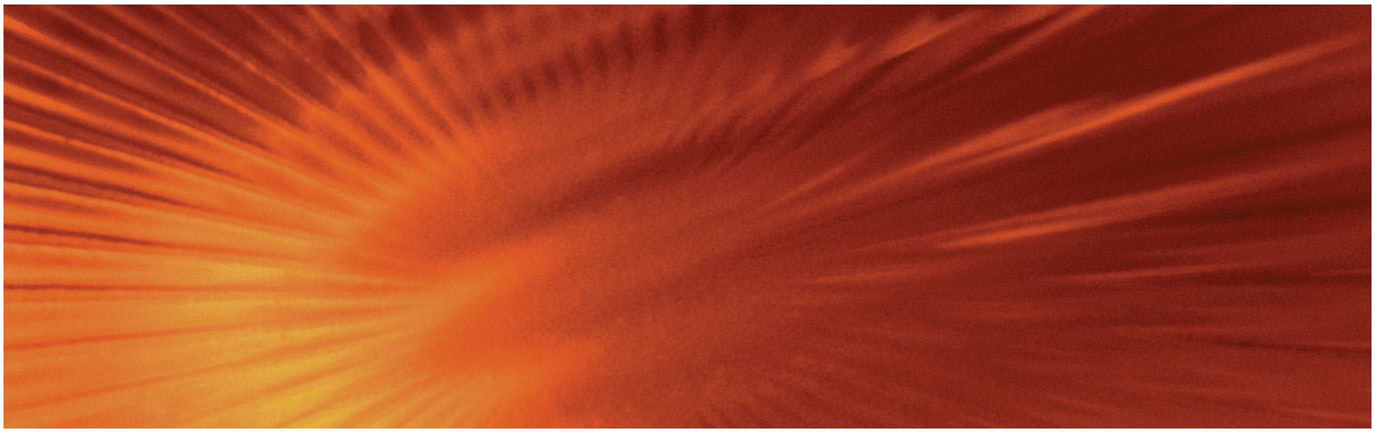


Setting and Using Specifications — An Overview

Michael Dobbert

Abstract: This paper presents techniques that manufacturers can use to set specifications and then describes how a metrologist can use those specifications in calibration. Specifications describe the warranted performance of a product or the expected accuracy of a measurement standard. From the manufacturer's perspective, specifications must describe performance that can be achieved cost effectively. This paper looks at several statistical issues related to setting specifications. Techniques for characterizing expected product performance, considerations for drift and performance variation due to external environmental conditions, and the significance of measurement uncertainty are covered. A framework is presented that relates the expected performance of a product to its specifications while maintaining metrology goals, producibility, and competiveness. For calibration, specifications are often used as Type-B uncertainty estimators. Drawing from the statistical issues related to setting specifications, this paper discusses the use of product specifications in an uncertainty analysis.

