

U.S. DEPARTMENT OF COMMERCE NATIONAL BUREAU OF STANDARDS

#### THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards<sup>1</sup> 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<sup>2</sup>—Radio Standards Laboratory,<sup>2</sup> 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<sup>2</sup>—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<sup>3</sup>-Materials Evaluation Laboratory-NBS/GSA Testing Laboratory.

<sup>&</sup>lt;sup>1</sup> Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D. C., 20234.

<sup>&</sup>lt;sup>2</sup> Located at Boulder, Colorado, 80302.

<sup>&</sup>lt;sup>3</sup> Located at 5285 Port Royal Road, Springfield, Virginia 22151.

UNITED STATES DEPARTMENT OF COMMERCE Alexander B. Trowbridge, Secretary NATIONAL BUREAU OF STANDARDS • A. V. Astin, Director

# Realistic Uncertainties and the Mass Measurement Process

An Illustrated Review

P. E. Pontius and J. M. Cameron

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



# National Bureau of Standards Monograph 103

Issued August 15, 1967

For sale by the Superintendent of Documents, U.S. Government Printing Office Washington, D.C. 20402 - Price 20 cents

Library of Congress Catalog Card Number: 67-60056

# Realistic Uncertainties and the Mass Measurement Process

# An Illustrated Review Paul E. Pontius and Joseph M. Cameron

This paper gives a review of the concepts and operations involved in measuring the mass of an object. The importance of viewing measurement as a production process is emphasized and methods of evaluating process parameters are presented. The use of one of the laboratory's standards as an additional unknown in routine calibration provides an accuracy check and, as time goes on, the basis for precision and accuracy statements.

Key Words: Measurement, measurement process, uncertainty, mass measurement, precision, accuracy, statistical control.

### Introduction

This paper is a condensed version of a lecture on "Error of Measurement" presented by Paul E. Pontius and Joseph M. Cameron at the Seminar on Mass Measurement, held at the National Bureau of Standards, Washington, D. C., November 30, December 1 and 2, 1964, and is essentially as presented by Paul E. Pontius at the 20th Annual ISA Conference held at Los Angeles, California, October 4-7, 1965.

It is a review of the mass measurement process from the initial basic concept to the statement of a measured mass value, examining in more or less detail certain important elements which are apt to be misunderstood, or perhaps misused. The importance of viewing measurement as a production process is emphasized and methods of evaluating process parameters are presented. The use of one of the laboratory's standards as an additional unknown in routine calibration provides an accuracy check and, as time goes on, the basis for precision and accuracy statements.

#### **Mass Measurement Requirements**

One role of the Bureau is to provide an extension of the mass measurement unit into the facilities of those who must use mass values to do other useful work. . . These large weights, for example, are for use by another part of the Bureau to calibrate force measuring devices.

The calibration service provides values for single, selected groups, and ordered sets of standards, the values being with reference to the national



standard of mass. These values, together with a value for their uncertainty, allow each user to determine, in combination with his measurement process, the uncertainty of his measurements.



The three photographs above started with a group of standards whose cumulative total mass was in excess of one million pounds, and ends with a micropound standard, a range in excess of ten to the twelfth power  $(10^{12})$ .



The accuracy requirements for a measurement are set partly by experience, partly by discussions with others, and partly by analysis. For a particular purpose, the accuracy requirement must be established with care, as it provides a point of departure for the entire measurement process. Frequently we tend to lose perspective in regard to what we are measuring, or what the measurements mean, particularly if we concentrate on routine procedures or are remote to the actual measurement.

The aiming point for our measurement is to establish the mass, or true value, of a particular object for it is, in concept at least, unique and invariant. If, for example, accuracy within .0] percent is sufficient for our purpose, the target center is the area within the next to the last circle Our measurements may group on either side of dead center, or may be randomly scattered across the center of the target, but as long as the spread is essentially within the target circle, the process is satisfactory for its intended use. Troubles arise when realistic requirements are divided by large arbitrary constants as specifications pass through various groups of people in a complex organization. Measurements accurate to better than .01 percent require attention to many details under more or less ideal conditions, and may not be obtainable under adverse conditions, consequently the entire measurement effort may be lost if the end use involves measurement processes of questionable precision. In the case of calibration, for example, in order to utilize the accuracy inherent in a good calibration, the user must work just as hard in his measurement process as the calibration facility did to determine the value of the standard originally.

