NBS TECHNICAL NOTE

288

Measurement Philosophy of the Pilot Program for Mass Calibration

P. E. PONTIUS

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] provides measurement and technical information services essential to the efficiency and effectiveness of the work of the Nation's scientists and engineers. The Bureau serves also as a focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To accomplish this mission, the Bureau is organized into three institutes covering broad program areas of research and services:

THE INSTITUTE FOR BASIC STANDARDS . . . provides the central basis within the United States for a complete and consistent system of physical measurements, coordinates that system with the measurement systems of other nations, and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. This Institute comprises a series of divisions, each serving a classical subject matter area:

—Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic Physics—Physical Chemistry—Radiation Physics—Laboratory Astrophysics[2]—Radio Standards Laboratory,[2] which includes Radio Standards Physics and Radio Standards Engineering—Office of Standard Reference Data.

THE INSTITUTE FOR MATERIALS RESEARCH . . . conducts materials research and provides associated materials services including mainly reference materials and data on the properties of materials. Beyond its direct interest to the Nation's scientists and engineers, this Institute yields services which are essential to the advancement of technology in industry and commerce. This Institute is organized primarily by technical fields:

—Analytical Chemistry—Metallurgy—Reactor Radiations—Polymers—Inorganic Materials—Cryogenics[2]—Office of Standard Reference Materials.

THE INSTITUTE FOR APPLIED TECHNOLOGY . . . provides technical services to promote the use of available technology and to facilitate technological innovation in industry and government. The principal elements of this Institute are:

—Building Research—Electronic Instrumentation—Technical Analysis—Center for Computer Sciences and Technology—Textile and Apparel Technology Center—Office of Weights and Measures —Office of Engineering Standards Services—Office of Invention and Innovation—Office of Vehicle Systems Research—Clearinghouse for Federal Scientific and Technical Information[3]—Materials Evaluation Laboratory—NBS/GSA Testing Laboratory.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D. C., 20234.

[2] Located at Boulder, Colorado, 80302.

[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

UNITED STATES DEPARTMENT OF COMMERCE • John T. Connor, Secretary
NATIONAL BUREAU OF STANDARDS • A. V. Astin, Director

NBS TECHNICAL NOTE 288

ISSUED MAY 6, 1966

(Reprinted January 1968, with minor corrections)

Measurement Philosophy of the Pilot Program for Mass Calibration

P. E. PONTIUS

Metrology Division
Institute for Basic Standards
National Bureau of Standards
Washington, D.C., 20234

NBS Technical Notes are designed to supplement the Bureau's regular publications program. They provide a means for making available scientific data that are of transient or limited interest. Technical Notes may be listed or referred to in the open literature.

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402
Price 30 cents.

FOREWORD

The National Bureau of Standards hopes to make the method described in this presentation available to various laboratories which may desire it as rapidly as arrangements can be made to put it into effect. The cost of operating the service must necessarily be borne by those laboratories, public or private, who advantage from its operation. It is clear that this service will be valuable and economical only where the workload of the laboratories justifies its use. It is also evident, and this is one of the good features about the method, that each laboratory must conform to the prescribed method in a calibration procedure.

One would expect, if experimental operation of the system in this field of measurement works out satisfactorily, that it would be extended to other fields of measurement. We now see how it could be applied in gage block calibration and it is likely that the procedure can be extended to other types of calibration, such as, electrical and temperature calibrations, particularly where the method of reduction of the data involves a complicated process which can be more readily carried out in accordance with a well devised computer program.

This suggests a new approach of the Bureau in fullfilling its responsibility as the nerve center for our national measurement system. It would be our hope that we can function better and more effectively, perform a greater service to the nation's commerce and industry in this way than serving merely as a routine laboratory to calibrate various instruments and standards as they are sent in to us for that purpose. It is our belief that the measuring system of the country must be as self-sufficient as possible; we want to help make it so.

A. G. McNish, Chief
Metrology Division
Institute for Basic Standards
National Bureau of Standards

Contents

The Measurement Philosophy of the Pilot Program for Mass Calibration

Paul E. Pontius

The Pilot Program for mass measurement is the result of a consideration in which the values produced are thought of as the products of a mass measurement process. The collective performance of elements of the mass measurement process results in establishing the process precision which, under certain conditions, can be described quantitatively by pertinent performance parameters. The uncertainty attached to the product of the process, the measured value, is computed from these parameters and reflects the total performance of the process rather than the immediate measurement which might have produced the value. Interpretations of uncertainty and surveillance tests are discussed. The Pilot Program in mass measurement, whereby suitable process performance parameters can be established for precise mass measurement processes in other facilities, is discussed.

Key words: Mass measurement process, process performance parameters, and uncertainty.

1. Introduction

In order to utilize the capabilities of a particular mass measurement process, it is necessary to have at least one mass standard of known value to establish the measurement unit and, equally important, to know quantitatively how well the process performs. The process produces mass values for a wide variety of objects and, in most instances, the objects and values pass on to others to serve many purposes. The uncertainty associated with values produced by the process establishes the suitability of these values for the intended usage, the amount of measurement effort necessary to meet the requirements with confidence, and the basis for agreement when the same measurement must be made with two different measurement processes. If the uncertainty is to be realistic, it must be formulated from process performance parameters which are established by all the data generated by the process to date. In addition, it must adequately reflect both the random variabilities and systematic errors associated with the process.

The activities of the Mass and Volume Section and the Statistical Engineering Laboratory have been directed, for the past several years, toward an objective evaluation of the mass measurement process and toward the establishment of suitable parameters of performance which can be used to compute realistic estimates of process uncertainty to be associated with the mass values produced. The success of these efforts provide the basis for the formulation of a different method for disseminating the mass measurement unit and for maintaining the standards of mass which are directly involved in measurement processes throughout the country. The resulting program, currently designated the Mass Measurement

Pilot Program incorporates, in each participating facility, the calibration procedures currently in use at the National Bureau of Standards which provide both a means to recognize and to utilize the maximum capabilities of the mass measurement processes.

The Mass Measurement Pilot Program, at the present stage of development, requires the participating facility to either have, or have access to, a pair of kilogram mass standards and suitable sensitivity weights which have been recently calibrated by NBS. The calibration of duplicates, subdivisions, and multiples of the kilogram are accomplished by using the equipment of the facility to make the observations in accordance with the prescribed procedures. The raw data is transmitted to NBS via teletype or other convenient means of communication. The data will be processed using an appropriate computer program. The monitoring function incorporated in the analysis will test the values obtained for the performance parameters against the appropriate parameters which represent the performance of the facility which produced the data. The mass values and appropriate uncertainties for the weights being calibrated are returned to the facility via teletype in a format suitable for inclusion in a report of calibration. The analysis sheets, which include, in addition to the statistical eavluation, a listing of supplementary information such as the equipment used, the operator, the weighing designs used and also a copy of all of the raw data listed essentially in the order it was taken, are forwarded by mail for evaluation and use as substantiating documentation. At the present time, the program is in limited operation at three facilities over a restricted range of nominal mass values. The success of the operation, to date, has been most gratifying.

The programs for data analysis, incorporated in the Pilot Program, strive to provide a service matched to the unique requirements of the total mass measurement process and to extract from the resulting data all possible information concerning the process performance. The procedures are designed to calibrate most ordered sets of mass standards, with few if any, extra observations over those required by other calibration procedures, and in addition, one obtains the statistical information necessary to assess the performance of the particular process that was used. The analysis of the data provides parameters relative to both short and long term process variability and it is possible to compute in advance, and verify the appropriateness of the uncertainty to be associated with each mass value determined. Facilities that can demonstrate a continuous "in control" operation through the use of the Pilot Program are, in essence, extensions of the NBS facilities and, as such, require only minimal calibration support.

2. The Measurement Process

An understanding of the analysis of the data and the significance of the resulting process performance parameters requires an understanding of the philosophy on which the Pilot Program is based. Following the principles suggested by Eisenhart (1)*, a mass measurement process, or system, consists of all the elements involved in making mass measurements; the standard, the equipment, the environment, the operator, the procedures used, the schedule of observations, and finally, the analysis or computations. The input to this process or system is the mass values of certain standards and the uncertainty assigned to the values relative to some "true" or defined unit and relative to some prior measurement process. The outputs or products of the process, are the mass values assigned to various objects and the associated uncertainties for these values. The uncertainty assigned to the resulting values must reflect the performance of both the local process and uncertainties in the transfer from the national standard. The local process must be considered adequate if the magnitude of the uncertainty is such that the values are satisfactory for the intended usage.

- - - - - - -

*Numbers in parenthesis refer to similarly numbered references at end of paper.

A characteristic of most precise measurement processes is that repeated measurements produce series of non-identical numbers. The desired values to represent the property in question are related to these numbers, sometimes directly and sometimes indirectly. The variability of the values obtained can be expressed quantitatively if the process is reasonably well behaved as evidenced by an analysis of large collections of appropriate data. One cannot assume, however, that one small series of repeated measurements from a particular process is truly representative of a very large group of repeated measurements made by the same process. All measurement systems respond to disturbances from many sources. The variability of repeated measurements may reflect the persistence of such disturbances in the form of a clustering of the data into groups or other forms such as trends or sharp discontinuities, each form generally being associated with certain types of disturbances.

Much of the development work associated with a measurement process is directed toward identifying the sources of significant disturbances and minimizing the associated variability either by modifying the process, by changing the design of the instrument, or by applying numerical corrections to the data. Since the largest variability is frequently associated with characteristic grouping in the collected data, the task of identifying the source of the disturbance and minimizing the effect is somewhat simplified. Unrelated events which occur in a more or less randomized order and which cause disturbances of nearly the same order of magnitude are much more difficult to identify and correct. Corrective actions must be chosen with care so that the resulting process remains compatible with the intended usage of the result. For example, certain difficulties encountered in mass calibration processes could be minimized if the density of weight material was restricted to one value or if all calibrations were made in a controlled air density environment. Such actions are not appropriate since they do not reflect the conditions under which the standards with assigned mass values must be used.

In order that the variability of a particular process be properly estimated, repeated measurements must be collected over the range of operating conditions and other variables i.e., objects under consideration, environmental changes, seasonal changes, changes in operators to which the process is likely to be subjected. If the process is free from significant systematic disturbances, at some point in time as the collection grows, the measurements at hand will suggest a band within which one is fairly certain that the value from the next measurement will lie (See figure 1a). Inasmuch as the results of the next measurement, when taken, will verify the existence of such a band, the remaining task is to describe the band quantitatively.

Several large sets of values, now in excess of 250 and continually growing, have been accumulated which are the products of the NBS mass measurement processes, these processes being nearly identical to many other mass measurement processes throughout the country. The measurements, extending over a considerable time interval, have produced values for the mass differences between a given pair of mass standards and values for selected mass standards of different nominal value through the calibration procedures. Studies indicate that the distribution of values for repeated measurements are essentially symmetric. Deviations from the normal or Gaussian distribution do not appear to be significant and the processes appear to be in a state of statistical control at the present time. Thus, the "best" value or accepted value for the mass difference between two standards or for the mass of a particular standard as established by calibration is defined accordingly as being the long term average of a large collection of appropriate repeated measurements.

One cannot normally repeat a particular routine measurement a sufficient number of times to establish the "accepted value" as defined. The optimum performance of a measurement process would be the ability to produce mass values, independent of time and location, which do not differ from "accepted values" as defined above in excess of predictable limits established by the appropriate performance parameters. Whereas one might normally think of the resulting value as being accurate within plus or minus the uncertainty, the proper statement reflects the fact that no statement can be made about a single value. It is either right or wrong i.e., either does or does not differ from the correct value by more than the stated uncertainty. It is the process to which the uncertainty applies - we can guarantee that a certain fraction of our results are within the uncertainty limits but cannot say which ones are in or out.

It is necessary to know that a particular process is capable of operating in a state of control, and that, at any given time, its operation remains in a state of control, and further, that the process performance parameters realistically describe the actual performance. Having met these conditions a measure of the agreement between the value computed from a given sub-set of data and the "accepted value" can be determined and is a function of the size of the sub-set and the appropriate process performance parameters. While all processes will not have identical performance parameters, values from sub-sets of measurements of a given mass difference produced by two different but similar processes, both in a state of control, will converge in practice to the same "accepted value" for the mass difference as the size of the sub-sets increases.

3. Process Precision

Process precision is a measure of the process performance with respect to time. It is not instrument precision but rather a measure of the agreement between the value obtained today with the value that would be obtained for the same measurement next week, next month, or next year. To express process precision quantitatively requires the process under study to be operating continuously in a state of statistical control, in the sense of the previous section. Realistically, a process in a state of control is never completely free from the effects of extraneous disturbance. The effect from a disturbance can be identified by unexpected trends on large "jumps" (relative to the normal range of data), which occur in a periodic (non-random) pattern. After identifying the source of the disturbance, corrective action based on theory can reduce the magnitude of the effect only to the point that it is no longer identifiable in the data. After all such corrections have been made, the process variability reflects the residue of effects from many small disturbances such as the response of the instrument to small changes in the local environment, the small variabilities of the mass of the standards and objects being compared with them which are environment dependent, and perhaps a host of other sources in which the effects are small but random in magnitude and which occur at varying time intervals.

