Analytical Metrology SPC Methods for ATE Implementation[a]

Howard Castrup, Ph.D.
Integrated Sciences Group

ABSTRACT

Although technology is growing in complexity and requires more careful monitoring and control than ever before, market dynamics are urging a constriction of calibration and testing budgets. To mediate between these opposing forces, decision makers need better and cheaper methods for controlling their measurement assurance processes. Various analytical approaches have been discussed for providing this capability. Among these is statistical process control (SPC). Unfortunately, "traditional" SPC tools (i.e., control charts), while useful in certain contexts, do not yield results which are especially useful for automated decision making.

In this paper, probability theory is employed to develop an analytical metrology SPC methodology which is amenable to implementation in automated testing environments. The methodology can be used to obtain in-tolerance probability estimates and bias estimates for both test systems and units under test without recourse to external measurement standards. This makes it particularly applicable in remote environments where measuring instruments are expected to function without calibration for extended periods of time.

1. Introduction

In 1687, Isaac Newton published his monumental Principia Mathematica in which he formulated three laws of motion. The third of these laws states that for every action there is an equal and opposite reaction. Since its publication, the third law has been verified experimentally to apply to all physical systems (at least all those accessible to classical mechanics).

Stretching our imaginations a little, we can see that a kind of third law applies to economic systems as well. Cost cutting in one area may result in increased costs in another, investments in one activity may yield benefits elsewhere, etc.

Over the past decade, we have been playing Russian Roulette with this economic third law in the management of periodic calibration and testing. Test and calibration programs are under persistent pressure to streamline operating budgets. A certain degree of streamlining can be accomplished by better workload planning, better personnel utilization, reduction of nonessential expenses, etc. Sooner or later, however, the axe is raised over such cost drivers as frequency of calibration.

Pushing for extended calibration intervals is becoming a favorite management pastime. Interestingly, the push to extend intervals seems to be peaking at a time when accuracy ratios between test and measuring equipment (TME) and units under test (UUTs) are shrinking. This means that measurement assurance is being diminished just when it's needed most. Beware the third law!

Economic reactions to cost cutting measures, such as extension of TME calibration intervals, may take the form of undesirable or even disastrous results. Saving money by reducing measurement assurance may cost a great deal in the final analysis. Such costs are manifested in increased warranty expenses, failed or compromised missions, loss of corporate reputation, loss of standing in international markets, etc.

Knowing this is not sufficient. Until a direct connection can be made between reduced measurement assurance and these cost factors, management will continue to push for extended TME calibration intervals. To make matters worse, there are increasing numbers of applications where calibration intervals are out of management's hands altogether. We

---

[a] Presented at the NSCL Workshop & Symposium, August 1991, Albuquerque, NM.

couldn't control them if we wanted. Examples include testing performed on board ships with extended deployment cycles and on board orbiting satellites. Although the inaccessibility of these systems to external calibration is sometimes alleviated by the application of on-site check standards, these standards are themselves subject to fluctuation and drift and require periodic calibration. Use of terrestrial and astronomical references helps for certain applications, but is often limited in scope and accuracy. Obviously, we have a problem.

2. Statistical Process Control to the Rescue

Clearly, what we want is to have our cake and eat it too. We want both effective measurement assurance *and* liberation from external periodic calibration. Recently, investigators have been experimenting with statistical methodologies to try and bring this about.[1,2] Prominent among these is *statistical process control* (SPC).

In traditional SPC applications, the results of UUT tests are used to monitor the testing process.[3-5] This is done by first computing *process control limits*. Process control limits consist of UUT tolerance limits expanded to include test process uncertainty contributions. These contributions are arrived at by multiplying test process uncertainties by statistical confidence multipliers. The multipliers are determined in accordance with the degree of confidence (e.g., 95%) desired in monitoring the test process. Tested (measured) UUT values are plotted against these control limits. The resulting plot is called a "control chart." The occurrence of an "out of control" value on a control chart signifies an out of control process. Since the control limits are determined statistically, identifying out of control processes using control charts is referred to as statistical process control.

From this brief synopsis, it may appear at first glance that traditional SPC is basically no different from ordinary tolerance testing, except that the usual tolerance limits are widened to reflect process uncertainties and the results of tests are plotted on charts. This view is correct to some extent. However, the key element of SPC, missing in ordinary UUT tolerance testing, is that test results are used to monitor the test process. This is the aspect of SPC that suggests possibilities for reducing dependence on external calibration without compromising measurement assurance. Presumably, if test process accuracy can be monitored statistically as described above, then dependence on external calibration of test systems can be reduced or even eliminated.

With its reliance on control chart monitoring, however, traditional SPC is difficult to fully implement in automated environments. This is due in part to the fact that the occurrence of a test result outside control limits is not in itself a particularly informative event. Such occurrences can be due to nonconforming UUTs, to test process instabilities or biases, or to some combination of these effects. Unraveling the causes of "out of control" test results typically requires human judgment and analysis. If SPC methods are to be automated, what is needed are more revealing and less ambiguous measures of process integrity than simple out of control occurrences.

Such measures are available through the application of new methods which will be collectively referred to in this paper as statistical measurement process control (SMPC). Historically, SMPC was invented[1] to answer a question posed by U.S. Navy technical representatives conducting a training audit. The question was this:

> "We have three instruments with identical tolerances of ±10 units. One instrument measures an unknown quantity as 0 units, the second measures +6 units and the third measures +15 units. According to the first instrument, the third one is out-of-tolerance. According to the third instrument, the first one is out-of-tolerance. Who's out-of-tolerance?"