The importance of incorporating the properties of the measurement process in setting up requirements or specifications is illustrated by the problem of adjustment tolerances for different classes of weights.

	TYPICAL	CLASS ADJ. TOL.						
NOMINAL VALUE	UNCERTAINTY (SYS. ERROR) OF STD.VALUE	S.D. OF SINGLE MEAS,	SINGLE MEAS. PROCESS UNCERTAINTY*	CLASS M (mg)	CLASS S (mg)			
10 g 5 g 1 g 500 mg 100 mg 10 mg	.0087mg .0050 .0047 .0024 .0009 .0008	.0074mg .004 .004 .0007 .0007 .0007	.031mg .017 .017 .005 .003 .003	.050 .034 .034 .010 .010 .010	.074 .054 .054 .025 .025 .014			

\* 3 S.D. + SYS. ERROR

The Class M and Class S adjustment tolerance limits for selected weights are shown in the two right hand columns. The uncertainty associated with the stated value for standards of the same nominal value is shown in the 2d column and the precision for a single measurement is shown in the 3d column. If one tries to establish the compliance with Class M adjustment tolerances by a single weighing against a known standard, the uncertainty of the process would be as shown in the 4th column. This uncertainty, compared with the quantity we are trying to detect, is such that in the first 4 cases the measurement uncertainty is a large fraction of the tolerance so that only those items well inside of tolerance have a good chance of of being passed. A measurement procedure more sophisticated than a single comparison with a e-known standard may be desirable.

	TYPICAL PR	CLASS ADJ. TOL.			
NOMINAL VALUE	UNCERTAINTY OF CLASS M (WITHIN TOL.)	S. D. OF SINGLE MEAS.	S. D. OF SINGLE MEAS. SINGLE PROCESS MEAS. UNCERTAINTY		CLASS S-1 (mg)
10 g 5g 1g 500 mg 100 mg 10 mg	.050 .034 .034 .010 .010 .010	.0074 .004 .004 .0007 .0007 .0007	.072 .048 .048 .012 .012 .012 .012	.074 .054 .054 .025 .025 .014	.18 .18 .10 .08 .05 .03

We would be in greater difficulties if we were to try to establish compliance with Class S adjustment tolerances in the same manner with reference to Class M standards, which are known only to be within the Class M tolerance limits. In 4 of the 6 examples, the process uncertainty is of the same order of magnitude as the quantity we are trying to check. These examples illustrate the necessity for a careful evaluation before venturing a commitment on the performance of a particular measurement process.

#### The Unit of Mass

INTERNATIONAL PROTOTYPE KILOGRAM
TRUE VALUE (EXACT BY DEFINITION)1kg
VOLUME AT 0 °C (HYDROSTATIC WEIGHING)46.40052 ml
VOLUMETRIC COEF. OF EXPANSION (BY MEASUREMENT ON PLATINUM-IRIDIUM
$\alpha = (25.863 + 0.00562\theta) \times 10^{6}$

By practically universal agreement, the mass of the International Prototype Kilogram is the basic unit for mass measurement. It is a particular object, defined to have an exact invariant mass of one kilogram, that is to say, the true value is one kilogram. The volume and the coefficient of volumetric expansion are necessary to determine the best estimate of the true value of other objects compared with this standard.



With the unit defined, we can logically construct a true value scale which has the property that some point on the scale will correspond to the mass of any chosen object. We call the major subdivisions of this scale nominal values. Other customary units, such as the pound, are not ambiguous if they have an exact definition relative to the basic unit. An intermediate point on the scale can be described either relative to the whole scale, as for example, 9.995 grams, or relative to the closest nominal value, in which case the point would be described as 10 grams minus 5 milligrams. The minus 5 milligrams may be called a correction or error, depending on one's viewpoint. The use of a nominal value and a correction is often convenient in computations, however, the word "correction", or "error", overly emphasizes the importance of the nominal value. Interpretation of tolerance limits on the value of the standard as the error automatically disregards the primary benefits of a good calibration. Only an ideal measurement method or process can produce true values of multiples and subdivisions of the basic unit which will exactly coincide with nominal values on the true value scale. It should be emphasized that, from a measurement standpoint, adjustment to nearly coincide with a nominal value is necessary only to assure an "on scale" condition when intercomparing equal nominal summations.

