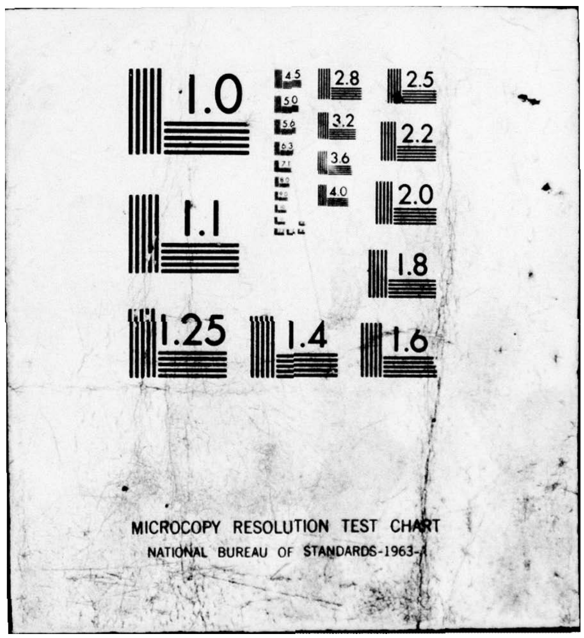
AD-A048 809 ROCKWELL INTERNATIONAL ANAHEIM CALIF AUTONETICS GROUP F/G 14/2
SYSTEMATIC MEASUREMENT ERRORS,(U)
1977 R B SCHUMACHER
UNCLASSIFIED X77-1125/101 NL

| OF |
AD
A048 809

END
DATE
FILMED
2-78
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

technical information from ...

Rockwell International

(1)

Systematic Measurement Errors

(6) SYSTEMATIC MEASUREMENT ERRORS

by

(10) Rolf B. F. Schumacher

(11) 1977

(12) 13p.

(14) X77-1125/101

THE DISTRIBUTION OF THIS REPORT IS UNLIMITED

X77-1125/101

390 713

SYSTEMATIC MEASUREMENT ERRORS

By Rolf B. F. Schumacher, Metrology Laboratory
Rockwell International, Autonetics Group
Anaheim, California

Abstract

Nature, origin, and treatment of systematic errors in measurements and calibrations are discussed. It is shown how systematic errors can vary, how random errors under one set of conditions will become systematic errors under another set of conditions and vice versa. Recommendations are made concerning the assignment of values to limits of systematic errors of measurements and standards.

I. Introduction

Statistical analysis of measurement errors has become an additional tool metrologists are using in an increasing number of applications and at an increasing rate in order to determine numerically the accuracy of measurements and the confidence attached to quoted accuracies. Methods of evaluating random errors are well known and in widespread use. This paper examines the main aspects of systematic errors and methods to evaluate them; it discusses the nature of the fluctuating boundaries between systematic and random errors in measurements and calibrations.

The systematic error of a measurement or calibration may be defined as the largest possible estimated difference between the true value of a measured quantity and the mean value towards which the measurement or calibration process tended as a limit at the time of the measurement and which difference could not be eliminated for technical or economic reasons; it is an estimate of the maximum limits of the effects of all error sources known or suspected to exist which tended to offset uniformly all results of repeated applications of the same measuring or calibration process at the time of the measurement. Thus, the "systematic error" is not an error in the accepted sense of the word "error", but rather an uncertainty; however, the term "error" will be used here for brevity and because it has been firmly established by custom.

II. Types of Measurement Errors

A. General

The error or uncertainty of a measurement can be divided into a fixed part and a variable part, where it should be understood that the fixed part remains approximately fixed only for the duration of the measurement process. For shorter durations, the variable part is likely to be smaller, and over longer time intervals, the phenomena causing the "fixed" error can be expected to vary also, thus making the variable part of the error larger at the expense of the "fixed" part. As the time interval under consideration becomes very large, any initially fixed error will probably vanish and all errors are likely to become variable.

The errors or uncertainties of a measurement can also be categorized by a random part and a systematic part. The random error is part of, but not necessarily identical with, the variable error; and the systematic error includes the fixed error, but may also include a portion of the variable part of the error.

All determinable errors which occur during one measurement process and which do not appear* to be part of the random error must be accounted for, otherwise the process may not be under statistical control. But ". . . until a measurement operation . . . has attained a state of statistical control it cannot be regarded in any logical sense as measuring anything at all."[4] – "Capability of control means that either the measurements are the product of an identifiable statistical universe or an orderly array of such universes, or if not, the physical causes preventing such identification may themselves be identified and, if desired, isolated and suppressed."[1]

No general statement concerning the relative magnitude of random and systematic errors can be made. In some disciplines, the best measurements that can be made have systematic errors very much smaller than random errors. In others, the systematic errors far outweigh the random errors, even for the best measurements which the present state of the art permits.

B. Systematic Errors

A systematic error is that value which, when added to, or subtracted from, the limiting mean value of a measurement process of a quantity, produces a range in which the "true value" of that quantity is believed to lie. We hope that the systematic error is always equal to or larger than the difference between the "true value" of the quantity and that value toward which the measuring process of the quantity tends, this tendency not being caused by chance fluctuations of any part of the measurement system.

The systematic error is a bias, but generally unknown in magnitude and direction, because the "true value" is an abstract concept which cannot be physically realized. More pragmatically ". . . the systematic error . . . of a measuring process refers to its tendency to measure something other than what was intended. . . On first thought, the 'true value' of the magnitude of a particular quantity appears to be a simple straightforward concept. On careful analysis, however, it becomes evident that the 'true value' of the magnitude of a quantity is intimately linked to the purposes for which knowledge of the magnitude of this quantity is needed, and cannot, in the final analysis, be meaningfully and usefully defined in isolation from these needs."[4]

A systematic error is always an estimate of a range, and estimating it requires a profound understanding of the measurement process if a realistic figure is to be arrived at. A quoted figure for the estimate of a systematic error means that the metrologist believes that he would have detected – and, therefore, been in a position to reduce – any error larger than that quoted as the systematic error. It implies that the actual, unknown error could be anywhere within that range. And ". . . the wiser and more careful the experimenter's search for systematic errors, and the more completely he has eliminated them, the less likely is it to lie near the limits of the range."[3]

Dorsey[3] describes the concept of the systematic error as being ". . . used to cover all those errors which cannot be regarded as fortuitous, as partaking of the nature of chance. They are characteristics of the system involved in the work; they may arise from errors in theory or in standards, from imperfections in the apparatus or in the

*The word "appear" is used, ". . . because, as is always the case in trying to find a law controlling a phenomenon, we can never be sure that we have discovered the law. Obviously such appearance is not sufficient in the logical sense although it must be in the practical sense."[11]

observer, from false assumptions, etc .... They are frequently called 'constant errors,' and very often they are constant throughout a given set of determinations, but such constancy need not obtain . . .