Answering this question proved to be beyond the reach of established analytical methodologies available at the time. A new tool was needed. Development of this tool was nurtured by the following thought problem: In this problem, we desire to measure the pressure of an automobile tire. We have three inexpensive gages of equal accuracy which we apply to the task. We also have some knowledge of the tire's condition and some expectations of its performance. Such knowledge includes the tire's age, how long ago it was inflated and to what pressure (as measured by a reliable and precise gage at the corner tire shop), and, last but not least, the appearance of the tire itself, i.e., the extent of its lateral bulge. These tidbits lead us to believe that the tire's pressure lies somewhere in the neighborhood of 207 kPa[b] (30 psi).

Suppose that the first gage measures 193 kPa (28 psi), the second measures 228 kPa (33 psi) and the third measures 690 kPa (100 psi). Obviously, the third gage is defective. We don't need any elaborate and sophisticated analytical machinery to tell us this. Moreover, we don't need to calibrate each gage using an external reference to corroborate our conclusions. All this is very nice, but what if the third gage had indicated 276 kPa (40 psi)? We would probably still conclude that it was defective; but what if its reading was 248 kPa (36 psi)? We now begin to approach a set of measurement results that call for a more sophisticated decision tool than mere intuition.

In considering this problem, the feeling was that, if we could confidently pronounce a particular gage defective, given an extreme set of measurement results, we should, in principle be able to draw some conclusions from any set of measurement results. This belief led eventually to answering the Navy training audit question.

3. The SMPC Methodology

Answering the training audit question first required deciding what form the answer should take. After some initial ruminations, it was decided that the most meaningful quantities to extract from an analysis of the problem were estimates of in-tolerance probabilities for the three instruments.

Solving for in-tolerance probability estimates involves finding statistical *probability density functions* (pdfs) for the quantities of interest and calculating the chances that these quantities will lie within their tolerance limits. Specifically, if $f(x)$ represents the pdf for a variable x, and $+L$ and $-L$ represent its tolerance limits, then the probability that the variable is in-tolerance is obtained by integrating $f(x)$ over $[-L, L]$:

$$P = \int_{-L}^{L} f(x)dx \qquad (1)$$

Performing the integration is usually pretty easy. Most often, the trick is conjuring up the pdf from what we have to work with. Taking stock of the training audit question, we see that what we have so far are three measurements of an unknown quantity and three sets of tolerance limits, as shown in Figure 1. We need more information than this to construct pdfs, but this information can be acquired by doing a little supplemental homework (see Appendix A).

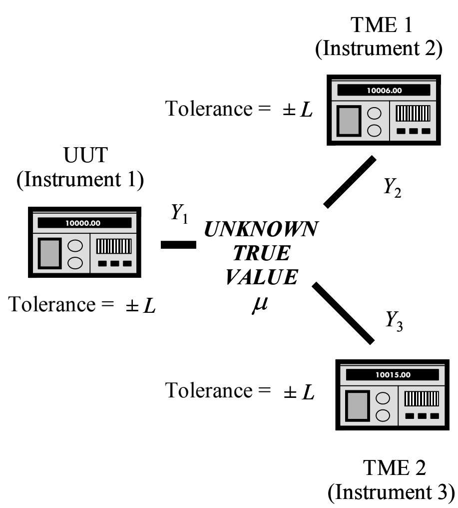

Figure 1. Training Audit Example. Three instruments measure an unknown true value. This true value may be external to all three instruments or generated by one or more of them. Instrument 1 is arbitrarily labeled the UUT. Instruments 2 and 3 are employed as TME.

[b] KiloPascals. One kPa = 1000 N/m^2, or about 0.145 psi.

For now, for discussion purposes, let instrument 1 serve as the UUT and label its indicated or "declared" value as Y_1 (see Figure 1). Likewise, let instruments 2 and 3 function as TME and label their declared values as Y_2 and Y_3, respectively.

We compute the quantities

$$X_1 = Y_1 - Y_2 ,$$

and

$$X_2 = Y_1 - Y_3 ,$$

which we will use to solve for the UUT (instrument 1) in-tolerance probability estimate.

3.1 Solving for the In-Tolerance Probability of Instrument 1

In probability theory, the notation $P(x\,|\,y)$ is used to denote the probability that an event x will occur, given that an event y has occurred. For example, x may represent the event that we gain weight and y may represent the event that we just ate six banana splits. In this case, $P(x\,|\,y)$ is the probability that we will gain weight given that we have just eaten six banana splits.[c]

$P(x\,|\,y)$ is a *conditional probability*. We can also form conditional pdfs. For instance, we can form the conditional pdf $f(\varepsilon\,|\,X_1, X_2)$. This is the pdf for an error ε being present given that we have measured the values X_1 and X_2 above. If we have $f(\varepsilon\,|\,X_1, X_2)$, we can estimate an in-tolerance probability for instrument 1 by using $f(\varepsilon\,|\,X_1, X_2)$ as the pdf in Eq. (1).

Readers inclined to tackle the necessary mathematics will find the method for determining $f(\varepsilon\,|\,X_1, X_2)$ described in Appendix B.

3.2 Solving for the In-Tolerance Probabilities of Instruments 2 and 3

In reviewing the Navy training audit question, it becomes apparent that there is nothing special about instrument 1 that should motivate our calling it the UUT. Likewise, there is nothing special about instruments 2 and 3 that should brand them as TME. Alternatively, we could have labeled instrument 2 the UUT and instruments 1 and 3 the TME, as in Figure 2. If we do this, we can calculate the in-tolerance probability for instrument 2 just as we have done for instrument 1. This involves computing the quantities

$$X_1' = Y_2 - Y_1 ,$$

and

$$X_2' = Y_2 - Y_3 ,$$

and forming the pdf $f(\varepsilon\,|\,X_1', X_2')$. Using this pdf in Eq. (1) yields the in-tolerance probability for instrument 2.

Similarly, if we compute

$$X_1'' = Y_3 - Y_1 ,$$

and

$$X_2'' = Y_3 - Y_2 ,$$

construct the pdf $f(\varepsilon\,|\,X_1'', X_2'')$, and use this pdf in Eq. (1), we get the in-tolerance probability for instrument 3.

