Differences in Guard Banding Strategies

A Beginner's Guide

by Marcello Lucano, Agilent Technologies Italia S.p.A.

Introduction

Setting test limits different than specification limits influences the risk of accepting defective units and rejecting conforming units. Much has been written about setting limits [2,3,4,5,6,7,17] to accomplish various strategies to maintain controlled False-Accept Risk[1] and False-Reject Risk[2]. Different standards and guides [8,9,10,11,12,13,14,15,16] have been written to help the users and the calibration laboratories to set the correct guard band in order to allow a correct statement of compliance with specification. Aim of this paper is review the different guard banding strategies presented in the last years and help the reader understand some of the criteria to consider when selecting a guard band. This comparison must be intended as a *beginner's guide*. Definitely further technical aspects must be carefully deepened before apply any of the guard band approaches currently existent.

Confidence in measurements

Guard band and measurement uncertainty are some of the most challenging aspects of calibration. The measurement uncertainty is the one of the factors that mainly affects the risk of incorrectly declaring a device as in-tolerance or out-of-tolerance. As smaller is the uncertainty as higher is the confidence in the measurements and in assessment of compliance with specification. Historically, for electrical work and still for mechanical the recommended Test Uncertainty Ratio (TUR)[3] is equal to or greater than 4:1. This maintains False-Accept Risk and False-Reject Risk under relatively small probability. False-accept and false-reject occurrences have financial consequences, and therefore, minimizing both is often a worthwhile objective. Lower TURs increase the probability of these occurrences and lead to many issues. They can yield impossible demonstrate the compliance, or non-compliance, with a specification at given confidence level. Thus efforts should be applied to ensure TUR of at least 4:1 when possible, reducing the uncertainty, possibly by applying corrections to the result, using more accurate equipment or by taking the mean of a larger amount of readings. If this is not practical or desirable, then it may be possible to evaluate compliance, or non compliance, at a different level of confidence. The application of guard band could be another solution, even if it should really be considered a choice of last resort.

---

[1] False-Accept Risk is the probability that measuring an out-of-tolerance device will indicate an in-tolerance condition due to measurement error. In statistics terminology it is also called *Probability of a False Acceptance (PFA)* or *Consumers' Risk*, because it affects the quality of the measurements done with that equipment or the calibration service provided to the final consumer.

[2] False-Reject Risk is the probability that measuring an in-tolerance device will indicate an out-of-tolerance condition due to measurement error. This risk does not affect the users' quality since those items would not be released. This risk is also entitled *Probability of False Rejecting (PFR)* or *Producers' Risk*. It may be evaluated from the standpoint of additional costs stemming from such factors as unnecessary calibration and adjustment rework.

[3] The TUR, Test Uncertainty Ratio, is the ratio between the performance limit (specification of unit being tested) and the calculation of total uncertainty in the measurement considering all contributors. This term often differs significantly from the older Test Accuracy Ratio (TAR) which was simply the ratio of the performance limit (specification of unit being tested) to the specification(s) of the equipment being used to measure its performance. For example, a given measurement might have a TAR of 5:1 but a TUR of only 2.5:1.

Measurement Uncertainties & Guard Banding

Guard bands and measurement uncertainties are often confused because in many applications they are the same value. But fundamentally they are different issues. While the measurement uncertainty is a statistical limit on the probable error, the guard band is part of the calibration process designed to alter the results on a statistical basis. A guard band is the offset from the performance limit (specification) to the acceptance test limit that is used for the testing decision criteria.

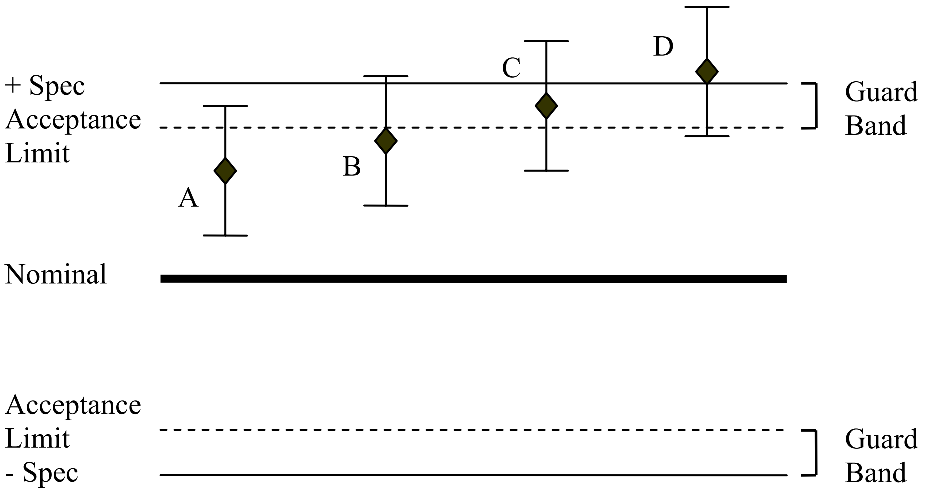
Use of a guard band has these purposes:

- To reduce the False-Accept Risk, which increases the confidence that an equipment really is in-specification by making allowance for measurement uncertainty.
- To provide confidence that the product will remain in-specification during the next calibration interval.