In our previous example, we elected to interpret the adjustment tolerance limits associated with our Class M set as the uncertainty of the value. While this may be appropriate with respect to the nominal value, such an interpretation raised serious doubts as to our ability to test the Class S weight set. If we had used the actual value and its uncertainty as a basis for our tests, the doubt essentially disappears. With minor modification at the 10 g level, the uncertainty of the values established for the Class S weights by our single measurement is clearly suitable for the task at hand. It must be emphasized that our apparent increase in measure-

	TYPICAL	CLASS ADJ TOL			
NOMINAL VALUE	UNCERTAINTY (SYS. ERROR) OF STD. VALUE	S.D. OF SINGLE MEAS.	SINGLE MEAS. PROCESS UNCERTAINTY *	CLASS S (mg)	CLASS S-I (mg)
10 g 5 g 1 g 500 mg 100 mg 10 mg	.0087 mg .0050 .0047 .0024 .0009 .0008	.0074 mg .004 .004 .0007 .0007 .0007	.031 mg .017 .017 .005 .003 .003	.074 .054 .054 .025 .025 .025 .014	.18 .18 .10 .08 .05 .03

\*3 S.D. + SYS. ERROR

ment capability did not require any change in our process hardware. It has been achieved, for the most part, by a change in philosophy.



Our access to the true value scale as established by the international standard is through prototype kilogram number 20. The estimated true value of number 20 is 1 kilogram minus 19 micrograms, based on several measurements. We can construct an accessible true value scale by setting off from the value of kg 20 an amount equal to the correction. Practically, the stated value is assumed to be exact, the uncertainty of the value introducing only a slight systematic error in our reconstructed scale.

By comparing other objects with kilogram 20, either singly or in combination, we can assign values relative to our accessible scale. A sufficient number of well calibrated standards which can be intercompared, and which may occasionally be compared with our prototype standard, serve to maintain our scale with perhaps a greater precision than was available in the starting measurements. All mass values on NBS Reports of Calibration are with reference to a minimum number of selected mass standards. For example, practically all sets



of metric weights are calibrated with reference to a pair of 1 kg or a pair of 200 g or a pair of 100 g weights. The national reference standards group

does not include weights of all denominations.



A practical measurement method is easy to visualize in the form of a broad outline of the elements of the method such as, the concept of the quantity to be measured, pertinent physical laws, various instruments, standards, the operators, procedures to be used, the environment in which the measurements are to be made, the computations which are to be made, and a means of establishing some parameters of performance. As we briefly review some of these elements, we will find that every mass measurement facility has many things in common.

Mass is an inertial property of an object, which, within the framework in which our measurements apply, is considered to be proportional to the amount of material. Mass is generally thought of as being measured through some application of



Newton's law of gravitational attraction, however, it is perhaps more precise to say that measurements are made by comparing the forces attracting suspended bodies toward the earth-that is the net vertical forces including the effects of G, air buoyancy, rotation of the earth, etc.



variety of requirements. Modern computation equipment ranging from desk calculator to electronic computer are now widely available so that laborious long hand computations are no longer necessary.



The environment in which the measurements are made does not vary substantially between calibration facilities. Weighing rooms are almost universally clean, with restricted access, and relatively free of vibration. With the possible exception of freedom from vibration, these desirable features are easily obtained.

People operate the equipment, following prescribed procedures. Operator skill increases with practice, and in time, operators in a given group approach a uniform level of skill.

Each comparison, or weighing, consists of a sequence of operations, more or less formalized. Detailed procedures and weighing designs, ranging from simple to complex, are available for a wide



751.42	-	Single Pan Da OBSERVA	mped Balances TIOH SHEET	list (Cr. & Dille)	
PMC Balance	M-4	10-9-6	4 Set 3 # 6235 NAS Tes	No. P7074	
Load	Dial Setting	Scale Reading	Computations		
0	000 0	18.52	48.78 ~ a= 49.59	0 6 mg	
1235	085.0	67.30	19.66 -		
235+120,	085.0	86.96	48.66 -	······································	
0 + 120.	000.0	38.30		CHFCK ON	
07074 SET	303 = 01	236		SENSITIVITY	
0	000.0	18.87	-49.33	OLIVOITIVITI	
2236	085.0	67.20	20.72		
2236+L20,	085.0	87.92	-49.17	CHECK ON	
0 + 120,	000.0	38.75		DIFFERENCE	
P7074 SET	- 303 #0	237			

While perhaps not generally considered so, analysis is a part of the measurement method. Whether done by machine . . .