3.3 Solving for Instrument Biases

The bias or "error" of an attribute can be found by solving for the attribute's *expectation value*. This expectation value is equal to the attribute's mean value. The mean value is obtained by multiplying the attribute's conditional pdf by the error ε and integrating over all values of ε. With this prescription, the biases of instruments 1, 2 and 3 are given by

---

[c] For me, $P(x\,|\,y) = 1$.

$$\text{Instrument 1 bias} = \int_{-\infty}^{\infty} f\left(\varepsilon \mid X_1, X_2\right)\varepsilon d\varepsilon ,$$

$$\text{Instrument 2 bias} = \int_{-\infty}^{\infty} f\left(\varepsilon \mid X_1', X_2'\right)\varepsilon d\varepsilon ,$$

and

$$\text{Instrument 3 bias} = \int_{-\infty}^{\infty} f\left(\varepsilon \mid X_1'', X_2''\right)\varepsilon d\varepsilon .$$

The details of the calculations are given in Appendix C.

As will be discussed in Section 4.2, these bias estimates can be employed as measuring attribute correction factors.

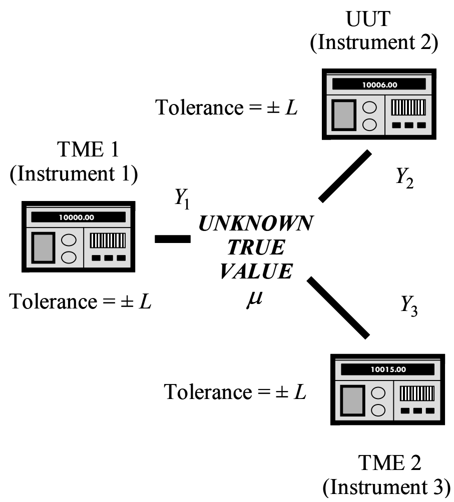

Figure 2. Exchanging UUT and TME Roles. Instruments 1 and 2 of the training audit example exchange roles as UUT and TME 1, respectively. This role swapping is done to estimate instrument 2 in-tolerance probability.

4. Application to ATE

To summarize the foregoing, we reversed the usual testing scenario by using a workload item to evaluate the in-tolerance probabilities of its TME. In other words, we regarded the UUT as the TME and the TME as the workload items. This about face contains the seeds of adapting SMPC methodology to ATE environments.

In adapting analytical methods to ATE environments, the assumptions which usually apply are the following:

1) ATE are periodically calibrated off-line, either directly or indirectly.

2) ATE attributes may or may not be supported by on-site or embedded check standards.

3) Software corrections, based on embedded standard values and/or on computed drift, may be applied to ATE attribute values during use.

So as to not introduce too many conceptual distractions initially, we will hold off on discussing software corrections until Section 4.2 and will defer consideration of embedded standards until Section 4.3.

4.1 Evaluating ATE Attribute In-Tolerance Probabilities

Consider an ATE attribute that tests n independent workload items over a span of time which is short relative to the ATE off-line calibration interval.[d] The result of each test is a pair of declared values; the ATE attribute's declared value and a workload item attribute's declared value. Either the pairs of values or their differences are maintained in ATE internal memory.

In the customary view of ATE testing, the ATE is the TME and the workload items are the UUTs. From the SMPC perspective, the ATE is the UUT and the workload items are the TME. We label the ATE attribute's declared value as Y_0 and the workload item attribute declared values as Y_i, $i = 1,2,\cdots,n$.

In Figures 1 and 2, UUT and TME comparisons are based on the measurement of a single underlying true value μ. In typical ATE usage, however, workload item testing may occur at different times and may involve different true values. This is not a problem in applying SMPC methodology to the evaluation of ATE attribute in-tolerance probabilities, since the quantities of interest are the *differences* in declared values $X_0 \equiv Y_0 - Y_i$, $i = 1,2,\cdots,n$.. In taking these differences, the precise true values pertaining at each test are subtracted out.

Thus, using the methodology of Section 3, the set of comparisons compiled from ATE workload tests can be used to compute an in-tolerance probability estimate for the testing ATE attribute. From the in-tolerance probability estimate, a decision can be made whether to submit the attribute for calibration or to implement some other corrective action, such as replacement with a standby unit. Note that no use of an external standard is involved in making these decisions.

SMPC evaluation of ATE attribute in-tolerance probabilities is a function that can be performed throughout the ATE's usage period. It can be incorporated as a user directed activity or can be implemented periodically in accordance with an automated procedure.

4.2 Computing ATE Attribute Correction Factors

In Section 3.3, it was shown that using SMPC can provide estimates of the biases of instrument attributes. In ATE applications, these estimates can be employed as software correction factors.

Suppose, for example, that instrument 1 of the Navy training audit problem is an ATE (i.e., is the UUT), and instruments 2 and 3 are workload items (i.e., the TME). Then, following "testing" of the ATE attribute by instruments 2 and 3, the attribute could be assigned a correction factor of β, where β would be calculated as in Section 3.3. The attribute could be compensated or corrected for "in software" by automatically subtracting the value β from subsequent ATE measurements.

It should be remarked that test system designers may be tempted to confine ATE applications of SMPC solely to computation of attribute biases for use as software correction factors. This restriction is not advisable. Evaluation of in-tolerance probabilities is also recommended. This point is demonstrated by example in Appendix C.

4.3 Accommodation of Check Standards

If on-site or embedded check standards are used to spot calibrate ATE attributes during deployment, in-tolerance probability estimates and bias estimates can be improved considerably. In applying SMPC with an ATE check standard, the check standard merely takes on the role of an additional workload item, albeit a comparatively accurate one.

With SMPC, not only can the in-tolerance probabilities and biases of the attributes of ATE assets and corresponding workload items be more accurately estimated, but in-tolerance probability and bias estimates can also be determined for resident check standards. Since check standards are subject to drift and fluctuation, using ATE assets and associated workloads to sanity check them in this way helps ensure that consensus with the "external world" is maintained.

