Using Reliability to Meet Z540.3's 2% Rule

Scott M. Mimbs
NASA Metrology and Calibration Program
John F. Kennedy Space Center
Phone: 321-861-5184
scott.m.mimbs@nasa.gov

Abstract

NASA's Kennedy Space Center (KSC) undertook implementation of ANSI/NCSL Z540.3-2006 in October 2008. Early in the implementation, KSC identified that the largest cost driver of Z540.3 implementation is measurement uncertainty analyses for legacy calibration processes. NASA, like other organizations, has a significant inventory of measuring and test equipment (MTE) that have documented calibration procedures without documented measurement uncertainties.

This paper provides background information to support the rationale for using high in-tolerance reliability as evidence of compliance to the 2% PFA quality metric of ANSI/NCSL Z540.3-2006 allowing use of qualifying legacy processes. NASA is adopting this as policy and is recommending NCSL International consider this as a method of compliance to Z540.3.

Topics covered include compliance issues, using EOPR to estimate test point uncertainty, reliability data influences within the PFA model, the validity of EOPR data, and an appendix covering "observed" versus "true" EOPR.

1. Introduction

NASA's Kennedy Space Center placed ANSI/NCSL Z540.3-2006 [1] on the Institutional Services Contract (ISC) that went into effect October 2008. In October 2009, KSC's ISC Standards & Calibration Laboratory achieved compliance to the new standard. A key component to KSC's compliance is using end-of-period reliability (EOPR) as evidence of conformance to Z540.3's probability of false acceptance (PFA) requirement for legacy calibration processes that have high instrument in-tolerance reliability.

The Z540.3 quality metric for conformance-test calibrations is a decision rule that states the "*...probability that incorrect acceptance decisions will result from calibration...shall not exceed 2%...*" This metric is known as the probability of false acceptance (PFA), false accept risk (FAR), and in older literature, consumer risk (CR). A detailed engineering review of the PFA model reveals the existence of discrete input values that dominate the model for a specified target value, such as 2% PFA. The PFA model (discussed in detail in section 3) utilizes measurement-process uncertainty and in-tolerance reliability as input variables, in conjunction with the specified tolerance of interest. The engineering review shows there is a threshold value for each of the two input variables that, when exceeded, ensures the target PFA is met, regardless of the value of the second variable. For example, when the ratio of the specified tolerance to measurement process uncertainty is 4.6:1 or greater, the PFA is constrained to 2% or less, independent of changes in the in-tolerance reliability. Likewise, when the in-tolerance reliability, also known as end-of-period-reliability (EOPR), is observed to be 89% or greater, the

PFA is constrained to 2% or less, independent of the measurement-process uncertainty. Thus, compliance to the Z540.3 2% PFA requirement is achievable with either measurement process uncertainty or observed EOPR alone, as long as that single variable is in its region of dominance.

The reason behind this particular behavior of the PFA model lies within the probability theory for false acceptance and the mathematical models used to calculate the risks. NASA, in conjunction with U.S. Navy and industry experts, performed the engineering review of the PFA model, looking into the factors that affect measurement uncertainty. The impetus for the review was to mitigate some of the costs associated with the initial implementation of Z540.3 for organizations having adequate legacy calibration procedures. In general, when a process has high reliability, most error sources are under control. This led to one "focal question."

> *What additional value, or useful information, will uncertainty analyses add to legacy calibration processes that have high EOPR?*

Based on the joint engineering review, NASA concluded that performing uncertainty analyses on legacy calibration processes with qualifying observed EOPR would not be required for meeting the PFA requirement and therefore the measurement uncertainty associated with the calibration processes would be adequate for that purpose. NASA's new policy states that observed EOPR at, or above, 89% are considered acceptable evidence of compliance to Z540.3's PFA requirement (sub-clause 5.3b) and measurement uncertainty requirements (sub-clause 5.3.3).

NASA is adopting this as policy and is recommending NCSL International consider this as a method of compliance for <u>transitioning</u> to Z540.3. It is essential to note that this recommendation and paper applies only to <u>documented legacy calibration procedures</u> with associated observed EOPR data. This method will eventually become obsolete due to the replacement of legacy equipment or change in calibration processes.

It is important to note that the description of observed EOPR in the context of this paper is also applicable to all usage of in-tolerance reliability within the PFA model. This includes four of the six compliance methods outlined in NCSL International's *Handbook for the Application of ANSI/NCSL Z540.3-2006* [2] that use in-tolerance reliability data as an input.

This report is broken into five main sections and an Appendix.

1. Introduction
2. Compliance issues/concerns
3. The PFA Model
4. Ensuring EOPR data is valid
5. Summary and Conclusions
6. Appendix - The concept of "observed" versus "true" EOPR

These topics are not new and documentation is readily available. The uniqueness of this specific application is the use of in-tolerance reliability as evidence for acceptable PFA and adequate measurement uncertainty. A detailed literature search indicated other proposed uses of "true" EOPR [3], but none documented the specific application discussed in this paper.

2. Compliance issues/concerns for Z540.3

In conjunction with the technical review of the PFA model, NASA also looked at Z540.3 compliance from a Quality perspective, specifically to determine if NASA needed to provide documentation that, in effect, "tailored" Z540.3 requirements. Although the technical review indicated no problems, the Quality review considered two sub-clauses as potential "audit-traps." As a precaution, NASA updated its policy to cover sub-clauses 5.3 b) and 5.3.3. The first sub-clause sets the acceptance criteria for conformance-test calibrations and the second establishes the requirements for the use of measurement uncertainty within the calibration system.

Sub-clause 5.3 b)

Sub-clause 5.3 b) of Z540.3 establishes a decision rule as the quality metric for conformance-test calibrations, where the probability of "*…incorrect acceptance decisions…*" from calibration tests will be less than 2 percent. Since compliance to this requirement depends entirely upon the type of probability expression used, in 2007 NASA requested an interpretation from the ANSI Z540 writing committee. They provided a written interpretation stating that an unconditional probability is the basis of compliance to the Z540.3 PFA requirement. This means for a population of like instruments, evaluation of compliance to the PFA requirement is <u>prior</u> to a specific calibration event. Therefore, the PFA estimation is reflective only of the characteristics of the calibration process for a population of like instruments. This interpretation is crucial for using in-tolerance reliability as evidence of compliance because it establishes the required probability model, and in-turn, the value of the reliability metric.

The potential "audit-trap" for this sub-clause comes from the NCSLI Z540.3 Handbook [2] rather than the Standard. Although the Handbook is non-interpretive, the belief was that auditors would use the Handbook for guidance on acceptable methods of compliance to Z540.3 sub-clauses. While the Handbook addresses six methods for achieving compliance to this sub-clause, and recognizes other methods exist, it does not directly address using <u>reliability alone</u> as a compliance method. Therefore NASA added the "89% Rule" as an acceptable method for legacy equipment.

Sub-clause 5.3.3

Sub-clause 5.3.3 of Z540.3 provides the requirements for measurement uncertainty and has two parts. The first part requires that all calibration measurement results and processes "*shall meet the requirements of their application.*" The second part of sub-clause 5.3.3 states that measurement uncertainty estimates include all components of measurement uncertainty that could influence the measurement result. These two parts together mean that evidence of compliance to 5.3.3 needs to reflect adequacy of measurement uncertainty for specific calibration processes.

From the technical perspective, to be effective, a calibration measurement process must account for and control any potential sources of measurement error[1] that might adversely influence the calibration result. Measurement uncertainty analysis is the method to evaluate these potential error sources as components of the overall measurement uncertainty, thus providing insight into the quality of the measurement data. The traditional method of providing evidence of

[1] Measurement error is not a mistake or a failure to follow a process (i.e., production error).

compliance to uncertainty requirements is a documented measurement uncertainty analysis with an "uncertainty budget" and a quantitative value. In-tolerance reliability can be mathematically demonstrated to provide evidence of compliance to sub-clause 5.3 b), thus, compliance to the first requirement of sub-clause 5.3.3 for conformance-test calibrations.