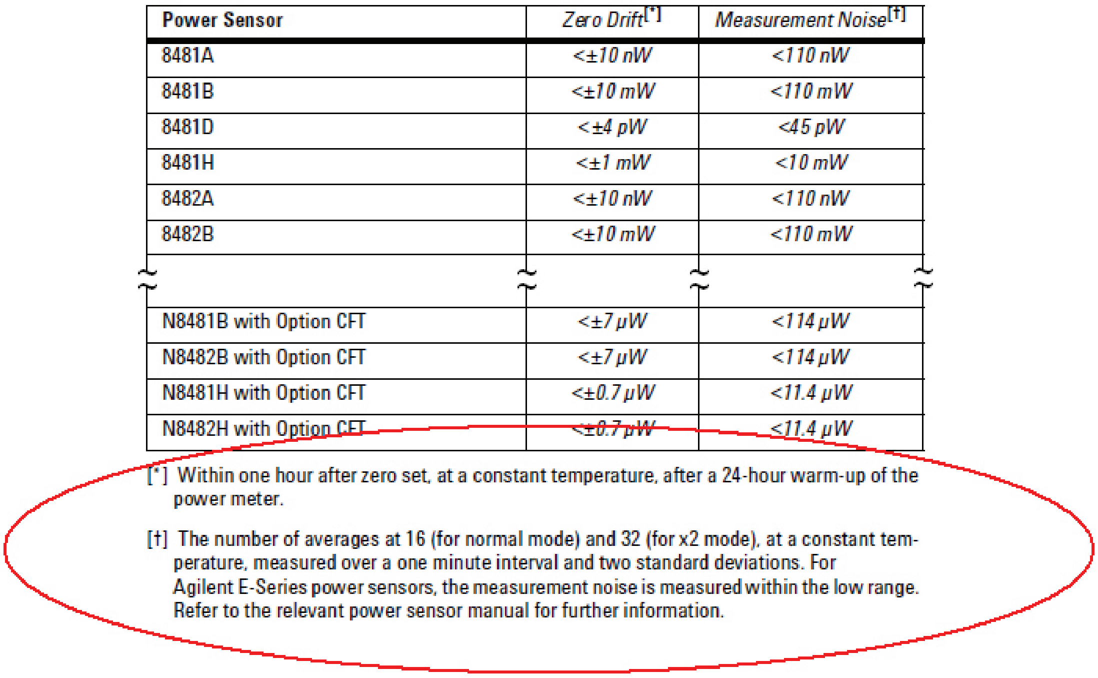
1. Introduction

Specifications convey critical information about a product or a device. A review of specifications is crucial when assessing the applicability of a product to an application, and ultimately to the decision of whether or not to purchase the product. In this sense, specifications indicate expected performance in addition to other important product characteristics. For the metrologist, specifications take on an expanded technical role. Using specifications to derive Type-B estimates for uncertainty is very convenient and widely exercised. In this case, the metrologist assumes that specifications describe product performance that can be maintained over time through calibration. The metrologist also needs to know additional details related to specifications, such as vital information for determining Type-B measurement uncertainties, or the operating conditions necessary to achieve the specified performance. Finally, specifications typically fall into one of two classifications, those that are warranted and those that are not, with warranted specifications being those that are central to the use of the product and are most likely the subject of periodic calibration. The warranted specifications define the fitness for use criteria that the manufacturer is willing to back with a warranty.

Michael Dobbert

Agilent Technologies
1400 Fountaingrove Parkway, Mail Stop: 3USH
Santa Rosa, CA 95403 USA
Email: dobbert-metrology@agilent.com

The process for setting specifications must attempt to meet the various needs of those that rely on the information the specifications communicate. It is important that the specifications accurately describe performance. An overly conservative specification that underestimates actual product performance may lead to rejecting the product for a specific application when, in fact, it is suitable. For the manufacturer, an overly aggressive specification can lead to increased manufacturing and warranty costs. Ideally, the specification setting process produces specifications that accurately describe performance,



Power Sensor	Zero Drift[*]	Measurement Noise[†]
8481A	<±10 nW	<110 nW
8481B	<±10 mW	<110 mW
8481D	<±4 pW	<45 pW
8481H	<±1 mW	<10 mW
8482A	<±10 nW	<110 nW
8482B	<±10 mW	<110 mW
N8481B with Option CFT	<±7 µW	<114 µW
N8482B with Option CFT	<±7 µW	<114 µW
N8481H with Option CFT	<±0.7 µW	<11.4 µW
N8482H with Option CFT	<±0.7 µW	<11.4 µW

[*] Within one hour after zero set, at a constant temperature, after a 24-hour warm-up of the power meter.

[†] The number of averages at 16 (for normal mode) and 32 (for x2 mode), at a constant temperature, measured over a one minute interval and two standard deviations. For Agilent E-Series power sensors, the measurement noise is measured within the low range. Refer to the relevant power sensor manual for further information.

Figure 1. Specification information important for designing a calibration procedure.

support making product comparisons, are measurable and are valid over variations in the external environment. Agilent Technologies uses the methods described in this paper when setting specifications for most products as they are introduced into the electronic measurement market.

2. Specification Definitions

Specifications, as described in this paper, relate specified tolerances to the expected performance of a product. Characterizing a sample of products, usually the first products built just before full-scale production of the product begins, provides an estimate of expected performance. Characterizing product performance requires a calibration procedure.

The role of calibration, however, is much larger than characterizing a product to set specifications. Calibration is necessary for monitoring performance over time, for inter-laboratory comparisons and for estimating product reliability. Additionally, the measured result from calibration may be used to correct or adjust the performance and is central when assessing conformance. Therefore, it is necessary that a product's performance be measurable, either directly or indirectly, and the results must be repeatable and reproducible.

