

The Calibration of Measurement Systems

The art of using a consistency chart

Donald J. Wheeler

Who can be against apple pie, motherhood, or good measurements? This is why everyone stands up and salutes when we are told to maintain our measurement systems in good calibration. But what is good calibration? By what method will we achieve it? And how will we know when we have it?

One day I found the following statement about measurement consistency and calibration in a well-known company's Laboratory Procedure Document: "It is the testing department supervisor's responsibility to make sure that the testing equipment is current and in good calibration before testing any samples." Nothing else was included. This was the end of the sentence, the end of the paragraph, and the end of the topic. While statements like this serve to identify who is the scapegoat when things go wrong, they give no guidance on what good calibration is, how to achieve it, or how to know it when you see it.

Rather than this "cover your anatomy" approach which identifies who is to blame, what is needed is a clear and easily understood definition of when a measurement system can be said to be operated consistently. In 1963, Churchill Eisenhart, a statistician working at the Bureau of Standards, wrote: "Until a measurement process has been 'debugged' to the extent that it has attained a state of statistical control it cannot be regarded, in any logical sense, as measuring anything at all." And the only way to determine if this is the case is by using a consistency chart.

CONSISTENCY CHARTS

A consistency chart provides an operational definition of a measurement system. According to W. Edwards Deming, an operational definition consists of three parts:

1. a criterion to be satisfied,
2. a test procedure for determining conformance to the criterion, and
3. a decision rule for interpreting the results of the test procedure.

When you have only the first part all you have is the basis for an argument. It is only when you have all three parts that you have progressed from wishful thinking to an operational definition. As we saw in last month's column, a process behavior chart provides an operational definition of how to get the most out of any process. And it can be used to determine when a measurement system is operating consistently.

1. A measurement procedure is consistent if and only if repeated measurements of the same item result in a sequence of values that are homogeneous (consistently similar).
2. To determine whether or not a sequence of values is homogenous we shall place those values on a Chart for Individual Values and compute the three-sigma limits for this chart according to the usual formulas based on the two-point moving ranges.

3. a. Whenever one of the individual measurements falls outside the three-sigma limits we will say that the measurement procedure is inconsistent.
- b. If this Consistency Chart shows no evidence of inconsistency and the limits are based on at least 17 values, then the measurement process can be said to display at least a minimal degree of consistency.
- c. In either case, the continued use of the Chart for Individual Values for repeated measurements of the same thing can be used to monitor the ongoing consistency of the measurement procedure and to identify occasions when the measurement system is operated inconsistently.

It is always a mistake to adjust or recalibrate a measurement process unless you have clear indication from a Consistency Chart that an adjustment is needed. In the absence of such a signal the measurements are already as consistent as they can be, and any adjustments to the measurement process will merely increase the variation in the measurements.

When you have evidence that the measurement procedure has changed in some way then you should not only seek to adjust or recalibrate the measurement system in order to remove the inconsistency, but you should also seek to find the assignable cause that resulted in the change in the measurement process in the first place.

By changing your measurement process so that it is no longer subject to the effects of an assignable cause you will be preventing future upsets in the measurement system. Moreover, in many instances, you will find that you have also reduced the measurement error. By persistently seeking to identify assignable causes of measurement inconsistency and removing the effects thereof, many practitioners have ended up with measurement processes that were better than they thought possible.

This use of a Consistency Chart as a long-term strategy for improvement has been thoroughly proven. It may not be as flashy as buying a sexy new measurement device, but it is often more effective, and it is always cheaper.

MULTIPLE WEIGHINGS OF NB10

Around 1940 the National Bureau of Standards obtained a ten-gram weight made out of chrome-steel alloy. This weight was designated NB10 and was weighed once each week. The values shown below come from for the first 30 weeks of 1963. The values shown in Figure 1 are the number of micrograms by which the measurement exceeded 9.999000 grams.

<i>X</i>	<i>mR</i>	<i>X</i>	<i>mR</i>	<i>X</i>	<i>mR</i>
591	—	602	5	592	6
600	9	597	5	601	9
594	6	593	4	601	0
601	7	598	5	598	3
598	3	599	1	601	3
594	4	601	2	603	2
599	5	600	1	593	10
597	2	599	1	599	6
599	2	595	4	601	2
597	2	598	3	599	2

