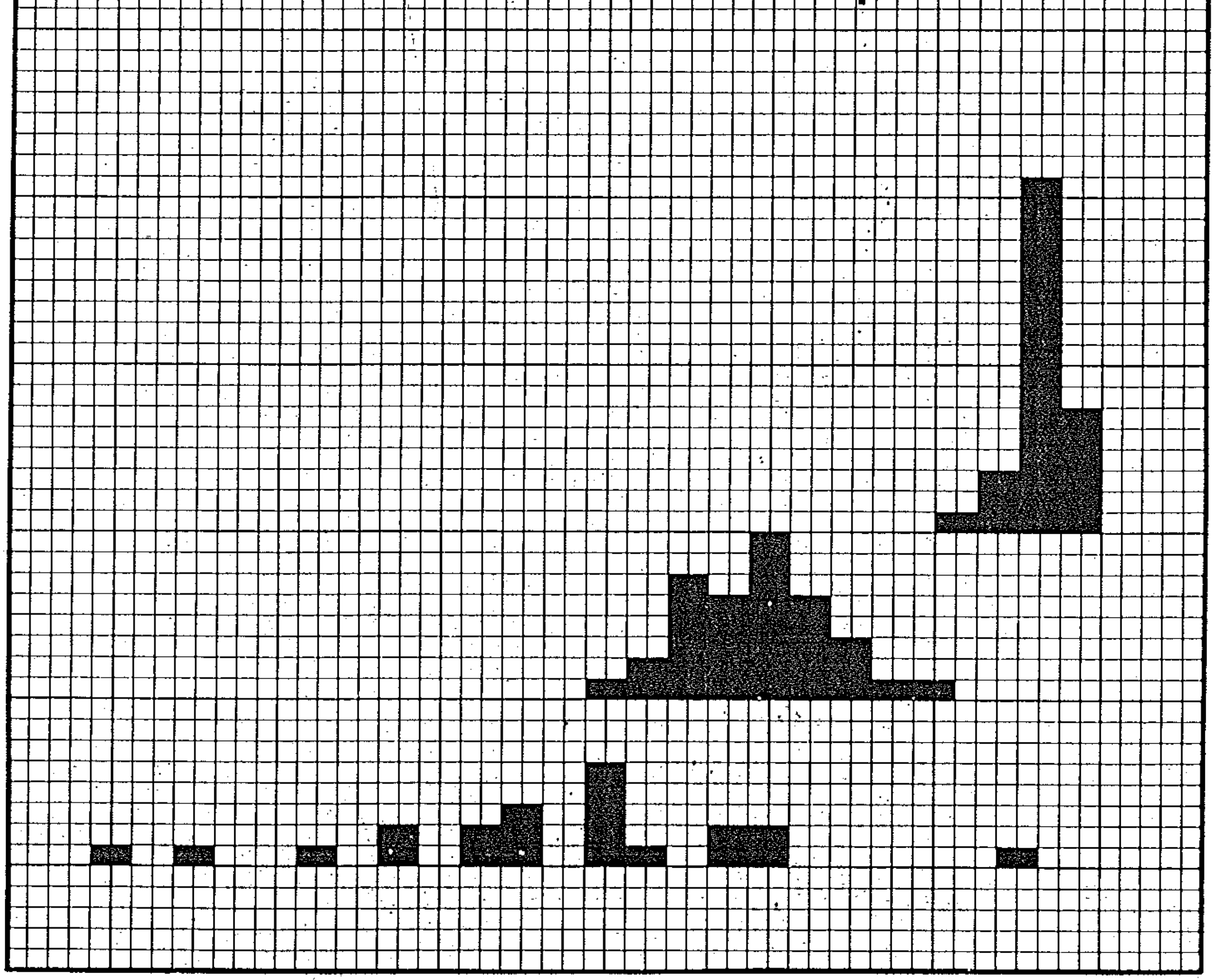
NBS SPECIAL PUBLICATION 644

U.S. DEPARTMENT OF COMMERCE/National Bureau of Standards

Expression of the Uncertainties of Final Measurement Results: Reprints



Expression of the Uncertainties of Final Measurement Results: Reprints

Churchill Eisenhart

Harry H. Ku

R. Collé

National Bureau of Standards
Washington, DC 20234

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, Secretary

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued January 1983

Over the intervening years since the publication of Eisenhart's and Ku's articles, it has become apparent that a few additional comments may be useful. It is equally apparent that a complete revision is neither necessary nor desirable inasmuch as the major thrust and content of the articles remain as valid and as appropriate as when first written. For this reason, these comments are made as a postscript.