---

[d] This allows us to regard the ATE attribute as fairly stable over the span of time covering the *n* tests.

5. Conclusion

Obviously, SMPC implementation in automated measuring systems, such as ATE, can enhance the measurement assurance of measuring and testing attributes in the field. Moreover, when one realizes that ATE testing of workload items in the field is analogous to product testing in a production environment, it can be seen that the use of SMPC can enhance measurement assurance in product testing applications as well.

SMPC implementation can also have a cost saving side. Although implementation of SMPC will incur some initial expense, significant cost savings may be realized by reducing dependence on external calibration, and, as a consequence, by extending calibration intervals. If recalibration expenses are viewed in the long term, this reduced dependence may more than justify the SMPC implementation investment. The third law still applies. The reaction to the expense will indeed be opposite, but it will be opposite in our favor and may be decidedly more than equal.

Acknowledgements

The author is indebted to Ms. Suzanne Griston and Mr. Jerry Hayes for providing critical reviews of draft material and for making key suggestions which improved both the paper's content and its readability. Acknowledgement is also due Mr. Del Caldwell and the U.S. Navy Metrology Engineering Center for motivating the creation of SMPC (by posing the Navy training audit question) and for sponsoring the initial research.

Appendix A

Methodology Development

A1. Introduction

The SMPC method discussed in this paper derives in-tolerance probabilities and attribute biases for both a unit under test (UUT) and a set of independent test and measuring equipment (TME) items, without recourse to external standards. The derivation of these quantities is based on measurements of the UUT attribute value made by the TME set and on certain information regarding UUT and TME uncertainties. The method accommodates arbitrary accuracy ratios between TME and UUT attributes and applies to TME sets comprised of any number of instruments.

To keep the discussion as uncomplicated as possible, the treatment in this appendix focuses on restricted cases in which both TME and UUT attribute values are normally distributed and are maintained within two-sided symmetric tolerance limits. Despite these restrictions, the methodological framework is entirely general. Extension to cases involving one-sided and asymmetric attribute tolerances merely requires additional mathematical brute force.

We also simplify the mathematics by restricting our present discussion to scenarios in which each TME makes only a single measurement of the UUT attribute. Accommodation of multiple measurements is discussed in Appendix D.

A2. Computation of In-Tolerance Probabilities

A2.1 UUT In-tolerance Probability

Whether a UUT provides a stimulus, indicates a value or exhibits an inherent property,[e] the declared value of its output, indicated value or inherent property is said to reflect some underlying "true" value.[f] The UUT declared value is

---

[e] A frequency reference is an example of a stimulus, a frequency meter reading is an example of an indicated value and a gage block dimension is an example of an inherent property.

[f] Suppose for example that the UUT is a voltmeter measuring a (true) voltage of 10.01 mV. The UUT meter reading (10.00 mV or 9.99 mV, or some such) is the UUT's “declared” value. As another example, consider a 5 cm gage

assumed to deviate from this true value by an unknown amount. We let Y_0 represent the UUT attribute's declared value and define a random variable ε_0 as the deviation of Y_0 from the true value. To simplify the discussion, ε_0 is assumed a priori to be normally distributed with zero mean and variance σ_0^2. The tolerance limits for ε_0 are labeled $\pm L_0$, i.e., the UUT is considered in-tolerance if $-L_0 \le \varepsilon_0 \le L_0$.

A set of n independent measurements are taken of the true value using n TME. Let Y_i be the measurement of the true value obtained using the *ith* TME. The measured discrepancies between UUT and TME declared values are labeled according to

$$X_i \equiv Y_0 - Y_i \,,\; i = 1,2,\cdots,n \,. \tag{A1}$$

The quantities X_i are assumed to be normally distributed random variables with variance σ_i^2 and mean ε_0.[g] Designating the tolerance limits of the *ith* TME attribute by $\pm L_i$, the *ith* TME is considered in-tolerance if $\varepsilon_0 - L_i \le \varepsilon_0 + L_i$.

We define dynamic accuracy ratios r_i according to

$$r_i = \frac{\sigma_0}{\sigma_i} \,,\; i = 1,2,\cdots,n \,. \tag{A2}$$

The adjective "dynamic" is used to distinguish these accuracy ratios from their usual static or "nominal" counterparts, defined by L_0 / L_i, $i = 1,2,\cdots,n$. The use of the word "dynamic" underscores the fact that each r_i defined by Eq. (A2) is a quantity which changes as a function of time elapsed since the last calibrations of the UUT and of the *ith* TME. This dynamic character is due to the fact that, in general, both UUT and TME standard deviations (bias uncertainties) grow with time since calibration. Computation of σ_0 and σ_i is described in Section A3 and in Appendix D.

Let P_0 be the probability that the UUT is in-tolerance at some given time since calibration. Using the foregoing definitions, we can write

$$P_0 = F(a_+) + F(a_-) - 1 \,. \tag{A3}$$

where $F(\cdot)$ is the distribution function for the normal distribution defined by

$$F(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\zeta^2/2} d\zeta \,, \tag{A4}$$

and where

$$a_{\pm} = \frac{\sqrt{1 + \sum r_i^2}\left(L_0 \pm \dfrac{\sum X_i r_i^2}{1 + \sum r_i^2} \right)}{\sigma_0} \,. \tag{A5}$$

In these expressions and in others to follow, all summations are taken over $i = 1,2,\cdots n$. The derivation of Eqs. (A3) and (A5) is presented in Appendix B.

A2.2 TME In-Tolerance Probability

Just as the random variables $X_1, X_2, \cdots, X_n$ are TME measured deviations from the UUT declared value, they are also UUT measured deviations from TME declared values. With this in mind, it is easy to see that, by reversing its role, the UUT can act as a TME, i.e., any of the n TME can be regarded as the UUT, with the original UUT performing

---

block. The declared value is 5 cm. The true value of the gage block dimension may be 5.002 cm or 4.989 cm, or whatever.