It must also be emphasized that a quantitative expression for process precision without defining the measurement procedure is meaningless. The concept requires that all of the future measurements be made insofar as possible following the procedures used in making the measurements which form the basis for a prediction statement. One could define, for example, the measurement to be the value as determined by a double substitution weighing in which a particular standard is compared with a specific object. A time plot of the values obtained by repetitions of this procedure would exhibit a certain variability. If, with all other things equal, one would decide that the value should be the average of three double substitution weighings, the continued plot of values obtained would show a marked decrease in the variability exhibited. Given the process precision and the defined process, one can generally compute the precision to be expected by a change in process definition. Precision statements alone may tend to make a particular measurement process look better than some other process when in fact it really is not.

Most measurement processes exist to provide a supporting service. Stated in another way, the resulting measured values are not in themselves the final product. As a consequence, the most common measurement philosophies reflect the practical requirement that a particular measurement effort need only produce values which are good enough for the purpose at hand. The user, primarily interested in making the necessary measurements, as quickly and as conveniently as possible, relies strongly on others to minimize the variabilities of the process and may as a consequence accept any of a variety of statements concerning instrument precision as estimates for process precision. Ingenuity and a substantial investment in engineering design has provided a variety of weighing equipment which is easy to use and which will produce good measurements. This state of affairs often compensates for errors which may be present in over-simplified descriptions of measurement processes. It is inevitable, however, at some time or another, the user will be confronted with either the need for more precise measurements or the need to verify that the performance is actually as it has been assumed to be.

Reasonably precise mass measurement processes now in existence have many elements in common. The mass standards used to disseminate the mass measurement unit, and many of the so-called working sets of standards, are of good quality and capable of remaining reasonably constant for periods of years. The instruments used are of good design, well engineered, and capable of high quality performance. The environments, particularly those used for mass calibration work, are almost without exception clean and with limited access. Because of the similarity, it is not unreasonable to expect that, for uniform procedures, the process precisions from a large number of facilities should be of nearly equal magnitude comparable to, and perhaps in some cases smaller, than the precision available at the National Bureau of Standards. The availability of such precision is not generally recognized because the more or less traditional procedures in general usage do not provide obvious checks on process performance.

Perhaps by far the majority of measurements are made with a technique to be designated the "direct reading" mode of operation. It is, indeed, convenient to use an instrument which has been calibrated or adjusted so that a one-to-one correspondence exists between the observed number and the mass of a particular object. The instrument, however, after original calibration and/or adjustment, must assume the requirements for long term stability normally associated with high quality standards. While every measurement made in this manner has some uncertainty component associated with the stability of the instrument, the degree to which satisfactory stability has been achieved is evidenced from the fact that operation in the "direct reading" mode is adequate for a wide variety of measurement requirements. If the systematic error or bias associated with the lack of stability of the instrument must be minimized, in the absence of another instrument which is less sensitive to disturbances which affect the long term variability, the "direct reading" mode of operation must be replaced by more sophisticated procedures.

The short term precision of an instrument is generally so much better than the long term stability that frequently several orders of magnitude improvement can be achieved by merely changing to a comparative mode of operation. By including one or more standards of known value in the operational procedures, the requirements on the instrument are changed. The measurements are, in effect, the difference between the unknown and the standard so that the instrument response need be reasonably continuous only over the period of time necessary to make the comparisons, and reasonably linear only over the range of differences between the standard and the unknowns. Two standards are frequently used in the mass measurement process, one of which has a nominal value sufficiently close to that of the unknown so that the indications will be on-scale, and a small standard, called a "sensitivity weight" which is used to establish the correspondence between the instrument indicating scale and the mass unit. The procedures of the Pilot Program are comparative procedures using either double substitution or single transposition weighing methods.

4. Performance Parameters

The uncertainty associated with a measured value is in essence a prediction of the band in which, in the absence of real change, one would expect the value from a repeated measurement at some time in the future would most likely fall. As such, the uncertainty must be formulated from parameters which describe the process performance. Large collections of data establish that a given process can be made to operate in a state of control and that parameters computed from small groups of data are useful for describing the process performance. The same parameters, computed from small groups of data for other, but similar, processes are also valid for use in describing the performance of these processes, the existence of a state of control being verified as the data collections grow. Periodic recomputation, based on a larger collection of data, increase confidence in the fact that the parameters do describe the real process performance. From the previous discussions, two parameters are of importance, the short term or within-group variability and the long term variability.

A major difficulty in the application of statistical methods to the analysis of measurement data is that of obtaining suitable collections of data. The problem is more often associated with conscious, or perhaps unconscious, attempts to make a particular process perform as one would like it to perform rather than accepting the actual performance. For example, repeated measurements of the difference between A and B can be shown symbolically as:

$$(A-B) = K\, f\, (0_1, 0_2, 0_3, \ldots, 0_n)$$

where K is a factor to transform the observed 0_i into mass units. In the usual measurement, the "best" value for (A-B) is considered to be the average of all of the observed differences, that is:

$$(\hat{A-B}) = K \left(\frac{\sum_{i=1}^{n} 0_i}{n} \right) = K\, \bar{0}$$

The variability of a particular group of data is usually expressed by an estimate of the standard deviation of an observation 0_i

$$s_{(A-B)} = K \left(\frac{\sum_{i=1}^{n} \left(0_i - \bar{0}\right)^2}{n-1} \right)^{1/2}$$

Executed properly, this procedure will produce good values and a valid estimate of the variability. However, there are several sources of difficulty. Some well established procedures, for example, are open ended with reference to the number of observations which are taken. As before, the measurement is repeated n times until the values of a group of some arbitrary size, j, is obtained that lies within some predetermined limit:

$$\text{Lower Limit} < 0_{n-j}, \ldots, 0_{n-1}, 0_n < \text{Upper Limit}$$

Sometimes the desired grouping is obtained early in the series of observations and at other times, many consecutive measurements must be made before the required conditions are satisfied. Rejection of data on the basis of arbitrary performance limits severely distorts the estimate of the real process variability. Such procedures defeat the purpose of the Pilot Program. Realistic performance parameters require the acceptance of all data that cannot be rejected for cause.

It is well known that consecutive repeated measurements should be avoided where possible. Even the most experienced operators have difficulty in remaining unbiased in observing a series of measurements which produce nearly the same indication. More suitable data is obtained if the consecutive indications occur on different parts of the reading scale, a condition which can be induced artificially in some measurement procedures. The Pilot Program procedures specify the use of either double substitution weighings, such as:

$$A \rightarrow O_1$$
$$B \rightarrow O_2$$
$$B + \Delta \rightarrow O_3$$
$$A + \Delta \rightarrow O_4$$

where

$$(\hat{A-B}) = g(m, O_1, O_2, O_3, O_4),$$

and single transposition weighings such as:

$$A - B \rightarrow O_1$$
$$B - A \rightarrow O_2$$
$$B + \Delta - A \rightarrow O_3$$

where

$(\hat{A-B}) = h(m, O_1, O_2, O_3)$, and "m" is the mass value assigned to the sensitivity weight Δ.

Consecutive indications do not normally occur at the same point on the indicating scale in either of these weighing methods. Each sequence is considered as a unit, that is, if one observation is rejected for cause, such as a door slam or equipment malfunction, the entire sequence must be repeated. Operators must be cautioned against using the redundancy of these methods as a means to pre-judge the data.

Having defined the weighing methods, a simplified notation for the mass difference is introduced as follows:

$$(A-B) = d_i$$

where d_i is the difference in mass units as computed from a series of observations made in accordance with the methods above. (Note: The details of the computations are omitted for clarity in this discussion.) To obtain realistic uncertainty values for (A-B) by means of repeated measurements, the sequence of measurements such as:

$$(A-B) = f(d_1, d_2, \ldots, d_n)$$

should be made over a significant time interval, and preferably, each measurement should be made under different environmental and procedural conditions. Except in rare instances, this is not a practical procedure and other means must be used to provide assurance that the measurements are essentially independent.

The procedures of the Pilot Program provide a means to complete a series of measurements in a minimum of time while retaining the features necessary to provide realistic estimates of within-group variability. To illustrate with a simple example, suppose A is a mass standard of known value M, B and C are unknowns which must be calibrated. With d_i the mass difference, a set of measurements can be made as follows:

$$\begin{array}{lllll} A & - B & & = d_1 \\ A & & - C & = d_2 \\ & B & - C & = d_3 \end{array}$$

While the first and second relations above will provide values for B and C, the "best" values to use are those which best fit all three equations. Such a fit is obtained if the set of equations is solved by the method of least squares, subject to the restraint, that A have the assigned value, M. The "best" values for B and C become:

$$\hat{B} = f(M, d_1, d_2, d_3) = M - 1/3(2d_1 + d_2 - d_3)$$

$$\hat{C} = g(M, d_1, d_2, d_3) = M - 1/3(d_1 + 2d_2 + d_3)$$

where functions f and g are linear combinations of M and d_i. It is not necessary to continually resolve for the coefficients, for once known the linear equations for $\hat{B}$ and $\hat{C}$ are appropriate for all measurements made according to this design.

A measure of the within-group variability is obtained by looking again at the original equations to determine how closely the computed "best" values actually satisfy the observation equations. In the previous example, for instance, if $\hat{B}$ and $\hat{C}$ satisfied the observation equations exactly, then:

$$\begin{aligned} A - B &= d_1 \equiv M - \hat{B} & \qquad & d_1 - (M - \hat{B}) \equiv 0 \\ A - C &= d_2 \equiv M - \hat{C} & \text{or} \qquad & d_2 - (M - \hat{C}) \equiv 0 \\ B - C &= d_3 \equiv \hat{B} - \hat{C} & \qquad & d_3 - (\hat{B} - \hat{C}) \equiv 0 \end{aligned}$$

in practice these differences are not zeros because of random errors in the three measurements.

$$d_1 - (M - \hat{B}) = \delta_1$$

$$d_2 - (M - \hat{C}) = \delta_2$$

$$d_3 - (\hat{B} - \hat{C}) = \delta_3$$

where δ_i are the residuals representing the differences between observed and fitted values. The estimate of the standard deviation of a comparison expressing the within-group agreement is computed as follows: (See figure 1b).[1]

$$s = \sqrt{\frac{\sum_{i=1}^{n} \delta_i^2}{n - p}}$$

The δ's are also linear functions of the observed differences and can be computed from the appropriate coefficients for the particular design. If a specified procedure is used routinely m times on a particular instrument and at or near the same nominal load, the long term estimate of the standard deviation for the within-group variability is:

$$\hat{\sigma}_w = \sqrt{\frac{\sum_{j=1}^{m} \left(\sum_{i=1}^{n} \delta_i^2 \right)_j}{m(n-p)}}$$

- - - - - - - - -

[1] p, as used here, is the number of constants fitted and (n-p) the degrees of freedom from error. In general, the degrees of freedom are a function of both the design and the number, n, of measurements made. Before computation the user should become familiar with the appropriate details of the weighing designs.

The product m (n-p) is the cumulative degrees of freedom for error where m is the number of times a particular series has been repeated. Both s and $\hat{\sigma}_w$ are important process parameters used in the Pilot Program.

The proper evaluation of the long term process performance implies the existence of a collection of repeated measurements over a long interval of time. Again, except in rare instances, it is seldom practical to establish a suitable collection of data just by repeating a measurement such as (A-B) in the previous example. Just as a modification to the weighing design provided a realistic means to determine a measure of the within-group variability, a further modification can be made which will allow the collection of suitable data simultaneously with the routine operation of the measurement process. This is accomplished by introducing an additional "unknown" into each of the groups sent in for calibration. The same "unknown" is always used in a given procedure so that one obtains the desired repetitions along with the regular work. If the "unknown" is nearly identical to the unknowns being calibrated, the properties of the collection of data concerning it can be ascribed to the other objects in the particular group. The results are the ideal simulation of the process - it is done under real life conditions and no assumptions of appropriateness are required. This can best be illustrated by describing a weighing design or series which is used in the Pilot Program.

Changing the notation slightly from the previous examples, the weights under consideration are designated X_1, X_2, X_3 and X_4, where X_1 is a standard of known value, X_2 is an "unknown" or check standard, and where X_3 and X_4 are objects which, together with the values established by the procedure, pass on to others to serve a wide variety of purposes. The mass difference, as determined by some prescribed method, is designated $d_1, d_2, \ldots, d_i$. Again, for clarity, many of the details will be omitted. The weighing design to be described requires difference measurements between all combinations of the objects concerned, thus the observation equations become:

$$\begin{array}{llllll}
X_1 & -X_2 & & & = & d_1 \\
X_1 & & -X_3 & & = & d_2 \\
X_1 & & & -X_4 & = & d_3 \\
 & X_2 & -X_3 & & = & d_4 \\
 & X_2 & & -X_4 & = & d_5 \\
 & & X_3 & -X_4 & = & d_6
\end{array}$$

This array of equations, in matrix form, becomes

$$\begin{bmatrix} +1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ +1 & 0 & 0 & -1 \\ 0 & +1 & -1 & 0 \\ 0 & +1 & 0 & -1 \\ 0 & 0 & +1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{bmatrix}$$

and in matrix notation becomes:

$$\underline{A}\,\underline{X} = \underline{D}$$

where $\underline{A}$ is the design matrix and the relation between X_1 and the differences d_i, in terms of d_i become:

$$\hat{\underline{X}} = (\underline{A}' \underline{A})^{-1} \underline{A}' \underline{D}$$

at this point, the equations cannot be solved because all of the measurements are difference measurements thus the matrix is singular. If X_1 is a standard of known value M, the value can be used to provide a restraint, or ground zero, so to speak. Having established such a restraint, the equations can be solved to provide the following linear equations for the values of the other objects:

$$\hat{X}_2 = f(M, d_1, \ldots, d_6) = N$$

$$\hat{X}_3 = g(M, d_1, \ldots, d_6)$$

$$\hat{X}_4 = h(M, d_1, \ldots, d_6)$$

If X_2 is the "unknown" which has been inserted in the group, and which will be used in every such group of weighings, the values N_i provides the collection upon which the estimate of the long term process variability is based. As the collection of values X_2 increase, the accepted value becomes the long term average of all of the values at hand, and the process variability is evident in the deviations of the individual values, from this accepted value. The measure of this agreement is another important performance parameter included in the Pilot Program. The statistical tests incorporated[2] are based on the standard deviation of the collection of values N_i, computed as follows:

$$\hat{\sigma}_T = \sqrt{\frac{\sum_{i=1}^{n} \left(N_i - \frac{\sum_{i=1}^{n} N_i}{n} \right)^2}{n-1}}$$

It is important that $\hat{\sigma}_T$ be computed from the total collection of values for the "check" standard.