Repeatability and reproducibility are often a function of the calibration procedure, but also the laboratory environment and the care and experience of the person performing the calibration. The unit under test can also contribute to repeatability. The unit under test contribution may be unavoidable, but careful definition of specifications is necessary to minimize what the GUM refers to as *intrinsic uncertainty*[1].

Intrinsic uncertainty is uncertainty due to the incomplete definition of a measurand. There is, therefore, a direct relationship between the definition of a specification and uncertainty of the calibration procedure.

To illustrate this relationship, consider the following example found in the GUM. Assume the parameter of interest is the thickness of a sheet of material and that the thickness is measured using a micrometer. The measured result depends on factors such as the temperature of the material, the force applied by the micrometer and others. Moving the micrometer to a second location on the material may give a different measured result. That difference may be due, in part, to the non-uniform thickness of the material. The variation in thickness of the material, in this case, represents an intrinsic uncertainty. However, the magnitude of the intrinsic uncertainty is dependent on the definition of the measurand. If the calibration procedure specified a particular location, or a series of locations for which multiple measurements are then averaged, the resulting intrinsic uncertainty would be less given a single measurement made at an arbitrary location.

As a second example, assume the parameter of interest is the level of noise sidebands on signal from a signal source. Whereas with the sheet of material in the previous example, thickness varied with location, with measurements involving noise, the measured value varies with time. The amount of variation depends upon the amount of filtering applied by the measuring device, or by the amount of averaging of data results. Again, the definition of the measurand affects the uncertainty of measurement even though the variation is a characteristic of, in this example, the signal source.

---

[1] See Annex D of the GUM. [1]

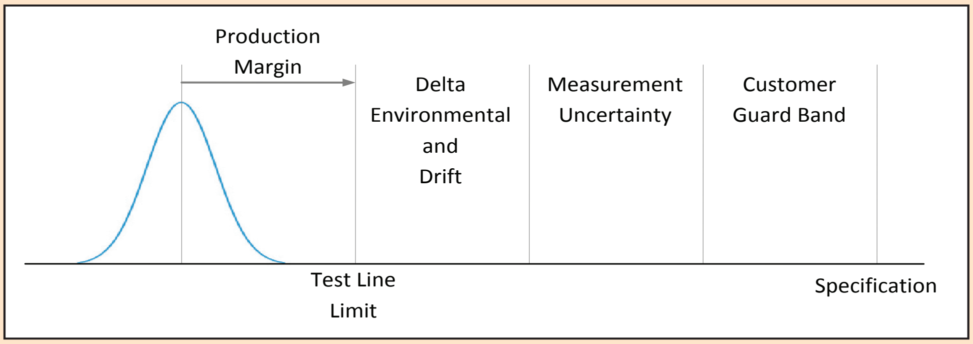


Figure 2. Specification model.

When setting a specification, therefore, attention must be given to the potential calibration procedures used to maintain the specification over time. It is important that specifications are defined with sufficient detail so that a single item, under more or less the same conditions, exhibits essentially the same performance when measured at different calibration laboratories. Figure 1 shows an example of a specification[2] with necessary detail for the calibration procedure.

3. Specification Model

This paper relies on relating production margins, test line limits (TLL) and specifications as originally put forward by Read and Read. [2] Figure 2 gives a graphical representation of this relationship for a single sided specification. For two-sided specifications, the indicated regions extend on both sides of the distribution.

The specification model shows a test line limit set at a point to achieve manufacturing yield goals based upon the expected performance of the manufactured product. The test line limit is the limit used by the manufacturer for the pass/fail criteria. The area between the test line limit and the specification is guard band. Guard band is a safety margin and accounts for possible changes in performance as a function of environmental conditions, expected drift during the calibration interval, measurement uncertainty and any additional guard band necessary to ensure products are within the specification.

4. Test Line Limits

In order to set a test line limit and predict what the manufacturing yield might be, manufacturers must estimate the future performance of a product. Commonly, this is accomplished by measuring a representative sample of items. If a Gaussian distribution describes the sampled items' performance, then a set of convenient tools exist for predicting performance and manufacturing yields.