[g] In other words, *populations* of TME measurements are not expected to be systematically biased. This is the usual assumption made when TME are either chosen randomly from populations of like instruments or when no foreknowledge of TME bias is available. Conversely, individual unknown TME biases *are* assumed to exist. Accounting for this bias is achieved by treating individual instrument bias as a random variable and estimating its variance. Estimating this variance is the subject of Section A3.

the service of a TME. For example, focus on the first (arbitrarily labeled) TME and swap its role with that of the UUT. This results in the following transformations:

$$\begin{aligned} X_1' &= Y_1 - Y_0 \\ X_2' &= Y_1 - Y_2 \\ X_3' &= Y_1 - Y_3 \\ &\vdots \\ X_n' &= Y_1 - Y_n \end{aligned}$$

where the primes indicate a redefined set of test results. In the above scheme, it is clear that no additional measurements need be made. Using the primed quantities, the in-tolerance probability for the first TME can be determined just as the in-tolerance probability for the UUT was determined previously. The process begins with calculating a new set of dynamic accuracy ratios. First, we set

$$\begin{aligned} \sigma_0' &= \sigma_1 \\ \sigma_1' &= \sigma_0 \\ \sigma_2' &= \sigma_2 \\ &\vdots \\ \sigma_n' &= \sigma_n \end{aligned}$$

Next, we compute the required set of ratios using Eq. (A2), i.e., we compute $r_i' = \sigma_0' / \sigma_i'$, $i = 1, 2, \cdots, n$. In addition, we set $L_0' = L_1$ and $L_1' = L_0$.

Designating P_1 as the probability that the first TME is in-tolerance and performing the foregoing transformations, we can write

$$P_i = F(a_+') + F(a_-') - 1$$

where

$$a_\pm' = \frac{\sqrt{1 + \sum r_i'^2}\left(L_0' \pm \dfrac{\sum X_i' r_i'^2}{1 + \sum r_i'^2} \right)}{\sigma_0'} .$$

Applying similar transformations yields in-tolerance probabilities for the remaining n-1 TME.

A3. Computation of Variances

A3.1 Variance in Instrument Bias

Computing the uncertainties in UUT and TME attribute biases involves establishing the relationship between attribute uncertainty growth and time since calibration. A number of models have been employed to describe this relationship.[6] To illustrate the computation of bias uncertainties, the simple negative exponential model will be used here. With the exponential model, if t represents the time since calibration, then the corresponding in-tolerance probability $R(t)$ is given by

$$R(t) = R(0)e^{-\lambda t} , \tag{A6}$$

where the parameter l is the out-of-tolerance rate associated with the instrument in question, and R(0) is the in-tolerance probability at time of calibration.[h]

[h] The parameters λ and $R(0)$ are usually obtained from analysis of a homogeneous population of instruments of the same model number or type.[6]

With the exponential model, for a given end of period in-tolerance target, R^*, the parameters λ and $R(0)$ determine the calibration interval for a population of instrument attributes according to

$$t = -\frac{1}{\lambda}\exp\left\{\frac{t}{T}\ln\left[\frac{R^*}{R(0)}\right]\right\}. \qquad \text{(A7)}$$

Rearranging Eq. (A7) and substituting in Eq. (A6) gives

$$R(t) = R(0)\exp\left\{\frac{t}{T}\ln\left[\frac{R^*}{R(0)}\right]\right\}. \qquad \text{(A8)}$$

For an instrument attribute whose values are to be maintained within tolerance limits [Picture], the in-tolerance probability can also be written, assuming a normal distribution, as

$$R(t) = \frac{1}{\sqrt{2\pi\sigma_b^2}}\int_{-L}^{L} e^{-\zeta^2/2\sigma_b^2}\,d\zeta\,, \qquad \text{(A9)}$$

where σ_b^2 is the expected variance of the attribute bias at time t. After some manipulation, Eqs. (A8) and (A9) yield the attribute bias standard deviation according to

$$\sigma_b = \frac{L}{F^{-1}\left\{\frac{1}{2}\left[1 + R(0)\exp\left(\frac{t}{T}\ln\left[\frac{R^*}{R(0)}\right]\right)\right]\right\}}, \qquad \text{(A10)}$$

where $F^{-1}(\cdot)$ is the inverse of the normal distribution function defined in Eq. (A4).

Substituting L_i, T_i, t_i, $R_i(0)$ and R_i^*, $i = 0, 1, \cdots, n$, in Eq.(A10) for L, T, t, $R(0)$ and R^* yields the desired instrument bias standard deviation.[i]

A3.2 Accounting for Bias Fluctuations

Each attribute bias standard deviation is a component of the uncertainty in the attribute's value. Bias uncertainty represents long-term growth in uncertainty concerning our knowledge of attribute values. Such uncertainty growth arises from random and/or systematic processes exerted over time. Another component of uncertainty stems from intermediate-term processes such as those associated with ancillary equipment variations, environmental cycles, diurnal electrical power level cycles, etc.

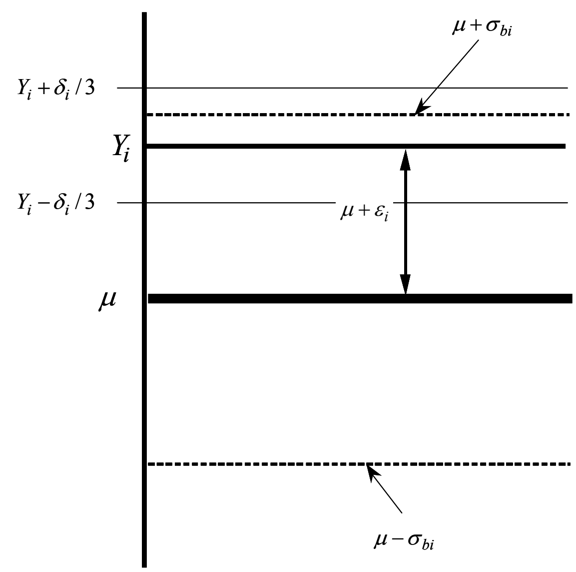

Figure 3. Measurement Uncertainty Components. The standard deviation [Picture] provides an indication of the uncertainty in the bias of the *ith* instrument's attribute. The variable δ_i is a measure of intermediate-term random fluctuations in this bias.