The second requirement of sub-clause 5.3.3 focuses specifically on those components of uncertainty that have an influence on the measurement results. This means to use reliability as evidence of compliance, components of uncertainty need to be reflected in the reliability data and constrained within known bounds. Although the technical review indicated two cases where measurement uncertainty could have an influence on in-tolerance reliability, neither case would have an adverse influence on the reliability. A more detailed discussion covers this topic in section 3, *The PFA Model.*

It is counter-intuitive for high reliability to occur when the measurement process uncertainty constitutes a significant portion of the specified tolerance for a given instrument, yet it occurs frequently. The following are three scenarios that provide rationale for high EOPR to occur when the ratio of measurement process uncertainty to tolerance (i.e., test uncertainty ratio, TUR) is small:

1. The EOPR data is in error due to a mistake in the collection of reliability data or in the documentation of measurement procedure (e.g., misapplied unit-under-test tolerance).
2. The Reference Standard out-performs its assigned accuracy specifications. Generally, instruments specifications cover a broad range of conditions, including variations in conditions of use. Instruments used in a controlled environment, such as a calibration laboratory, will normally perform well within the allowed tolerance limits. Effectively, the instrument (Reference Standard) consistently operates within a fraction of its tolerance limits, and the measurement processes in which it is used will, in reality, have a higher TUR than was estimated using the Reference Standard's accuracy specifications.
3. The ratio between the resolution and specified accuracy of the unit-under-test (UUT) is very low (e.g., below 2:1). The instrument's resolution dominates the measurement uncertainty for these "resolution-limited" instruments, resulting in TUR values below 2:1. In cases where the inherent physical characteristics of the UUT are significantly better than its resolution, the instrument will have high in-tolerance reliability regardless of the low TUR. For example, caliper micrometers often have high in-tolerance reliability coupled with low resolution-to-accuracy ratios. This is possible because design tolerances for key mechanical components, such as the lead screw, are smaller than the instrument's resolution, often by an order of magnitude.

Additional Considerations

As mentioned earlier, NASA has recommended NCSLI incorporate using in-tolerance reliability as a Z540.3 compliance method for legacy calibration processes. Members of NCSLI's 171 and 174 committees raised several questions on this compliance method concerning probability of false reject (PFR), the bounding of measurement uncertainty, and the confidence in the reliability data. Although considered early in its review, NASA did not believe any of these areas would create an issue to achieving compliance, based on the following technical rationale.

Although sub-clause 5.3 requires measurement decision risk be addressed, a specific probability of false reject (PFR) value or limit is not a direct Z540.3 requirement; however, it can influence compliance to several sub-clauses. In general, PFA and PFR are interrelated in that changes in one affect the other. From a pragmatic viewpoint, the use of an in-tolerance reliability target, such as 89%, bounds PFR. For example, with an EOPR of 89%, there is an 11% rejection rate. The amount of the rejections that are incorrect (i.e., PFR) are based on the ratio of the

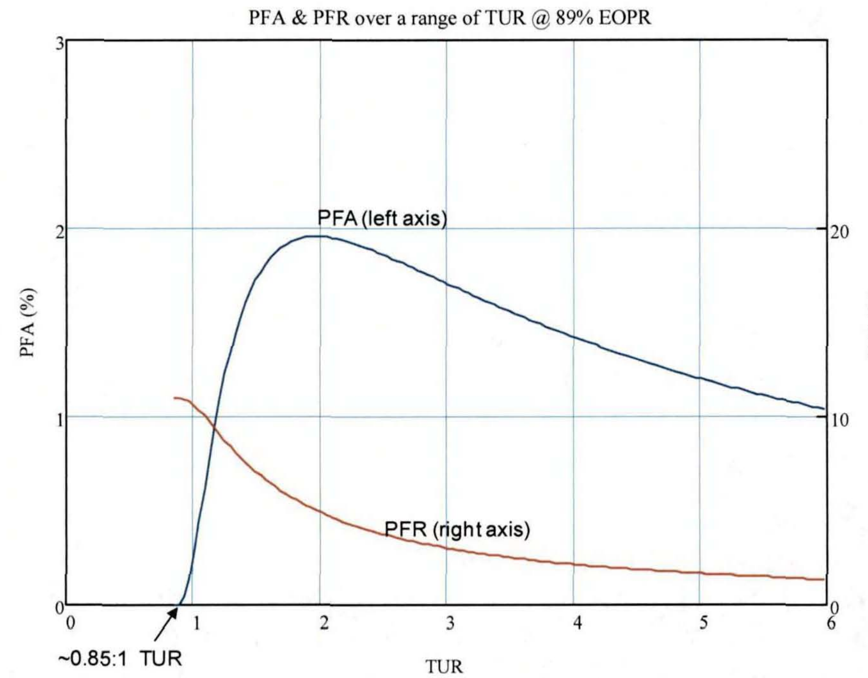


Figure 1: PFA and PFR graphed aver a range of TUR values for 89% EOPR.

measurement process uncertainty to the tolerance limits, known as the test uncertainty ratio (TUR). Figure 1 illustrates as the TUR decreases, the PFR increases to a point where nearly all rejections are false rejections. Therefore, when using 89% EOPR for a compliance method, PFR will not exceed 11% even in the worst-case scenario.

When observed in-tolerance reliability for a calibration process is high, it indicates that the measurement uncertainty is somehow constrained or bounded, even when not quantified. In general, if the calibration process is reliable, all the uncertainty sources are either insignificant or have been addressed through the design, implementation, and control of the calibration process. Assuming the data is valid, observing high EOPR means either:

1. An extremely good UUT, and an acceptable Reference Standard, providing a miniscule PFA (e.g., 0.01%), or
2. A relatively good UUT, and a good Reference Standard, providing an acceptable PFA (e.g. < 2.0%).

Without one of these combinations, observing high in-tolerance reliability is not possible.

In addition to the PFR, Figure 1 plots the PFA over a range of TUR values for 89% observed in-tolerance reliability. It illustrates that when observing 89% reliability, PFA decreases in the lower TUR regions, starting at approximately 2:1. Although counter-intuitive, this is indicative that when observing high reliability, measurement uncertainty is constrained, as discussed earlier.

The last topic raised by the NCSL committee members concerns the uncertainty in the measurement reliability estimate, typically expressed as a confidence interval about some mean value, and the resulting impact on parameters such as PFA, PFR, etc. It is essential to note that any discussions concerning the use of observed in-tolerance reliability data pertains to all compliance methods for meeting Z540.3's PFA metric and not just the proposed method using the 89% rule.

There has been a suggestion to consider using the lower confidence limits of the binomial probability in lieu of the mean estimate of the observed EOPR data for PFA estimation. The recommendation is one means of managing the effects of large confidence intervals caused by, for example, small sample sizes. Although the intent of this recommendation is to provide for better estimates of measurement reliability, the consequence of this action would be to necessitate increased sample populations for applications using EOPR data to estimate PFA. This is especially true of instruments with Test Uncertainty Ratios (TUR) of 2:1 or lower which could not meet Z540.3's 2% PFA metric without extremely large sample populations (in excess of 6,000 calibrations). Section 5 of this paper provides additional detail on the validity of in-tolerance reliability for PFA estimation.

3. The PFA Model

The probability of false acceptance (PFA) is also known as false accept risk (FAR), consumer risk, or Type 2 risk. To avoid confusion with other NASA risk initiatives, such as Probabilistic Risk Assessment (PRA), this paper will favor the term PFA over the more traditional FAR. PFA and FAR estimation is identical mathematically, thus they are interchangeable terms.

PFA traces its roots to the consumer risk of the 1940's and 1950's. The mathematical concepts of the probability theory and false acceptance calculations used in this paper are contained in NASA-HANDBOOK-8739.19-4, *Estimation and Evaluation of Measurement Decision Risk* [4], or the NCSL International *Handbook for the Application of ANSI/NCSL Z540.3-2006* [2]. This paper concentrates on concepts using charts and graphs and attempts to limit the use of direct mathematical expressions except where they add clarity to the discussion.