"In searching for systematic errors, the logical procedure is to make a series of measurements, then to change something and to make another series, and to compare the means of the two groups. This will be repeated as often as may seem necessary. None of the series can be long, for an extended delay offers opportunity for unanticipated changes to occur. If the two means being compared do not differ by more than the sum of their technical probable errors, their difference is of no physical significance – it proves nothing. Hence, *the presence of a systematic error that does not exceed the sum of the technical probable errors of the two groups of observations used in the search cannot be established without great difficulty, if at all . . .*

"In the absence of such a search, the worker can do no more than hope that all is going well. The fact that he sees no reason for suspecting the presence of an unknown systematic error is of no importance at all, no matter who the observer is. The really troublesome errors are exactly those that are not suspected. The expected ones can usually be to some extent eliminated."[3]

Sometimes economics dictate the use of a coarse measurement system whose major component of systematic error can be determined by *comparison with a more accurate measurement system*. This component can then sometimes be accounted for by a correction. The remaining systematic error will thereby be reduced; it may even become negligibly small, but it will theoretically never become zero. However, the systematic error of the coarser system may vary, and it may not be practical to determine it each time a measurement is to be made. In this case, no correction may be applicable, and the comparison with a more accurate system may only serve to estimate the range of the systematic error of the coarser system more accurately than would be possible without recourse to a more accurate system. This is the case, for instance, with a working instrument with a rated accuracy or uncertainty of 1% routinely calibrated by working standards with 0.1% uncertainty *limits*. The 1% *uncertainty of* the working instrument is its systematic error contribution to all measurements made with it, unless corrections are provided, in which case the systematic error may be less. The calibration against the standards then serves to assure that this systematic error does not exceed 1%.

III. The Changing Nature of Errors

The boundaries between systematic and random errors are fluctuating. What appears as a systematic error under one set of circumstances may appear as a random error under another and vice versa, just as a sculpture changes its appearance to a viewer walking around it. The consequences of this fact, as they affect the evaluation of uncertainties of standards in a laboratory, are illustrated by a few examples.

A. A Random Error Becoming Systematic Error

When a standards laboratory A, say the National Bureau of Standards, calibrates a reference standard for us, laboratory A reports to us the value of the standard, as it has determined it by its measurements, and the limits of uncertainty of that measured value. Part of laboratory A's uncertainty originates from its random error, part from its systematic error. Random error and part of the systematic error can be expected to add a small variable component to laboratory A's measured value in such a way that each time laboratory A measures the standard, the resulting measured values will differ slightly from the ones obtained earlier or later. However, once a value and its limits of uncertainty are reported to us, they are fixed, no more variable, least of all "random variables", and the entire uncertainty, including random error, must be considered by us as a systematic error when we need to apply it. The necessity of the foregoing was discussed in detail by Youden in Reference 15. (See also Ref. 10.)

In part III.B., under "varying Systematic Errors of Standards", we shall see how trend charts can help us "unfreeze" the random error and the variable part of the systematic error of laboratory A in the long run to reduce the total uncertainty accompanying the value of our standard. But without such techniques, the total uncertainty of a higher echelon measurement becomes a systematic error in its lower echelon application. (See also Reference 15.)

B. Varying Systematic Errors of Standards

The certification uncertainty of our standard is not the only uncertainty introduced by the standard. In general, another uncertainty term must be found which combines all those fluctuations in the value it represents as may be caused by changes in the standard itself, usually due to drift, instability, or use (wear), or by external influences which temporarily affect the value of the standard, such as temperature, barometric pressure, air ionization, solar activity, etc., and which are not controlled – and may not even be known to have an effect – at the time the value of the standard is being determined. Some of these effects may at least partly be accounted for in the random error of the measurement process employed when determining the value of the standard. But, unlike that part of the random error which is entirely caused by the measuring process and the effect of which will remain constant once the measuring process is terminated and the value of the standard reported, the latter effect will continue to change the value of the standard, adding to the uncertainty.

In general, we would not know whether differences in subsequent values of our standard as certified by Labora*tory A are caused by the random error of Laboratory A* only, by changes in Laboratory A's systematic error, by changes in the value of the standard itself, or by a combination of these causes. To be able to assign to a standard an uncertainty which includes all experienced effects of variation from the unknown and unknowable true value of the standard, we must analyze the history of that standard.

A valuable tool in analyzing and displaying the history of a standard is a Trend Chart (see Figures 1 and 2). The dots and x's on the Trend Charts represent values reported by the National Bureau of Standards (NBS) for the standards in question at the indicated dates. Their dispersion pattern suggests a Trend Line which represents the best value known of each standard. The Trend Line value of the standard is *most likely a more accurate value of the standard than the* latest NBS reported value, and the Trend Chart has on occasion served to correct errors in the NBS reported values. The dispersion of the points shows graphically the effect of all the variables influencing the known value of the standard over the long run, such as possible long term variations in the systematic error of the standardizing Laboratory (NBS in this case), its random error, and actual changes of the standard. In fact, the Trend Charts of the gage blocks indicated a shrinkage of the blocks before the reasons for such shrinkage were known (wear effects could be excluded in the case of the blocks represented by Figure 2).

Furthermore, the calculated envelope around the points at two or three sigma of the points' dispersion around the Trend *Line represents the total uncertainty of the value of* the standard, excluding the systematic error of the standardizing laboratory. The systematic error of the standardizing

APPENDIX

Examples on Accumulating Systematic Errors

A. Linear Accumulation

Where component magnitudes are plainly added to arrive at an overall magnitude, as in the stacking of gage blocks or the series connection of resistors, systematic errors are also added. An example is the series of resistors, Figure A, with values as in Table 4.

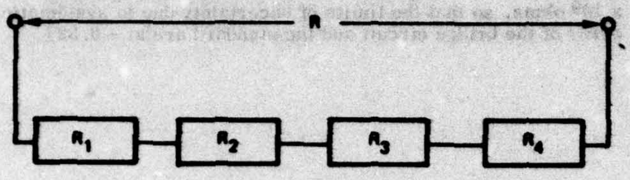


Figure A. Example of Linear Accumulation of Systematic Errors, Series Connected resistors or stacked gage blocks

Table A
Values of Resistors, Figure A.

Element	Measured Value	Total Uncertainty of Measured Value: Symbol	in %	in Ohms
R_1	10 MΩ	e_{R1}	1%	100 kΩ
R_2	5 kΩ	e_{R2}	5%	250 Ω
R_3	130 Ω	e_{R3}	5%	6.5 Ω
R_4	10 Ω	e_{R4}	5%	0.5 Ω
R =	10 005.14 kΩ	e_R =		100.257 kΩ

The total resistance R of this series of resistors is 10 005 140 ohms with an uncertainty of 100 257 Ohms or 1%.

Or, applying the propagation of error equation (1):

The equation for R is

$$R = R_1 + R_2 + R_3 + R_4$$

$$= 10\ 005\ 140\ \Omega$$

$$R = f(R_1, R_2, R_3, R_4)$$

$$e_R = \left|\frac{\partial R}{\partial R_1} e_{R_1}\right| + \left|\frac{\partial R}{\partial R_2} e_{R2}\right| + \left|\frac{\partial R}{\partial R_3} e_{R3}\right| + \left|\frac{\partial R}{\partial R_4} e_{R4}\right| \qquad (A)$$

$$\frac{\partial R}{\partial R_1} = \frac{\partial R}{\partial R_2} = \frac{\partial R}{\partial R_3} = \frac{\partial R}{\partial R_4} = 1$$

hence $e_R = e_{R_1} + e_{R_2} + e_{R_3} + e_{R_4}$

$$= 100\ 257\ \Omega = 1\%\ \text{of}\ 10\ M\Omega$$

B. Non-Linear Accumulation

a. Series-Parallel Resistors

A network of series-parallel connected resistors may serve to illustrate the propagation of errors in a non-linear case. The network is as in Figure B with values as in Table B.