A common assumption is that product specifications describe 95 % of the population of product items. From the mean, μ, and standard deviation, σ, an interval of $[\mu - 2\sigma, \mu + 2\sigma]$ contains approximately 95 % of the population. However, when manufacturers set product specifications, the test line limit is often set wider than 2σ from the population mean. This is due to several factors.

The quantities μ and σ pertain to the population of product items. However, the data available for setting specifications is from a limited sample of pre-production or early-production units. Therefore, it is necessary to approximate μ and σ with the sample mean, $\bar{x}$, and the sample standard deviation, s, determined from the sample data. Because $\bar{x}$ and s are approximations to μ and σ, the interval $[\bar{x} - 2s,\ \bar{x} + 2s]$ may, or may not, represent 95% of the population. This is due to sampling error. To address this, manufacturers can rely on *tolerance intervals* to set test line limits.

A tolerance interval is an interval that contains *at least* a desired proportion of a distribution at a stated confidence. That is, the interval $[\bar{x} - k_2 s,\ \bar{x} + k_2 s]$ contains at least proportion, p, of the population with confidence, γ, where the factor, k_2, is dependent on p, γ, and n, the number of samples used to determine $\bar{x}$ and s. For a closed interval (in the case of a two-sided specification), an approximate value for the *k factor* [3] is,

$$k_2 = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right)(z_\alpha)^2}{\chi^2_{\beta,n-1}}}\ , \qquad (1)$$

where

k_2 = the two-sided *k factor*,

z_α = the inverse standard normal function for probability $\alpha = \frac{1-p}{2}$,

$\chi^2_{\beta,n-1}$ = the inverse chi-square function for probability $\beta = 1 - \gamma$ with $n - 1$ degrees of freedom.

For example, assuming the number of items in a sample is 30, to find the *k factor* for a two-sided tolerance interval that includes at least 95 % of the population with a 98 % confidence, then $z_{0.025} = -1.96$ and $\chi^2_{0.02,29} = 15.57$ and,

---

[2] From the Agilent N1913A/N1914A EPM Series Power Meters User's Guide, December 7, 2009. [Available at http://cp.literature.agilent.com/litweb/pdf/N1913-90001.pdf]

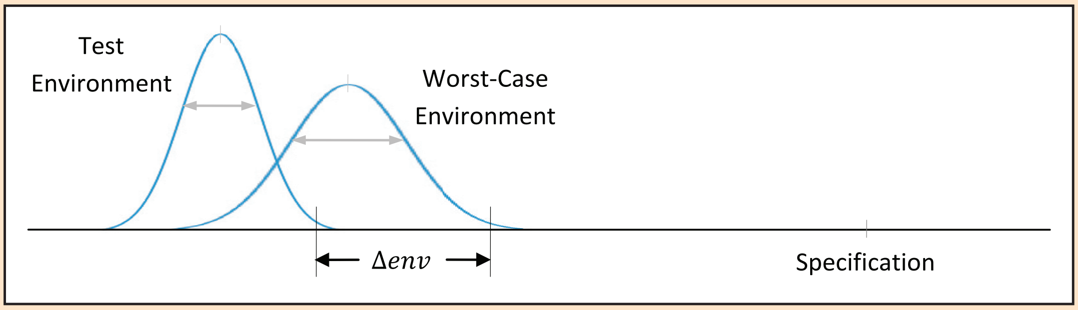


Figure 3. Changes in the population due to changes in the environment.

$$k_2 = \sqrt{\frac{(30-1)\left(1+\frac{1}{30}\right)(-1.96)^2}{15.57}} = 2.72 \quad (2)$$

For open intervals (for one-sided specifications), an approximation to the k factor is,

$$k_1 = \frac{(z_{1-p}) + \sqrt{z_{1-p}^2 - ab}}{a} , \quad (3)$$

where $a = 1 - \frac{z_{1-\gamma}^2}{2(n-1)}$ and $b = z_{1-p}^2 - \frac{z_{1-\gamma}^2}{n}$.

To achieve desired manufacturing yield goals, the test line limit is set, at a minimum, so that $TLL = \bar{x} \pm ks$, where k is set by either equation (1) or equation (3) for either two-sided or one-sided specifications. Equations (1) and (3) rely on data from a Gaussian distribution. Therefore, it is necessary to verify the Gaussian distribution assumption[3].