"All models are wrong, some are useful." This saying, credited to the famous statistician George Bell, recognizes that theoretical models simulate reality only when the underlying assumptions are satisfied. Although this rarely happens, models can be very useful with an understanding of how far and why the model deviates from reality. This holds true for the methodologies used to calculate PFA. Like all models, deviations to the model assumptions will affect a PFA estimate. The trick is to know how useful the PFA estimate may be considering deviations, known, and unknown. Some PFA model assumptions are:

1. Measurement processes are ideal
2. Tolerance specifications are ideal
3. Measurement process uncertainty estimate is ideal
4. Uncertainty distributions are Gaussian with a mean of zero
5. The standard deviation estimate of the measured test-point of the unit under test (UUT) is ideal
6. In-tolerance reliability used to estimate the test-point standard deviation is ideal

As used in Z540.3, PFA is a quality metric for calibration processes, thus providing a quantifiable measure of confidence that the calibrated equipment meets specified requirements, such as the manufacturer's tolerance. Although the model expects "perfect" processes, tolerances, and estimates, this is an unreasonable expectation. Therefore, an understanding of the model limitations helps PFA become a *reasonable* and useful calibration quality metric.

Elements of the PFA Model

In general, there are three variables used to estimate PFA: unit-under-test (UUT) tolerance, calibration process uncertainty, and test-point uncertainty. The first two are a part of the calibration process, thus are relatively fixed, while the in-tolerance reliability can vary based on factors outside of the calibration process, such as the interval between calibrations and equipment usage.

Although the mathematics behind PFA calculation uses integral calculus, this discussion will focus on the functional elements of the model, as described below.

$$PFA = f(Tol, u_{mp}, \sigma_{tp})$$

1. Tol - is the specified tolerance for the subject test-point of the unit under test (UUT). A conformance-test calibration verifies this tolerance. The specified tolerance may be either the manufacturer's tolerance or a user-defined performance-based tolerance.
2. u_{mp} - is the calibration measurement-process uncertainty. This is the combined standard uncertainty (one standard deviation) of the calibration measurement process for a measured test-point. The estimate should include all pertinent error sources from the measurement process, including the reference standard, and unit under test.
3. σ_{tp} - is the standard deviation of the *a priori* population distribution of the subject test-point. This is the bias[2] of the UUT measured test-point, expressed as the standard uncertainty. This can be calculated using "Type A" analysis (a statistically valid number of repeat measurements from a population) or it can be estimated using EOPR data. Ideally, this value is bounded by the specified test-point tolerance to some confidence level or coverage factor, *k*.

A brief discussion of each component follows in conjunction with a description of its relationship to the other components of PFA model.

Tolerance - *Tol*

The first element is the tolerance of the unit-under-test (UUT) for the subject test-point. The objective of this type of calibration is verification of conformance to the tolerance. Although, usually derived from the UUT manufacturer's specifications, it may also be a user-defined performance requirement. The tolerance influences the PFA model by its size relative to the other two variables. Poorly developed specifications, and/or improperly applied tolerances will influence the reliability data. This, in turn, influences the PFA results when using in-tolerance reliability data in the model.

[2] "A systematic discrepancy between an indicated or declared value of an attribute and its true value." [5]

Poorly specified tolerances affect the EOPR by affecting the calibration decisions. An overly conservative tolerance (larger than required) can reduce the number of false acceptance and rejection decisions, thus potentially increasing the EOPR. In contrast, an under-specified tolerance (smaller than required) can increase the number of false acceptance and rejection decisions, thus potentially decreasing the EOPR. In either case, the EOPR data can be considered valid, since the calibration provider is not in direct control of the specified tolerance.

If the specification is misunderstood or misapplied, the EOPR data may not be valid. This is not a reflection of the measurement uncertainty's influence on EOPR, but rather a mistake with validation of the calibration process. It can manifest in two ways:

1. The tolerance used in the calibration procedure is **smaller** than specified. A smaller than specified tolerance most likely will increase the probability of false rejection (PFR) for the process, thus decreasing EOPR. If this type of error exists in combination with high EOPR, it is indicative of a conservatively specified UUT. Adjusting the calibration process to the specified tolerance limits should increase the EOPR, if all other factors remain the same.
2. The tolerance used in the calibration procedure is **larger** than specified. A larger than specified tolerance most likely will increase the EOPR. If the larger tolerance is an error in calibration process or set-up, the EOPR is probably not valid as evidence of compliance to measurement uncertainty requirements for that process.

   If the larger tolerance is intentional, such as "limited calibration," then the EOPR data is valid for that application only.

Measurement Process Uncertainty - u_{mp}

Measurement uncertainty is the doubt that exists about the value of a measurement. This doubt is the result of the combined effect of all the error sources that may affect a measurement process, in this case a calibration process. The error sources most often encountered in making calibration measurements include, but are not limited to the following:

- Reference standard accuracy
- Repeatability
- Resolution Error
- Operator Bias
- Environmental Factors Error
- Computation Error

Evaluation of these potential error sources as components of the overall measurement uncertainty provides information as to the "goodness" of the calibration process. One facet of NASA's engineering review was the identification of those sources of measurement process uncertainty that could erroneously increase EOPR. The examination uncovered only two uncertainty components that might cause EOPR to be erroneously high:

1. Insufficient reference standard resolution - Although this component could lead to erroneous EOPR data, it would be a failure of the calibration process design, by misapplication of a standard.

An extreme example would be the calibration of a gage block with an optical scale. Assume the gage block has a tolerance of four micro-inches and the optical scale a tolerance of one micro-inch, with the scale's minor division at 1,000 micro-inches. Even though the scale has the accuracy at the etched divisions, it would be impossible to resolve within the gage block accuracy.

2. Reference standard uncertainty – This is an issue with the accuracy specification of the standard and follows the same rationale as the discussion under the "Tolerance" heading.
   a. If the reference standard uncertainty is in reality larger than its tolerance, then calibration rejections should increase, thus decreasing the EOPR.
   b. If the reference standard uncertainty is in reality smaller than its specified tolerance, then the EOPR will be high, thus indicating a conservatively specified tolerance. This can happen, for example, with reference standard tolerances that must accommodate environments outside of a laboratory.

A common relationship between the measurement process uncertainty and the tolerance is known as the Test Uncertainty Ratio (TUR). It is defined in Z540.3 and represents the ratio of the span of the UUT tolerance to twice the 95% expanded uncertainty of the calibration measurement process. The TUR is useful while discussing the PFA model, because it keeps the relationship of the UUT tolerance in perspective to the calibration measurement process uncertainty.

$$TUR = \frac{\pm Tol}{2 \cdot U_{95}} \qquad U_{95} = k \cdot u_c \qquad k = 1.96$$

Where U_{95} is the 95% expanded uncertainty, k is the coverage or confidence factor, and u_c is the combined standard uncertainty, as defined in the ISO *Guide to the Expression of Uncertainty in Measurement* (GUM) [6].

The standard deviation of the *a priori* population distribution - σ_{tp}

This is the UUT test-point uncertainty and influences the PFA model similar to the measurement process uncertainty, in that it represents the standard deviation in one of the two Gaussian distributions. One way of obtaining this data is through many repeat measurements of a population of instruments and use of proper statistical tools to calculate the standard deviation. For a calibration provider with thousands of instruments, totaling tens-of-thousands of test points, this is not economically feasible. An alternative method is to use the in-tolerance reliability to estimate the standard deviation. End-of-period-reliability (EOPR) is the probability of a unit being in-tolerance at the end of its normal calibration interval. Although in-tolerance probability is binomial (number of successes divided by total trials), EOPR is assumed to be a Gaussian (normal) distribution when estimating the standard deviation of the population. The following estimates test-point uncertainty:

$$\sigma_{tp} \approx u_{tp} = \frac{Tol}{\Phi^{-1}\left(\frac{1+p}{2}\right)}$$

Where $\Phi^{-1}()$ is the inverse normal distribution function and p is the EOPR.