The "four one's" series just described is the identical schedule of the intercomparisons used at the 1 kg level in the Pilot Program. The restraint condition, however, is the sum of the values, M_1 and M_2, for the two kilogram standards X_1 and X_2. In this case, the solution provides values for all four objects.

- -

[2] While $\hat{\sigma}_T^2$ is frequently called the between-group variability, it is really the total process variability, that is

$$\sigma_T^2 = \sigma_W^2 + \sigma_B^2$$

where σ_T^2 = total variance obtained as above

σ_W^2 = with-in group variance obtained as above

σ_B^2 = between-group variance

In an ideal measurement process, one would expect that σ_W and σ_T would be very nearly equal thus σ_B should accordingly approach zero. This appears to be the case in some mass measurement processes which have been studied in detail.

$$\hat{X}_1 = K\,f\,(M_1, M_2, d_1, d_2, \ldots, d_8)$$

$$\hat{X}_2 = K\,g\,(M_1, M_2, d_1, d_2, \ldots, d_8)$$

$$\hat{X}_3 = K\,h\,(M_1, M_2, d_1, d_2, \ldots, d_8)$$

$$\hat{X}_4 = K\,q\,(M_1, M_2, d_1, d_2, \ldots, d_8)$$

subject to the condition that the values obtained for the two standards must sum to the value of the starting restraint. The difference between two computed values for the standards, generally considered to be a check on the constancy of the standards, or the computed value for one of the standards provides the same information concerning the long term or variability of the process. In other series, a separate "unknown" or check standard is included in each decade for this purpose.

In matrix notation, the residuals take the following form:

$$[\underline{I} - \underline{A}(\underline{A}'\underline{A})^{-1}\underline{A}']\underline{D} = (\underline{D} - \underline{D}_c) = \underline{\Delta}$$

where $(\underline{D} - \underline{D}_c)$ is the difference between the observation for a particular comparison and the computed difference for the same comparison as explained before. The solution of this set of linear equations can be used easily once the coefficients have been established for a given design. The use of computer programs to solve the arrays of equations has greatly facilitated the preparation and evaluation of a wide variety of intercomparison procedures.

The long term estimate of the standard deviation for the within-group variability and the long term process performance based on the accepted value of the check standards provide the basis for stating that a particular mass measurement process is operating in a state of control. If, for each data set, the within-group standard deviation, s , does in fact substantiate the correctness of the long term within-group standard deviation, $\hat{\sigma}$, and if the value obtained for the check standard substantiates the accepted value of the check standard, the process is considered to be operating in a state of control. A picture of the total performance of a measurement process can be obtained by plots such as shown in figure 1a and 1b.

Plotting the standard deviation against the value of the check standard for each run such that each point represents the value obtained for the check standard, as indicated by its position with reference to the horizontal scale and the standard deviation of the data group that produced that value, as indicated by its position with reference to the vertical scale, one obtains a figure such as shown in figure 2. It is immediately obvious that the internal agreement of a given set of data as reflected by the standard deviation, is not in itself enough to describe the performance of the process. The values obtained for the check standard agree with the accepted value for both large and small standard deviations. The scatter of the points, which should be circular or elliptical in the ideal situation, suggest control limits which form a rectangle which will encompass most of the data points. Similar evidence could be produced for measurement processes in most facilities if their results are free from significant systematic disturbances.

Because of the similarity of these processes amongst themselves and with the NBS, it is not unreasonable to expect that plots such as that in figure 2 are characteristic of each mass measurement process, the only variable being the range of the values on each plot. Facilities which use procedures similar to those which have been discussed can collect data concerning the performance of their respective processes as they perform certain routine measurements, such as the calibration of certain types of standard. Useful estimates of the process performance parameters can be established with relatively few measurements. In the absence of major difficulties, and if the process is allowed to establish its own limits, only minor adjustments to the initial estimates would be expected as the collection of data increases. The Pilot Program analysis provides the necessary information to update the estimate of the standard deviation for the within-group variability, the values for the check standards, and to establish the agreement of the values obtained in a particular data set with the long term estimates of the parameters for the process.

5. Uncertainty Computation

The uncertainty statements attached to the values produced by the mass measurement processes are objective measures of the consistency of the total mass measurement system. In the absence of an independently reproducible standard mass unit, the uncertainty of measurement at any given location is dependent upon certain prior measurement, many of which were performed in other facilities. Problems in the formulation of realistic uncertainty statements are those associated with the propagation of error, as measured items pass from one facility to another. It is important therefore to state the manner in which the uncertainties are carried forward. Propagation of error rules take many forms, some of which are substantially subjective, and some of which are objective. To the extent that simple subjective rules adequately and economically serve the intended purpose, there can be no serious argument; however, it must be emphasized that the decision to use these rules is based on factors other than an objective evaluation of the process. Serious efforts to achieve measurement agreement must be based on realistic uncertainty statements.

The uncertainty associated with a mass value produced by a given measurement process consists of two components, one of which is a function of the random errors and the presence or absence of systematic errors in the local measurement process, and one of which is a function of the uncertainty of the measurement processes which have gone before, such as those used to provide the value for the starting standards for the local process. In elementary form, the uncertainty of value X are often expressed as:

$$\text{Uncertainty of } X = \text{Uncertainty of } P_1 + \text{Uncertainty of } P_2 + \ldots..$$

where P_1 is the contribution of the local process and P_2 to P_n the contribution of previous appropriate measurements. Such a combination is frequently used to make rough estimates as to the maximum expected limits of inaccuracy. For example, a precision estimate from an equipment catalog, expressed as a percentage of load, might be used as the uncertainty P_1. The adjustment tolerance of a particular class of weights, expressed as a percentage deviation from nominal value, might be used for the uncertainty P_2. If the combination produces an uncertainty for X_1 sufficiently small when compared to the requirement for the task at hand, the measurement process would receive no further attention. There are several fundamental problems associated with this approach. The "not-to-exceed" philosophy seldom permits the utilization of the process to its maximum capabilities. The confidence may have a false basis, for example, the precision obtainable in comparing weights is not directly applicable to other weighing problems. Other problems develop because most equipment is sufficiently precise to see effects which are ignored, in such an analysis.

The contribution to the uncertainty of the final value produced by one measurement process can be, in a series as shown above, expressed, for the moment, as:

$$\text{Uncertainty of } X = f(\sigma) + KE$$

where σ is an appropriate measure of the randomness of the process, KE is a measure of the systematic error introduced in the process, and the function f is not as yet specified. All uncertainty discussions relate to the form of the function, and how one might combine the performance parameters of several measurement processes. The formulation of realistic uncertainty statements presumes the performance of the measuring systems involved is reasonably well behaved. The particular processes must be operating in a state of statistical control and numerical parameters must be available which describe the variability. The uncertainty statement is a measure of the demonstrable consistency of the total measurement system and insofar as the basic concepts of the measurement are absolute, uncertainty is a quantitative statement of inaccuracy.

The variability exhibited in a collection of data and the associated standard deviation are functions of the process definition. For example, a collection of data involving the same object and the same instrumentation treated in several different ways, is shown in figure 3. It is immediately apparent that the variability and thus the standard deviation associated with each treatment is markedly different even though the total measurement effort is the same. The uncertainty of the value established for X relative to S relates the correctness of the long term average of the respective total collection of data while the precision of each treatment refers just to the variability of the collection of data about their central values. The uncertainty of the average in each case should be the same so the function $f(\sigma)$ must include provisions to reflect the differences between the various defined measurement procedures.

One element in common with all measurement procedures is the method by which the intercomparisons are made. Thus a particular weighing method or methods must be specified in all appropriate procedures. The Pilot Program specifies either a double substitution weighing or a transposition weighing depending upon the equipment used. Other weighing methods can be evaluated with reference to the defined method. For example, the double substitution method is, in essence, the average of two comparisons thus one would expect the variability of the method would be somewhat less than that for a single comparison such as in a single substitution weighing. This is clearly shown in figure 3a and 3b. It should be pointed out however, that the variability of the single substitution method is somewhat in excess of that which would be expected on the basis of statistical comparisons alone because the double substitution method tends to minimize certain extraneous effects which have an influence on the variability.

While the variability of the defined process is frequently referred to as the standard deviation of a single measurement, one does not normally make repeated measurements according to the prescribed procedure to establish the magnitude of the standard deviation. Rather, the collection of measurement is obtained through the choice of an appropriate weighing design, as shown in figure 3c. While all of the intercomparisons indicated in the schedule are made by double substitution weighing, the values for the objects in question are those which provide the "best" solution for all of the intercomparisons, as described before. The data shown in figure 3c are the values determined for one of the "unknowns", say X_1. The reduction in variability associated with the design is, in essence, the same as one would expect if measurement has been defined to be the value which would be obtained from the average of several double substitution weighings. The design of measurement schedules should incorporate a valid estimate of precision and a check on the long term variability of the process. This latter can be achieved by addition of a check standard. The selection of a design should accomplish these purposes with a minimum amount of effort.

Having defined the weighing method, the weighing design, and having used a particular set of equipment, each measurement produces one data point such as shown in figure 3c. If the process is sufficiently stable, the average of a collection of such data points will tend to a limiting mean value which represents the mass difference between the standard of known value, S, and a particular unknown, X_1. The function $f(\sigma)$ must then include a factor to reflect the agreement of the average of a group of n independent measurements. As the group size increases, the long term average would be called the accepted value.

The function of $f(\sigma)$ appropriate to the task of determining the uncertainty to be associated with an echelon in the chain can now be expressed as:

$$f(\sigma) = 3\sigma_T = 3\sqrt{d^2\sigma_W^2 + \sigma_B^2}$$

where

d is a factor determined by the weighing design used

σ_w is a within-group estimate of the standard deviation [3]

σ_B is the estimate of the standard deviation between groups which at the present state of evaluation appears to be practically zero in the normal calibration processes. If, for some unexplained reason, the long term variability exhibits unclear evidence of grouping, σ_B may be the largest term under the radical sign

If the reported result is the average of n repeated independent runs with the defined measurement procedure, then of course

$$f(\sigma) = 3 \frac{\sigma_T}{\sqrt{n}} = 3 \sqrt{\frac{d^2 \sigma_w^2 + \sigma_B^2}{n}}$$

Having established an expression for the random component of uncertainty associated with a process, P_1, the manner in which the uncertainty can be established for a value assigned to some object by means of a series of measurements can be considered in more detail. Suppose, for a typical example, the value for kilogram K_d is the product of the chain of events shown in figure 4. The first group of weighings, 0_1 to 0_6, consists of intercomparisons between K_{20}, K_4 and the national reference standard Kg_1 and Kg_2; 0_7 through 0_{12} are intercomparisons between the national reference standard and K_a and K_b, the latter perhaps being the starting standards as required in the Pilot Program; 0_{13} to 0_{15} is a "three-one's" series which is used to establish the value for K_c; and finally, 0_{16} to 0_{18} are three independent comparisons between K_c and K_d to establish the value for K_d. By replacing the plus and minus signs in the schedule of comparisons with 1 and -1, and zeros elsewhere, the design matrix mentioned earlier is obtained. This matrix can be solved subject to the restraint that K_{20} have a fixed value, the solution providing the values for all of the other weights.

In practice, such an array is seldom considered as a unit. Separating such an array into groups or series (See figure 5) such as 0_1 to 0_6; 0_7 to 0_{12}; 0_{13} to 0_{15}; and 0_{16} to 0_{18}, does not affect the result. The value or values obtained in one series is carried forward as a restraint for the solution of the next series. The value obtained for K_d is a linear function of the starting value assigned to K_{20} and the values obtained in each series. Since the systematic error associated with the value of K_{20} is zero, and assuming all processes to be in a state of control with negligible between-group variability and negligible introduction of systematic error, the uncertainty of K_d is essentially due to random errors only and can be expressed as:

$$\text{Uncertainty } K_d = 3 \sqrt{\frac{d_1^2 \sigma_{w_1}^2}{n_1} + \frac{d_2^2 \sigma_{w_2}^2}{n_2} + \frac{d_3^2 \sigma_{w_3}^2}{n_3} + \frac{d_4^2 \sigma_{w_4}^2}{n_4}}$$

assuming double substitution weighings are used in all comparisons. n is one in all except the last series which has three repeated independent measurements so that n = 3. The coefficient d has a value other than one in all except the last series.
Thus:

- - - - - - - - -

[3] σ_w^2 is the long term within standard deviation for a particular defined weighing method e.g., double substitution. Now suppose we change our weighing method to that of double transposition, then, we could compute its standard deviation from σ_w obtained for the double substitution method by a suitable multiplier, 1/2 in this case.

$$\text{Uncertainty } K_d = 3\sqrt{\frac{d_1^2\sigma_{w_1}^2 + d_2^2\sigma_{w_2}^2 + d_3^2\sigma_{w_3}^2 + \sigma_{w_4}^2}{3}}$$

The hypothetic chain of events in figure 4 and 5 imply that all mass measurements must start with K_{20}, a procedure which at the present time is neither practical or desirable. In order to keep K_{20} out of the chain, it is necessary to assign a fixed value at some intermediate point in the chain. This is the situation in regard to the national reference standard kilograms, and may also be the situation at other points in the chain. To illustrate, the first series (See figure 5) of the hypothetic situation of figure 4 will be used to establish fixed values for Kg_1 and Kg_2, the national reference standards. These values are used repeatedly as restraints, thus the uncertainty resulting from .r calibration introduces adjustment errors in all calibrations with reference to the national reference standards. This systematic error is a factor to be considered in all mass measurements.