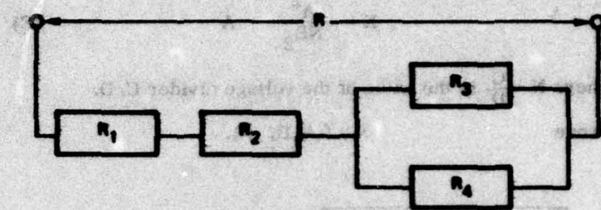


Figure B. Series-Parallel Connected Resistors

Table B.
Values of Resistors, Figure B

Element	Measured Value	Total Uncertainty of Measured Value: Symbol	in %	in Ohms
R_1	1000 Ω	e_{R_1}	1%	10 Ω
R_2	500 Ω	e_{R_2}	3%	15 Ω
R_3	350 Ω	e_{R_3}	5%	17.5 Ω
R_4	150 Ω	e_{R_4}	5%	7.5 Ω

The equation for R, Figure B, is

$$R = R_1 + R_2 + \frac{R_3 R_4}{R_3 + R_4} = 1605\ \Omega \qquad (B)$$

$$R = f(R_1, R_2, R_3, R_4)$$

Substituting R, Equation B, for Q in Equation (1), the result is again as in Equation (A). The partial derivatives of Equation (A) are here:

$$\frac{\partial R}{\partial R_1} = \frac{\partial R}{\partial R_2} = 1$$

$$\frac{\partial R}{\partial R_3} = \frac{R_4^2}{(R_3 + R_4)^2} = \left(\frac{150}{500}\right)^2 = 0.09$$

$$\frac{\partial R}{\partial R_4} = \frac{R_3^2}{(R_3 + R_4)^2} = \left(\frac{350}{500}\right)^2 = 0.49$$

Thus Equation A becomes

$$e_R = 10\ \Omega + 15\ \Omega + 0.09 \times 17.5\ \Omega + 0.49 \times 7.5\ \Omega$$

$$= 30.25\ \Omega = 1.9\%$$

b. Measuring a High Resistance

This example illustrates the calculation of the systematic error incurred in measuring the value of a 1 gigaohm resistor (X) against a 100 kiloohm standard resistor (A) whose value was determined earlier with a total uncertainty of 0.018%. The apparatus used in the measurement is a differential high resistance bridge shown in Figure C.

This bridge is first balanced with resistor X connected to terminals 1 and 2 and with the calibrated decade resistor set to zero (B_1=0). Resistor X is then disconnected and the bridge balanced again by increasing the resistance of arm B by an amount B_2, using the decade resistor. The measured value of X is then

$$X = \frac{A^2}{NB_2} - A \qquad \text{(C)}$$

where $N = \frac{C}{D}$ is the value of the voltage divider C/D.

Hence $X = f(A, B_2, N)$.

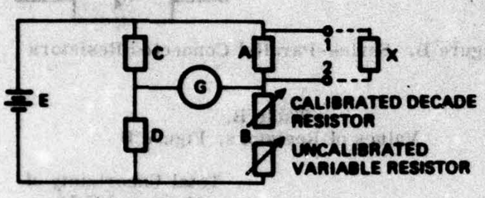


Figure C. Differential High Resistance Bridge

The values and uncertainties of the individual elements of Equation (C) are found as tabulated in Table C.

Table C
Values of Elements, Equation (C)

Element	Measured Value	Total Uncertainty of Measured Value: Symbol	in %	Magnitude
A	100 005 Ω	e_A	0.018%	18 Ω
B_2	9.905 Ω	e_B	0.27%	0.027 Ω
N	1.000	e_N	0.011%	0.000 11

Applying equation (B) where Q now must be replaced by X of Equation (C), the systematic error of the resistance measurement due to the uncertainty of the bridge elements then becomes:

$$e_X = \left|\frac{\partial X}{\partial A} e_A\right| + \left|\frac{\partial X}{\partial B_2} e_B\right| + \left|\frac{\partial X}{\partial N} e_N\right| \qquad \text{(D)}$$

The partial derivatives of X with respect to A, B_2, and N are then, respectively,

$$\frac{\partial X}{\partial A} = \frac{2A}{NB_2} - 1$$

$$\frac{\partial X}{\partial B_2} = -\frac{A}{NB_2^2}$$

$$\frac{\partial X}{\partial N} = -\frac{A}{N^2 B_2}$$

so that (D) becomes

$$e_X = \left|\left(\frac{2A}{NB_2} - 1\right) e_A\right| + \left|\frac{A^2}{NB_2^2} e_B\right| + \left|\frac{A^2}{N^2 B_2} e_N\right|$$

$$= \left(\frac{2 \times 100\,005}{1 \times 9.905} - 1\right) 18 + \frac{(100\,005)^2}{1 \times (9.905)^2} \times 0.027$$

$$+ \frac{(100\,005)^2}{1^2 \times 9.905} \times 0.000\,11$$

$$= 3\,226\,834 \text{ ohms or approximately } 3.2 \times 10^6 \text{ ohms.}$$

The value of X, using equation (C), calculates to be 1.009 6 x 10^9 ohms, so that the limits of uncertainty due to systematic error of the bridge circuit and the standard are at ± 0.32%.

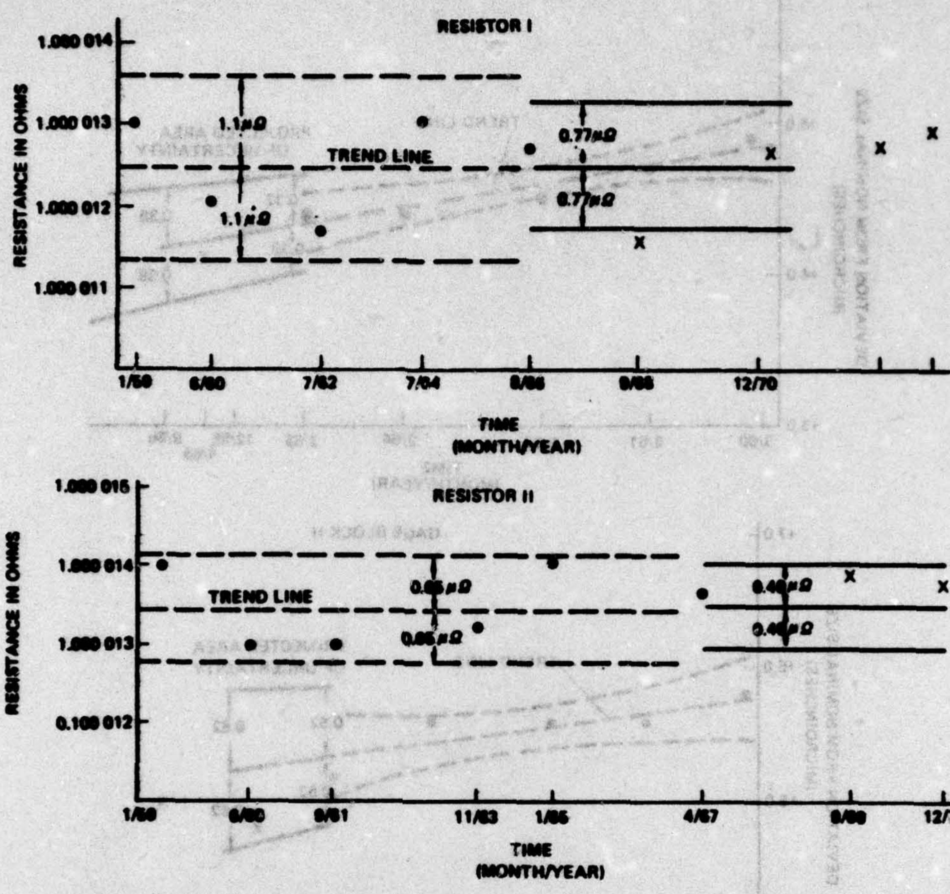


Figure 1. Trend Charts of Two Thomas Type Standard Resistors

laboratory is possibly in part included in a Trend Chart spanning a long period, but we don't know how much of it is included, if anything. Thus, to estimate the total uncertainty of the standard's value at any time, we must add the reported systematic error of the standardizing laboratory to the width of the envelope on the Trend Chart. And this total uncertainty of the standard is part of our systematic error when we use the standard as a basis for measurements. Note that this systematic error includes random variables which have been frozen for the time being and must be considered as being part of the systematic error.