Uncertainty contributions due to intermediate-term random variations in attribute values usually need to be estimated heuristically on the grounds of engineering expectations.[j] Youden, for example, provides a graphical method for qualitatively evaluating contributions from human factors, laboratory processes and reference standards.[8] Development of a quantitative method is a subject of current research. For now, heuristic estimates are the best we have to work with. Heuristic estimates should represent upper bound (i.e., 3-sigma) limits for process uncertainty magnitudes. Experienced metrologists can often provide reasonable guesses for these limits. Thus, if we denote upper bounds for heuristically estimated contributions by δ_i, $i = 1, 2, \cdots, n$, the standard deviation for the attribute of the *ith* TME is given by

[i] The variable t_i is the time passed since calibration of the UUT ($i = 0$) or of the ith TME ($i = 1,2, \cdots, n$).

[j] In the parlance of the ISO/TAG4/WG3,[7] such estimates are referred to as Type B estimates.

$$\sigma_i = \sqrt{\sigma_{bi}^2 + \delta_i^2 / 9} \tag{A11}$$

The relationship of uncertainty variables to one another is shown in Figure 3.

A4. Example

The training audit problem described in the text of this paper provides an illustrative example of the use of SMPC. In this example, we set $R(0) = 1$ and process uncertainty equal to zero, for simplicity. Designating instrument 1 as the UUT, instrument 2 as TME 1 and instrument 3 as TME 2, we have

$$\begin{aligned} L_1 &= L_2 = L_3 = 10 \\ X_1 &= Y_0 - Y_1 = -6 \\ X_2 &= Y_0 - Y_2 = -15. \end{aligned}$$

Unless otherwise indicated, we can assume that the in-tolerance probabilities for all three instruments are approximately equal to their average over period values. Since the three instruments are managed to the same R^* target, have the same tolerances, and are calibrated in the same way using the same equipment and procedures, their standard deviations at the time of the audit should be about equal. Thus, according to Eq. (A2), the dynamic accuracy ratios are

$$r_1 = r_2 = 1.$$

Then, using Eq. (A5), we get

$$\begin{aligned} a_\pm &= \frac{\sqrt{1+(1+1)}\left[10 \pm \dfrac{-6-15}{1+(1+1)}\right]}{\sigma_0} \\ &= \frac{\sqrt{3}(10 \mp 7)}{\sigma_0}. \end{aligned}$$

In calculating the standard deviation σ_0, we make use of the fact that the end of period in-tolerance probability target for Navy general purpose equipment is $R^* = 0.72$. As stated above, we assume that we can use average over period in-tolerance probabilities for $R(t)$. With the exponential model, if $R(0) = 1$, the average in-tolerance probability is roughly equal to the in-tolerance probability half way through the calibration interval. Thus setting $t = T / 2$ in Eq. (A10) yields

$$\begin{aligned} \sigma_0 &= \frac{10}{F^{-1}\left\{\dfrac{1}{2}\left[1 + \exp\left(\dfrac{1}{2}\ln 0.72\right)\right]\right\}} \\ &= \frac{10}{F^{-1}(0.92)} \\ &= 10 / 1.43 \\ &= 6.97\,. \end{aligned}$$

Substituting in the expression for $a_\pm$ above gives

$$\begin{aligned} a_\pm &= \frac{\sqrt{3}(10 \mp 7)}{6.97} \\ &= 2.49 \pm 1.74\,. \end{aligned}$$

Thus, the in-tolerance probability for instrument 1 (the UUT) is

$$
\begin{aligned}
P_1 &= F(0.75) + F(4.23) - 1 \\
&= 0.77 + 1.00 - 1 \\
&= 0.77 .
\end{aligned}
$$

To compute the in-tolerance probability for instrument 2, we use the quantities

$$
\begin{aligned}
X_1' &= -X_1 \\
&= 6 \\
X_2' &= X_2 - X_1 \\
&= -9
\end{aligned}
$$

Substituting in Eq. (A5) and recalling that $\sigma_0' = \sigma_0$ in this example gives

$$
\begin{aligned}
a_\pm' &= \frac{\sqrt{1+(1+1)}\left[10 \pm \dfrac{6-9}{1+(1+1)}\right]}{\sigma_0'} \\
&= \frac{\sqrt{3}\left(10 \pm 1\right)}{6.97} \\
&= 2.49 \pm 0.25 .
\end{aligned}
$$

Thus, by Eq. (A3), the in-tolerance probability for instrument 2 (TI 1) is

$$
\begin{aligned}
P_1 &= F(2.24) + F(2.73) - 1 \\
&= 0.99 + 1.00 - 1 \\
&= 0.99 .
\end{aligned}
$$

To compute the in-tolerance probability for instrument 3, we use the quantities

$$
\begin{aligned}
X_1'' &= X_1 - X_2 \\
&= 15 \\
X_2'' &= -X_2 \\
&= 9 .
\end{aligned}
$$

Substituting in Eq. (A5) and recalling that $\sigma_0'' = \sigma_0$ gives

$$
a_\pm'' = 2.49 \pm 1.99 .
$$

Thus, by Eq. (A3), the in-tolerance probability for instrument 3 (TI 2) is

$$
\begin{aligned}
P_2 &= F(4.48) + F(0.50) - 1 \\
&= 1.00 + 0.69 - 1 \\
&= 0.69 .
\end{aligned}
$$

Summarizing the results, we estimate a roughly 77% in-tolerance probability for instrument 1, a 99% in-tolerance probability for instrument 2, and a 69% in-tolerance probability for instrument 3. In the Navy program, instruments are candidates for recalibration if their in-tolerance probabilities fall below 72%. Thus the decision to result from this exercise would be to send instrument 3 in for calibration.