As previously discussed, the PFA model expects a perfect, or "true," standard deviation to represent the UUT test-point uncertainty for a population of instruments. As collected, EOPR data includes the effects of drift, wear, abuse, different standards, different technicians, recalibrated standards, varying ambient conditions, and other factors. In addition to these effects, the calibration events that generate EOPR data affect the calibration decisions that, in turn, can dominate EOPR data. Measurement-process uncertainty influences the in-tolerance reliability through the false acceptance and rejection decisions associated with calibration. In other words, measurement process uncertainty "taints" the EOPR data during its initial collection, making the raw, or "observed," EOPR data appear worse than a perfect, or "true," end-of-period-reliability.

Because a perfect standard deviation does not exist, variance addition rules provide a method for removing the influence of measurement-process uncertainty, thereby adjusting the "observed" standard deviation to the "true" standard deviation. The observed test-point variance is the sum of the inherent ("true") variance of the test-point and the variance due to the measurement process uncertainty, thereby allowing estimation of the "true" test-point uncertainty.

$$\sigma_{tp(obs)}{}^2 \approx u_{tp(obs)}{}^2 = u_{tp(true)}{}^2 + u_{mp}{}^2$$

$$\sigma_{tp(true)} \approx u_{tp(true)} = \sqrt{u_{tp(obs)}{}^2 - u_{mp}{}^2}$$

For this report, the term "observed" will indicate as-collected EOPR data that contains measurement-process uncertainty, and the term "true" indicates EOPR data without the measurement-process uncertainty.

The Appendix at the end of this paper provides additional information on the concept of "observed" versus "true" EOPR.

Dominant variables in the PFA model

The PFA model is a complex interplay of two Gaussian probability distributions over the range of the specified tolerance. The measurement process uncertainty and the test-point uncertainty estimated by EOPR represent the standard deviations for these two distributions. As discussed previously, there is a threshold value for each of these variables that, when exceeded, ensures the target PFA is met, regardless of the value of the second variable. Compliance to the Z540.3 2% PFA requirement is achievable with either measurement process uncertainty or observed EOPR alone, as long as that single variable is in its region of dominance. Up to this point, the discussion has focused on when in-tolerance reliability is the dominant variable at 89% observed EOPR.

Due to the non-linear nature of the Gaussian distribution, the influences of measurement process uncertainty and EOPR on the PFA model are also non-linear. For a fixed EOPR value, as the measurement process uncertainty increases (i.e., decreasing TUR), the PFA will increase proportionally until it reaches a maximum probability and then start decreasing rapidly to zero. The converse is also true – as the measurement process uncertainty decreases (i.e., increasing TUR), the PFA will decrease toward zero. At a point where the measurement process uncertainty is sufficiently small, it becomes the dominant variable in the PFA model.

For example, when the measurement process uncertainty decreases to a point the TUR is 4.6 or larger, the PFA will always be 2% or less, independent of the EOPR value.

Figure 2 illustrates the thresholds where measurement process uncertainty and EOPR dominate the model for a predetermined PFA value. Using the PFA model, the TUR value is an iterative result over a range of observed EOPR values, for the specified PFA. Figure 2 graphically shows that for PFA values of 2% and 2.7%, the TUR threshold is 4.6:1 and 3.33:1 respectively. At these TUR values, measurement uncertainty is the dominant variable for the given PFA value. The choice of an EOPR value of 85% is due to its popularity as an in-tolerance reliability target for many organizations.

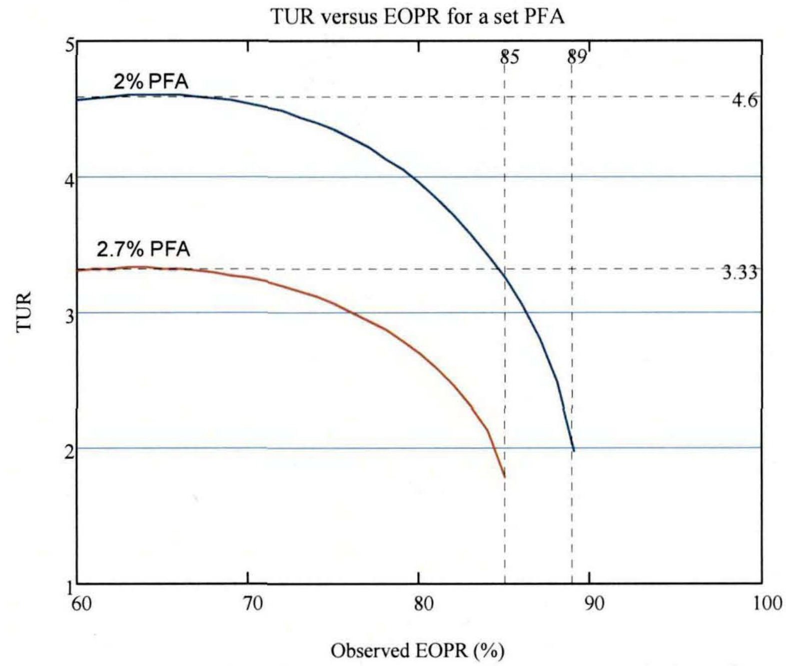


Figure 2: A graphical representation of TUR values over a range of EOPR for a predetermined PFA value.

An important concept illustrated in Figure 2 is that for any given EOPR value, there is a corresponding maximum PFA. This fact could help organizations, which have EOPR data for their legacy inventories, transition to Z540.3. Although their target EOPR value may not be the 89% needed to achieve the default 2% PFA, Z540.3 sub-clause 5.3 allows organizations to establish a suitable measurement decision risk metric. This allows these organizations to transition to Z540.3 with existing in-tolerance reliability data, because the resulting PFA adds no additional risk to the organization's customers. This applies only to legacy equipment that meets the organization's EOPR target value.

4. Ensuring EOPR data is valid

In-tolerance reliability (EOPR) is a measure of the ability of an instrument to hold its accuracy for the duration of its normal calibration interval. The value of in-tolerance reliability is that it is empirical data, containing actual information on the UUT calibration and its usage throughout the calibration cycle. As mentioned earlier, this data may include the effects of drift, wear, abuse, different standards, different technicians, recalibrated standards, varying ambient conditions, and other factors.

Normally, a unit-under-test (UUT) is declared in-tolerance only if all test-points are found in-tolerance. Conversely, an instrument is considered out-of-tolerance even if one test-point is found to be out-of-tolerance. This distinction is important to the usage of in-tolerance reliability in estimating test-point uncertainty. Ideally, the EOPR data collection is at the test-point level. However, for most organizations, in-tolerance reliability data is available only as a percent in-tolerance at the UUT item (serial number) level or higher. Figure 3 illustrates the hierarchy of measuring and test equipment (MTE) from the nomenclature to the test point level.

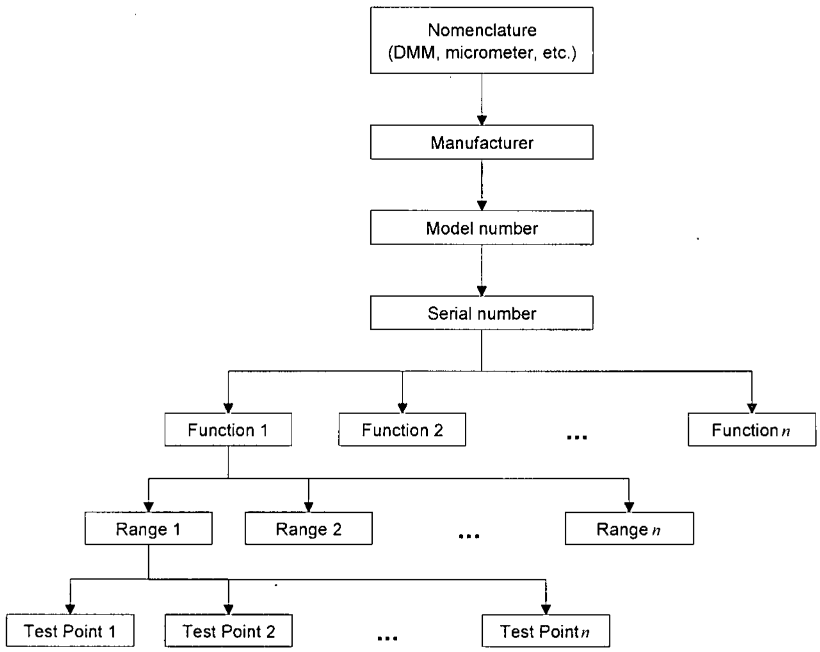


Figure 3: MTE hierarchy from nomenclature to test-point level.