For a detailed description on how to construct Trend Charts, see References 8* and 9.

Figure 1 shows the Trend Charts of two standard resistors constructed in July 1964 and January 1965 respectively (dashed lines). At that time, the best values for these two resistors were 1.000 012 43 and 1.000 013 44 (Trend Lines). In the absence of any information to the contrary, the Trend Lines were believed to be horizontal, assuming *no* detectable change in the values of these resistors for the near future (two to three years). Because of the scarcity of the points available, 2-sigma control limits were drawn based on the scatter of the points around their average (the Trend Line in this case); this gives a 95 percent confidence that the "true value" (i.e. the limiting mean value of all experienced variables) lies between those control limits, disregarding for the moment any systematic error of NBS. As more points become available 3-sigma control limits would reduce the risk of the limiting mean value being outside the control limits to an insignificant 0.3 percent.

As the Trend Line value of these standards is the best known value, this is the value assigned to the resistor at any time, regardless of the latest calibration value of the resistor reported to us by NBS. Note that the reported values have been jumping up or down by as much as one part per million from one calibration to the next. If we would always assign the latest calibration value to these standards, our entire measurement system would experience these ripples, ripples which in the end would cancel each other around the limiting mean value, the Trend Line value. Instead, we compute a new Trend Line value, a new mean value in this case, every time a new calibration value is obtained. Thus, every new calibration value will change the new mean only insignificantly, if at all. Hence the Trend Line value assigned to the standard will remain stable over the years.

The control limits represent the limits of uncertainty of the Trend Line value and hence, represent the limits of uncertainty due to variable error components of the value which we assign to the standard. They encompass many more variables and sources of uncertainty than those experienced by NBS in the calibration of the standards. The limits of uncertainty about the Trend Line of the first five values of Resistor I, Figure 1, are at 0.77 microohms from the Trend Line; adding to that the estimated limits of the systematic error by NBS of about 0.5 microohms, the total uncertainty of the value of Resistor I (1.000 012 48 ohms) is 1.27 microohms. This is the systematic error introduced

*Reference 8 should only be used with an errata sheet available from the author since it contains numerous typographical errors.

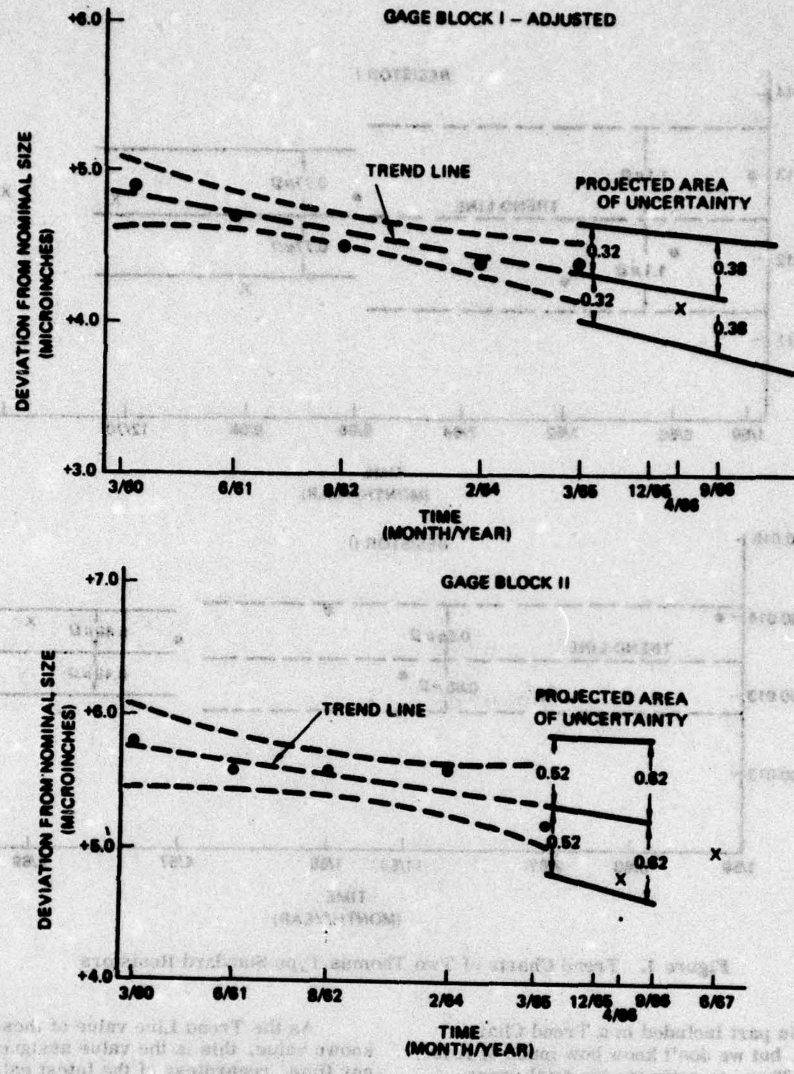


Figure 2. Trend Charts of Two Gage Blocks

by the standard to any measurement made with it. Note that the 0.77 microohm control limits represent only 95 percent confidence that this range brackets the "true value of the standard. If we want to increase the confidence limits to 99.73 percent (3-sigma equivalent), the control limits would have to be widened to 1.77 microohms about the Trend Line value for a total uncertainty of 1.77 + 0.5 = 2.27 microohms. In order to narrow these limits of uncertainty, we need a longer history with more values as exemplified by the Trend Chart of Resistor II, Figure 1.

The first five values of Resistor II yielded a Trend Line (average) of 1.000 013 44 ohms with 95 percent confidence limits of ± 0.65 microohms. A sixth value obtained in April 1967 of 1.000 013 6 ohms, as reported by NBS, gave rise to a new assigned value of Resistor II of 1.000 013 47 ohms with 95 percent confidence limits narrowed down to 0.49 microohms and a total uncertainty of the standard's value of 0.49 + 0.5 (estimated NBS systematic error) = 1 microohm.

Many a laboratory is using the latest calibration value of their standard as the value assigned to it until the next calibration of that standard and use the uncertainty reported by the calibrating laboratory (NBS in our case) as the total uncertainty of the standard's value. The fallacy of that practice becomes evident from the above discussion. These laboratories in fact experience a far higher uncertainty than they are aware of.

Going back to Resistor I, Figure 1, we note that the next reported calibration value (September 1968) was 1.000 011 6 ohms and fell, therefore, slightly outside the control limits. We rejected the value as being in error, realizing that we stood a 5 percent change of rejecting a perfectly valid value. But a new measurement confirmed that the September 1968 value was erroneous.