Uncertainty Assessments Must Be Complete

The uncertainty of a reported value is meant to be a credible estimate of the likely limits to its actual *error*, i.e., the magnitude and sign of its deviation from the truth. As such, uncertainty statements must be based on as nearly complete an assessment as possible. This assessment process must consider every conceivable source of inaccuracy in the result.

A measurement process generally consists of a very complicated sequence of many individual unit operations or steps. Virtually every step in this sequence introduces a conceivable source of inaccuracy whose magnitude must be assessed. These sources include:

- Inherent stochastic variability of the measurement process;
- Uncertainties in standards and calibrated apparatus;
- Effects of environmental factors, such as variations in temperature, humidity, atmospheric pressure, and power supply voltage;
- Time-dependent instabilities due to gradual and subtle changes in standards or apparatus;
- Inability to realize physical model because of instrument limitations;
- Methodology procedural errors, such as incorrect logic, or misunderstanding what one is or should be doing;
- Uncertainties arising from interferences, impurities, inhomogeneity, inadequate resolution, incomplete discrimination, etc.;
- Metrologist errors, such as misreading of an instrument;
- Malfunctioning or damaged apparatus;
- Laboratory practice including handling techniques, cleanliness, etc.; and
- Computational uncertainties as well as errors in transcription of data, and other calculational or arithmetical mistakes.

This list should not be interpreted as exhaustive, but rather as illustrative of the most common generic sources of inaccuracy that may be present.

The various sources of inaccuracy are generally classified into sources of *imprecision* (random components) and sources of *bias* (fixed offsets). To which category a particular source should be properly assigned is often difficult and troublesome. In part, this is because many experimental procedures or individual steps in the overall measurement process embody both systematic and stochastic (random) elements. (For an alternative discussion that questions the need for a clear cut distinction between random and systematic components of uncertainty, see [7].) One practical approach is to classify the sources of inaccuracy according to how the uncertainty is estimated. In this way, sources of imprecision are considered to be those components which *can be* and *are* estimated by a statistical analysis of replicate determinations. For completeness, the *systematic uncertainty components* can be considered to be the residual set of conceivable sources of inaccuracy that are biased and not subject to random variability, and those that may be due to random causes but *cannot be* or *are not* assessed by statistical methods. The systematic category includes sources of inaccuracy other than biases in order to obtain a complete accounting of all sources of inaccuracy in the measurement process. Hence, it is meaningful to report a random uncertainty contribution, only if one has a computed statistic for the magnitude of its imprecision or random variation. Many sources of inaccuracy may exist consisting of several components from both the random and systematic categories and can be assessed only after consideration of the more fundamental processes involved. The uncertainty in the calibration of an instrument with a standard reference material, for example, would have not only components from the uncertainty in the standard itself, but also uncertainty components arising from the use of the standard in performing the calibration.

also applicable to measurement intercomparisons

Assessment of Imprecision (Random Uncertainties)

Although the treatment and expressions of reporting the imprecision of measurement results were adequately covered in the original article, a number of points are of sufficient importance to deserve reemphasis.