. . . or by hand, the analysis verifies that such parameters continue to be applicable.



A particular measurement method is like a specification for a particular measurement. The specific instrument, the standards to be used, the specific operations to be performed and the planned sequence in which they are to be carried out, the operator, the location, and the method of computation and analysis, collectively define a particular measurement method. Until the measurement has actually been made and analyzed, the performance is only "on paper" and therefore ideal.

A MEASUREMENT PROCESS PRODUCES: I.A USEFUL MEASURED VALUE 2. AN ESTIMATE OF UNCERTAINTY FOR THAT VALUE

A measurement process involves the actual physical operation of the specified equipment following the procedures as closely as possible. It is subject to the many variations that can and do occur during the operation. The end result is an estimated best value, which, in order to be useful, must be accompanied by the uncertainty with reference to known performance parameters.

Changes in any one or in a group of elements of the method constitutes, in effect, a different particular method and a different process which will in turn produce a different result and a different uncertainty. Small changes can make the difference between a useful value or a wasted effort.

NSTRUMENT /
ARE DIFFERENCES IN INDICATION TO BE INTERPRETED AS A DIFFERENCE IN MASS? OR PROCESS VARIABILITY?

Because we must establish the mass of the object in question by measuring the mass difference between it and some known standard, the comparator is a vital element in the process. The inherent characteristic of the comparator is precision – not accuracy. The fundamental question is whether the indicated difference is really a mass difference, or an indication of some other variability. While we may be able to identify large sources of variability, in the limit, we cannot differentiate between instrument precision, variability from extraneous sources, or variability of the standard.



We start by determining the indicated difference between two objects that are nearly alike.



From our first comparison, it appears that the round knob weight on the left is clearly heavier than the flat knob weight by one scale division. If we stop here, we would simply state the value of one object in terms of another, however, we have no way of knowing the uncertainty to associate with this value.



If we repeat the comparison at some other time, we are quite likely to obtain a different result. This raises a serious question—which of the two results is correct?

We repeat the comparison again . . .

... and again. Now there are four different values, none of which alone can be considered the best measure of the difference, but considered as a group they can tell us something about the





instrument. Continuing to record the indicated difference between two similar objects, and preferably making the comparisons in the environment in which the instrument is to be used, a plot is made against time of the differences which may look like this.



The first plot indicates a severe rounding off, which may be from several causes. Such a response clearly lacks the appearance of randomness. The second plot at least appears to be random. The third plot, while perhaps appearing to be random, obviously lacks the precision of the second plot. The range of the differences as plotted gives us an idea of the smallest mass difference that can be detected with assurance, and is obviously related to the requirements our measurements must meet. Repeated independent measurements of the same mass difference are essential to the evaluation of the instrument.



The operator, or manufacturer, must search for cause and effect until repeated indications for the same load, or differences are reasonably consistent. Effects which are periodic in nature, but with a period significantly longer than the period of the instrument, can be minimized in the design of the weighing method.

One additional requirement, generally beyond the control of the operator, is that of linearity. An instrument, used as a comparator rather than a direct reading device, requires linearity only in the neighborhood of the actual load.

# **PROBLEM:**

# OBSERVED DIFFERENCES TO MASS DIFFERENCES METHOD: 1. SUBSTITUTION 2. TRANSPOSITION

3."DIRECT READING"

The problem of establishing the correspondence between observed differences and mass differences is a part of the weighing method. The first two methods, substitution and transposition, are comparative methods. That is to say, the method requires observations relative to a suitable standard along with the unknown. With these methods, the measurement equipment need be continuous only over the time interval required for making a group of observations and linear only over the range of the difference between the standard and the unknown. Most direct reading equipment is in a sense a substitute standard, that is, at some point in time it is calibrated with reference to a standard, and from that point until recalibration, it is generally assumed to have a long term constancy approaching that of the standard. Most mass measurement equipment can be used either way. The smallest uncertainties invariably will be associated with the comparative mode of operation.