Figure 1: Thirty Weekly Weighings of NB10 January thru August 1963

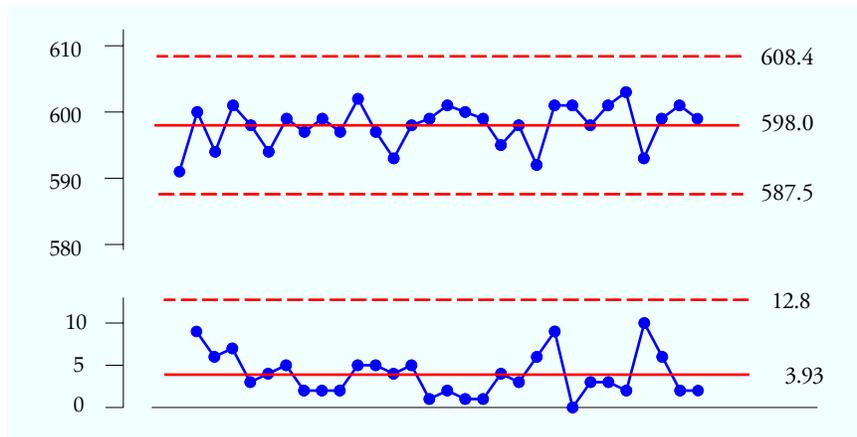


Figure 2: Consistency Chart for Weighings of NB10

The consistency chart for these data is shown in Figure 2 where we find that these measurements display a reasonable degree of consistency. When this ten-gram standard is weighed to the nearest microgram (that one part in ten-million) even the Bureau of Standards does not always get the same value! Since presumably the standard is not changing from week to week, and since gravity is also fairly steady, we interpret the variation in these readings as measurement error.

PRECISION

Precision has to do with the degree that a measurement system can reproduce the same value when it is applied to the same object—a property that is also known as repeatability, test-retest error, measurement error, and replication error. The National Institute for Standards and Technology today recommends that the precision of a measurement system be reported using the estimated standard deviation of the measurement system. Such an estimate is readily found from the Consistency Chart used to establish and maintain the consistency of the measurement process.

The moving range chart in Figure 2 characterizes the measurement error. When we divide the average moving range of 3.93 micrograms by the bias correction factor of 1.128 standard deviations we get 3.48 micrograms per standard deviation. This estimate is based on 30 weighings and is said to have 18 degrees of freedom (because the average moving range has $[1 + 0.605 * (30 - 2)] = 18$ degrees of freedom).

An equivalent way of characterizing the uncertainty in a measurement is to use the Probable Error of a measurement. This quantity is estimated by simply multiplying the estimated standard deviation by the constant 0.675. For our example here this is $0.675 (3.48) = 2.3$ micrograms. The Probable Error is the median amount by which a measurement will err: Half the time the measurement will err by 2.3 micrograms or less, and half the time it will err by 2.3 micrograms or more. Thus, while these measurements are made to the nearest microgram, they are only good to within about 2 micrograms.

These three quantities, the estimated standard deviation, the degrees of freedom in the estimate, and the Probable Error will characterize the uncertainty in the measurements. Of

course, the Probable Error and the estimated standard deviation will only make sense when the measurement system is consistent—that is when the consistency chart shows the measurement system to be predictable. With an inconsistent measurement process the estimated standard deviation and the Probable Error only characterize the hypothetical potential of what could be, if and when the measurement process is operated consistently (i.e., predictably).

BIAS

When operating with a known standard, which has an accepted value, we can use the average of the consistency chart to check for bias in our measurement system. Now it is important to note that the accepted value for a known standard is not the “true value,” but merely the value obtained using some measurement system that is accepted or designated as the “master measurement system.” Bias can only be determined relative to such a master measurement system. So, while in theory there may be a true value for some item, in practice we can never know this value. We can only know the value obtained by some measurement system, and every measurement system will contain some uncertainty.

NB10 was designated as a ten-gram standard. With an average weight of 9.999598 grams, we have a discrepancy of 402 micrograms. Since this National Bureau of Standards scale that measures such small weights with an uncertainty of 0.23 parts per million is presumably one of the master measurement methods, it is difficult to say if it is the scale or the standard that is biased here. In fact, with these data alone it is impossible to figure this out.

If we take 9.999598 grams as the accepted value for NB10, then we could use NB10 to evaluate another measurement system relative to the NBS scale. Repeated measurements of NB10 on a second scale, when placed on a consistency chart, would allow us to test for detectable bias between the measurement systems.

When a known standard is not available, we cannot discuss bias relative to a standard. However, with a designated standard we can compare two or more measurement systems for bias relative to each other. (I will illustrate this in January’s column.)

ADJUSTING THE MEASUREMENT PROCESS

The lab at a plant in Kentucky was very careful to keep their measurement processes calibrated. One piece of equipment that was routinely used to test production samples was checked by testing a known standard each day. Following this test the instrument was adjusted to compensate for the difference between the observed value and the accepted value for the standard. While I tried to explain the concept of a consistency chart to the lab director, he simply would not change the way he operated this piece of equipment.