If the indicated first series is done once, with great care, the value obtained is like the first value in figure 3c and, under the previous assumptions, the uncertainty could be stated as:

$$\text{Uncertainty } Kg_1 = 3\sqrt{d_1^2\ \sigma_{w_1}^2}$$

A recalibration at a later date might produce the second point. One must now make a choice between the two values. If the points were as shown in figure 3c, all things being equal, it would probably be wise to retain the original value on the basis that, with reference to its higher precision, no change is indicated. If the points were of equal precision then the average of the two would be recommended. If only one value is to be retained, either one will do, no reduction in the overall uncertainty being achieved by using the new value or alternatively in retaining the old. As the collection of data points increases, the situation changes. If the process is stable, the average of the group is a better estimate of the limiting mean than any single point. The uncertainty of the average is expressed as:

$$\text{Uncertainty } Kg_1 = 3\sqrt{\frac{d_1^2\ \sigma_{w_1}^2}{n}}$$

The uncertainty of the values of the national reference standards which in turn becomes a source for systematic error in other measurements, is computed in this manner. The uncertainty of the starting standards used in the Pilot Program is also computed in this manner on the basis of at least two calibrations rather close in time and one calibration perhaps a month or two later.

Kilogram standards alone are not very practical. Therefore every calibration laboratory must consider the problem of establishing value for sets of standards which are multiples or subdivisions of the kilogram. While there are several approaches to this problem, the procedures of the Pilot Program are the same as those used in the National Bureau of Standards calibration program. These procedures, generally credited to Hayford(2) and Benoit (3) have been recently modified by Cameron and Raybold (4) to provide a means to establish the values for ordered sets of standards relative to two selected starting standards with known values. The factors considered essential to any calibration process; a means to monitor the starting standards and the process, and sufficient redundancy to allow the computation of within-group standard deviations, are incorporated. In essence, the first or starting series of weighings include, in addition to the starting standards, a group or summation of weights, treated as a unit. The value for the summation established in the first series is used as the restraint for the second series in which the objects of the summation are considered individually. A typical procedure is shown in figure 6.

Before proceeding to the calibration of subdivisions and multiples of the kilogram, the uncertainty statement can be generalized in the following manner (ignoring the possibility of between-run variability):

$$\text{Uncertainty } (\bar{X}) = 3\sqrt{\frac{d^2 \sigma_w^2}{n} + \sum_i \frac{k_i^2 d_i^2 \sigma_{w_i}^2}{n_i}} + KE$$

where

d, σ_w, n, are associated with the local measurement, or the last series used to obtain the desired values.

d_i, σ_{w_i}, n_i ($i=1$ to m) are the series which stand between the local measurement or series and the series in which fixed starting values provide the restraint.

k_i is the ratio of the nominal value of the restraints produced by each of the previous series.

K is the ratio of the nominal value of the weight or object in question to the nominal value of the weight or object which is assumed to have a fixed value (that which is used for starting restraint).

E is the uncertainty of the starting restraint.

In the case of a chain of measurement at the kilogram level, such as the previous examples, $k_1 = k_2 = k_3 = 1$ and $E = 0$, thus the expression reduces immediately to that previously given.

Again, the array is separated into several series, for example, O_1 to O_6, O_7 to O_{15}, and O_{16} to O_{24}. The series are related by including in a given series weights which are in both the previous series and the following series. In the first series, O_1 to O_6, for example, the 500g, 300g and 200g are always used together so that the value produced by this series is the value for the summation. In the following series, each of these weights is treated individually, subject to the restraint that the summation have the value produced by the preceding series. While the example stops at 10g, closing off with an extra 10g weight designated C, this could be replaced by a 10g summation, continuing on as appropriate.

The uncertainty computation can be illustrated by considering the uncertainty to attach to the value produced for the 10g weight. With the assumption as before, there are three series to consider, the local series which produced the value, O_{16} to O_{24}, and the two previous series which introduced restraints. The uncertainty statement becomes:

$$\text{Uncertainty 10g} = 3\sqrt{d^2 \sigma_w^2 + k_1^2 d_1^2 \sigma_{w_1}^2 + k_2^2 d_2^2 \sigma_{w_2}^2} + KE$$

where k_1 is the ratio of 10g to the summation 100g restraint from the second series or 0.1, and k_2 the ratio of 10g to the summation 1 000g from the first series or 0.01. The values for d are obtained from the design analysis for the nominal values in question. The restraint is the sum of the assumed values for Kg_1 and Kg_2 so that E becomes the sum of the associated uncertainties ($E_1 + E_2$) and K is the ratio of 10g to 2 000g. It is fairly obvious that in "working down" from the kilogram, the third term under the radical sign can usually be neglected so that the uncertainty reduces to:

$$\text{Uncertainty 10g} = 3\sqrt{d^2 \sigma_w^2 + 0.01\, d_1^2 \sigma_{w_1}^2} + 0.005\,(E_1 + E_2)$$

Similar procedures are used to calibrate weights which are multiples of the kilogram, a typical procedure being outlined in figure 7. Again, the array is divided into series, 0_1 to 0_{11} being the series which includes the starting kilograms, and 0_{12} to 0_{20} being the series which includes the largest weights of the example set. Using the generalized uncertainty formulation, the uncertainty for the 20kg weight of the set would be computed as follows: The random contribution of the local series 0_{12} to 0_{20}, is straight forward, subject to the restraint that the summation 10kg have the value of the sums of the individual values obtained in the previous series. The random contribution from the restraint must be evaluated with care. In the previous examples, the summation used as a restraint was first treated as a single weight, then treated as individual weights in the following series. "Working up" from the kilogram level the procedure is reversed. The starting series, 0_1 to 0_{11}, provides individual values for the 2kg, the 3kg, and the 5 kg weights of the set. The uncertainty of the summation is not only based on the sum of the individual variances, but also must include a covariance term because all values were determined in the same series.

The uncertainty for the value assigned to the 20kg weight can be written as:

$$\text{Uncertainty 20kg} = 3\sqrt{d^2\sigma_w^2 + k_1^2\,[d_2^2 + d_3^2 + d_5^2 + 2(d_{2,3} + d_{2,5} + d_{3,5})]\,\sigma_{w_1}^2} + KE$$

As before, k is the ratio of the nominal value of the weight in question to the nominal value of the restraint, or 2, and K is the ratio of the nominal value of the weight in question to the nominal value of the starting restraint, or 10. The uncertainty statement then becomes:

$$\text{Uncertainty 20kg} = 3\sqrt{d^2\sigma_w^2 + 4\,[\quad]\,\sigma_{w_1}^2} + 10\,(E_1 + E_2)$$

where the value between the square bracket depends on the design. It should be noted that the contribution of the local measurement to the total uncertainty series is significantly different in "working up" from "working down". In the former instance, the contribution from prior series may often be significantly larger than that from the local series, while in the latter case, the reverse is true.

6. Interpretation of the Uncertainty

Perhaps the best known mass standards, relative to the national reference standards, are the various check standards used in the National Bureau of Standards mass calibration processes. The accepted values for the check standards are the averages of more than 100 independent calibrations. It is obviously not practical to accumulate such a vast amount of data for every mass standard to be calibrated, however, the characteristics of such a procedure can be simulated by any check standard. There is nothing magic about the choice of a check standard; in fact, to properly reflect the process performance it should be similar in all respects to most of the weights which are being calibrated routinely. It then follows that if a similar standard is removed from any given set and inserted in place of the check standard, in time, one would accumulate a set of data which would have essentially the same characteristics as the set of data from the NBS check standard and from which an appropriate long term average value could be computed.

The value determined for a particular object as a result of a normal calibration is like one chosen at random from the collection of values determined for the check standard. It then follows that such a single value will rarely deviate from the as yet undetermined accepted, or long term average, value for the same object in excess of the uncertainty for that value as determined from the collection of data from the check standard. From this viewpoint, the accepted or "true" value for the object at hand will lie within the band established by the computed value plus or minus the uncertainty, say, at least 99 times out of 100. This is true, not only for the National Bureau of Standards mass calibration processes but for other processes which are in a state of control and from which the uncertainty has been computed generally in accordance with the previous discussions. It then follows that the comparison of the results of measurements on a particular object made by different processes must be judged in terms of overlapping uncertainty bands as applied to the appropriate computed values and not by the lack of agreement of the computed values alone.

One might argue that a pessimistic estimate of uncertainty will always provide overlapping uncertainty bands. While the use of excessive uncertainty limits in lieu of correcting sources of error occurs frequently, one cannot argue with such a philosophy as long as the claimed uncertainty is adequate for the measurement requirements which are to be satisfied and as long as it can be demonstrated that the actual process performance is within the claimed uncertainty. More stringent measurement requirements will eventually force more realistic process evaluations and the acceptance of realistic uncertainties. If, on the other hand, the results of measurements on the same objects from any two facilities do not produce, as a minimum, overlapping uncertainty bands, it is a clear indication of either the optimistic assignment of uncertainty by one or both facilities or the existence of a significant uncorrected systematic error in one or both measurement processes. The procedures of the Pilot Program test the above premise in the following way.

After the initial calibration of suitable starting standards, and after the participating facility has assured itself that all of the necessary equipment is in proper working order, the participating facility performs three separate independent calibrations of a selected set of weights in accordance with the prescribed procedures. Each data set is computed and analyzed by NBS to establish an estimate of the appropriate process performance parameters and values for the test set. A report of calibration which states the average value of the three values obtained for each object in the test set, and the appropriate uncertainty of the average is forwarded to the participating facility.

The test set and the starting standards are then returned to the National Bureau of Standards for re-calibration with reference to the national reference standards. The values obtained for the starting standards are combined with those of previous calibrations to provide a better estimate of the appropriate values and uncertainties. The test is considered satisfactory if the uncertainty bands associated with the National Bureau of Standards measurement processes, centered on the values obtained by this process, overlap the uncertainty bands obtained by the participating facility centered on the values obtained by their measurement process.[4] The results obtained in some of these tests are shown in figures 8 and 9.

The initial estimates of the appropriate values for the process performance parameters, being based on three calibrations or approximately twelve to fifteen degrees of freedom for error, may require some adjustment after additional performance data becomes available. This data is provided for the participating facility by the use of the Pilot Program procedures in performing normal measurement tasks such as the calibration or recalibration of appropriate mass standards for their own use or for use by others. If the process is stable, perhaps, after approximately twenty independent sets of data have been accumulated, no further significant changes in the process performance parameters would be expected. Additional data sets provide the assurance that the process remains in a state of control and allow updating the parameter on a yearly or some other periodic basis.

The magnitude of realistic uncertainty estimates provide a measure of the degree of consistency obtainable in the total mass measurement system. It is unfortunate that far too much emphasis is placed on the magnitude of the uncertainty to be associated with a measured value rather than on the uncertainty relative to the manner in which the calibrated objects are to be used. The magnitude of an uncertainty is a function of the manner in which the measurement process is defined. Procedures, such as those of the Pilot Program, provide realistic uncertainty estimates for a reasonable measurement effort, the magnitude of which can be reduced by increasing the measurement effort, or increased by using simplified procedures. While it is necessary to establish some point of departure, the "best" measurement processes are not necessarily the ones with the smallest uncertainties but rather those which meet the particular measurement requirements with a minimum of effort. Fundamental to the task of matching the measurement effort to meet realistic requirements is confidence in the performance of the local measurement process, such as provided by the control chart techniques of the Pilot Program, and an understanding of the manner in which the uncertainty statement is formulated. All measurements based on standards for which the assigned values are assumed to be fixed have uncertainty statements of the form:

$$\text{Uncertainty} = f(\sigma) + \text{Systematic Error}$$

where $f(\sigma)$ refers to the local measurement process, and the systematic error term is the cumulative effect of all of the measurements which have gone before to establish the assumed value of the standard. The relative magnitude of the two components is of interest.