The only way to assess realistically the overall imprecision is to make direct—or preferably, when possible, indirect—replicate determinations [1] and calculate an appropriate statistic such as the standard error of the mean. It is extremely important to be definite on what constitutes a "replicate determination" because the extent to which conditions are allowed to vary freely over successive "repetitions" of the measurement process determines the scope of the statistical inferences that may be drawn from measurements obtained [2, sec. 4.1]. When measurements of a particular quantity made on a single occasion exhibit closer mutual agreement than measurements made on different occasions so that differences between occasions are indicated, the value of the computed standard error of the mean of all the measurements obtained by lumping all of the measurements together will underestimate the actual standard error of the mean. A more realistic value is given by taking the arithmetic means of the measurements obtained on the respective occasions as the *replicate determinations* and calculating the standard error of their mean in the usual way [3, sec. 3.5].

In many situations, it may not be possible or feasible because of time and cost constraints to perform a sufficient number of completely independent determinations of the measurement result. For results derived from several component quantities, the individual imprecision estimates must be propagated to obtain the imprecision of the final result. It must be emphasized, however, that these estimates of imprecision should not be based exclusively on the information derived from just the present measurements. Presently derived information should be added to the information accumulated in the past on the imprecision of the measurement process. In this way, more realistic and reliable canonical values of the imprecision statistics may be established over time. Ideally, every major step or component of the measurement process should be independently assessed. This would include not only the variability inherent in the particular measurement of concern, but also the imprecision arising from corrections, calibration factors, and any other quantities that make up the final result.

Assessment of Systematic Uncertainties

Although a general guideline for the approach to the assessment of systematic uncertainties can be formulated, there are, unfortunately, no rules to objectively assign a magnitude to them. For the most part, it is a subjective process. Their magnitudes should preferably be based on experimental verification, but may have to rely on the judgment and experience of the metrologist. In general, each systematic uncertainty contribution is considered as a quasi-absolute upper bound, overall or maximum limit on its inaccuracy. Its magnitude is typically estimated in terms of an interval from plus to minus δ about the mean of the measurement result. By what method then should the magnitude of these maximum limits be assigned? It may be based on comparison to a standard, on experiments designed for the purpose [4], or on verification with two or more independent and reliable measurement methods. Additionally, the limits may be based on judgment, based on experience, based on intuition, or based on other measurements and data. Or the limits may include combinations of some or all of the above factors. Whenever possible, they should be empirically derived or verified. The reliability of the estimate of the systematic uncertainty will largely depend on the resourcefulness and ingenuity of the metrologist.

The Need for an Overall Uncertainty Statement

Without deprecating the perils of shorthand expressions, there is often a need for an overall uncertainty statement which combines the imprecision and systematic uncertainty components. Arguments that it is incorrect from a theoretical point of view to combine the individual components in any fashion are not always practical. First, an approach which retains all details is not amenable for large compilations of results from numerous sources. And second, this approach shifts the burden of evaluating the uncertainties to users. Many users need a single uncertainty value resulting from the combination of all sources of inaccuracy. These users believe, and rightly so, that this overall estimate of inaccuracy can be most appropriately made by the person responsible for the measurement result. It must be emphasized, however, that there is no one clearly superior appropriate method for reporting an overall uncertainty, and that the choice of method is somewhat arbitrary. Several methods are commonly employed [5,6].

One method is to add linearly all components of the systematic uncertainty and linearly add the total to the imprecision estimate. Since the individual systematic uncertainties (δ_j) are considered to be maximum limits, it logically should be added to an imprecision estimate at a similar confidence level. That is, for example, the overall uncertainty u may be given by

$$u = [t_\nu(\alpha)]s + \sum_{j=1}^{q} \delta_j$$

where s is the computed standard error based on ν degrees of freedom, $t_\nu(\alpha)$ is the Student-t value corresponding to a two-tail significance level of $\alpha=0.05$, 0.01, or 0.001 (depending on the practice in the measurement field concerned), and δ_j is the magnitude of the estimated systematic uncertainty for each of the identified q systematic uncertainty components. This approach probably overestimates the inaccuracy, but can be considered as an estimate of the maximum possible limits. For example, if someone estimated that five contributions of about equal magnitude made up the total systematic error, that person would have to be very unlucky if all five were plus, or all five were minus. Yet, if there was one dominant contributor, it might be a very valid approximation.