If a unit is inside the acceptance test limit, then it is deemed as passed. If it is outside the acceptance test limit (even if within the specification), it is deemed as having failed.

The issues of measurement uncertainty and guard banding are much easier to illustrate than to explain.[4]

In the drawing below there is a measurement uncertainty intervals shown above and below each measurement approximated at a 95% level of confidence. Also shown with dotted lines is an acceptance limit set by the guard band from the specification limit, illustrated slightly smaller than the uncertainty.



In this illustration measurements B, C, and D would be flagged to indicate that conformance or non-conformance cannot be stated at a 95% level of confidence. Measurements A and B represent a passed condition since the measured value falls within the acceptance limit. The measurements C and D represent a failed condition because they do not meet the acceptance limit and corrective action (e.g., adjustment) is required.

---

[4] There are many more potential conditions than are shown here, and in fact the entire set of potential conditions around the –Spec are left off. But the four choices shown are sufficient to illustrate the parameters being described here.

Choosing a guard band

Choosing a guard band is not easy. Fifty years ago many military programs would use a guard band of 25% of the performance limit. The reasoning was that the measurement was presumably designed to have a TUR of 4:1 or better, and thus such would provide an extremely high confidence that no faulty units would be passed as acceptable. But in the meantime it would provide a worse False-Reject Risk. Note that a TUR of 4:1 doesn't imply that the risk is 25 %.
Actually, assuming both the specification and uncertainty have Gaussian distribution at 95% confidence level, the False-Accept Risk for a TUR 4:1 is only 0.79 % and the False-Reject Risk is 1.5%. The values under the curves are probabilities.

The statistics involved to choose a specific percentage of the measurement uncertainty as a guard band to achieve a target False-Accept Risk and/or False-Reject Risk are complicated. It requires you to know the priori probability that both the calibration system and the device under test are in-tolerance. A good estimate of the a priori probability pay be difficult to obtain. Historical or device population information for estimating a priori probability may not be readily available and may not represent the specific device under test.

Given this difficulty various approaches are taken to approximate the desired results. Various methods for defining a guard band exist.

Guard band equal to uncertainty

A common practice demanded under accreditation schemes based on ISO17025:2005[1] is to use a guard band equal to the measurement uncertainty. This is suggested by ISO 14253-1[8], ILAC-G8[9], SIT/Tec-015/07[10] and comes from the old revisions of ISO Guide 25, drafts 5 and 6, which were explicit about claiming an in-tolerance condition:

> "When ... parameter(s) are claimed to be within specified tolerances, the measurement value(s), extended by the estimated uncertainty of measurement, shall fall within the appropriate specification limit."

Thus, the acceptance test limits (TL) are established by subtracting the uncertainty from the product's published specification (SL).

$$TL = 1 - \frac{1}{TUR} \times SL \qquad \text{Eq. (1)}$$

In the following table are reported different values of TL depending of the TUR

TUR	*SL*	*U*	*TL*
10 : 1	10	1.00	9.00
5 : 1	10	2.00	8.00
4 : 1	10	2.50	7.50
3 : 1	10	3.33	6.67
2 : 1	10	5.00	5.00
1.9 : 1	10	5.26	4.74
1.8 : 1	10	5.56	4.44
1.7 : 1	10	5.88	4.12
1.6 : 1	10	6.25	3.75
1.5 : 1	10	6.67	3.33
1.4 : 1	10	7.14	2.86
1.3 : 1	10	7.69	2.31

This method over-aggressively controls the False-Accept Risk while ignoring the calamitous results of a tremendously increased False-Reject Risk. This is well explained in the calculations of [3].

If we assume a STD[5] used to test product has a fourth of the uncertainty of the UUT[6] (TUR of 4) and we accept parts that measure in-tolerance and reject those that measure greater than the specifications (thus we let the acceptance test limit equal the specification limit) and assuming both, the specification of UUT and the uncertainty of the STD, have Gaussian distribution at 95% confidence level, as mentioned before, we can expect to accept 0.8% of bad parts and reject 1.5% of the good parts.