Assuming for the moment that $f(\sigma)$ is satisfactory for the job at hand, one must consider the magnitude of $f(\sigma)$ relative to the systematic error. Consider, for example, the problem of establishing the value for a 20g weight relative to a 20g standard which has an assigned value and an attached uncertainty. As has been stated earlier, the uncertainty band associated with the value would encompass, in the absence of change, almost all of the values which would have been obtained if the calibration process had been repeated a great many times. If $f(\sigma)$ for the local measurement, such as in figure 3a, is large in comparison to the uncertainty obtained by a process, such as in figure 3c, the contribution of the systematic error to the total uncertainty is insignificant. With regard to possible changes in the standards, perhaps from wear or damage, only those of sufficient magnitude to effect the local process uncertainty are of interest. While periodic recalibration could easily detect changes on the order of the magnitude of the reported uncertainty, such changes would still be insignificant with reference to the manner in which the standards are used. On the other hand, the use of a different procedure, such as shown in figure 3c, all other things being equal, will frequently provide sufficient precision to monitor standards for possible changes which might be significant relative to the routine measurement process.

It may well be that $f(\sigma)$ for the local process is of the same order of magnitude, or perhaps smaller, than the reported uncertainty for the value established by another laboratory, for example, a calibration laboratory. In this instance, there is little to be gained

- - - - - - -

[4] The degree of overlap required for an exact statistical test depends on allowance for systematic error, the process parameters, etc.

by routine recalibration in the same manner as was used to establish the original value. The task at hand becomes that of reducing the magnitude of the systematic error associated with the value for the standard at hand. This requires the minimization of the contribution of prior measurement processes which use fixed values for starting standards as restraints. This includes, in the case of the 20g standards, the original "working down" to the 20g level from the kilograms as well as all one-to-one measurement processes leading to the value of the standard at hand. If sufficient precision is available, the local facility could be handicapped by not using procedures such as those incorporated in the Pilot Program to "work down" from selected kilogram standards in their own facility. This is not to say that all measurements must be made in this manner for such precision may be far smaller than necessary for most routine measurement. In such cases, however, the reliance on routine recalibrations by others becomes unnecessary when the local process is perfectly capable of both monitoring and recalibrating as necessary with no significant increase in uncertainty.

While the precise measurement process considers the closeness of value to some even nominal value, the closeness is only a requirement to assure "on scale" value. Generally many users are only concerned about the extent of the deviation from nominal. This situation causes considerable confusion as to the role of the calibration laboratory, the role of the weight and weighing equipment manufacturer, and the role of the existing class adjustment tolerance structure. For example, many calibration requests ask for the establishment of values for the objects submitted so that someone can establish the compliance or non-compliance with appropriate class adjustment tolerances by looking at a report of calibration. Such practice directs the attention to the assignment of a number generally without regard to the uncertainty of the number and without any intention of ever using the number after the compliance requirement has been satisfied. The mass calibration procedures of the National Bureau of Standards and the procedures of the Pilot Program are designed to determine the "best" mass values for individual objects, and to provide realistic estimates of the uncertainties of these values. Any weight of reasonable quality which has been calibrated by the use of these procedures is adequate for establishing a point of reference with sufficient precision to determine the compliance with any established class adjustment tolerance.

To further illustrate, figure 10 lists the accepted adjustment tolerances for the various class weights in use in this country. The rows are in order of increasing allowed deviation from nominal with the class designation in the columns as appropriate. While the lack of order between the established classes may not be particularly attractive from an esthetic point of view, these classes were established many years ago and are accepted as adequate for a wide variety of uses. The first column lists typical precisions of measurements for a set of instrumentation and a defined process (essentially a double substitution weighing using the equipment available at the National Bureau of Standards). The second column lists the uncertainty that might be expected from a single calibration using the procedures of the Pilot Program. It should be apparent that a calibrated value, which is "on scale", will provide a reference point from which the "correction" can be stepped off in terms of scale divisions per mass unit to establish the "on scale" location of the nominal value far more precisely than can be established by weights from any class which are only known to be "within tolerance". Such practices might also be advantageous in weighings other than those used to test weights.

The class adjustment tolerance structure provides a mechanism whereby certain types of information can be obtained through the use of greatly simplified measurement practices. The formulation of such a structure requires many compromises which are not always appropriate to the measurement task at hand. There are many subtle relationships concerning the economics of the manufacturing of weights and weighing equipment, the generally available precision of the equipment and processes in current use, and the factors which the user can safely ignore and still accomplish his assigned task. It may appear desirable, from the viewpoint of a calibration laboratory for example, to establish a universal basis for stating apparent mass values, or to establish new classes of weights with smaller allowed deviations from nominal, or perhaps to restrict the material for all weight sets to one carefully controlled alloy. While such actions might simplify the tasks of the particular calibration laboratory, they are of no particular benefit to the users, and may perhaps be disastrous to the manufacturers. The user has no control over the objects which must be weighed in the course of accomplishing a certain task, therefore, the basis for stating the value, the density of the material, and so on, are immaterial as long as the information is specified correctly. The manufacturer, first and foremost, must serve the user by marketing convenience at a competitive price. It is a matter of economic necessity that performance claims and compliance with class adjustment tolerances be demonstrable by a wide variety of measurement processes.

Certain weight classes have two tolerance limits, one of which is designated an acceptance tolerance, and one of which is designated a maintenance tolerance. While the manufacturer will, in general, try to adjust the weight as close to nominal as practical, later testing with a process of questionable precision may produce a value near or beyond the acceptance tolerance limits. Re-adjustment is absurd if the process performance parameters are not well known. While maintenance tolerances are generally associated with some sort of "wear" allowance, it is now apparent that such tolerances also serve to compensate for the variability of measurement processes used to test the items in question. It seems clear that refinement by compromise will only benefit isolated measurement areas, general improvement in the consistency of mass measurements in all areas must be based on a complete understanding of the measurement process. This is particularly acute since the advent of the precise rapid weighing now generally available. The deviations from nominal of the weights built into these instruments are far smaller than allowed by the class adjustment tolerance structure. Further, the precision of the instruments is such that differences in the basis for stating values is clearly discernible. Procedures which approach the evaluation and usage of such equipment through the class adjustment tolerance structure are doomed to failure from the start.

The uncertainty as determined by the Pilot Program procedures is applicable to apparent mass correction as well as the mass value. The observed differences in the various weighings include the buoyant effects of the local environment. This effect is treated as if it were a small "negative weight" being carried along with the object under test and with a mass proportional to the displacement volume of the test object and the density of the local environment. It can be shown that this assumption is valid relative to the precision of the local measurement process as long as the changes in pertinent variables in the course of an intercomparison do not exceed certain predetermined limits. Since practical measurement processes can never effect a complete separation between the mass of the test object, and the buoyant effect of the local environment, some point of reference must be clearly established to provide the basis for measurement agreement. In the early 1800's, based on the best practices of the time, the practice was introduced whereby the mass of a test object would be given a value, for example 100g, if it "exactly balanced" a brass standard of 100g known value in the prevailing environment. Over the years this practice has been refined to a more precise specification for the apparent mass value, with the added complication that every object has an infinite number of apparent mass values depending upon choice of reference material and defined environments. The mass, or true mass values, in the Pilot Program are defined on a basis which is independent of choice of material and from which the apparent mass value on any basis can be readily calculated.

The values obtained from the Pilot Program are those which account for the "negative weight" whose value is computed from the values provided for the density of the weight material and from the computed density of the local environment. These values are fictitious in the same sense as the apparent mass values in that they can never be verified in a practical measurement process. These values contain a systematic error which originates in the value used for the density of the weight material, however, the magnitude of the error becomes real only through evidence of changes in measured mass differences which appear to relate with the computed environment density values. In most measurements this effect is a second order one if the density values of the objects concerned have been established with reasonable care. A feature of this procedure, the ability to recompute values on the basis of better knowledge of the density of the material, without repeating the measurements, may be a source of confusion relative to the claimed uncertainty of the value.

Suppose, for example, the value for a given mass standard has been computed on the basis of an assumed density A. As long as future calibrations also use the same value for the density, agreement within the uncertainty band would be expected. If, however, on the basis of further study, it was decided that B is a better value for the density of the material, such further study perhaps being the completion of a hydrostatic weighing, or the analysis of many measurements which exhibit a correlation between computed values and environmental variables, the new value may or may not differ from the original value in excess of the stated uncertainty. Again repetitions would result in a series of values which would be encompassed by the process uncertainty limits. Since the particular effect is only one of many which collectively establish the process performance parameters and, in turn the uncertainty limits, it is rather unlikely that the refinement of the value used for the density of the material would significantly affect the uncertainty limits immediately. If, on the other hand, the process precision becomes significantly better, or the difference in density of the objects being intercompared increases, or the range of environmental conditions over which the measurements must be made increases, it may be necessary to refine both the values used for the density of the materials, and the manner in which the density of the environment is computed to obtain an "in control" measurement process.

7. Surveillance Tests

Surveillance test procedures, while not directly a part of the Pilot Program, are a part of the Pilot Program philosophy. For those who look to others for calibration service, the problem of the continuing validity of the values contained in a report of calibration is ever present. Recalibration on a more or less fixed time schedule does not necessarily provide assurance that all is as it should be. If, on recalibration, a significant change has occurred, all measurements which have been completed in the interval are suspect. It is important to establish some sort of monitoring procedure so that gross changes can be detected as soon as possible. It is recommended, for example, that immediately upon receipt of a newly calibrated set of weights, intercomparisons should be made to verify the values reported. This practice is perhaps contrary to policies which seemingly conserve the accuracy of the standards through careful storage and infrequent use. While careful usage must be emphasized, frequent usage is also necessary if a real confidence in the values is to be established.

The calibration of an ordered set of weights by means of the Pilot Program procedures results in a set of values which are largely based on measurements between various summations within the set. The uncertainty of the values is a measure of the ability of the calibration process to produce identical values on recalibration. In other words, a recalibration, in the absence of real change will produce another set of values wholly within the established uncertainty bands, and if the values are considered individually, the values from the second calibration are no better or worse than those from the first calibration. Under certain circumstances, it may be desirable to average the two values and compute an appropriate uncertainty, however one would have considerably more assurance in the average of a larger group. If the values from the second calibration are as expected, the first values might well be retained until there is evidence of change, or until a sufficient number of values have been established so that the uncertainty of the average reflects a significant improvement over the uncertainty of the individual value.

The surveillance test procedures are, generally speaking, the calibration procedures in reverse. Given an ordered set of mass standards and the reported values, verify by intercomparison of selected summations that the values are "correct" within the precision of the local measurement process. The uncertainty of the stated value is of little consequence unless the precision of the local measurement process is on the same order of that used to establish the initial values for the standards. As a rule the first task is to establish a measure of the precision of the local measurement process. In the absence of information as provided by the Pilot Program, immediately upon receipt of a newly calibrated set of weights, the local facility should start a daily measurement of the mass difference between one weight of the set and a summation from the set of equal nominal value, using a procedure such as the double substitution method previously described. The difference as determined by the local measurement can be subtracted from the expected difference as computed from the reported values to establish a residual which in turn can be plotted in the order taken following the usual control chart techniques. As soon as it has been established that the process is reasonably well behaved, the time between intercomparisons might be extended somewhat, perhaps finally being done only as an operation prior to the use of the standards for other purposes.

A relatively few measurements can be devised which will include all of the weights of the set. For example, a 100g weight can be compared with the summation 100g, a 10g with the summation 10g and so on. If after a few measurements have been made it appears that there might be gross discrepancies, these should be brought immediately to the attention of the calibration facility. Such occasions should be quite rare. After it has been established that the local process is operating in a state of control as determined by the random appearance of the collection of repeated measurements, any intercomparison which produces a difference significantly different from that which is expected is clear evidence of change. It is possible to identify the suspect by varying the intercomparison procedure. For example, if the 100g does not agree with the 100g summation, one can test the 50g against the summation 50g, and the 30g against a 30g summation, and the 20g against a 20g summation. From these weighings it should be possible to identify the weight which has changed and to establish some idea of the magnitude of the change. If the change persists after cleaning, a new value can be assigned on the basis of this work for temporary use until it is convenient to have the set checked by the calibration laboratory.

While the surveillance tests must be made with care, they can be made in a short period of time. Additional data can be accumulated with little effort by adopting a routine procedure whereby selected intercomparisons are made immediately prior to the use of the calibrated weights for other purposes. Control charts provide a continuous record to verify the constancy of the standards and to provide a basis for decisions concerning recalibration. It should be emphasized that surveillance tests do not verify that no changes have occurred but rather that changes, if any, which might have occurred are not of sufficient magnitude to be of concern relative to the local measurement process. In essence, the same procedures are used in surveillance tests conducted in the mass laboratory at the National Bureau of Standards.

With good estimates of the magnitude of the process performance parameters, it is possible to compute definite limits for surveillance test procedures. Surveillance tests at the Bureau are made with reference to these limits, and with reference to one or more checks against other standards to guard against undetected "drifting" of the entire set. Many sets which are submitted for recalibration are first checked by surveillance tests. If it is established that no significant change, relative to the precision of the Bureau mass measurement processes, has occurred, the previous values are continued in the new report. If one, or perhaps two, items have changed, new values are generally reported for these items, with the previous values being continued for other items. Frequently, changes are only apparent in the smaller weights of the set, for which new values are established by the normal calibration procedures. If surveillance tests show that several changes have occurred, it becomes more efficient to recalibrate the entire set.