When choosing tolerance interval probability, p, and confidence, γ, it is necessary to consider several factors. Setting the confidence for a tolerance interval establishes the risk that the test line limit will contain less than the desired proportion of the population of manufactured product items. To illustrate, assume for a moment that a manufacturer produces a multi-parameter product that requires 50 independent measurements. If all 50 test line limits have been set with a 98 % confidence for the tolerance interval, it is expected that one of the test line limits is set too tight (i.e., 98 % of the 50 test line limits). The manufacture has one of several options in this case. The manufacturer may accept a lower production yield, loosen the test line limit and the corresponding specification, or modify the product or relevant processes. The tolerance interval confidence sets the risk that a manufacturer might need to take one of these actions.

For choosing the tolerance interval probability, a generally accepted minimum value is 95 %. However, manufacturers may choose a probability other than 95 % for different reasons. Consider again a multi-parameter product. Manufacturers wish to have high yields for the entire product so that the yield considering *all* parameters meets the respective test line limits. If the product parameters are statistically independent, the overall yield, in this case, is the product of the probability for each parameter. For a product with just three independent parameters, each with a test limit intended to give 95 % probability, the product would only have a $(95\ \%)^3$ or 85.7 % chance of meeting all test line limits, which is perhaps unacceptable to the manufacturer. For this reason, manufacturers select tolerance interval probabilities greater than 95 % so that the overall probability is acceptable. Furthermore, not all test line limits are necessarily based on the same probability. If a multi-parameter product has some parameters that perform very well with respect to market requirements, manufactures may choose to set wide test line limits for those parameters to virtually remove the possibility that they will present yield problems in manufacturing, and yet, still meet customer requirements.

5. Environmental Factors

Product specifications normally indicate an operating range for environmental conditions. For example, a specification may stipulate an operating temperature range of 0 °C to 55 °C. The implication being that the product meets the specification over this range. However, testing the product over the full environmental range, on an ongoing basis, is frequently cost prohibitive. Therefore, manufacturers may test over a restricted range of environmental conditions but rely on a guard band to ensure the product meets all specifications over the full operating temperature range.

A population of product items likely experiences a shift in the mean, a change in the standard deviation, or both, when operating under various environmental conditions. Figure 3 illustrates an example in which the mean shifts towards the specification and the standard deviation increases when comparing parameter performance measured in a worst-case environment against the environment used for production testing. In this case, the test environment is the environment in which normal production testing occurs while the worst-case environment is the environment with conditions causing the greatest shift in performance. Figure 3 denotes the change in performance due to environmental conditions as Δenv. Specifically,

---

[3] Histograms and quantile-quantile plots provide a graphical method for assessing the Gaussian assumption. Numerical tests, such as the D'Agostino's K-squared test or the Shapiro-Wilk test, are valuable and can be built into software for automating the analysis tasks.

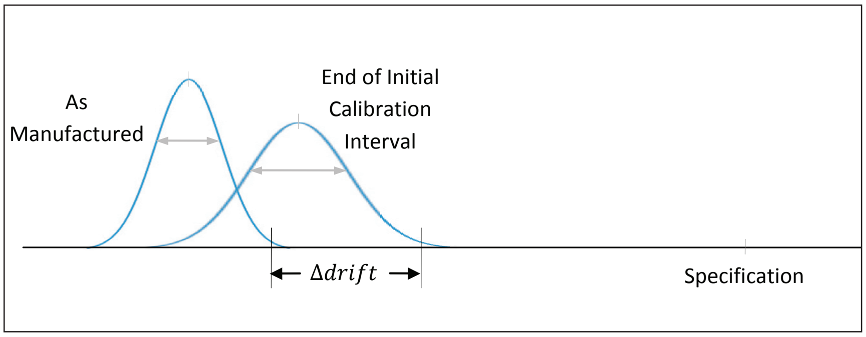


Figure 4. Changes in the population due to drift.

$$\Delta env = (\bar{x}_w + ks_w) - (\bar{x}_t + ks_t) = (\bar{x}_w - \bar{x}_t) + k(s_w - s_t), \tag{4}$$

where

Δenv = performance shift due to environmental conditions,

$\bar{x}_w$ = the sample mean from the worst-case environment,

$\bar{x}_t$ = the sample mean from the test environment,

s_w = the sample standard deviation from the worst-case environment,

s_t = the sample standard deviation from the test environment,

k = the *k factor* (determined by equation (1) for two-sided specifications, or by equation (3) for one-sided specifications).