Ideally, to estimate test-point uncertainty, reliability data comes from the test-point level. When EOPR at levels above the test point (e.g., range, function, serial number, and/or model) are used in the estimation of PFA, the resulting PFA will be larger or more "conservative." This is because, for instruments with multiple test points, ranges, and/or functions, the in-tolerance probability for each test-point is inherently greater than the observed EOPR at the instrument level. From a compliance perspective, this means using EOPR at levels above the test-point is acceptable for achieving the 2% PFA requirement, because the reported PFA will be greater than any given test-point PFA.

In all PFA estimations, valid EOPR data is essential to good results. In essence, EOPR is the number of in-tolerance devices (successes) divided by total calibrations of like devices (trials) for as-received instruments with like resubmission intervals. As such, the collected EOPR data is valid for the time-of-test, assuming the adequacy of data collections rules. EOPR validity depends on data capture rules such as in/out-of-tolerance coding, calibration process stability, and homogeneity of data.

As a binomial probability, EOPR is reflective of past performance, as well as an estimate of future performance. EOPR data is collected over time for single items or populations of the same instrument make/model. As with all data sampling, the larger the sample size, the more confidence in the information inferred from the EOPR data, specifically future performance. Confidence limits rely on sample size and the number of successes versus the number of trials.

For example, it takes 22 successes out of 22 trials to provide a 90% confidence level, yet if the 23rd trial results in a failure (e.g., out-of-tolerance), the lower confidence level drops to 84%. This equates to observed in-tolerance reliabilities of 100% and 95.7% respectively.

Although essential to adjusting calibration intervals, confidence limits do not provide the full measure of the validity of the EOPR data at the time-of-test. Observed EOPR is reflective of the conditions at time-of-test based on past performance, therefore using it to estimate the PFA for the calibration cycle would be valid, assuming all collection policies are adequate.

To ensure EOPR data validity, collection polices must be adequate and well controlled. Although not exhaustive, the following are basic data collection requirements needed to ensure the validity of EOPR data.

1. The subject calibration process (procedure) needs to be documented and validated.
2. The subject calibration process has to be consistent and stable over the EOPR collection time-period. Investigate all major changes to the calibration process that could possibly have a negative influence on the EOPR data. This includes items such as test point changes, reference standards, as well as homogeneity of the data such as parameters, tolerances, and calibration intervals. Investigate any changes to the calibration process to verify that the EOPR is still valid. Not all changes will negatively influence the EOPR data.
3. Document the data capture policy to include proper identification of as-received conditions, and accept/reject rules, which also covers the in-tolerance and out-of-tolerance coding policy. Establish data filters to include calibrations that are received early and late within the specified interval.
4. Establish and document the minimum sample size for an instrument population. This is crucial to potential changes in the calibration interval that will affect the reliability value.
5. Establish and document a policy for aggregating sample populations when the model level is too small to provide an adequate sample size. Equipment groupings must be reasonably homogenous in terms of function, range, accuracy, and calibration process. NASA Reference Publication 1342 [7] provides guidance in this area.

5. Summary and Conclusions

When NASA's Kennedy Space Center began transitioning to ANSI/NCSL Z540.3:2006, an engineering review was initiated to examine ways to mitigate some of the costs associated with achieving compliance to the new standard. The largest cost driver identified for implementation was measurement uncertainty analyses on legacy calibration processes without documented uncertainties. NASA concluded from this review that legacy calibration processes with in-tolerance reliability above 89% met the Z540.3 PFA metric; therefore, the associated measurement uncertainty would be adequate. In essence, if the calibration process is reliable, all the uncertainty sources are either insignificant or have been addressed through the design, implementation, and control of the calibration process. A thorough engineering review provided the rationale that this can be true under certain circumstances:

1. Observed EOPR values equal to or greater than 89% provides objective evidence of compliance to Z540.3's PFA requirements (sub-clause 5.3b).
2. In addition, it provides that the measurement uncertainty associated with the calibration processes would be adequate for that purpose (sub-clause 5.3.3).

These conclusions are predicated on the validity of the EOPR data, which is essential to all PFA estimates. To ensure the validity and integrity of EOPR data, adequate collection methods are crucial.

NASA's policy for legacy equipment requires a full analysis of the measurement process if the EOPR falls below 89%, or if the reliability is affected by changes to the instrument specifications or the measurement process.

With NASA's adoption of ANSI/NCSL Z540.3-2006, the 89% rule will help phase-in Z540.3 across the organization. This method could help other organizations phase-in Z540.3 with an understanding of three key limitations of using in-tolerance reliability for evidence of compliance:

1. It is limited to legacy equipment.
2. It is limited to organizations with the requisite reliability data.
3. It will eventually become obsolete.

References

1. *Requirements for the Calibration of Measuring and Test Equipment,* ANSI/NCSL Z540.3-2006, NCSL International, 2006
2. *Handbook for the Application of ANSI/NCSL Z540.3-2006,* NCSL International, 2009
3. Ferling, John, *The Role of Accuracy Ratios in Test and Measurement Processes*, Proc. Measurement Science Conference, Long Beach, 1984
4. *Estimation and Evaluation of Measurement Decision Risk*, NASA-HANDBOOK-8739.19-4, 2010
5. *Measurement Uncertainty Analysis Principles and Methods*, NASA-HANDBOOK-8739.19-3, 2010
6. *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization (ISO), 1995
7. *Metrology – Calibration and Measurement Processes Guidelines*, NASA Reference Publication 1342, NASA, 1994
8. Dobbert, Michael, *Understanding Measurement Risk*, Proc. NCSL International Workshop & Symposium, St. Paul, August 2007

Appendix

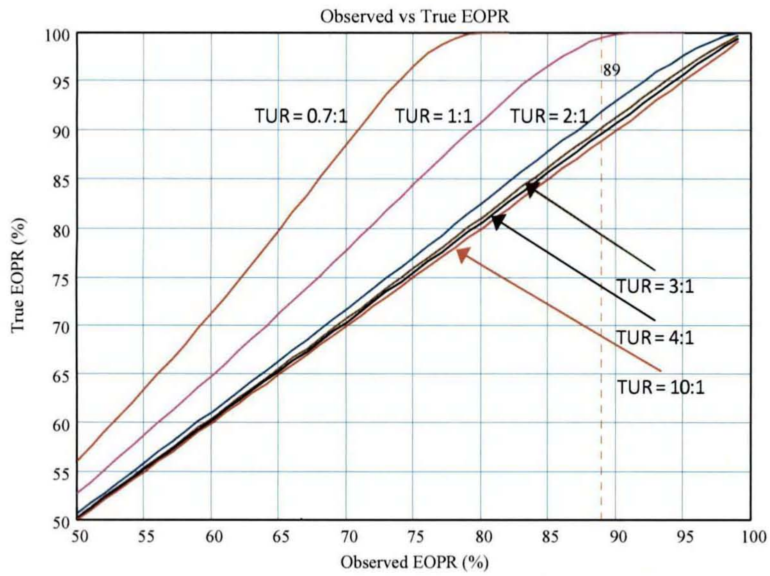
The concept of "observed" versus "true" EOPR

True versus Observed EOPR

With EOPR's strong influence on PFA results, the topic of "true" versus "observed" EOPR warrants additional emphasis. This is especially important in light of the potential cost savings to implementing organizations.

In general, mathematical models will return a result regardless of the input source, unless the source violates a mathematical principle. The PFA model is no exception. The objective of this Appendix is to look at the effect of using "observed" versus "true" End of Period Reliability (EOPR) in the calculation of measurement decision risk.