8. Pilot Program in Operation

The Pilot Program is still essentially an experimental operation subject to limited expansion until the development of more efficient means for processing the data is completed. While formal requirements for participation have not been established, certain factors must be considered in the acceptance of additional participating facilities. Inasmuch as the participating facilities are essentially extensions of certain portions of the mass calibration facilities of the National Bureau of Standards, these facilities should be those which provide such services to others or those which have certain unique measurement requirements that justify the use of such procedures to successfully accomplish the task at hand. The participating facilities should be those who can apply the knowledge and process performance data that is obtained from the Pilot Program procedures to the establishment of realistic measurement requirements. It is expected that the participating facilities will assist in the refinement of the program by being completely objective in the interchange of information with the Bureau and other participating facilities concerning measurement requirements and process performance. Each facility may be asked from time to time to provide certain calibration services or to evaluate certain procedures which will contribute to the establishment of a consistent, largely self-supporting, mass measurement system which will meet the requirements of the nation.

There are three more or less defined phases in the adoption of the Pilot Program procedures. The initial phase is essentially devoted to training, testing of equipment, and establishing a line of communications. Those facilities which are not familiar with the intercomparison methods of calibration will be asked to become familiar with these methods through the use of somewhat simplified versions which must be hand-computed. During this phase, the starting standards must be obtained and calibrated. The suggested equipment tests should be performed to establish that all is in order prior to starting the next phase. The second, or verification phase, consists of the several calibrations performed by the participating facility and repeated by the mass laboratory of the National Bureau of Standards. The data from these tests provide the starting estimates of the process performance parameters, a well calibrated ordered set of mass standards in addition to the starting standards, and the starting data for suitable control charts. The last phase is the operational phase in which the participating facility uses the Pilot Program procedures as necessary to conduct his daily business. The data forwarded to the Bureau is computed and analyzed. If all is in order, the body of a report is returned as soon as possible, followed by a copy of the analysis. If there are indications of loss of control, the facility will be directed to repeat certain series, or perhaps an entire calibration, the body of the report being released only when all is in order.

A typical report might be as shown in figure 11. The first part of the report, in addition to identifying the transmission number and participating facility, identifies the objects under test, states the basis for computing the volumes of the test objects, and identifies the starting standards and the source for the values used as the restraint. The body of the report states the mass value, the uncertainty, and the volume. The apparent mass correction is listed as a convenience for certain usage. While it is not explicitly stated, the apparent mass value is with reference to normal brass (density 8.4 grams per cubic centimeter at 0°C) and to a defined environment with air density of 0.0012g per cubic centimeter at 20°C. The last part of the report contains the explanation of the uncertainty The participating facility is free to transcribe this report on a format of choice subject only to the restriction that the continuity of documentation be retained through suitable references.

A typical analysis might consist of two sheets such as shown in figures 12 and 13. The first part of the analysis states who the calibration is for, the date of the observations, and identifies the series, the operator, the instrument, the type of weights, and so on. The weighing design is shown in detail. The observations are listed essentially in the order in which they were taken. The computed observed differences for each weighing is listed together with the residual, delta. The computed standard deviation for this calibration, the long term average standard deviation, the value obtained for the check standard and the accepted value of the check standard are listed together with the data used to compute the air density. The second sheet lists the values obtained for the weights in the particular series. The F ratio provides a test for the significance of the difference between the standard deviation computed for the particular series of weighings and the accepted long term standard deviation. The DIFF/S.D. OF DIFF provides a t-test for the difference between-the-value obtained for the check standard and the long term accepted value. Other data might also be listed to facilitate correlation studies to determine the effect of various variables on the total process variability.

Inasmuch as certain elements of the program require a very close relation with the production effort of the participating facilities, every effort will be made to provide prompt service for computation and analysis, and for the recalibration of starting standards should accidental damage occur. In the event of gross difficulties, consulting services will be made available. Copies of data, computations, and analysis sheets will be maintained at the Bureau to provide revised estimates for process performance parameters, and to provide information pertinent to the evaluation of the mass measurement process performance in various locations. From time to time, conferences will be held to discuss the results of studies which indicate desirable changes in procedure, to discuss areas requiring further study, and to discuss problems concerning the acceptance of the program in lieu of conflicting subjective instructions.

The major difference between the Pilot Program procedures and other widely used calibration procedures is a clear differentiation between taking the data, or the operational phase, and the computing of the values, or the analysis phase. The data is considered acceptable (or the repeat of an entire series is called for) on the basis of the consistency of the entire group of measurements. The performance of the group is compared to that of other groups generated by the process, as contrasted to one's taking action item by item while making the measurements. One of the hardest problems in adopting the Pilot Program procedures is an almost universal tendency to prejudge data in accordance with personal opinion frequently formulated from incomplete analysis of insufficient data sets, or in accordance with arbitrary rules which are based on vague generalities. All data must be accepted, except that which can clearly be rejected for cause such as an unusual local disturbance or obvious instrument malfunction, so that the performance parameters can reflect the actual performance rather than the expected performance. Once realistic performance parameters have been established for a fixed procedure, they form a basis for the evaluation of procedural changes, environmental effects, and many other elements of the measurement process. Obviously the process is not sensitive to changes from any source which does not significantly affect the magnitude of the performance parameters.

While the use of procedures, such as are incorporated in the Pilot Program, involves an increased effort in the measurement process, the benefits are many. The procedures provide a means to utilize the full capabilities of the mass measurement processes now available. Measurement decisions can be based on realistic estimates of the uncertainty of the stated value. The accumulative records in notebooks or control chart form provide irrefutable evidence of process performance. If the correlation between the records and the process is ever in question, the validity of the records can be demonstrated almost at will. All facilities which adopt the procedure will have a common basis for discussion in mass measurement problems. The Pilot Program procedures should not lead to a race for the smallest uncertainty. Values for the same object from two sources in which the uncertainties do not overlap is a clear indication of lack of realistic control by one or both of the facilities involved. The emphasis of the program is on the total performance of the measurement process, and if one is not able to repeat values for his own standards such as the check standards within the predicted limits, he is not in a position to do such work for others.

This paper is presented as a general description of an objective approach to the measurement problem. Many points which are mentioned will be discussed in much greater detail in papers to follow, written by those who are considerably more knowledgeable in their respective areas than I. The Pilot Program for mass measurement, having been conceived something less than a year ago, reflects the efforts of many people across the country who are actively interested in an objective approach to measurement problems. The assistance of J. M. Cameron, extending over several years, has made such a program possible. The participation of the Calibration Laboratory of White Sands Missile Range, the Navy-Eastern Standards Laboratory and the Calibration Laboratory of Redstone Arsenal is acknowledged. The cooperation and suggestions of Mr. Gordon Anderson of White Sands Missile Range are particularly acknowledged. At the present state of development, the Pilot Program is a joint product of the Statistical Engineering Laboratory and the Mass and Volume Section of the National Bureau of Standards.

References

(1) Eisenhart, Churchill - Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems, Journal of Research of the National Bureau of Standards, Vol. 67C, No. 2, April-June, 1963.

(2) Hayford, John A. - On the Least Square Adjustment of Weighings, U. S. Coast and Geodetic Survey Appendix 10, Report for 1892 (1893).

(3) Benoit, M. J. R. - Travaux et Memoirs du Bureau International des Poids et Mesures 13, 1, (1907).

(4) Cameron, J. M. and Raybold, R. C. - Weighing Designs for Transposition or Double Substitution Weighings (in preparation).

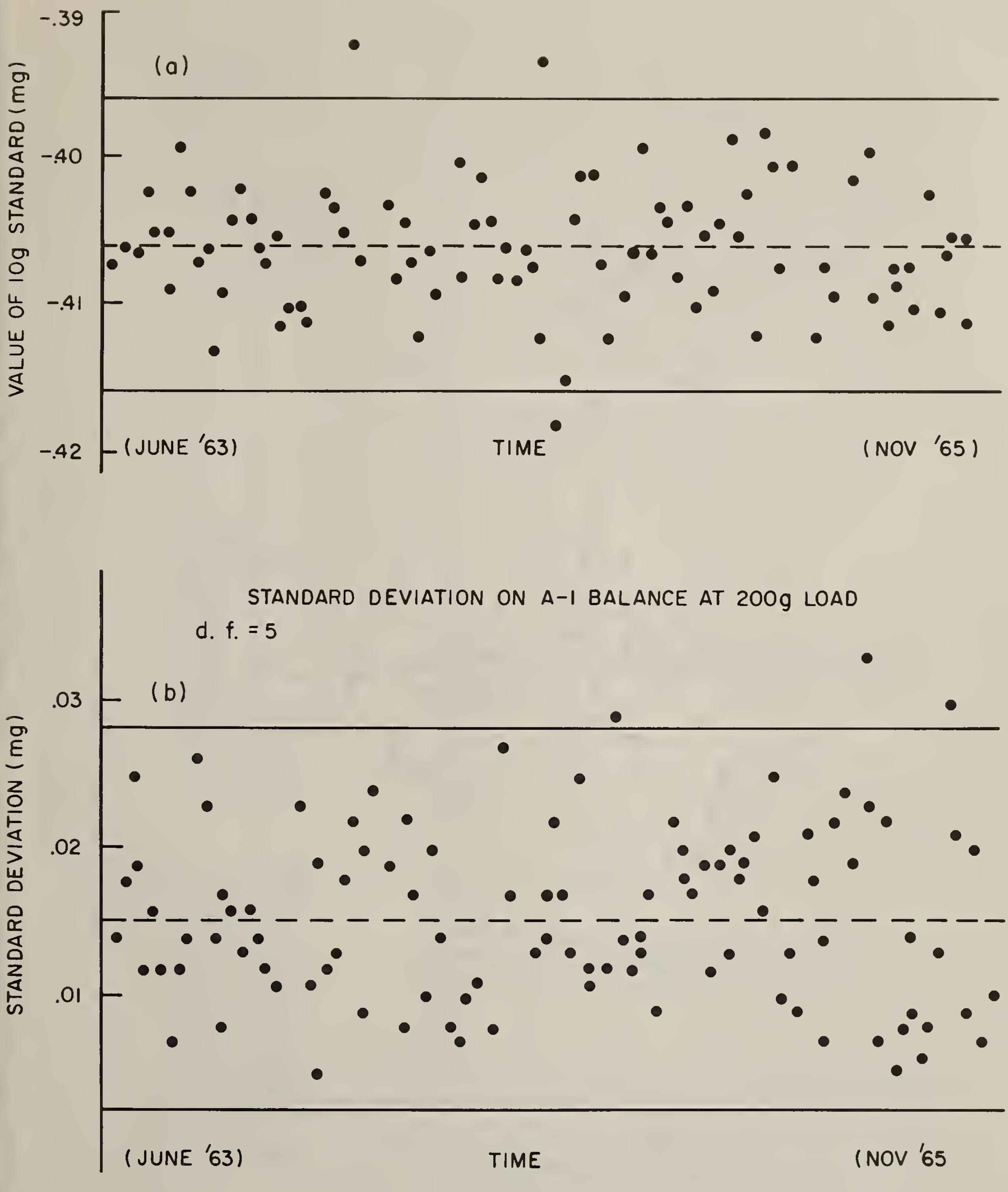

FIGURE 1 Typical control charts for the mass measurement process performance parameters.

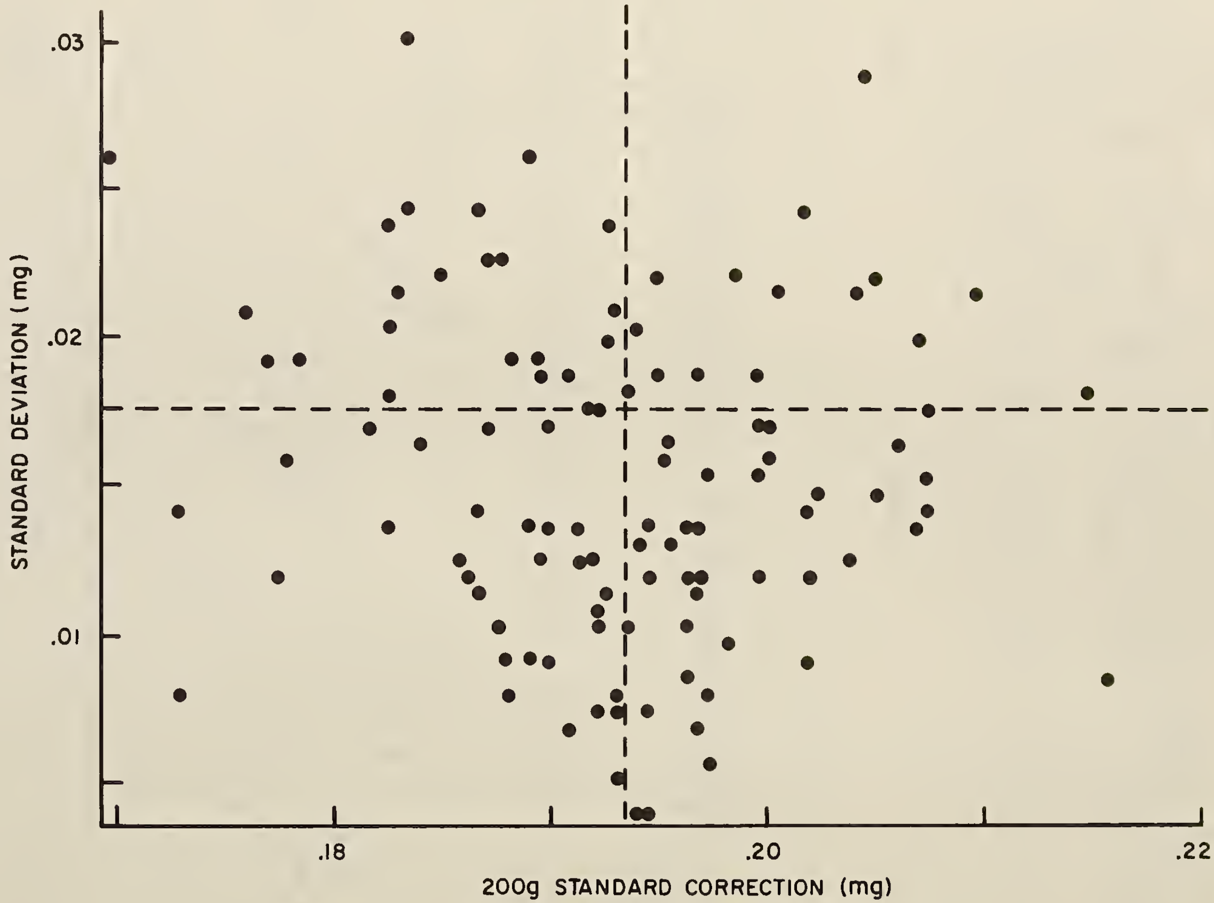

FIGURE 2 Cross plot of the values obtained for a "check" standard and the standard deviation of the appropriate observations.