Characterizing change in performance due to environmental conditions requires taking measurements while exposing product items to each environment. The *k factor*, in this example, is the same for the samples taken from the test environment and the worst-case environment. This implies that the proportion, confidence and, importantly, the number of samples from each environment are the same. Including *Δenv* as a component of the guard band ensures that products meet all specifications over the full range of environmental conditions.

6. Drift

So far, this discussion assumes the change in performance due to environmental conditions to be non-permanent. That is, when returned to the test environment, product performance is the same as it was prior to exposing the product to the worst-case environment. However, that is not necessarily the case. Stress due to environmental change, as well as everyday use, transport, aging and other factors may induce small changes in performance that accumulate over time. In other words, products drift. The effect of drift is that from the time of manufacture to the end of the initial calibration interval, it is likely that performance has shifted. This is illustrated in Fig. 4.

Similar to the impact of environmental conditions, a population of product items also experiences a shift in the mean, a change in the standard deviation, or both, due to the mechanisms associated with drift. Normally, drift is managed by setting an end-of-period reliability target, that is, the likelihood that a product is in tolerance (within specification) at the end of the calibration interval, and adjusting the calibration interval as necessary to achieve the target. Evaluation of the calibration interval is possible as calibration history builds. To ensure products meet specification over the initial calibration interval, manufacturers may include an additional guard band between the test line limit and the specification. It is possible to determine a value for *Δdrift* (shown in Fig. 4) as was done as for *Δenv* in equation (4). However, this may not be practical due the slow, accumulative characteristics associated with drift. For that reason, to estimate *Δdrift*, manufacturers may alternatively rely on accelerated aging tests, reliability data from similar products, engineering analysis or reliability models.

7. Measurement Uncertainty

Measurement uncertainty of the manufacturing test process adds variability to the measured test results. Measurement uncertainty also affects the process for setting the test line limits. The test line limit setting process attempts to characterize product-to-product variation. However, depending upon the extent of the data collecting exercise[4], the observed data from a sample of product items will also include variation from measurement uncertainty. The test line limits, when set based on observed data, account for all variation present while collecting data, including measurement uncertainty. Nonetheless, manufacturers may add additional guard band due to measurement uncertainty. Doing so allows margin for accommodating various guard banding requirements that may be applied during recalibration of a product.

For example, many calibration laboratories adhere to *ILAC-G8* [4], when reporting compliance to specifications. In this instance, compliance is stated when the measured value, plus the 95 % expanded uncertainty, does not extend past the specification. When the measured value plus the 95 % expanded uncertainty does extend past the specification, compliance is not stated. When possible, the customer expectation is that the product can be adjusted so that compliance can be stated. Including guard band to account for measurement uncertainty ensures that it is possible to adjust a product to meet specification, even when assessing compliance as stated by *ILAC-G8*.

8. Guard Band

The total guard band between the test line limit and the specifications is simply the sum of the individual guard band components

---

[4] Some sources of uncertainty are constant during the data collecting exercise, such as the uncertainty of a correction that is applied equally to all measurements. An uncertainty such as this does not contribute to the observed variation.

for environmental factors, drift, measurement uncertainty and customer guard band. For example,

$$spec = tll + \Delta env + \Delta drift + \Delta unc + gb_{customer}, \quad (5)$$

where Δunc is the measurement uncertainty component of guard band and $gb_{customer}$ is customer guard band. Note that the terms Δenv, $\Delta drift$, Δunc and $gb_{customer}$ are not standard deviations. Therefore, they cannot be combined in a root-sum-square fashion.

The customer guard band is any additional guard band the manufacture may deem convenient or necessary. For example, consider a product in which the performance varies somewhat over a range of conditions, yet, for clarity, the manufacturer chooses a single value for the specification over the entire range. This, in effect, adds guard band to the better performing regions in order to produce the desired specification. Consider as a second example where a manufacturing test procedure achieves a better measurement uncertainty than the procedure used for re-calibration. In this case, an added amount of guard band may be necessary to ensure end-of-period reliability objectives.