#### **Weighing Method**



SUBSTITUTION METHOD



To illustrate the principle, the double substitution method is performed as follows: We start with a simulated equal arm balance, a tare weight – the white cylinder near the base of the balance, a sensitivity weight of known value immediately in front of the dark weight near the center, and two nearly equal brass weights, one with a flat knob in the center and one with a round knob on the left. The scale indication is in arbitrary numbers and the tare weight is necessary to establish an "on scale" condition.



The first observation is that produced with the round knob weight on the pan.

 $(1) \land \rightarrow \circ \circ_1$ (2)  $B \rightarrow O_2$ 

The second observation is that produced with the flat knob weight, which might be a standard, replacing, or substituted for, the round knob weight.

The third observation is that produced by repeating the previous step and adding the sensitivity weight to the pan load.





, The fourth observation is a repetition of the first step including the sensitivity weight.

 $A - B - K \left\{ \frac{f(1) - (2) + (4) - (3)}{2} \right\}$  $\underline{\mathcal{P}} K \left( \frac{0_1 - 0_2 + 0_4 - 0_3}{2} \right)$ В+т-В≏К {(3)-(2) }  $m \mathfrak{L} K (0_3 - 0_2)$  $A-B = \frac{m}{(0_3-0_2)} \left( \frac{0_1 - 0_2 + 0_4 - 0_3}{2} \right)$ 

Using the requirement for continuity, a relation can be established for A minus B from the average of the two sets of differences as shown. Using the linearity requirement, the constant of proportionality K, or the mass value of the indicating scale division can be determined from the second and third observation. Finally, the difference A minus B is expressed as a function of the observations, in ratio form and the value of the sensitivity weight.

DOUBLE SUBSTITUTION  $A-B = \frac{m}{(0_3-0_2)} \left[ \frac{0_1-0_2+0_4-0_3}{2} \right]$ SINGLE TRANSPOSITION  $A-B \stackrel{\text{\tiny \ensuremath{\square}}}{=} \frac{m}{(O_2-O_2)} \left[ \frac{O_1-O_2}{2} \right]$ DOUBLE TRANSPOSITION  $A-B = \frac{m}{(0_3-0_2)} / \frac{0_1 - 0_2 + 0_4 - 0_3}{4}$ 

All usual methods result in very similar relations expressing the difference between two objects being compared. In all cases, A minus B is expressed as a ratio between sets of observations multiplied by the value of the sensitivity weight. Obviously requirements for knowledge of the value of M are minimized when the size of the ratio involving the observation is small. The constant of proportionality, K, is really the ratio in front of the bracket terms which we call the value of the division. The strange equal sign is used to indicate that the relations shown are observational equations and not mathematical identities. With the measurement method agreed upon, let us now discuss its performance—we put it into production and see how it works out as a measurement process.

#### Measurement as a Process

MEASUREMENT PROCESS OUTPUT..... MEASUREMENT PROCESS AVG....LIMITING MEAN VARIABILITY.....PRECISION BIAS .....SYSTEMATIC ERROR PROCESS LIMITS.. UNCERTAINTY OR ACCURACY

A measurement process is essentially a production process, the "product" being numbers, that is, the measurements. A characteristic of a measurement process is that repeated measurements of the same thing result in a series of non-identical numbers. To specify a measurement process involves ascertaining the limiting mean of the process; its variability due to random imperfections in the behavior of the system, that is, its precision; possible extent of systematic errors from known sources, or bias; and overall limits to the uncertainty of independent measurements.





The chart shows measurements on a 200 g weight, plotted in the order in which they were taken. Despite the presence of one or two stragglers, the measurements tend to cluster around the central line—the process average or limiting mean. Our confidence that the process has settled down to a single limiting mean is strengthened as the length of the record is increased. We may have satisfied ourselves regarding the mean but what about the next measurement?