Due to the effects of the measurement process uncertainty, true EOPR is always larger than observed EOPR. This becomes more pronounced as the uncertainty increases in relation to the test tolerance (i.e., decreasing TUR). Figure 4 illustrates the relationship of a fixed observed-EOPR value to the true-EOPR for a decreasing TUR. Note that the difference between true and observed EOPR becomes more significant for TUR values below 4:1. This may be more evident in the table than the associated graph.



TUR	Observed EOPR	True EOPR
10:1	89%	89.1%
4:1	89%	89.7%
3:1	89%	90.3%
2:1	89%	92.0%
1:1	89%	99.4%
0.7:1	89%	100%

Figure 4: True versus observed EOPR. The affects become more pronounced as the TUR drops below 4:1

True and Observed EOPR in the PFA Model

As with any model, PFA estimation is dependent on the quality of the input data to provide the best possible results. The influence of calibration-process uncertainty on EOPR data compromises the PFA model and results in very "conservative" PFA estimates. Some may view "conservative" results as acceptable, although the results are wrong. Using "conservative" PFA results can lead to performing measurement uncertainty analyses on legacy calibration processes that, in reality, are meeting Z540.3's requirements.

Figure 5 illustrates the PFA results when substituting observed for true EOPR and correcting the observed EOPR. The differences in the PFA results are more noticeable below a TUR of 3:1.

Appendix

The concept of "observed" versus "true" EOPR

The right graph of Figure 5 indicates that the probability of a false accept (PFA) is no greater than 2% for all TUR values when the observed EOPR is at or above 89%.

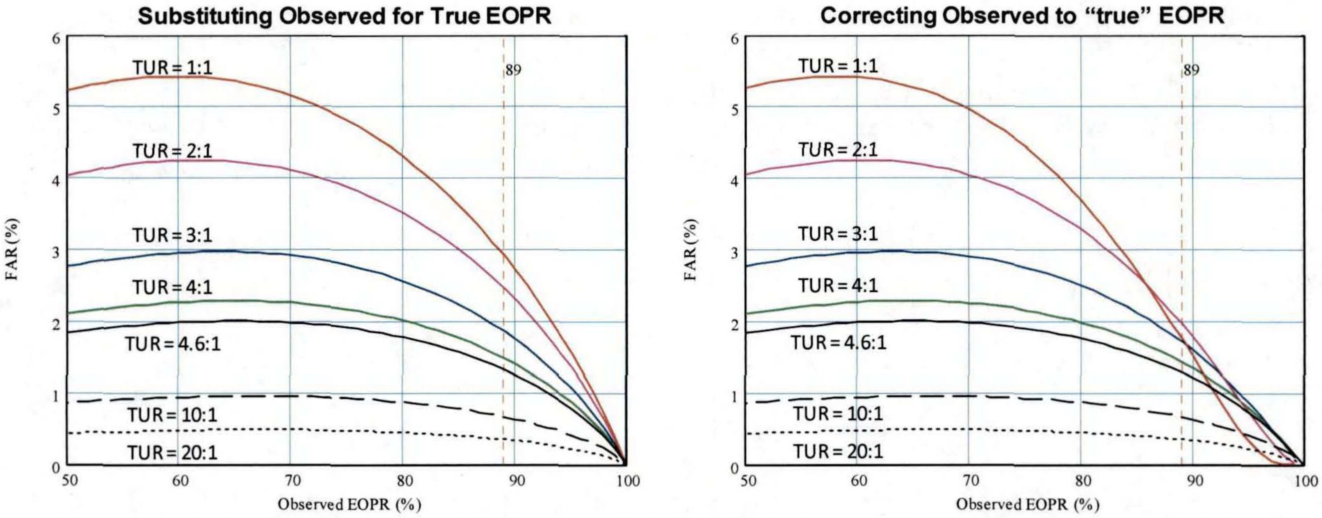


Figure 5: Using the same PFA model, the left graph treats Observed EOPR as True. The right graph "corrects" Observed to True. Both graphs plot PFA against the Observed EOPR.

Note that in Figures 5, the y-axis label is FAR (false accept risk) in lieu of PFA.

Figure 5 also illustrates where the dominant-variable changes between measurement process uncertainty and EOPR. As just noted, above 89% observed EOPR, the PFA remains 2% or less for all measurement uncertainty values as they relate to the tolerance. It can also be seen that as the measurement process uncertainty decreases, the probability of making a wrong calibration acceptance or rejection decision decreases. For all TUR values 4.6:1 and greater, the probability of incorrect acceptance decisions cannot exceed a 2% PFA, regardless of the EOPR value.

Modeling PFA with Monte Carlo Simulations

The PFA model normally uses integral equations such as in Figure 5. Monte Carlo simulation uses random sample modeling techniques in lieu of the more direct equations, thus provides a different perspective of observed and true EOPR within the PFA model. Mike Dobbert used this technique his 2007 NCSLI paper, *Understanding Measurement Risk* [8].

The following Monte Carlo plots are generated using two Gaussian distributions, each with a mean of zero. The Monte Carlo simulations use the following expression.

$$y = e_{uut} + e_{mp}$$

Where,

y = the calibration result and is plotted on the y-axis.

e_{uut} = the UUT error and is the standard deviation for one distribution, as estimated by EOPR and is plotted on the x-axis.

e_{mp} = the measurement-process uncertainty and is the standard deviation for the other distribution.

Appendix

The concept of "observed" versus "true" EOPR

For all Monte Carlo plots, the red regions are false accept or reject (labeled accordingly) and the green regions are the corresponding correct accept or reject regions (unlabeled).

Figure 6a and 6b represent the extreme limits of the PFA model as illustrated by the Monte Carlo simulation. The reason for examining the limits is to understand how the simulation works, as well as to illustrate the functional behavior of the PFA model. To reach the extreme limits, the PFA model requires either a "perfect" UUT or a "perfect" measurement process, both of which do not exist. As shown by the fact that neither line enters into the false acceptance regions, we

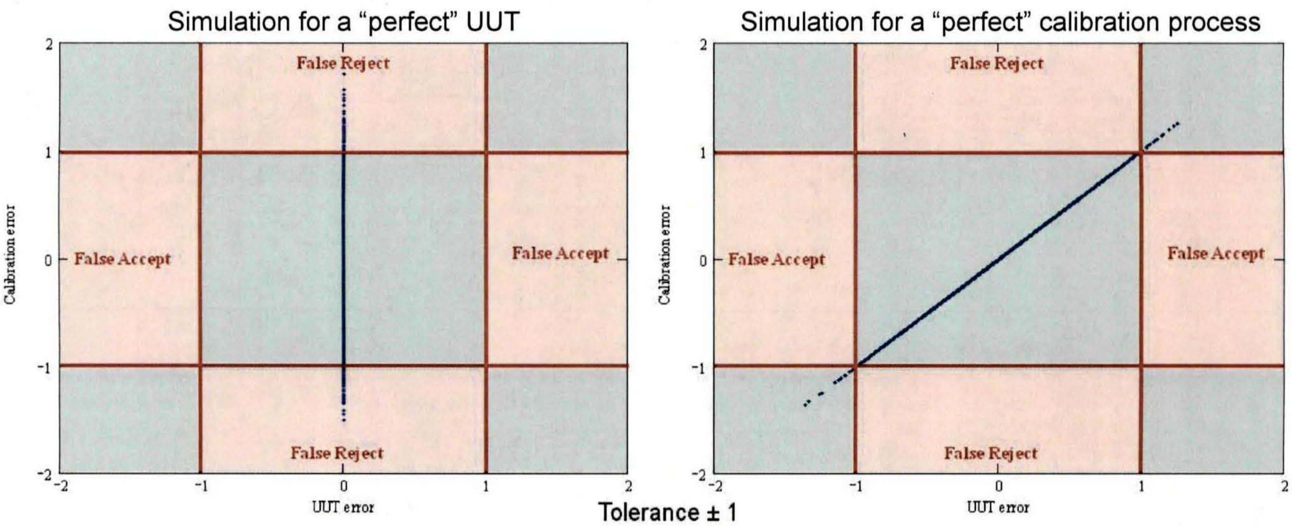


Figure 6a: Monte Carlo simulation when the unit under test (UUT) is "perfect," with no bias ($e_{uut} = 0$), thus $y = e_{mp}$.