Equation (5) relates the test line limit, which is set based on product performance, to the specification for a product. Setting the test line limit with the flexibility of choosing tolerance interval proportion, confidence and sample size, allows the manufacturer to balance production yield and production costs with quoted performance and marketability of the product.

9. Using Specifications for Type-B Evaluation of Standard Uncertainty

The GUM provides guidance for evaluation of standard uncertainty and specifically includes manufacturer's specifications as a source of information for Type-B estimates. Relying on manufacturer specifications for determining measurement uncertainty is very convenient. That is, assuming a product meets its specifications (an assumption verified through calibration) and developing uncertainty based on those specifications is easily manageable. For this reason, it is common.

To evaluate a Type-B uncertainty, the GUM gives specific advice[5] when an uncertainty is quoted at a given level of confidence. In this instance, an assumption can be made that a Gaussian distribution was used to determine the quoted uncertainty. The standard uncertainty can then be determined by dividing by the appropriate factor given the stated level of confidence. Various manufacturers state a level of confidence[6] for product specifications and applying this GUM advice to product specifications quoted at a level of confidence is common and accepted by various accreditation bodies.

When a manufacturer states a level of confidence for a specification, this implies that the end-of-period measurement reliability for the specified parameter is at least the quoted level of confidence (e.g., 99 %). Claims of high measurement reliability are possible given a process for setting specifications designed for high manufacturing yields and employing guard bands for the reasons discussed in this paper. However, it is important to keep two points in mind when relying on a specification stated at a level of confidence. First, specification level of confidence applies to individual parameters and not to the overall measurement reliability of a multi-parameter product. Second, a manufacturer cannot guarantee an end-of-period reliability. The level of confidence claim can only mean that it is possible to achieve the stated end-of-period measurement reliability, but the responsibility for ensuring that is with the owner of the product. Specifically, the owner must operate the product consistent with the assumptions manufactures make regarding drift, must calibrate and adjust appropriately, and must monitor end-of-period reliability.

Without a statement of level of confidence, or other information for a specification, the GUM directs us to use the uniform distribution when determining standard uncertainties. Yet, doing so is likely a conservative estimate of the uncertainty. Regardless, the methods discussed in this paper produce product specifications for which high measurement reliability is readily attainable by the owner of the product.

10. Conclusions

Manufacturers set specifications on products to enable meeting manufacturing cost objectives while providing performance demanded by customers. This is accomplished by carefully characterizing product performance and relying on statistical analysis. The statistical analysis provides information that manufacturers publish on product data sheets. That information is provided as statements of confidence, supplemental data that accompanies the specifications or notes and statements applying to specific specifications. This information is a benefit to metrologists when using the product as a laboratory standard and for estimating overall test system accuracy. Regardless of the information provided, product specifications set using the methods discussed in this paper result in products capable of meeting or exceeding published specifications and that are maintainable over the product's useful life.

11. References

[1] "Guide to the Expression of Uncertainty in Measurement (GUM)," BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, *International Organization for Standardization (ISO)*, Geneva, Switzerland, 1995. [Available at www.bipm.org/en/publications/guides/gum.html]

[2] S.L. Read and T.R.C. Read, "Statistical Issues in Setting Product Specifications," *Hewlett-Packard Journal*, vol. 39, no. 3, pp. 6-11, June 1988. [Available at www.hpl.hp.com/hpjournal/pdfs/IssuePDFs/1988-06.pdf]

[3] NIST/SEMATECH, "e-Handbook of Statistical Methods," February 16, 2010. [Available at www.itl.nist.gov/div898/handbook/prc/section2/prc263.htm]

[4] *ILAC-G8*:03/2009, "Guidelines on the Reporting of Compliance with Specification," *International Laboratory Accreditation Cooperation (ILAC)*, 2009. [Available at www.ilac.org/guidanceseries.html]

---

[5] See subsection 4.3.4 of the GUM.

[6] Level of confidence for a specification is not the same as level of confidence for a tolerance interval. When setting test line limits, the confidence level of the tolerance interval relates to the risk of the tolerance interval not including the desired proportion of a population. Confidence level for a specification relates to measurement reliability.