It seems clear that we cannot give an exact answer but will have to content ourselves with a statement that allows for the scatter of the results. Our goal is to make a statement with respect to a new measurement that is independent of all those that have gone before. As indicated in the chart, if we had a sufficiently long record of measurements we could set limits within which we were fairly certain that the next measurement would lie. Such a statement should be based on a collection of independent determinations, each one similar in character to the new observation, that is to say, so that each observation of the collection and also the new observation can be considered as random drawings from the same probability distribution. These conditions will be satisfied if the collection of points is independent, that is free of patterns, trends and so forth; and provided it is from a sufficiently broad set of environmental and operating conditions to allow all the random effects to which the process is subject, to have a chance to exert their influence on the variability. Suitable collections of data can be obtained by incorporating an appropriate measurement into daily routine weighing procedures, for example, a daily measurement of the difference between two laboratory weights, or in the regular calibration of the same weight.

If the measurements tend to cluster when taken close together in time, like the results shown on the chart, some systematic effect is present and certainly the results are not independent. This may



be due to some as yet undetermined cause, and the group means may have the appearance of randomness of the previous chart.



The group means may tend to a limit and the process may have all the properties of a good measurement system, once the allowance is made for the grouping. It is important that grouping be properly handled in determining the precision of the process. By modifying the process or changing the schedule of measurements to give the effect of independent measurements, we can arrive at a situation like the values on the 200 g standard. The shaded band is meant to suggest a limit, not an artistic slide.

From a study of a sequence of such independent measurements, we can use control chart techniques to set up limits within which the next value should lie. In the case where we have an extremely long sequence, a bar, as illustrated in the chart, can be marked off on either side of the mean so that some suitable fraction, say 99 percent, of the observations are within the interval represented by its length.





We can reverse the process and say that the probability is 99 percent, that the true value, or limiting mean, will not be more than the width of the bar from any observation chosen at random. This will be true of the next observation as well, provided it is an independent measurement from the same process. The probability statement attaches to the sequence of such statements. For each individual new observation the statement is either true or false but in the long run 99 percent of such statements will be true.

Assuming that the limits on the chart are based on large numbers of observations, we would find that very nearly the intended percentage of all such bars, centered on the observed values, would in fact overlap the mean. Only in those cases, such as the points in the area outside of the control limits, will the bar fail to overlap the mean. This is expected in only 1 percent of the cases. More frequent occurrence is a clear indication of either loss of control or that the limits were not properly set. Once we are satisfied that the process has a limiting mean value and is stable enough to permit



prediction we turn our attention to evaluating its precision.

#### **Process Precision**

Let us now take a look at the situation in weighing to see what is involved in the study of the precision of the process.



A characteristic of a measurement process is that it produces non-identical results. In our previous charts we had measurements of a 200 g weight, here are shown four measurements of the difference in mass. Through the redundancy here 3 extra measurements—we get our grip on precision. In weight calibration we do not rely on repeated measurements of the same quantity but achieve the same result in another way.

When we intercompare four objects, for example, four 1-kg standards, we could use six observations. Weight S is compared with A for  $a_1$ , S with B for  $a_2$  and so on. If S were a standard and the rest





inknowns, we again have 3 more measurements han we need and these serve to tell us of the precision of the process.



S IS AN ESTIMATE OF O, THE LONG-RUN STANDARD DEVIATION

known weights do. (The quantity, k, is the number of unknowns in the system.)



A simple example, using only three of the observations of the previous series, with S as the standard, A as the unknown, and B as the check standard, might give rise to the values shown. If everything were perfect, all equations representing the weighings would be satisfied exactly. Their lack of agreement would give a measure of the variability.

In general, for such weighing, there will be a discrepancy between the observed value and the best value calculated from the data, "best" meaning in most cases the value obtained in the method of least squares. If all is going well, none of these deviations will be too large, and also certain combinations of them, such as the sum of the squares, will also be well behaved. For statistical analysis the standard deviation, S, is used as the measure for describing variability. The quantity, S, is a function of the observational errors and will change with each set of data just as the values for the un-



If the process is in a state of control these values of *s* will scatter about some value which is the true or long run standard deviation of the process.



The argument that the uncertainty should be based on the internal agreement of today's values on the grounds that each day is unique or that weighing conditions are better on one day than on another may well be true. However, it will be expensive to make enough measurements on a given day to be sure that the variability has indeed changed from its long run average or to provide a reliable enough value to represent today's results. If the process did not change, using today's value would be analogous to keeping the last value of a sequence rather than using the mean represented by the dotted line. It is a sign that weighing conditions are not being reproduced, i.e., that the process is not in control if the standard deviation does not stay within predicted limits. Let us now look again at the check standard.