A year or so later I was teaching a class at a competitor’s plant in South Carolina. When I walked into the lab, before I could even say hello, the lab director said: “You don’t know how much grief you have already saved us. Have you ever seen this machine?” He then pointed to the same machine that was in the lab in the Kentucky plant. “The maker of this machine says to recalibrate the machine each day. We used to do that. But now we use a consistency chart on the tests of the standard, and we only recalibrate whenever the chart shows evidence of a lack of consistency. *And when we started doing this, suddenly the whole plant started running better.*”

There is a time to recalibrate, and there is a time to refrain from recalibration. If you don’t

know the difference you are definitely going to mess things up by making needless adjustments and by also failing to make needed adjustments. And the inevitable outcome will always be increased variation.

An experiment I perform in class uses a bead board to illustrate this point. First I let Technician One mimic the Kentucky lab director. I collect one observation from the bead board. If that value differs from the target value of 10, I make an adjustment to the funnel of the bead board that is proportional to the difference between the observed value and the target value. (That is, if I get a value of 8, then I adjust the funnel up two units.) Following this adjustment I collect another 20 or so beads before taking a new sample and making a new adjustment. I continue until my bead board is full of beads.

Next I let Technician Two mimic the South Carolina lab director. I leave the funnel wherever Technician One left it, take some samples and only make an adjustment when the consistency chart indicates the need to do so. Between these periodic checks I run batches of approximately 20 beads, and continue until my bead board is full of beads. The histograms for 1000 values produced by Technicians One and Two are shown in Figure 3.

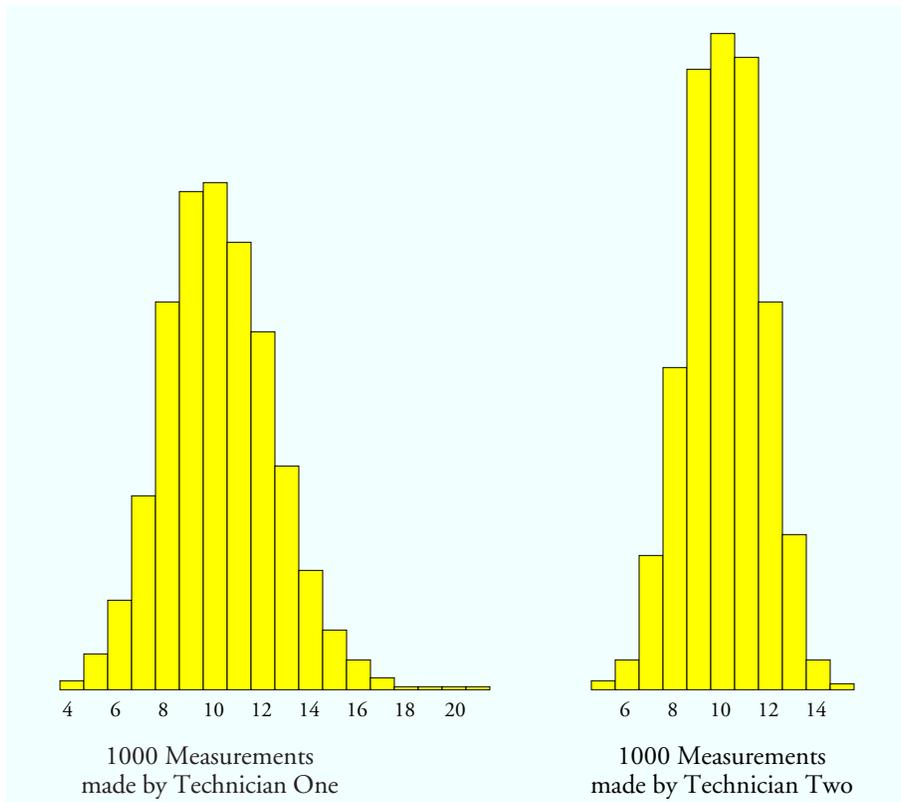


Figure 3: Repeated measurements of a standard by Technicians One and Two.

Technician One has an average of 10.181 and a variance statistic of 5.499. Technician Two has an average of 10.064 and a variance statistic of 2.688. The needless adjustments made by Technician One doubled the uncertainty in the measurements. I have done this experiment in class after class, week after week, something like a thousand times, and the result is always the

same. No matter how good Technician One does, Technician Two always does better. About the only time Technician Two makes an adjustment is at the start of his shift (when Technician One has left the funnel off-target).