Appendix B
Derivation of Eq. (A3)

Let the vector **X** represent the random variables X_1, X_2, $\cdots$, X_n obtained from n independent TME measurements of ε_0. We seek the conditional pdf for ε_0, given **X**, that will, when integrated over [-L_0, L_0], yield the conditional probability P_0 that the UUT is in-tolerance. From basic probability theory,

$$f\left(\varepsilon_0 \mid \mathbf{X}\right) = \frac{f\left(\mathbf{X} \mid \varepsilon_0\right) f\left(\varepsilon_0\right)}{f\left(\mathbf{X}\right)} \,, \tag{B1}$$

where

$$f\left(\varepsilon_0\right) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\varepsilon_0^2/2\sigma_0^2} \,. \tag{B2}$$

Since the components of **X** are s-independent, we can write

$$f\left(\mathbf{X} \mid \varepsilon_0\right) = f\left(X_1 \mid \varepsilon_0\right) f\left(X_2 \mid \varepsilon_0\right) \cdots f\left(X_n \mid \varepsilon_0\right), \tag{B3}$$

where

$$f\left(X_i \mid \varepsilon_0\right) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\left(X_i - \varepsilon_0\right)^2/2\sigma_i^2}, \quad i = 1, 2, \cdots, n \,. \tag{B4}$$

Combining Eqs. (B1) through (B4) gives

$$\begin{aligned} f\left(\mathbf{X} \mid \varepsilon_0\right) f\left(\varepsilon_0\right) &= C \exp\left\{-\frac{1}{2}\left[\frac{\varepsilon_0^2}{\sigma_0^2} + \sum_{i=1}^{n} \frac{\left(X_i - \varepsilon_0\right)^2}{\sigma_i^2}\right]\right\} \\ &= C \exp\left\{-\frac{1}{2\sigma_0^2}\left[\varepsilon_0^2 + \sum_{i=1}^{n} r_i^2 \left(X_i - \varepsilon_0\right)^2\right]\right\} \\ &= C e^{-G(\mathbf{X})} \exp\left\{-\frac{1}{2\sigma_0^2}\left[\left(1 + \sum r_i^2\right)\left(\varepsilon_0 - \frac{\sum X_i r_i^2}{1 + \sum r_i^2}\right)^2\right]\right\}, \end{aligned} \tag{B5}$$

where C is a normalization constant. The function G(**X**) contains no ε_0 dependence and its explicit form is not of interest in this discussion.

The pdf of **X** is obtained by integrating Eq. (B5) over all values of ε_0. To simplify the notation, we define

$$\alpha = \sqrt{1 + \sum r_i^2} \tag{B6}$$

and

$$\beta = \frac{\sum X_i r_i^2}{1 + \sum r_i^2} \tag{B7}$$

Using Eqs. (B6) and (B7) in Eq. (B5) and integrating over ε_0 gives

$$\begin{aligned} f\left(\mathbf{X}\right) &= C e^{-G(\mathbf{X})} \int_{-\infty}^{\infty} e^{-\alpha^2\left(\varepsilon_0 - \beta\right)^2/2\sigma_0^2} d\varepsilon_0 \\ &= C e^{-G(\mathbf{X})} \frac{\sqrt{2\pi}\sigma_0}{\alpha} \,. \end{aligned} \tag{B8}$$

Dividing Eq. (B8) into Eq. (B5) and substituting in Eq. (B1) gives

$$f\left(\varepsilon_0 \,|\, \mathbf{X}\right) = \frac{1}{\sqrt{2\pi}\left(\sigma_0/\alpha\right)} e^{-\left(\varepsilon_0-\beta\right)^2/2\left(\sigma_0/\alpha\right)^2} \tag{B9}$$

The in-tolerance probability for the UUT is obtained by integrating Eq. (B9) over [-L_0, L_0]. With the aid of Eqs. (A5), (B6) and (B7), this results in

$$\begin{aligned} P_0 &= \frac{1}{\sqrt{2\pi}\left(\sigma_0/\alpha\right)} \int_{-L_0}^{L_0} e^{-\left(\varepsilon_0-\beta\right)^2/2\left(\sigma_0/\alpha\right)^2} d\varepsilon_0 \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\left(L_0+\beta\right)/\left(\sigma_0/\alpha\right)}^{\left(L_0-\beta\right)/\left(\sigma_0/\alpha\right)} e^{-\zeta^2/2} d\zeta \\ &= F\left(a_-\right) - F\left(a_+\right) \\ &= F\left(a_+\right) + F\left(a_-\right) - 1\,, \end{aligned} \tag{B10}$$

which is Eq. (A3).