Figure 2 shows Trend Charts of two special gage blocks to illustrate the fact that some standards undergo gradual changes in time. In such cases, a Least Squares Trend Line can be constructed, depicting with sufficient accuracy the behavior of the standard and permitting a forecase of its most likely value for the near future. Since the slope of such a Least Squares line is also burdened with some uncertainty, the band of uncertainty around the forecast future values of such standards continues to widen as time progresses as shown in Figure 2, expressing numerically our uncertainty about the exact value of the slope. This widening bandwidth of uncertainty also tells us when we need to have the standard recalibrated, lest its uncertainty grows intolerably large for the requirements of our products.

The dots represent the values known at the time at which the Trend Charts were constructed and on which the Trend Charts are based. The x's represent values obtained later and are shown to illustrate the predictive value of the Trend Charts. The Trend Charts are, of course, updated each time a new value is obtained.

As time progresses, however, technology, the methods used in calibrating the standards, and frequently the accuracy requirements, change. There come times then when we can no longer continue the Trend Charts as before. New values, obtained by different measurement methods, are no longer readily comparable with old ones. As an example, refer to Figure 3, the Trend Chart of Resistor II in the 1970's. The uncertainty values attached to the calibration values of this resistor in the 1960's (Figure 1) were at best within 1 microohm; the uncertainty values quoted by NBS for the values charted in Figure 3 are typically around 0.09 microohms. The new Trend Line value assigned to this resistor is now 1.000 013 934 ohms, and 3-sigma equivalent control limits are at ± 0.50 microohms from that Trend Line. (Two-sigma control limits are also shown as light dotted lines at ± 0.18 microohms for comparison purposes, showing how the closer agreement between the plotted points results in a much narrower band in which the "true value" of the resistor is likely to lie.) NBS's systematic error is now reduced to about 0.03 microohms, so that we can now say that the best known value of this resistor is "1.000 013 93 ohms with a total uncertainty of 0.54 microohms, derived from bounds of ±0.03 microohms on the systematic error and a computed random error of 0.50 microohms based on a two-tail 0.0027 probability value for 3 degrees of freedom."

C. Varying Systematic Errors of Measurements

The magnitude of a systematic error of a measurement process is established by the capability and willingness of the metrologist to measure those phenomena which cause systematic errors, to find their quantitative effects (i.e., what quantity of the phenomenon is required to cause a measurement error of a given magnitude), and to account for them or eliminate them. If it is known how much of a phenomenon causes a given amount of a measurement error and if we are capable – technically and economically – of measuring the phenomenon and making the required adjustment, the systematic error is determinable and should be eliminated. If the state of the technology or economics forbid a determination and subsequent elimination of the systematic error, it remains unknown and must be estimated. The following example is intended to illustrate the transitory nature of the limits between systematic and random errors of a measurement process.

An unknown standard resistor with a nominal value of 1000 ohms was compared against a known standard resistor one day between 10:00 a.m. and 10:20 a.m.. A ratio set was used for comparing the two resistors in air. The five values obtained were plotted on a chart, Figure 4. The ambient temperature was 25.4 degrees Celsius, varying by less than 0.1 degree during that time. The temperature coefficients of resistance of the two resistors were known, and all plotted values were corrected for a base temperature of 25 degrees Celsius.

The effect of temperature on the entire measuring system was not known. The ambient temperature varied between approximately 25.5 and 24.5 degrees C at a typical rate of approximately 0.5 degrees C per hour. It could, therefore, be assumed that the system was not at thermal equilibrium at any time; and the effect of this inequilibrium on the result of a single set of measurements would have to be estimated and a numerical value assigned to this effect as part of the systematic error if the value of the unknown resistor had to be determined on the basis of the first set of measurements. The nearly perfect agreement of the values obtained in the first set indicates that the resolution of the ratio set was insufficient for the measuring system to respond significantly to random fluctuations. An analysis of the random error of this set based on more than an inspection of the data, where four values are identical and only one differs from the other four by the smallest unit possible to resolve, would be of little value.

The correct value of the standard resistor had been determined by a more accurate measurement process before and after the measurements described in this example and was 1000.0120 ohms with a constant systematic error and with a random error of considerably less than 1 ppm.

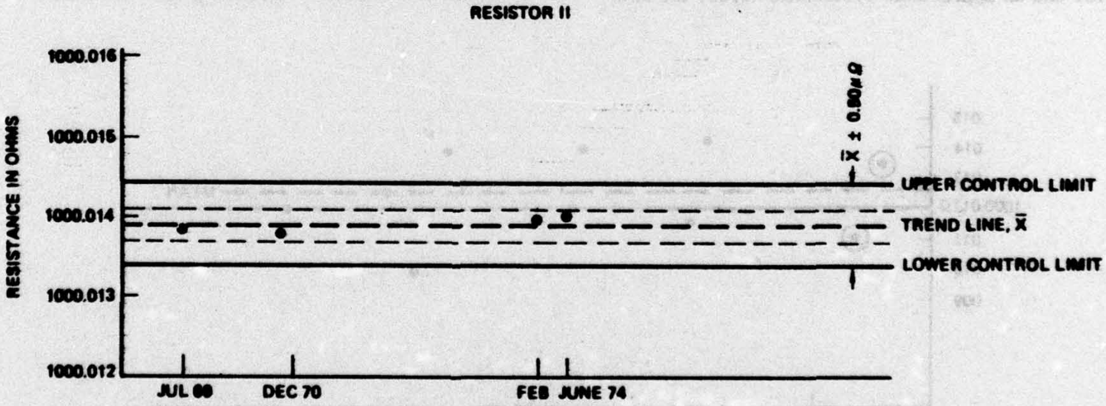


Figure 3. Trend Chart of Recent Values of Resistor II

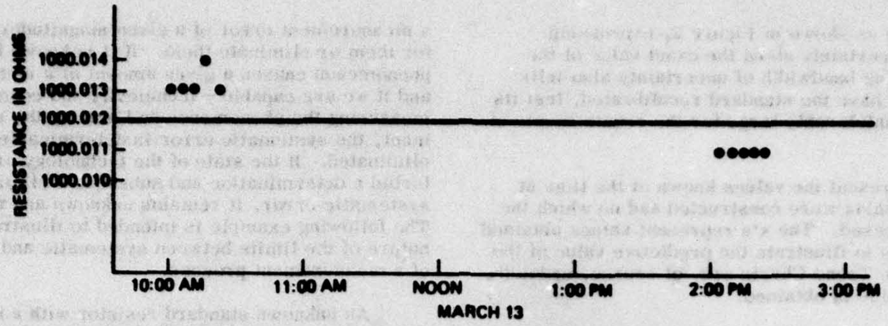


Figure 4. Two Sets of Measurements of a 1000-ohm Standard Resistor

Between 2:00 and 3:00 p.m. of the same day, another set of measurements was made and is also plotted on the control chart. The ambient temperature was 24.9 degrees Celsius. Again, the five measured values, corrected for the base temperature, agreed very well; perfectly, as a matter of fact. But they differed markedly from those obtained previously. It could be said that each set of measurements was under (simple) statistical control individually; but both sets together indicate an out-of-control condition, the two sets being offset from one another by a systematic error. Since no means were available to keep the ambient temperature more stable, and since the effect of temperature inequilibrium on the components of the measuring system was not known, this systematic error is unknown. If it had been possible to predict the systematic error due to temperature difference and inequilibrium or to maintain the ambient temperature more closely, the measured values could have been adjusted for the temperature inequilibrium, or the condition causing the inequilibrium could have been avoided. (The effect of temperature on the systematic error of the resistance measurement is given here as one example only. Many other sources of systematic error exist in practice and may influence the measurement uncertainty.)