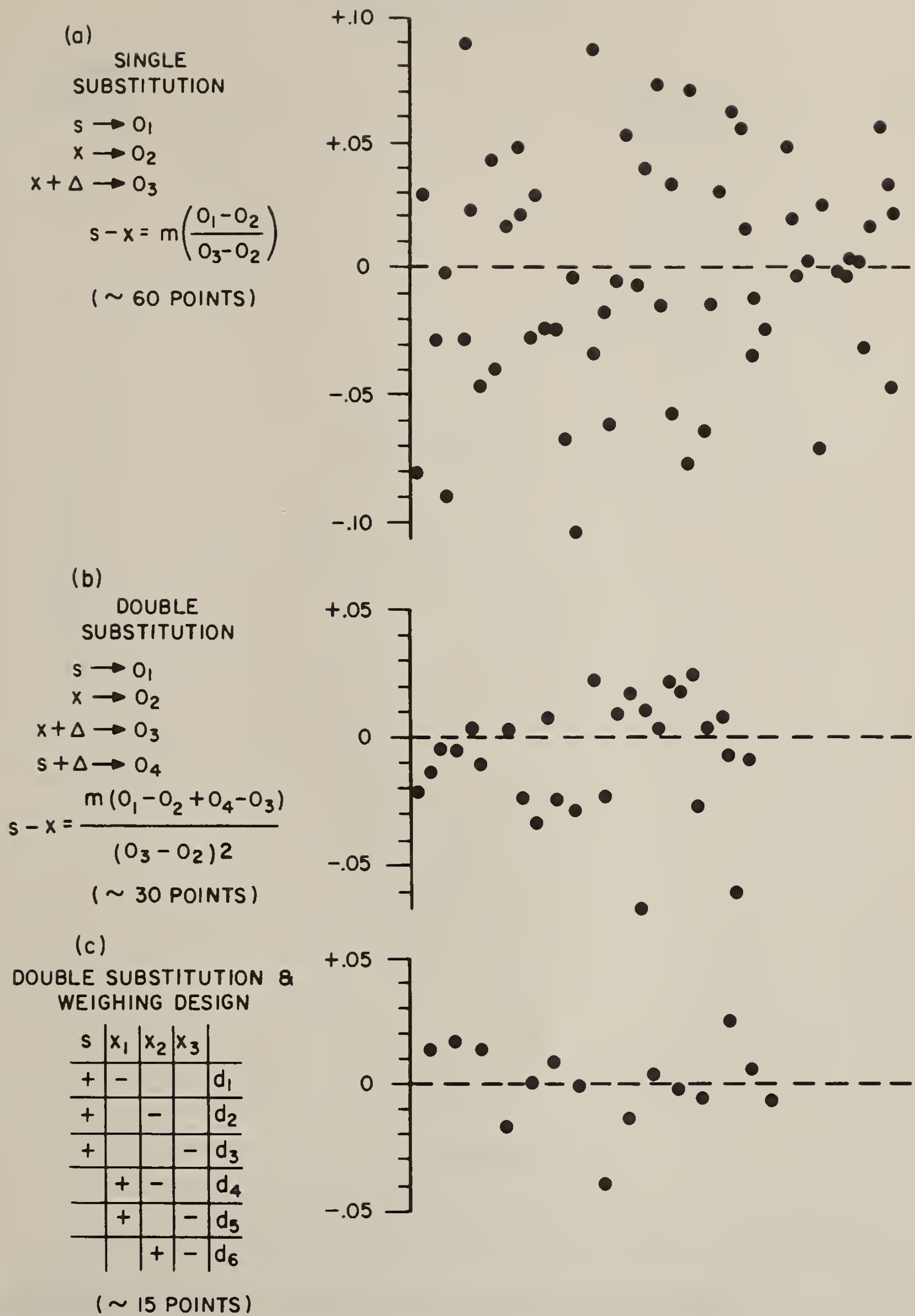

FIGURE 3 Process precision as a function of defined measurement methods.

Nominal Mass Values

K_{20}	K_4	Kg_1	Kg_2	K_a	K_b	K_c	K_d	OBSERVED DIFFERENCE
+	-							O_1
+		-						O_2
+			-					O_3
	+	-						O_4
	+		-					O_5
		+	-					O_6
		+	-					O_7
		+		-				O_8
		+			-			O_9
			+	-				O_{10}
			+		-			O_{11}
				+	-			O_{12}
				+	-			O_{13}
				+		-		O_{14}
					+	-		O_{15}
						+	-	O_{16}
						+	-	O_{17}
						+	-	O_{18}

FIGURE 4 Weighing schedule through four echelons at the one kilogram level

Nominal Mass Values

(1st Series)

K_{20}	K_4	Kg_1	Kg_2	OBSERVED DIFFERENCE
+	-			O_1
+		-		O_2
+			-	O_3
	+	-		O_4
	+		-	O_5
		+	-	O_6

(2nd Series)

Kg_1	Kg_2	K_a	K_b	
+	-			O_7
+		-		O_8
+			-	O_9
	+	-		O_{10}
	+		-	O_{11}
		+	-	O_{12}

(3rd Series)

K_a	K_b	K_c	
+	-		O_{13}
+		-	O_{14}
	+	-	O_{15}

(4th Series)

K_c	K_d	
+	-	O_{16}
+	-	O_{17}
+	-	O_{18}

FIGURE 5 Weighing schedule of figure 4 separated into appropriate series for each of four schedules.

Nominal Mass Values

Kg_1^*	Kg_2^*	1 kg	500g	300g	200g	100g	100g**	50g	30g	20g	10g	10g**	C	OBSERVED DIFFERENCE
+	-													O_1
+		-												O_2
+			-	-	-									O_3
	+	-												O_4
	+		-	-	-									O_5
		+	-	-	-									O_6
			+	-	-	+	-							O_7
			+	-	-	+		-	-	-				O_8
			+	-	-	-	+							O_9
			+	-	-		+	-	-	-				O_{10}
			+	-	-		-	+	+	+				O_{11}
				+	-		-	-	-	-				O_{12}
					+	-	-							O_{13}
					+	-		-	-	-				O_{14}
					+		-	-	-	-				O_{15}
								+	-	-	+	-		O_{16}
								+	-	-	+		-	O_{17}
								+	-	-	-	+		O_{18}
								+	-	-		+	-	O_{19}
								+	-	-		-	+	O_{20}
									+	-		-	-	O_{21}
										+	-	-		O_{22}
										+	-		-	O_{23}
										+		-	-	O_{24}

Notes:

1. Single asterik designates standards of known value.
2. Double asterik designates "check" standards.
3. C is extra 10g weight needed to "close" the design illustrated.
4. All other weights are from the set being calibrated.

FIGURE 6 Weighing schedule for one kilogram to ten grams

Nominal Mass Values

Kg_1^*	Kg_2^*	1 kg	500g	300g	200g	100g	100g**	50g	30g	20g	10g	10g**	C	OBSERVED DIFFERENCE
+	-													O_1
+		-												O_2
+			-	-	-									O_3
	+	-												O_4
	+		-	-	-									O_5
		+	-	-	-									O_6
			+	-	-	+	-							O_7
			+	-	-	+		-	-	-				O_8
			+	-	-	-	+							O_9
			+	-	-		+	-	-	-				O_{10}
			+	-	-	-		+	+	+				O_{11}
			+	-	-		-	+	+	+				O_{12}
				+		-	-	-	-	-				O_{13}
					+	-	-							O_{14}
					+	-		-	-	-				O_{15}
					+		-	-	-	-				O_{16}
								+	-	-	+	-		O_{17}
								+	-	-	+		-	O_{18}
								+	-	-	-	+		O_{19}
								+	-	-		+	-	O_{20}
								+	-	-	-		+	O_{21}
								+	-	-		-	+	O_{22}
									+		-	-	-	O_{23}
										+	-	-		O_{24}
										+	-		-	O_{25}
										+		-	-	O_{26}

Notes:

1. Single asterisk designates standards of known value.
2. Double asterisk designates "check" standards.
3. C is extra 10g weight needed to "close" the design illustrated.
4. All other weights are from the set being calibrated.

FIGURE 6 Weighing schedule for one kilogram to ten grams

Nominal Mass Values

Kg_1^*	Kg_2^*	$1k^{**}$	2k	3k	5k	$10k^{**}$	10k	20k	$20k^{**}$	OBSERVED DIFFERENCE
-	-		+							O_1
-		-	+							O_2
	-	-	+							O_3
+	-	-	-	+						O_4
-	+	-	-	+						O_5
-	-	+	-	+						O_6
-	-	-	-		+					O_7
			-	-	+					O_8
+		-	-	-	+					O_9
-	+		-	-	+					O_{10}
	-	+	-	-	+					O_{11}
			+	+	+	-				O_{12}
			+	+	+		-			O_{13}
						+	-			O_{14}
			+	+	+	+		-		O_{15}
						+	+	-		O_{16}
			+	+	+		+	-		O_{17}
			+	+	+	+			-	O_{18}
						+	+		-	O_{19}
			+	+	+		+		-	O_{20}

Notes:

1. Single asterisk designates standards of known value.
2. Double asterisk designates "check"standards or "fill-in" weights to complete series.
3. All other weights are from the set being calibrated

FIGURE 7 Weighing schedule for one kilogram to ten kilograms

Nominal Mass Values

Kg_1^*	Kg_2^*	1k**	2k	3k	5k	10k**	10k	20k	20k**	OBSERVED DIFFERENCE
-	-		+							O_1
-		-	+							O_2
	-	-	+							O_3
+	-	-	-	+						O_4
-	+	-	-	+						O_5
-	-	+	-	+						O_6
-	-	-	-		+					O_7
			-	-	+					O_8
-		-	-	-	+					O_9
-	+		-	-	+					O_{10}
	-	+	-	-	+					O_{11}
			+	+	+	-				O_{12}
			+	+	+		-			O_{13}
						+	-			O_{14}
			+	+	+	+		-		O_{15}
						+	+	-		O_{16}
			+	+	+		+	-		O_{17}
			+	+	+	+			-	O_{18}
						+	+		-	O_{19}
			+	+	+		+		-	O_{20}

Notes:

1. Single asterisk designates standards of known value.
2. Double asterisk designated "check"standards or "fill-in" weights to complete series.
3. All other weights are from the set being calibrated

FIGURE 7 Weighing schedule for one kilogram to ten kilograms

WSMR SUMMARY OF PILOT PROGRAM

Nominal Value	Range of 3 values	Average of 3 values	Uncertainty of the average	Avg - NBS	NBS 1963	Uncertainty of NBS value
1 kg	0.688 mg	11.213 mg	0.426 mg	0.302 mg	10.911 mg	0.110 mg
500 g	0.386 "	5.556 "	0.235 "	0.151 "	5.405 "	0.081 "
300 "	0.239 "	3.466 "	0.285 "	0.116 "	3.350 "	0.103 "
200 "	0.298 "	2.360 "	0.265 "	0.089 "	2.271 "	0.073 "
100 "	0.197 "	0.954 "	0.183 "	0.084 "	0.870 "	0.070 "
50 "	0.032 "	0.638 "	0.092 "	-0.016 "	0.654 "	0.009 "
30 "	0.038 "	0.362 "	0.058 "	0.007 "	0.355 "	0.013 "
20 "	0.0432 "	0.3076 "	0.0402 "	-0.0052 "	0.3128 "	0.0122 "
10 "	0.0124 "	0.1553 "	0.0227 "	-0.0012 "	0.1565 "	0.0087 "
5 "	0.0191 "	0.1021 "	0.0112 "	0.0027 "	0.0994 "	0.0050 "
3 "	0.0138 "	0.0764 "	0.0091 "	0.0009 "	0.0755 "	0.0070 "
2 "	0.0053 "	0.0893 "	0.0076 "	0.0016 "	0.0877 "	0.0066 "
1 "	0.0044 "	0.0036 "	0.0050 "	-0.0006 "	0.0042 "	0.0047 "
Σ1 "	0.0078 "	-0.1417 "	0.0050 "	-0.0035 "	-0.1382 "	0.0047 "

FIGURE 8 Summary of Initial Pilot Program Tests
(White Sands Missile Range Calibration Division)

NAVY-ESL PILOT PROGRAM (revised)