Appendix C
Estimation of Biases

Obtaining the conditional pdf $f(\varepsilon_0|\mathbf{X})$ permits the computation of moments of the UUT attribute distribution. Of particular interest is the first moment, or distribution mean. The UUT distribution mean is the conditional expectation value for the bias ε_0. Thus the UUT attribute bias is estimated by

$$\begin{aligned} \beta_0 &= E\left(\varepsilon_0 \,|\, \mathbf{X}\right) \\ &= \int_{-\infty}^{\infty} \varepsilon_0 f\left(\varepsilon_0 \,|\, \mathbf{X}\right) d\varepsilon_0\,. \end{aligned} \tag{C1}$$

Substituting from Eq. (B9), and using Eq. (B7) gives

$$\beta_0 = \frac{\sum X_i r_i^2}{1 + \sum r_i^2} \tag{C2}$$

Similarly, bias estimates can be obtained for the TME set by making the transformations described in Section A2.2. Thus, for example, the bias of TME 1 is given by

$$\beta_1 = E\left(\varepsilon_1 \,|\, \mathbf{X}'\right) = \frac{\sum X_i' r_i'^2}{1 + \sum r_i'^2} \tag{C3}$$

To exemplify bias estimation, we again turn to the Navy training audit problem. Using Eqs. (C2) and (C3), and recalling that $\sigma_0 = \sigma_1 = \sigma_2$, we get

Instrument 1 (UUT) bias: $\beta_0 = \dfrac{-6-15}{1+\left(1+1\right)} = -7$

Instrument 2 (TME 1) bias: $\beta_1 = \dfrac{6-9}{1+\left(1+1\right)} = -1$

Instrument 3 (TME 2) bias: $\beta_2 = \dfrac{15+9}{1+\left(1+1\right)} = 8\,.$

If desired, these bias estimates could serve as correction factors for the three instruments. If used in this way, the quantity 7 would be added to all measurements made with instrument 1, the quantity 1 would be added to all measurements made with instrument 2 and the quantity 8 would be subtracted from all measurements made with instrument 3.

Note that all biases are within the tolerance limits (±10) of the instruments. This might encourage respective users to continue to employ their instruments with confidence, i.e., to forego submitting them for recalibration. Recall, however, that the in-tolerance probabilities computed in Section A4 indicated only a 77% chance that instrument 1 was in-tolerance and an even lower 69% chance that instrument 3 was in-tolerance. Such results tend to provide valuable information from which to form a cogent baseline for making judgments regarding instrument maintenance. Such a baseline does not tend to emerge from considerations of attribute bias alone.

Appendix D

Treatment of Multiple Measurements

In previous discussions, the quantities $\mathbf{X}_i$ are based on single measurements of the difference between the UUT attribute and the ith TME's attribute. In most ATE applications, however, testing of workload items is not limited to single measurements. Instead, multiple measurements are usually taken. Thus, instead of *n* individual measurements, we will ordinarily be dealing with *n* sets or samples of measurements.

Let n_i be the number of measurements taken using the *ith* TME's attribute, and let $X_{ij} \equiv Y_0 - Y_{ij}$ be the *jth* of these measurements. We compute the quantities

$$X_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \tag{D1}$$

and

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(X_{ij} - X_i \right)^2 . \tag{D2}$$

The variance associated with the mean of measurements made using the *ith* TME's attribute is given by

$$\sigma_i^2 = \sigma_{bi}^2 + s_i^2 / n_i + \delta_i^2 / 9 ,$$

where the variables σ_{bi} and δ_i are as defined in Section A3.2. We use the square root of this variance to determine the quantities

$$r_i = \frac{\sigma_0}{\sigma_i}, i = 1, 2, \cdots, n \tag{D3}$$

used in Eqs. (B6), (B7) and (B9).

Note that including sample variances is restricted to the estimation of TME attribute variances. UUT attribute variance estimates contain only the terms indicated in Eq. (A11). This underscores what we hope to achieve in constructing the pdf $f(\varepsilon_0|\mathbf{X})$. What we are after are estimates of the in-tolerance probability and bias of the UUT attribute. In this, we are interested in the attribute as an entity distinct from any process uncertainties induced by its measurement.

It is important to keep these considerations in mind when the UUT and the ith TME switch roles. What we are after in that event is information on the attribute of the *ith* TME as a distinct entity. Accordingly, the relevant transformations are

$$\sigma_0' = \sqrt{\sigma_{bi\,i}^2 + \delta_i^2 / 9}$$

$$\sigma_1' = \sqrt{\sigma_{b1}^2 + s_1^2 / n_1 + \delta_1^2 / 9}$$

$$\vdots$$

$$\sigma_i' = \sqrt{\sigma_{b0}^2 + s_i^2 / n_i + \delta_i^2 / 9}$$

$$\vdots$$

$$\sigma_n' = \sqrt{\sigma_{bn}^2 + s_n^2 / n_n + \delta_n^2 / 9} \, , \qquad \text{(D4)}$$

and

$$\begin{aligned} X_1' &= X_1 - X_i \\ X_2' &= X_2 - X_i \\ &\vdots \\ X_i' &= -X_i \\ &\vdots \\ X_n' &= X_n - X_i \end{aligned} \qquad \text{(D5)}$$

All other expressions are the same as are used in treating single measurement cases.

References

1 Castrup, H.T., Intercomparison of Standards: General Case, SAI Comsystems Technical Report, U.S. Navy Contract N00123-83-D-0015, Delivery Order 4M03, March 16, 1984.

2 Jackson, D.H., Instrument Intercomparison: A General Methodology, Analytical Metrology Note AMN 86-1, U.S. Navy Metrology Engineering Center, NWS Seal Beach, January 1, 1986; and "Instrument Intercomparison and Calibration," Proc. 1987 MSC, Irvine, January 29, 30.

3 ASTM Manual on Presentation of Data and Control Chart Analysis, ASTM Special Technical Publication 15D, American Society for Testing and Materials, Philadelphia, October 1976.

4 Juran, J.M., ed. Quality Control Handbook, 3rd edition, McGraw-Hill, New York, 1974.

5 Belanger, B. Measurement Assurance Programs, Part I: General Introduction, NBS Special Publication 676-I, May 1984; and Measurement Assurance Programs Part II: Development and Implementation, NBS Special Publication 676-II, April 1984.

6 NCSL, Establishment and Adjustment of Calibration Intervals, Recommended Practice RP-1, November 15, 1989.

7 Guide to the Expression of Uncertainty in Measurement, 3rd Draft, ISO/TAG4/WG3, 16 February 1990.

8 Youden, W.J., "Uncertainties in Calibrations," IRE Trans. Instr. I-11, Nos. 3&4, pp 133-138, December 1962.