The resistance of the unknown standard resistor was then measured repeatedly under similar circumstances. The results of these measurements were entered on the chart shown in Figure 5. Each plotted point is the mean of five successive determinations of the resistance of the unknown corrected, as far as possible, for 25 degrees Celsius. The points now are randomly distributed about a mean of 1000.0124 ohms. Whereas each group of five resistance determinations was made with a very small random error and an appreciable systematic error, the new mean has a drastically reduced systematic error, but an appreciable random error. By taking the measurements over a prolonged period of time, the systematic error of each group was allowed to vary and to become part, the major part in fact, of the random error of the measuring process.

Reference 1 describes the phenomena of errors changing their nature in the following terms: "If the cause system is enlarged, then what was previously predictable bias may, in terms of the new cause system and process, vary in a random fashion and therefore be attributable now only partly, if at all, to bias. Furthermore, enlargement of the cause system requires re-evaluation not only of the bias involved in the stated accuracy but also of the stated precision."

Temperature changes, like many other causes of systematic errors, frequently occur in patterns, such as rough sinusoids or, in the case of closely controlled environments, in sawtooth patterns, not randomly. The random appearance of the points on the chart of Figure 5 is due to the random selection of the times at which the measurements were made. It is interesting to note that an overestimate of the limits of uncertainty due to random errors would be obtained if the points would follow exactly a sinusoidal or a sawtooth pattern and the calculations were made as if they were normally distributed. In fact, all points of a sinusoid are within a region bounded by $0 \pm \sqrt{2}\sigma$ *or within 1.41 standard deviations* from the mean, and all points of a sawtooth curve of isosceles triangles with height h are within a region bounded by

$$\frac{h}{2} \pm \sqrt{3}\sigma$$

or within 1.73 standard deviations from the mean.

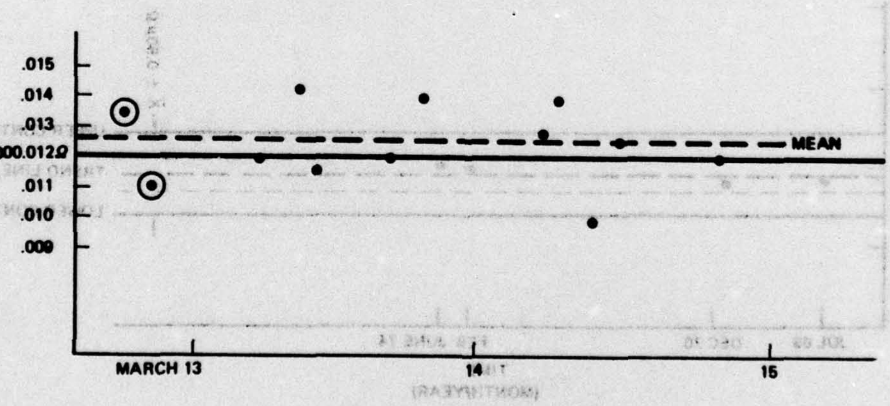


Figure 5. Twelve Means of Sets of Five Measurements

⊙ Averages of values from Figure 4.

The preceding examples illustrate, therefore, that:

1. A measuring process ordinarily under statistical control will get out of control if the systematic error of the process varies, but control may be reestablished if enough changes are allowed to occur and the changes in systematic error are allowed to join the chance system of variations.

2. If changes in systematic error are allowed to occur, the standard deviation of the measuring process will be larger than in the absence of such changes, and the systematic error will be reduced.

3. The time element will probably always change the "cause system" and transform constant errors to changing errors, and parts of systematic errors into random errors.

The last example, however, raises another question. Is the 0.4-milliohm difference between the mean and the certified value attributable to a systematic error or could it be caused entirely by the fluctuations which a mean from a limited number of measurements could be expected to display? Is there, in other words, a bias in the measuring process which should be eliminated? Section VI will show one way to answer this question.

IV. The Relative Importance of Systematic Errors

The systematic error frequently becomes the insurmountable barrier toward more accurate measurements, especially at lower echelon laboratories. But sometimes it can be disregarded, no matter what its magnitude.

For instance, the National Bureau of Standards certifies standards to values with uncertainties relative to the legal units as maintained at NBS. These uncertainties *reflect mainly the errors accrued in comparing a submitted* standard with the national reference group, mostly indirectly, i.e., through intermediate standards. But they make no allowance for the systematic errors which may be inherent in the national reference group; this part of the systematic error is the difference between the magnitude of the unit of measurement as represented by the national reference group and the absolute magnitude of the unit as the unit is defined. This additional systematic error must be taken into consideration when measurements based on the legal unit (volt, meter, ohm, kilogram, etc.) are to be expressed in absolute units or in the fundamental units of physical measurements (kilogram, meter, second, ampere, kelvin) from which the unit under consideration is derived. It must also be taken into consideration when measurements are compared internationally, *i.e.* when measurements made in terms of the U.S. legal unit are compared with measurements made by a foreign laboratory which had its standards certified by a different national laboratory. In the case of the unit of electromotive force, the volt, this additional systematic error is in 1976 estimated to be less than one part per million and probably less than one-half part per million.

However, in most electrical measurements made in the U.S., the systematic error of the national reference group is disregarded, for it is of no consequence as far as compatibility of measurements and production within the U.S. are concerned, as long as all measurements are based on the same reference group.

The same reasoning for disregarding some systematic errors can be applied to similar cases on a much smaller scale. The reference standard of Company C may have an uncertainty due to systematic errors of e_{ssc}. Since all *measurements made within* Company C have that common systematic error component, the uniformity of production and the compatibility of subsystem and hardware components can be assured even if e_{ssc} is neglected. The measurements within the company are then in units "as maintained at Company C." Such an arrangement could greatly facilitate the evaluation of relative uncertainties of the measurements made by the various departments and on the various production lines of Company C, especially if the products are of high precision. The previously neglected systematic error component e_{ssc} may have to be taken into account in evaluating the finished product. But in some instances it does not have to be reintroduced at all.

V. Accumulating Systematic Errors

The question now arises: Given the estimates for the maximum limits of all non-negligible systematic errors which must be taken into consideration, how is their effect on the measurement estimated? If the systematic error of a *measurement* is dependent upon several systematic errors whose estimated maximum limits are given, how are they combined to yield the systematic error of the measurement?

For a general *discussion of the treatment of systematic* errors, let a system measuring the Quantity Q be described as $Q = f_{(a,b,c,\ldots)}$, where a, b, c... are components of the system representing the known quantities A, B, C... The magnitudes of these systematic errors have been estimated to be not larger than e_A, e_B, e_C.... The propagation of error equation then gives the relative contribution of each systematic error term to the resulting systematic error as

$$\left|e_Q\right| = \left|\frac{\partial Q}{\partial a} e_A\right| + \left|\frac{\partial Q}{\partial b} e_B\right| + \ldots \left|\frac{\partial Q}{\partial i} e_i\right| + \ldots \left|\frac{\partial Q}{\partial n} e_n\right| \quad (1)$$

provided the errors are mutually independent.

The application of the propagation of error equation (Ref. 5) is illustrated in the Appendix with a few examples.