Nominal Value	Range of 3 values	Average of 3 values	Uncertainty of the average	Avg - NBS	NBS 1965	Uncertainty of NBS value
1 kg	0.03 mg	12.60 mg	0.16 mg	.156 mg	12.444 mg	.090 mg
500 g	0.03 "	5.57 "	0.09 "	.134 "	5.436 "	.058 "
300 "	0.03 "	3.16 "	0.09 "	.098 "	3.062 "	.101 "
200 "	0.03 "	2.31 "	0.08 "	.041 "	2.269 "	.098 "
100 "	0.019 "	.893 "	0.055 "	.047 "	.846 "	.070 "
50 "	0.034 "	.477 "	0.027 "	-.004 "	.481 "	.036 "
30 "	0.025 "	.258 "	0.018 "	-.003 "	.261 "	.025 "
20 "	0.013 "	.153 "	0.013 "	.003 "	.150 "	.019 "
10 "	0.002 "	.074 "	0.007 "	.008 "	.066 "	.012 "
5 "	0.0068 "	.0635 "	0.0039 "	-.0027 "	.0662 "	.0062 "
3 "	0.0047 "	.0082 "	0.0030 "	.0040 "	.0078 "	.0045 "
2 "	0.0010 "	.0163 "	0.0027 "	.0023 "	.0140 "	.0038 "
1 "	0.0016 "	.0154 "	0.0028 "	.0008 "	.0146 "	.0037 "

FIGURE 9 Summary of Initial Pilot Program Tests (Navy Eastern Standards Laboratory)

Nominal Value	NBS MASS MEASUREMENT		LIMIT FOR DEVIATION FROM NOMINAL VALUE (METRIC WEIGHT) ON APPARENT MASS BASIS (Class designations parenthesized)							
	Typical Precision at Nominal Load	Typical Uncertainty of Stated Value								
			(S)	(M)	(A,B)	(S-1)	(P)	(C)	(Q)	(T)
25 kg			62 mg	125 mg					1.2 g	4.5 g
20	2.5 mg	18.0 mg	50	100	120 mg	200 mg	400 mg	600 mg	1.0	3.8
10	1.5	12.0	25	50	80	100	200	400	500 mg	2.2
5	1.5	9.5	12	25	50	50	100	250	250	1.4
3	1.5	6.4	7.5	15	-	30	60	-	150	1.0
					(S-1)	(A,B)		(Q)	(C)	
2		5.0	5.0	10	20	30	40	100	150	750 mg
1	.060	.11	2.5	5	10	20	20	50	100	470
						(P)	(A,B)			
500 g	.060	.081	1.2	2.5	5.0	10	14	30	70	300
300	.060	.073	.75	1.5	3.0	6	-	20	-	210
200	.013	.073	.5	1.0	2	4	8	15	40	160
100	.013	.0154	.25	.5	1	2	6	9	30	100
50	.0074	.0090	.12	.25	.6	1.2	4	5.6	20	62
30	.0074	.0129	.074	.15	.45	.90	-	4.0	-	44
20	.0074	.0123	.074	.10	.35	.70	2	3.0	10	33
			(M)	(S)						
10	.0074	.0087	.050	.074	.25	.50	1.5	2.0	7.0	21
							(Q)	(A,B)		
5	.004	.0050	.034	.054	.18	.36	1.3	1.0	5	13
3	.004	.0070	.034	.054	.15	.30	.95	-	-	9.4
							(A,B)	(Q)		
2	.004	.0067	.034	.054	.13	.26	.6	.75	3	7
1	.004	.0047	.034	.054	.10	.20	.4	.5	2	4.5
500 mg	.0007	.0024	.010	.025	.08	.16	.3	.38	1.5	3
300	.0007	.0018	.010	.025	.07	.14	-	.3	-	2.2
200	.0007	.0014	.010	.025	.06	.12	.14	.26	.7	1.8
100	.0007	.0009	.010	.025	.05	.10	.10	.20	.5	1.2
			(M,J)			(A,B)	(P)			
50	.0007	.0006	.010	.014	.042	.07	.085	.16	.35	.88
30	.0007	.0011	.010	.014	.038	-	.075	.14	-	.68
20	.0007	.0011	.010	.014	.035	.04	.07	.12	.2	.56
10	.0007	.0008	.010	.014	.030	.030	.06	.10	.15	.40
					(A,B)	(S-1)				
5	.0007	.0006	.010	.014	.02	.028	.055	.08	.1	
3	.0007	.0011	.010	.014	-	.026	.052	.070	-	
				(A,B)	(S)			(C)	(Q)	
2	.0007	.0011	.010	.010	.014	.025	.050	.050	.060	
1	.0007	.0008	.010	.010	.014	.025	.050	.050	.050	
.5	.0007	.0006	.010	.010	.014	.025				
.3	.0007	.0011	.010	-	.014	.025				
.2	.0007	.0011	.010	.010	.014	.025				
.1	.0007	.0008	.010	.010	.014	.025				
.05	.0007	.0006	.010	.010	.014	.025				

FIGURE 10 Mass Standard Class Adjustment Tolerance Schedule

PILOT PROGRAM NUMBER 3047/212.31
TRANSMISSION NUMBER NBS 000
1400 APRIL 1

REPORT OF CALIBRATION NUMBER 3047/212.31
WSMR

ITEM: SET OF MASS STANDARDS 1 KG TO 100 G DESIGNATED SET SS-B, SET OF MASS STANDARDS 50 G TO 1 G DESIGNATED SET SS-A, ALL WEIGHTS STATED TO BE MADE OF STAINLESS STEEL, DENSITY 8.027 G PER CM3 at 20C.

THE ABOVE ITEMS HAVE THE VALUES SHOWN BELOW BASED ON COMPARISONS AT WSMR AGAINST NATIONAL REF STDS NS1 KG1 AND NS 1 KG2

MASS G	UNCERTAINTY G PLUS/MINUS	VOLUME CM3	APP MASS CORR TO NOM. MG
1 000.006 151	0.000 110	124.580 3	-0.334
500.002 388	0.000 066	62.290 0	-0.854
300.002 172	0.000 103	34.374 1	0.227
200.000 809	0.000 099	24.916 0	-0.488
100.001 027	0.000 071	12.458 1	0.378
50.000 446	0.000 036	6.229 03	0.120
30.000 244	0.000 032	3.737 42	0.050
20.000 196	0.000 028	2.491 61	0.066
10.000 139	0.000 019	1.245 81	0.074
5.000 029 8	0.000 009 7	0.622 90	-0.002 5
3.000 042 9	0.000 008 6	0.373 74	0.023 5
1.999 995 1	0.000 007 4	0.249 16	-0.017 9
1.000 012 5	0.000 005 0	0.124 58	0.006 0

THE UNCERTAINTY FIGURE IS AN EXPRESSION OF THE OVERALL UNCERTAINTY USING A 99% CONFIDENCE LIMIT FOR THE POSSIBLE EFFECT OF RANDOM ERRORS OF MEASUREMENT ADDED TO AN ALLOWANCE FOR THE UNCERTAINTY OF THE STARTING STANDARDS. SYSTEMATIC ERRORS FROM OTHER SOURCES ARE ASSUMED TO BE NEGLIGIBLE.

END OF REPORT

FIGURE 11 Typical Pilot Program Report

```
WHITE SANDS MISSILE  TEST NO.3047/212.3   SETS SS-B,SS-A
SERIES 2 62C  BAL.10KG  STAINLESS STEEL 5-12-65  STILES
    5     3     2     1   STD1  SUM1
    +     -     -     +     -               A1
    +     -     -     +           -         A2
    +     -     -     -     +               A3
    +     -     -           +     -         A4
    +     -     -     -           +         A5
    +     -     -           -     +         A6
          +           -     -     -         A7
                +     -     -               A8
                +     -           -         A9
                +           -     -         A10
OBSERVATIONS
    0.        14.900    0.        0.        5.200     0.        1.03088659
    0.        15.700    0.        0.        5.100     0.        0.94335849
    0.        15.100    0.        0.        4.700     0.        0.96150000
    0.        15.400    0.        0.        5.000     0.        0.96149999
    0.        15.200    0.        0.        4.800     0.        0.96149999
    0.        15.000    0.        0.        5.700     0.        1.07522580
    0.        12.300    0.        0.        4.100     0.        1.21946341
    0.        13.800    0.        0.        3.700     0.        0.99005941
    0.        14.000    0.        0.        3.200     0.        0.92588888
    0.        14.000    0.        0.        3.800     0.        0.98035294
    4.700     0.        14.700    5.200     0.        14.900    0.
    4.000     0.        15.500    5.100     0.        15.700    0.
    4.800     0.        15.100    4.700     0.        15.100    0.
    5.200     0.        15.300    5.000     0.        15.400    0.
    5.200     0.        15.400    4.800     0.        15.200    0.
    5.300     0.        14.900    5.700     0.        15.000    0.
    4.300     0.        12.300    4.100     0.        12.300    0.
    4.300     0.        14.300    3.700     0.        13.800    0.
    4.000     0.        13.200    3.200     0.        14.000    0.
    4.000     0.        13.600    3.800     0.        14.000    0.
         A(I)            DELTA
      -0.351741          0.099301
      -0.653233         -0.130407
       0.050249         -0.116050
       0.050249          0.264404
       0.301492          0.063409
      -0.251243         -0.180656
       0.100497         -0.
       0.552735          0.251243
       0.                -0.229708
      -0.100497         -0.021535
 SUM OF SQUARES        D.F.           OBSERVED S.D.      ACCEPTED S.D.
      0.26325387       5.00000000        0.22945756          0.32399999
 VALUE OF STD      ACCEPTED VALUE     DIFF/S.D.OF DIFF
      0.28257680       0.                2.20908606
      T                P                H.R.
      20.00000000      647.27599335      0.00830000
  AIR DENSITY          TV             VOL OF SUM
      1.02082933       294.62672806      12.82067633
```

FIGURE 12 Typical Pilot Program analysis - Sheet 1

WHITE SANDS MISSILE TEST NO.3047/212.3 SETS SS-B,SS-A
SERIES 2 62C BAL.10KG STAINLESS STEEL 5-12-65 STILES

WT	CORR	S.D.	V
500.00	5.73517227	0.13838308	64.10329914
300.00	3.47627735	0.19049205	38.46198416
200.00	2.40126657	0.18015993	25.64133334
100.00	0.89555192	0.12791570	12.82062769
100.00	0.28257680	0.12791570	11.91770649
100.00	1.27600610	0.12791570	12.82067633

OBSERVED S.D.	ACCEPTED S.D.	F RATIO	F.05
0.22945756	0.32399999	0.50	3.02

SUM OF SQUARES	D.F.	VARIANCE
0.26325387	5.	0.05265077

----------- ------- PREVIOUS TOTALS

SS= D.F.= VAR= NEW S.D.= ....

VALUE OF STD	ACCEPTED VALUE	DIFF/S.D.OF DIFF
0.28257680	0.	2.20908606

----------- PREVIOUS TOTAL N=......

----------- NEW TOTAL N+1=......

----------- NEW AVERAGE

T	P	H.R.
20.00000000	647.27599335	0.00830000
AIR DENSITY	TV	VOL OF SUM
1.02082933	294.62672806	12.82067633

FIGURE 13 Typical Pilot Program analysis - sheet 2

☆ U. S. GOVERNMENT PRINTING OFFICE : 1968—301-023/77

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, chemistry, and engineering. Comprehensive scientific papers give complete details of the work, including laboratory data, experimental procedures, and theoretical and mathematical analyses. Illustrated with photographs, drawings, and charts.

Published in three sections, available separately:

• Physics and Chemistry

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $5.00; foreign, $6.00*.

• Mathematics and Mathematical Physics

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $2.25; foreign, $2.75*.

• Engineering and Instrumentation

Reporting results of interest chiefly to the engineer and the applied scientist. This section includes many of the new developments in instrumentation resulting from the Bureau's work in physical measurement, data processing, and development of test methods. It will also cover some of the work in acoustics, applied mechanics, building research, and cryogenic engineering. Issued quarterly. Annual subscription: Domestic, $2.75; foreign, $3.50*.

TECHNICAL NEWS BULLETIN

The best single source of information concerning the Bureau's research, developmental, cooperative and publication activities, this monthly publication is designed for the industry-oriented individual whose daily work involves intimate contact with science and technology—*for engineers, chemists, physicists, research managers, product-development managers, and company executives.* Annual subscription: Domestic, $1.50; foreign, $2.25*.

*Difference in price is due to extra cost of foreign mailing.

NONPERIODICALS

Applied Mathematics Series. Mathematical tables, manuals, and studies.

Building Science Series. Research results, test methods, and performance criteria of building materials, components, systems, and structures.

Handbooks. Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Miscellaneous Publications. Charts, administrative pamphlets, Annual reports of the Bureau, conference reports, bibliographies, etc.

Monographs. Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

National Standard Reference Data Series. NSRDS provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated.

Product Standards. Provide requirements for sizes, types, quality and methods for testing various industrial products. These standards are developed cooperatively with interested Government and industry groups and provide the basis for common understanding of product characteristics for both buyers and sellers. Their use is voluntary.

Technical Notes. This series consists of communications and reports (covering both other agency and NBS-sponsored work) of limited or transitory interest.

CLEARINGHOUSE

The Clearinghouse for Federal Scientific and Technical Information, operated by NBS, supplies unclassified information related to Government-generated science and technology in defense, space, atomic energy, and other national programs. For further information on Clearinghouse services, write:

Clearinghouse
U.S. Department of Commerce
Springfield, Virginia 22151

Order NBS publications from:

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

U.S. DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20230

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE

37