Figure 6b: Monte Carlo simulation when the measurement process is "perfect," with no error ($e_{mp} = 0$), thus $y = e_{uut}$.

see that at both extremes ($e_{uut} = 0$ or $e_{mp} = 0$), the probability of false acceptance is zero.

Figure 6a assumes every device within the test population is "perfect," without any bias error. A "perfect" UUT means that the $e_{uut} = 0$. Figure 6a also assumes there is measurement process error, characterized by a distribution, thus the calibration result becomes $y = e_{mp}$. In this extreme case, the probability of false acceptance will always be zero, because no UUT is ever out-of-tolerance. However, the probability of false rejects will increase because all rejections are false.

Figure 6b assumes the measurement process error is "perfect," represented by $e_{mp} = 0$. Figure 6b also assumes every device within the test population has some bias error, represented by a distribution, thus the calibration result becomes $y = e_{uut}$. At this extreme limit, all the results fall on an infinitely narrow diagonal line, passing between the corners of the false-accept regions, where the PFA is again zero. In this case all rejections are correct, thus the PFR is zero.

Figure 6a and 6b illustrates that the vertical spread *about* the diagonal line is a function of the measurement-process uncertainty and the spread *along* the diagonal line is a function of the UUT test-point uncertainty as estimated by the EOPR. In other words, with the EOPR fixed, a change in the measurement process uncertainty causes the model to change along the y-axis, in effect, making the diagonal line appear thick or narrow. With measurement process uncertainty fixed, changes in the EOPR cause the model to change along the x-axis, in effect, shortening or lengthening the diagonal line.

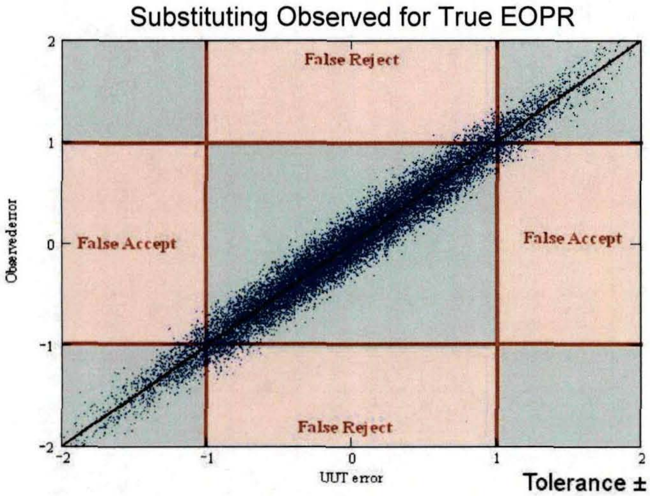
Figures 7 through 10 plot the results of Monte Carlo simulations with varying TUR values. They follow the same pattern as Figures 5, with the left graph substituting observed for true EOPR and

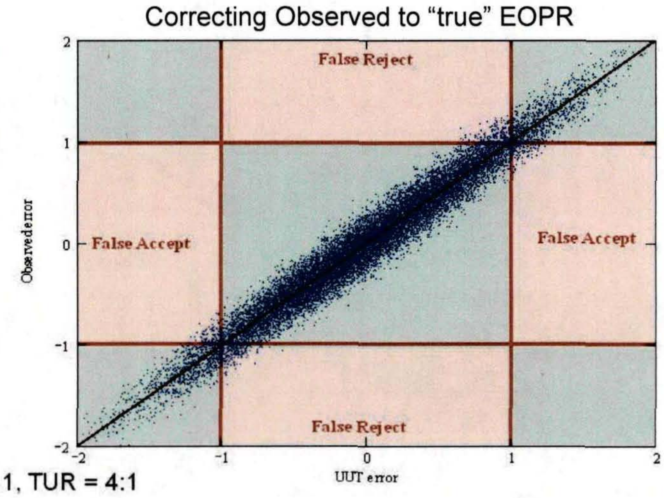
Appendix

The concept of "observed" versus "true" EOPR

the right graph correcting the observed to "true" EOPR, using variance addition to remove the measurement process uncertainty from the UUT error (e_{uut}). Although in all graphs, the EOPR is 89%, the EOPR in the right graph is being "corrected," thus causing the e_{uut} to approach zero as the TUR decreases (increasing measurement uncertainty).

Figure 7a and 7b illustrate the simulations for a 4:1 TUR, which confirms the previous discussion on when variables dominate the PFA model. In this case, the low measurement process uncertainty (high TUR) is beginning to dominate the FAR model, thus there is very little discernible difference between the two graphs.



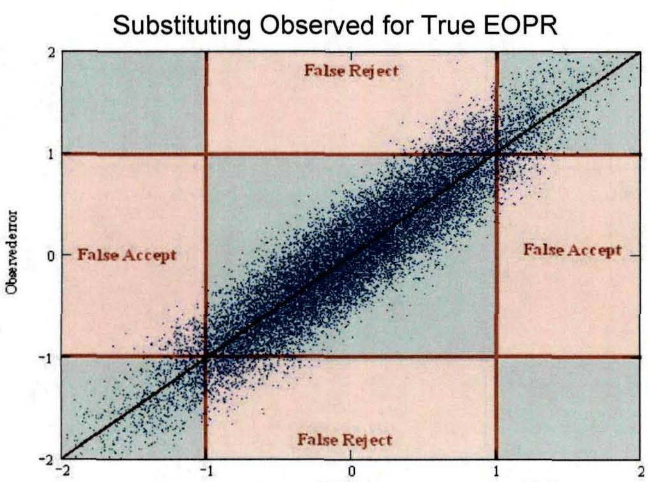


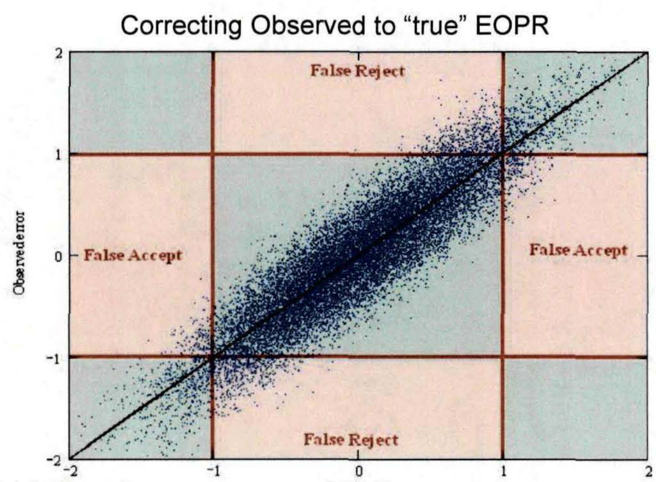
Tolerance ± 1, TUR = 4:1

Figure 7a: Monte Carlo simulation of the PFA model when substituting observed for True EOPR. True EOPR = 89% and PFA = 1.48%.

Figure7b: Monte Carlo simulation of the FAR model when correcting observed to "true" EOPR. Observed EOPR = 89%, "true" EOPR = 89.7%, and PFA =1.42%.