Mindful of the facts that systematic errors are usually quoted as "$\pm e_i$" or "$+e_{i1}$, $-e_{i2}$", i.e. with a positive and a *negative limit*, and that *they can be only positive or negative*, but never both, metrologists have tended to devise methods to reduce the purely additive effect of the individual terms of equation 1. The most commonly used method to reduce this purely additive effect of errors is the rss-method by which the individual terms are squared and the square-root of the sum of the squares used as the estimate of the overall uncertainty. The conscientious metrologist will in the vast majority of cases avoid such an arbitrary reduction and adhere strictly to the additive form of the individual terms of that equation as written.

Youden[15] has shown that the combination of error terms in quadrature yields erroneous results in a chain of laboratories because it is deficient in logic. (See also Ref. 10.) In *fact*, any method to reduce the combined effects of several systematic error terms is arbitrary and can seldom be justified logically for the following reasons:

1. Individual systematic error terms come from different populations, distributions, having different origins and means, and are unrelated.

2. In most practical problems, the number of separate, individual error terms is small, and the probability that all *error terms* have the same sign is appreciable; four error terms still have a 12.5 percent chance of ganging up with the same sign, and the consistent metrologist will not quote 3-sigma limits of uncertainty on the random error and take a 12.5 percent chance on the systematic error.

3. Although systematic errors are mostly believed to be overestimates, they may also be underestimates – and frequently are.

But this problem appears to be more a theoretical than a practical one. In accordance with Juran's principle of "the vital few and the trivial many", in most practical cases only a few individual component errors determine the magnitude of the resulting overall systematic error, while the others are of little or no consequence.

Thus remains to consider the theoretically possible but rare case of a large number, say ten or more, of systematic error terms of approximately the same magnitude where some mutual cancellation is likely to occur. In this case, the critical metrologist will accept any logical rationale to reduce the total error to something less than the sum of all individual terms. However, prudence and consistency would dictate that the resulting total error be not less than the sum of the nine largest individual terms when three-sigma limits are quoted for the random error and not less than the sum of the six largest individual terms when two-sigma limits are quoted for the random error.

VI. Keeping Measurement Errors Under Control in a Standards Laboratory

A standards laboratory involved in certifying standards for lower echelon laboratories must make considerable efforts to keep its errors under control. Ordinarily, it does not suffice that the measurement errors are determined periodically and that it is assumed the errors thus determined are typical and recurring at the same magnitude. A positive proof that the random error of a measurement was indeed typical and that no unusual systematic errors have occurred during the measurement would go a long way in enhancing the laboratory's confidence in its measurements and contribute significantly to an eventual reduction of realistically quoted limits of uncertainty.

Quality control practices, time proven and honored in production processes, provide us with excellent tools for the observation of the quality of measurements: Control Charts. "Measurement of some property of a thing is ordinarily a repeatable operation. This is certainly the case for the types of measurement ordinarily met in the calibration of standards and instruments. It is instructive, therefore, to regard measurement as a production process, the 'product' being the numbers, that is, the measurement processes in the laboratory with mass production processes in industry."[4] Reference 2 gives detailed instructions on establishment and maintenance of control charts.

Figure 6 is a hypothetical example of a control chart for a 1000-ohm standard resistor. The resistor serves as our standard and is periodically calibrated against a higher echelon standard. Its value is established by a Trend Chart. Each time we calibrate a 1000-ohm standard resistor for somebody else, we also measure one of our own 1000-ohm standards (in this case the one whose control chart appears in Figure 6) by the same process and using the same instruments, in a pilot measurement.

Each pilot measurement consists of a set of four measurements taken at different times. The mean of each set of four is plotted on the $\overline{X}$-chart, and the range, the difference between the largest and the smallest measured values, is plotted on the R-chart. Upper and lower control limits are calculated as outlined in References 2 and 7.

Although occasionally a point may fall outside the control limits as a matter of pure chance, each time this happens the existence of some abnormality in the measurement is suspected, and the measurement is repeated. If the point of the repeat measurement also falls outside the control limits, the existence of an abnormality is taken as being confirmed. The causes for this abnormality are then determined, removed, and the measurement is repeated. If the points on the $\overline{X}$ and the R-chart fall within the control limits, the measurement is considered valid. Since our customer's unknown standard was measured by the same process, its value thus determined is also considered valid.

Points of the $\overline{X}$ chart will reveal the existence of short term trends or cycles, and unusual systematic errors. The R-chart is predominantly an indicator of the quality of the measuring process, reflecting, among other things, the care of the operator and the control of the environment in which the measurement was performed. Since a control chart is the result of a compilation of a considerable amount of data concerning one measuring process, the average experienced range $\overline{R}$ can be used to determine the standard deviation of the measuring process. "Most experimental scientists have very good knowledge of the variability of their measurements, but hesitate to assume known σ without additional justification. Control charts can be used to provide the justification."[7]

The last point on the control chart for averages and that for ranges in Figure 6 represent the last calibration of 1000 ohm resistors performed by our lab. The average of the four measurements taken on the pilot resistor charted in Figure 6 was $\overline{X}$ = 1000.0121 ohms and the range of the four

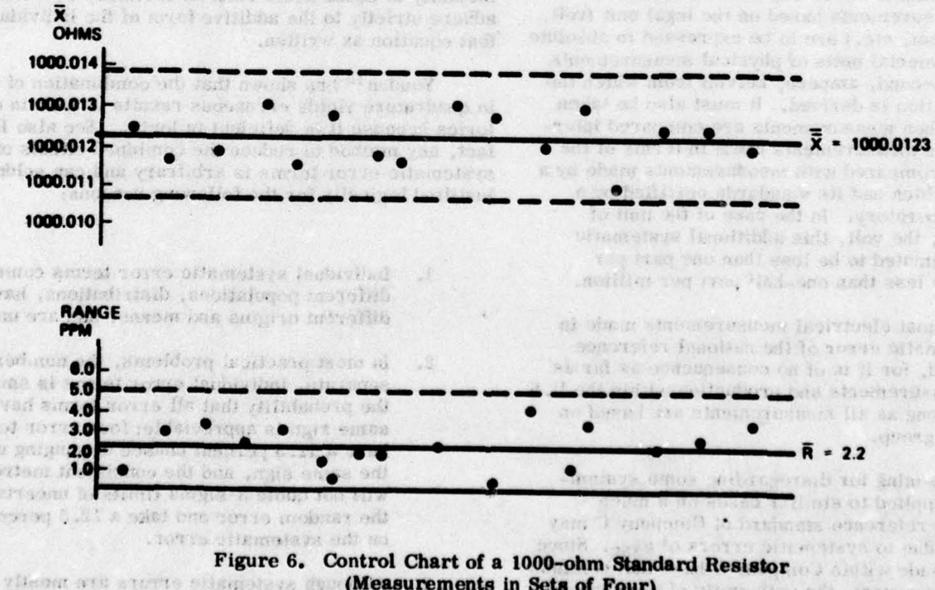


Figure 6. Control Chart of a 1000-ohm Standard Resistor (Measurements in Sets of Four)

measurements was then 3.8 milliohms. The recent grand average of pilot measurements made with this resistor is $\overline{\overline{X}}$ = 1000.0123 ohms and the average range is $\overline{R}$ = 2.2 milliohms as plotted. Upper and lower control limits on the $\overline{X}$-chart of the averages are at a distance of

$$A_2\overline{R} = 0.729 \times 2.2 = 1.6 \text{ milliohms}$$

from the grand average $\overline{\overline{X}}$. Upper and lower control limits of the R-chart of ranges are at

$$D_4\overline{R} = 2.282 \times 2.2 = 5.0 \text{ milliohms}$$

and

$$D_3\overline{R} = 0 \times 2.2 = 0.0 \text{ milliohms}$$

respectively, as plotted. The estimated mean standard deviation of the calibration process which yielded these results is

$$\delta = \frac{\overline{R}}{d_2} = \frac{2.2}{2.059} = 1.07 \text{ milliohms,} \qquad (2)$$

and the uncertainty of measured value of 1000.0121 ohms due to random error at 3-sigma levels is

$$3\delta_{\overline{X}} = \frac{3\overline{R}}{d_2\sqrt{n}} = \frac{3 \times 2.2}{2.059\sqrt{4}} = 1.6 \text{ milliohms}$$

The factors A_2, d_2, D_3, and D_4 are tabulated in references 2 and 7.