#### **Process Mean**



Each value obtained for the check standard serves not only a check on the process mean, but also can be used for evaluating the process variability. The same check standard, perhaps one of a group reserved for this purpose, is used consecutively in a given procedure until many independent values are obtained.

The importance of randomness cannot be overemphasized. As the collection of independent measurements on the check standard grows, it must be continually re-evaluated with reference to predicting the band within which the next point will lie. Slow drifts or sharp discontinuities are cause for concern until corrected, or satisfactorily explained.

If values return to normal after cleaning, one can rest easy, knowing the process is behaving properly. Indication of permanent changes are sometimes harder to explain, and even the most careful laboratories must occasionally repeat measurements because of troubles with foreign material adhering to or falling off the standard. If the new mean value persists over a sufficient number of measurements,



it is proper to assume the standard has changed for some reason.



A check on *just* the value of the standard or *just* the precision is not enough. It turns out that the value for the precision and the value for the check standard are generally independent, that is, when s is small the deviation of the value determined for the check standard from the accepted value is equally often big and small. For control we need both conditions.



For a given set of observations the precision must be proper as shown on vertical scale and we must have a check on a known weight to establish that the limiting mean has not changed as shown on horizontal scale. Until these conditions are fulfilled, we cannot be sure exactly what it is that we are measuring. These are necessary conditions, and in perhaps most cases, also sufficient conditions to proclaim that the measurement process is in a state of control, as indicated by points within the central rectangle.



Because the check on the standard is spread over a considerable time interval, the variability will

include the proper diversity of environmental and other factors and the sequence will, in the absence of seasonal or other systematic trouble, approximate a sequence of independent values. If the weighing conditions are reproducible, then the daily standard deviation, s, and the variability as computed from the values of the check standard will be in agreement, i.e., the long run average of the variability as estimated from the control chart on the standard deviation should approach the corresponding value from the control chart based on the variability of the values of the check standard. Frequently, one is not in as good a shape as that indicated on the slide. When the measurements are spread out in time or space, an additional component of variation enters so that the lower chart gives an overly optimistic view of the process. A realistic estimate of process variability has to be based on that from the upper chart which reflects the total variation to which the measurements are subject. One would still use the within occasion variability for checking on control of the process, of course.

DETERMINING THE MASS OF AN OBJECT AND THE ASSOCIATED UNCERTAINTY IS A CALIBRATION.

## ROUTINELY, THE CALIBRATION MUST BE LIMITED TO A FEW MEASUREMENTS.

If in calibration we could measure the difference between the standard and the unknown again and again we could make an uncertainty statement similar to those just discussed for the case of measurements of a fixed difference, but in fact, we cannot routinely make enough measurements of this type to permit reliable estimates of the uncertainties.

#### Process Parameters and Uncertainty of Calibration

If we could be sure that our measurements of the difference between the unknown and the standard came from a process in a state of statistical control, that is to say a stable process with a known variability, then we could transfer the properties of the process to the individual measurement and be correct a stated percentage of the time. THE MEASUREMENT PROCESS REMAINS, AND IS, IN A SENSE, A CAPITAL INVESTMENT.

# THE MEASUREMENTS, LIKE PRODUCTS, PASS ON TO OTHER DESTINATIONS.

All who weigh, or make other measurements, should concentrate on the properties of the measurement process—the degree to which the process re-creates the same value for its standards and exhibits the same level of variability. These are the properties that remain. The weights that are calibrated pass on to other destinations.



S, S<sub>A</sub> and S<sub>B</sub> can be nearly equal. If S0, then lab a and LAB B CAN CALIBRATE THIER OWN SET FROM SELECTED STANDARD WEIGHTS

At every stage in the extension of a measurement unit from an accepted standard to the ultimate user, there are three items of interest—a standard item, or items, with announced values and associated uncertainty, an assembly of equipment and procedures necessary for making the necessary comparisons, and the items which must be measured to accomplish some useful task. The uncertainty of the values established for the user are of paramount importance. This uncertainty has two components one associated with the value of the starting standard and one reflecting the contribution of the local measurement process. The total uncertainty at any particular place becomes the systematic error for those who must use the service provided.