Figures 8a and 8b illustrate the simulations for a 2:1 TUR. The difference between the right and left graphs is beginning to be discernable, due to the higher measurement process uncertainty (lower TUR). Although both graphs are beginning to show a counterclockwise rotation, Figure 8b is spreading less, with fewer points in the false accept regions. Counting the points in the





Tolerance ± 1, TUR = 2:1

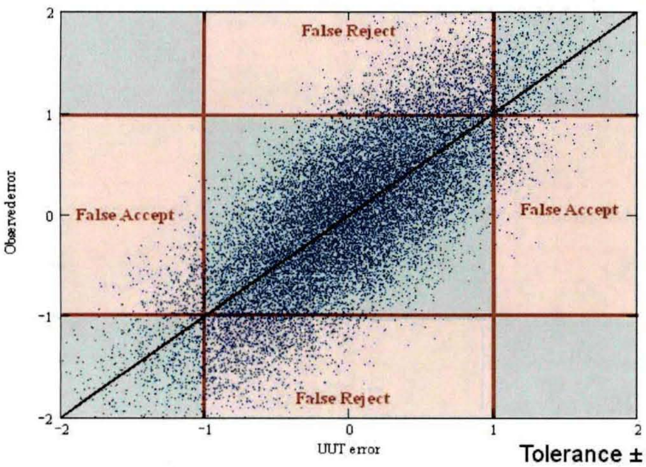
Figure 8a: Monte Carlo simulation of the PFA model when substituting observed for True EOPR. True EOPR = 89% and PFA = 2.45%.

Figure 8b: Monte Carlo simulation of the PFA model when correcting observed to "true" EOPR. Observed EOPR = 89%, "true" EOPR = 92.0%, and PFA =1.95%.

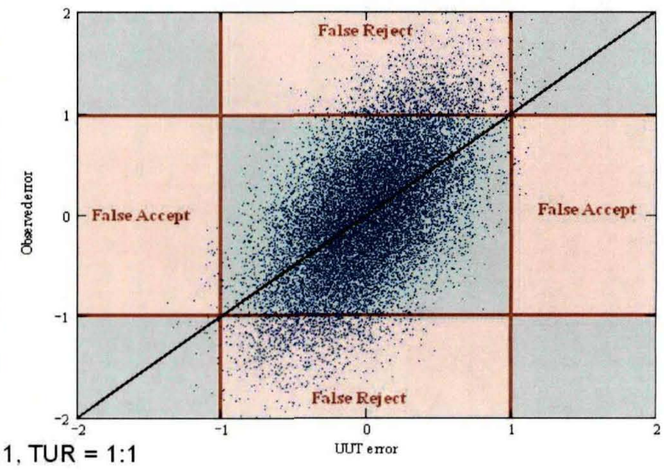
false accept regions and dividing by the total sample population would confirm the value calculated by the PFA integral equations.

Figure 9a and 9b illustrate the simulations for a 1:1 TUR. As the measurement process uncertainty increases in relation to the specified tolerance, the two graphs become markedly different. The rotation is now obvious for both graphs, although Figure 9b's rotation is more pronounced. In addition, as Figure 9a is spreading larger, the number of points in the false-accept region of Figure 9b has decreased dramatically. This is indicative of the e_{uut} becoming an ideal UUT, with little error. As indicated in the caption, the difference between the observed and "true" EOPR is more than 10% due to the measurement process uncertainty.

Substituting Observed for True EOPR



Correcting Observed to "true" EOPR



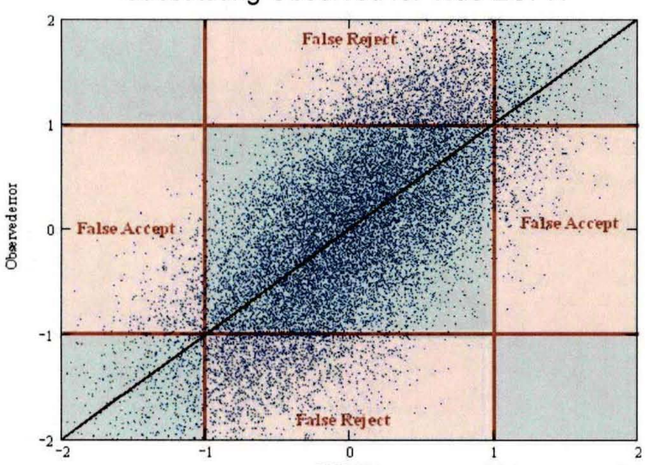
Tolerance ± 1, TUR = 1:1

Figure 9a: Monte Carlo simulation of the PFA model when substituting observed for True EOPR. True EOPR = 89% and PFA = 3.54%.

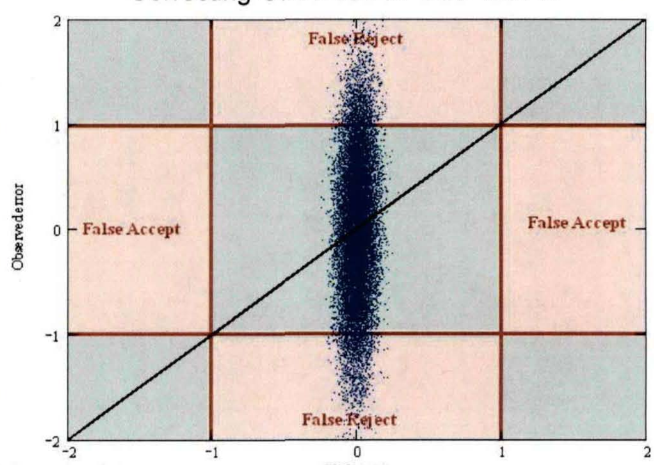
Figure 9b: Monte Carlo simulation of the PFA model when correcting observed to "true" EOPR. Observed EOPR = 89%, "true" EOPR = 99.4%, and PFA =0.24%.

Figure 10a and 10b illustrate the simulations for a 0.82:1 TUR, a case where the measurement process uncertainty is larger than the specified tolerance that the calibration is verifying. It is counter-intuitive that such a situation could have high reliability, yet Figure 10b illustrates that when it occurs, EOPR dominates the PFA model. Figure 10b represents the ideal UUT where

Substituting Observed for True EOPR



Correcting Observed to "true" EOPR



Tolerance ± 1, TUR = 0.82:1

Figure 10a: Monte Carlo simulation of the FAR model when substituting observed for True EOPR. True EOPR = 89% and FAR = 3.82%.

Figure 10b: Monte Carlo simulation of the FAR model when correcting observed to "true" EOPR. Observed EOPR = 89%, "true" EOPR = 100%, and FAR =0%.

the e_{uut} is approaching zero. At this point, there are no false accepts, and all rejects are false.

Figures 9b and 10b are illustrative of EOPR saturating the PFA model. They both are beginning to resemble the situation shown in Figure 6a where the unit under test bias is zero - in other words, a perfect UUT. Obviously, a perfect UUT is impossible, yet there are many cases where high reliability exists with low TUR. As discussed earlier, there are two situations where this will occur, assuming the EOPR is adequate.

1. The Reference Standard out-performs its assigned accuracy specifications. Generally, instruments specifications cover a broad range of conditions, including variations in conditions of use. Instruments used in a controlled environment, such as a calibration laboratory, will normally perform well within the allowed tolerance limits. Effectively, the instrument (Reference Standard) consistently operates within a fraction of its tolerance limits, and the measurement processes in which it is used will, in reality, have a higher TUR than was estimated using the Reference Standard's accuracy specifications.
2. The ratio between the resolution and specified accuracy of the unit-under-test (UUT) is very low (e.g., below 2:1). The instrument's resolution dominates the measurement uncertainty for these "resolution-limited" instruments, resulting in TUR values below 2:1. In cases where the inherent physical characteristics of the UUT are significantly better than its resolution, the instrument will have high in-tolerance reliability regardless of the low TUR. For example, caliper micrometers often have high in-tolerance reliability coupled with low resolution-to-accuracy ratios. This is possible because design tolerances for key mechanical components, such as the lead screw, are smaller than the instrument's resolution, often by an order of magnitude.

Comparing the right and left graphs of Figures 7 through 10 illustrates calibration-process uncertainty's influence on EOPR as the TUR decreases. Without adjusting for this influence, the PFA model results will show a higher probability of incorrect calibration decisions, especially in the low TUR regions. This is true in all cases of PFA estimation, but is more crucial for legacy systems, where it could lead to performing measurement uncertainty analyses on legacy calibration processes that in reality are meeting Z540.3's requirements, based on high reliability.