The two last points, like the others before, fell within the control limits, indicating that no abnormal systematic or random error had occurred during the measurement and that the measurement was, therefore, typical for this particular process.

The distribution of the points in Figures 4 and 5 is not necessarily typical for normal laboratory comparisons between standard resistors where it is frequently possible to control the factors influencing the measurement (e.g. temperature) more closely and attain narrower uncertainty bands, i.e. narrower regions bounded by control limits.

The difference between the value of the standard as determined by the Trend Chart and the grand average $\overline{\overline{X}}$ on the Control Chart must be considered a systematic error of known magnitude and sign if the difference is statistically significant. A significant difference must be removed by adjustment or by applying this difference as a correction to all measurements made by the process under consideration.

The difference is statistically different at a level α, if it is larger than

$$\mu = z_p \frac{\sigma_{\overline{X}}}{\sqrt{n}} \qquad (3)$$

where z_p is the standard normal variable or the ordinate of the normal curve for a cumulative area of p under the normal curve, and where $p = 1 - \frac{\alpha}{2}$

Using Equation (3), we can now answer the question at the end of Section III C. Let us assume that the average range of the five measurements (n=5), was R = 0.7 milliohms. Choosing α = 5% or 0.05, p = 0.975. We first calculate $\delta_{\overline{X}}$ from Equation (2) as

$$\delta_{\overline{X}} = \frac{\overline{R}}{d_2} = \frac{0.7}{2.326} = 0.30 \text{ milliohms.}$$

Then, from Equation (3),

$$\mu = 1.96 \frac{0.30}{\sqrt{5}} = 0.26 \text{ milliohms.}$$

Since the difference between our measured average and the certified value is 0.4 milliohms and, therefore, larger than μ, the difference is significant, and we conclude that a bias exists which calls for immediate correction.

To allow for changes in the systematic error contributed by the measuring process or the standard, the grand average must be updated periodically, at which time it may be necessary to omit some of the oldest points and recalculate the average on the basis of the latest points. Between 10 and 25 points should be available for calculating $\overline{\overline{X}}$ and $\overline{R}$ initially, and their values should be updated when additional groups of points become available.

VII. CONCLUSION

The majority of all sources of uncertainty which a metrologist normally encounters must be treated as systematic, even though many of them may have originated as random errors. The boundaries between systematic and random errors are fluctuating, and systematic errors may at times recover their random nature if we increase their "cause system", i.e. frequently the time period over which we consider them. The statistical treatments developed for the determination and analyses of random errors do not normally apply to systematic errors. We have seen how we can determine the magnitudes of systematic errors of standards and measurements, how we can control them, and how a control of the systematic errors, combined with a control of the random errors of measurements, can lead to a positive and definitive control of the entire measurement process. Trend Charts for standards and Control Charts for measurement processes are invaluable tools for the sophisticated metrologist to enhance his knowledge of the behavior and uncertainty of his standards and measurement processes.

The emphasis is on collecting and analyzing information. Proper handling of data can then help to convert available information into a different form, but data manipulation cannot be a substitute for knowledge.

VIII. ACKNOWLEDGEMENTS

The author is deeply indebted to Messrs. H. H. Ku and J. M. Cameron of the Statistical Engineering Laboratory of NBS as well as to their colleagues in the SEL not only for their efforts in reviewing an earlier draft of this paper and for their valuable comments pertaining to it, but also for the guidance which this group has given, and continues to give, to the author as well as to a great many metrologists throughout the nation in matters relating to the application of statistical methods to measurement data. A great many thanks also to Messrs. C. Peterson and Wells, then at the Resistance and Reactance Section of NBS, Messrs. Beers and Tucker of the Length Section, and Messrs. Belecki and Wells of the Electrical Reference Standards Section for providing important information without which some analyses presented here would not have been possible. The author is especially indebted to Mr. Kenneth Lund of Autonetics' Metrology Laboratory; his many suggestions and comments of a technical and editorial nature have greatly contributed to the improvement of this article. Any errors of commission or omission which may have inadvertently slipped into this paper, however, remain the sole responsibility of the author.

REFERENCES

1. American Society for Testing and Materials, Tentative Recommended Practice for Use of the Terms Precisions and Accuracy as Applied to Measurement of a Property of a Material, ASTM Designation: E-177-61-T, 1961.

2. American Society for Testing Materials, ASTM Manual on Quality Control of Materials, Special Technical Publication 15-C, January, 1951.

3. Dorsey, N. Ernest and Churchill Eisenhart, "On Absolute Measurement," The Scientific Monthly, Volume LXXVII, No. 2, August, 1953.

4. Eisenhart, Churchill, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," Journal of Research, National Bureau of Standards, C. Engineering and Instrumentation, Volume 67C, No. 2, April-June, 1963, pp. 161-187.

5. Ku, Harry Hsien, "Notes on the Use of Propagation of Error Formulas", Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation, October-December 1966.

6. Ku, Harry Hsien, editor, Precision Measurement and Calibration, NBS Special Publication 300 - Volume 1, U.S. Government Printing Office, Washington, D.C. 1969.

7. Natrella, M.G., Experimental Statistics, National Bureau of Standards Handbook 91, U.S. Government of Printing Office, Washington, 1963.

8. Schumacher, Rolf B. F., "Keeping Standards and Measurements under Control in a Standards Laboratory", Paper No. 68-607, presented at the 23rd ISA Annual Conference and Exhibit, New York, N.Y. 1968.

9. Schumacher, Rolf B. F., "Statistical Control in a Standards Laboratory", Measurements and Data, May-June 1969.

10. Schumacher, Rolf B. F., "Systematic Errors in Measurements and Calibrations", paper presented at the Sixth Measurement Science Conference at Cal-Poly, San Luis Obispo, CA, December 1976.

11. Shewhart, W.A., Economic Control of Quality of Manufactured Product, New York: Van Nostrand Co., Inc., 1931.

12. Youden, W.J., "Instrumental Drift", IRE Transactions on Instrumentation, Volume I-11, Numbers 3 and 4, December 1962.

13. Youden, W.J., "Measurement Agreement Comparisons", NBS Miscellaneous Publication 248, Proceedings of the 1962 Standards Laboratory Conference.

14. Youden, W.J., "Systematic Errors in Physical Constants", Physics Today, Vol. 14, No. 9 pp. 32-34, 36, 38, 40, 43, September, 1961.

15. Youden, W.J., "Uncertainties in Calibration", IRE Transactions on Instrumentation, Volume I-11, Numbers 3 and 4, December, 1962