Variation never subtracts. It always adds. The inherent variation of the measurement system cannot be reduced by making needless adjustments. It does not matter whether you make a lot of adjustments, or a few adjustments. Needless adjustments can only increase the variation. (Another example of this my July column "Process Monitor Charts" where only two out of 129 PID controllers did as well as a simple process behavior chart in spite of the fact that the PID controllers made over ten times as many adjustments as were called for by the process behavior chart.)

The trick is to only make adjustments when they are needed. To do that you need an operational definition of when to make an adjustment. And the consistency chart provides just such an operational definition. Everything else is just wishful thinking.

RUBBER RULERS

Some measurement systems are inherently variable over time. When this is the case the Consistency Chart for repeated measurements of the same thing will reveal this inherent variability.

Consider the case of a vision system created to measure the effective diameter of the steel insert for the rim of a steering wheel. The inserts were formed by bending a mild steel rod into a circle and welding the ends together. The vision system consisted of a back-lit plate with locating pins, a video camera mounted above the plate, and a computer that would count the pixels inside the image of the insert. Once the number of pixels was known the computer would convert the area into an effective diameter for the insert.

When shown this wonderful, new, fancy measurement system Richard Lyday decided to check it for consistency by repeatedly measuring the same piece of steel. Since positional variation was part of the measurement system, he would measure his part, take it off, reload it, and measure it again. After 30 such repeated measurements, carried out over the course of an hour, he got the chart in Figure 4. Since the insert could not have grown a quarter-inch in diameter, the trend on the X-Chart makes it clear that there are problems with this measurement system.

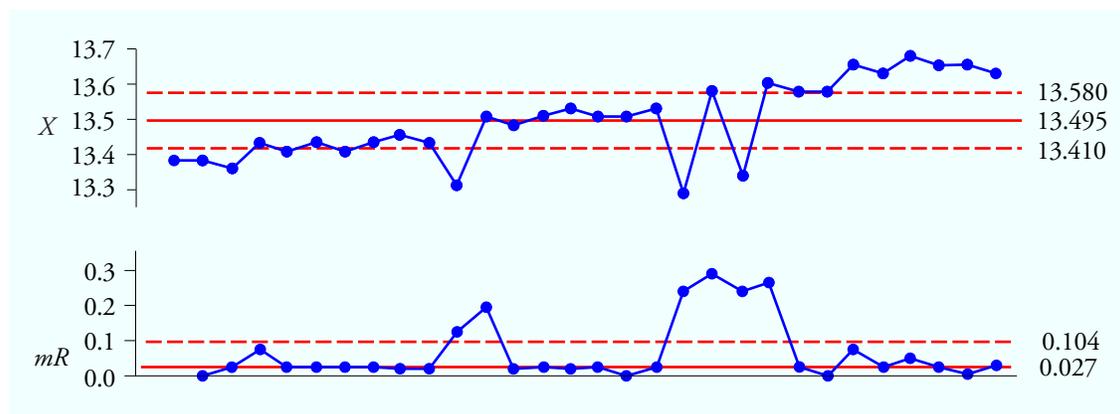


Figure 4: Consistency Chart for Vision System Measurements of One Part

The trend on the X-Chart was attributed to an inherent feature of the camera. The computer was obtaining progressively larger diameters for this one insert because the pixels in the image were shrinking as the camera warmed up. Since the computer did not actually determine the size of each pixel, it was “fooled” by this camera drift. The solution for this problem would be to reprogram the computer to measure both the insert and, at the same time, a fixed reference area on the back-lit plate. By adjusting the insert pixel count by the pixel count for the known reference area, the effect of pixel drift could be removed from this measurement system.

In addition to the trend, there was also the problem of the stalagmites and stalactites on the XmR Chart. These occurred when the vibration of the camera was sufficient to blur the image so that the computer lost count. With an 800-ton press in the building, and with the camera mounted on the roof truss, this vibration was going to be part of the operating environment for this measurement system.

Thus, when working with measurement systems that are rubber rulers, the traditional solution of “measure the sample then measure the standard” is appropriate. However, the use of this approach with a consistent measurement process is equivalent to Technician One. The only way to justify this “measure the standard then measure the sample” approach is to demonstrate the inconsistent nature of the measurement process by means of a Consistency Chart for repeated measures of the same thing.

SUMMARY

Checking the calibration of a measurement system using either a known standard or a designated standard is appropriate. The trick is in knowing what to do with the results of such an operation. Needlessly adjusting a measurement system in the name of recalibration will always degrade the quality of the measurements. Failing to make a needed adjustment will also degrade the quality of the measurements. So what is needed is an operational definition of when to adjust, and when to refrain from adjusting, a measurement system. The simplest such operational definition is a consistency chart. Use it or suffer the consequences.

Next month I will consider consistency issues for destructive tests.