Two other approaches have also been widely used. These methods add in quadrature all of the systematic uncertainty components, and either add the resulting quantity *linearly* to the standard error estimate,

$$s + \sqrt{\sum_{j=1}^{q} \delta_j^2} \quad ,$$

or add it *in quadrature* to the standard error estimate,

$$\sqrt{s^2 + \sum_{j=1}^{q} \delta_j^2}$$

These are frequently considered (erroneously) to correspond to a confidence level with $P=68\%$.

In another method, often termed the PTB approach [6], the component systematic uncertainties are assumed to be independent and distributed such that all values within the estimated limits are equiprobable (rectangular or uniform distribution) [8]. With these assumptions, the rectangular systematic uncertainty distributions can be convoluted to obtain a combined probability distribution for which the variance may be computed. This may then be combined in quadrature with that for the random uncertainty. In its simplest form, the uncertainty components are combined to form an overall uncertainty by

$$u = k\sqrt{s^2 + (1/3)\sum_{j=1}^{q} \delta_j^2} \quad ,$$

where k is customarily taken as 2 or 3. The above simple form is not appropriate when one of the component δ_j's is much larger than the others; in such a case it will be more informative to keep that component separate from the others and add it linearly.

A Concluding Thought

If there is one fundamental proposition for the expression of uncertainties, it is

> The information content of the statement of uncertainty determines, to a large extent, the worth of the final result.

This information content can be maximized by following a few simple principles:

BE EXPLICIT

PROVIDE DETAILS

DON'T OVERSIMPLIFY

When an overall uncertainty is reported, one should explicitly state how the separate components were combined. In addition, for results of primary importance, a detailed discussion and complete specification of all of the separate uncertainty components is still required. In this way, some users will benefit from having the metrologist's estimate of the overall uncertainty, while more sophisticated users will still have access to all of the information necessary for them to evaluate, combine, or use the uncertainties as they see fit.

REFERENCES AND NOTES

[1] Youden, W. J. Statistical aspects of analytical determinations. The Analyst 77: 874–878; 1952, Dec: Reprinted in Journal of Quality Technology 4: 45–49; 1972, Jan: Youden, W. J., Connor, W. S., Making one measurement do the work of two; Chemical and Engineering Progress 49: 549–552; 1953, Oct. Reprinted in Journal of Quality Technology 4: 25–28; 1972, Jan.

[2] Eisenhart, Churchill. Realistic evaluation of the precision and accuracy of instrument calibration systems. J. Res. Nat. Bur. Stand. (U.S.) 67C (2): 161–187; 1963; Reprinted as paper 1.2 in NBS Special Publication 300-1.

[3] Eisenhart, Churchill. Contribution to panel discussion of adjustments of the fundamental physical constants. Langenberg, D. N., Taylor, B. N., eds. Precision Measurement and Fundamental Constants, Nat. Bur. Stand. (U.S.) Spec. Publ. 343: 509–518; 1971.

[4] Youden's burette experiment, Journal of Quality Technology 4: 20-23 1972, Jan. Youden, W. J., Systematic errors in physical constants. Physics Today, 14, 32–34, 36, 38, 40, 43; 1961, Sept. Reprinted as paper 1.4 in NBS Special Publication 300-1. Youden, W. J., Enduring values, Technometrics 14: 1–11; 1972, Feb.

[5] Campion, P. J.; Burns, J. E.; Williams, A. A code of practice for the detailed statement of accuracy. National Physical Laboratory. London: Her Majesty's Stationery Office. 1953. II-57.

[6] Wagner, Siegfried R. On the quantitative characterization of the uncertainty of experimental results in metrology: PTB-Mitteilungen 89: 83–89; 1979, Feb.

[7] Müller, Jörg G. Some second thoughts on error statements. Nuclear Instruments and Methods 163, 241–251; 1979.

[8] A numerical comparison of uncertainty limits resulting from these assumptions with those implied by several alternative distributional assumptions is provided by table 1 on page 184 of [2], and discussed on the same and following page.

Churchill Eisenhart
Ronald Collé
July 1980