NATIONAL REPORT OF	BUREAU OF ST CALIBRATION	ANDARDS TEST NO. 189	160 U.	S. DEPARTMENT	IC STANDARDS OF COMMERCE				
FOR	DR LOCKHEZD-CALIFORNIA COMPANY BURBANK, CALIFORNIA								
ITEM	SET OF MASS MFR. SEEDER STATED DENS	STANDARDS 1 ER-KOHLBUSCH ITY 10KG TO	OKG TO 1G SERIAL NO. 7 1G 7.89 G	DESIGNATED SE B1163 PER CM3 AT 20	тс c				
THE NATIO	THE ABOVE I NAL STANDARD APPARENT MA	TEMS HAVE THE OF MASS. SE SS VALUE AND	MASS VALUES E ATTACHED SU UNCERTAINTY I	SHOWN WITH RE UPPLEMENT FOR FROM	PERENCE TO LIMITATIONS DENSITY.				
ITEM	NOMINAL	APP MASS CORR (MG)	TRUE MASS CORR (MG)	UNCERTA INTY (MG)	VOL AT 20 (CM)3				
1KG 500G 200G 100G 50G 30G 20G 10G 5G 3G 2G 1G	1000.00 500.00 200.00 100.00 50.00 20.00 10.00 5.00 3.00 2.00 1.00	0.045700 0.278974 -0.265581 -0.018969 -0.093402 -0.013024 -0.000559 0.075999 -0.118775 0.002680 0.015487 0.043916 -0.035294	9.126309 4.819581 2.458779 1.797272 0.814718 0.441037 0.271878 0.257624 -0.027964 0.048086 0.042730 0.062078	$\begin{array}{c} 0.110227\\ 0.069271\\ 0.063098\\ 0.052346\\ 0.064860\\ 0.032909\\ 0.020735\\ 0.01434\\ 0.010710\\ 0.005702\\ 0.004076\\ 0.003502\\ 0.004076\\ 0.00362\\ \end{array}$	$\begin{array}{c} 126.74386883\\ 63.37196684\\ 88.02312469\\ 25.34877014\\ 12.67437446\\ 6.33719146\\ 3.80231580\\ 2.5348687\\ 1.26742357\\ 0.63371965\\ 0.38023355\\ 0.25349329\\ 0.25349329\\ 0.12673927 \end{array}$				
UNCERTAIN OF RANDOM FROM KNOW	THE UNCERTA TY USING THR ERRORS OF M N SOURCES BE	INTY FIGURE I BE STANDARD D EASUREMENT, T ING NEGLIGIBL	S AN EXPRESSI EVIATIONS AS HE MAGNITUDE E.	ON OF THE OVE A LIMIT TO TH OF SYSTEMATIC	RALL E EFFECT ERHORS				
		3	4	TOK THE DIREC PAUL E. PONTI	TOK, US, CHIEF				
TEST COMP WASHINGTO	LETED AUGUST	17, 1966		MASS AND VOLU	ME SECTION				

Any report of calibration or report of test must state a realistic uncertainty based on actual process performance. All of the pertinent data must be included so that the local processes can minimize the introduction of additional systematic errors. The random component of the uncertainty is a function of the measurement effort in the local process, reflecting the actual performance of that particular measurement process.



There is no substitute for the evidence provided by the repeated calibration of the same object, over an extended time period, in demonstrating what the measurement process can do. These measurements should be independent repetitions, made under all the diversity of condition by which the method is affected so as to represent the set of conditions to which we wish our prediction to apply. The internally based precision estimate is applicable only to a narrower range of conditions, and it is only when the measurement conditions are highly reproducible that the two estimates of precision become equal.



The routine calibration of one of the laboratory's weights, used as check standard, tells us what the process can do—it is not just a simulation of the calibration process—it is the real thing—without the need for any assumptions. It provides the basis for the precision statement or gives us a check on any internally based statement. We can say to our clients: "If we calibrate your weight a large number of times the results would look like those on the chart. We did it only once so that your value is like one of these points. Which one, we cannot say but we are fairly certain that it is within the indicated uncertainty."

U.5. GOVERNMENT PRINTING OFFICE : 1967 OL-257-257



# U.S. DEPARTMENT OF COMMERCE

WASHINGTON, D.C. 20230

OFFICIAL BUSINESS

POSTAGE AND FEES PAID U.S. DEPARTMENT OF COMMERCE