Now, if we test to the guard band equal to uncertainty, acceptance test limit 75% of the specification limit in this case, the false accepts are a mere 0.02%. But we could expect to reject 10% of our good products.

With TUR 2:1 and the guard band equal to uncertainty, than acceptance test limit 50% of the specification limit, the false accepts are 0.03% and false rejects increase to 33%.

This guard banding method may be appropriate for the aerospace industry where the additional cost of testing (rejecting good products) is offset by higher mission reliability, but for many other situations, the economic impact to the calibration supplier (and inevitably the customer) may outweigh the benefit of this practice. The dramatic cost increase resulting from rejecting many more good products could be hardly justified.

Besides ISO17025:2005[1], which superseded the ISO Guide 25, at par. 5.10.4.2 states:

> "When statements of compliance are made, the uncertainty of measurement shall be taken into account".

This is less explicit than before and allows different guard banding philosophies co-exist.

---

[5] STD, Standard with known value and associated uncertainty.

[6] UUT, Unit Under Test.

Probability of Compliance

This method is suggested in UKAS M3003-M2 and adopted from SIT/Tec-015/07 also. This allow to declare compliance, or non-compliance at different confidence levels. For 95% confidence level, this set the TL inside the SL by 1.6448σ. The formula to calculate the test limit is:

$$TL = SL - (u_c \times 1.6448) \qquad \text{Eq. (2)}$$

In the following table are reported different values of TL depending of the TUR

TUR	*SL*	*U*	u_c	*TL*
10 : 1	10	1.00	0.50	9.18
9 : 1	10	1.11	0.56	9.09
8 : 1	10	1.25	0.63	8.97
7 : 1	10	1.43	0.71	8.83
6 : 1	10	1.67	0.83	8.63
5 : 1	10	2.00	1.00	8.36
4 : 1	10	2.50	1.25	7.94
3 : 1	10	3.33	1.67	7.26
2 : 1	10	5.00	2.50	5.89
1.9 : 1	10	5.26	2.63	5.67
1.8 : 1	10	5.56	2.78	5.43
1.7 : 1	10	5.88	2.94	5.16
1.6 : 1	10	6.25	3.13	4.86
1.5 : 1	10	6.67	3.33	4.52
1.4 : 1	10	7.14	3.57	4.13
1.3 : 1	10	7.69	3.85	3.67

This method maintains minor False-Accept Risk and a lower False-Reject Risk than before. As suggested by M3003 this approach could be applied when the specification limits have been treated as absolute, analogous to a rectangular probability distribution.

When the specification is characterised by a normal distribution the UKAS M3003 in section M3 and SIT/Tec-015/07 suggest the following method.

RSS[7] Guard band

It is stated in the GUM[12] that, when a specification is quoted for a given coverage probability, then a normal distribution can be assumed. Some manufacturers state confidence levels for their pecifications. If both the uncertainty U and the specification SL are stated at the same coverage probability, then the acceptance test limits (*TL*) are established by the square root of the subtraction of the square uncertainty from the square product's published specification (*SL*).

---

[7] In statistical terminology this means *Root Sum (or Summed) Square*, the square root of the sum of the squares.
RSS = $\sqrt{}\ (x_1^2 + x_2^2 + x_3^2 \ldots + x_n^4)$. It is used to calculate the aggregate accuracy of a measurement when the accuracies of the all the measuring devices are known. The average accuracy is not merely the arithmetic average of the accuracies (or uncertainties), nor is it the sum of them.
Appling to this case, we can say that the specification is met when $\sqrt{}(Error\ Measured^2 + U^2) \leq SL$. This provides the same results of Eq. (3)

$$TL = \sqrt{SL^2 - U^2} \qquad \text{Eq. (3)}$$

Another RSS method reported in [5] is expressed by the formula:

$$TL = 1 - \frac{1}{TUR^2} \times SL \qquad \text{Eq. (4)}$$

This second one is much more conservative than the first RSS method, with a lower False-Accept Risk but an higher False-Reject Risk.

In the following table are reported different values of TL depending of the TUR.

TUR	SL	U	$TL = \sqrt{SL^2 - U^2}$	$TL = 1 - \frac{1}{TUR^2} \times SL$
10 : 1	10	1.00	9.95	9.90
9 : 1	10	1.11	9.94	9.88
8 : 1	10	1.25	9.92	9.84
7 : 1	10	1.43	9.90	9.80
6 : 1	10	1.67	9.86	9.72
5 : 1	10	2.00	9.80	9.60
4 : 1	10	2.50	9.68	9.38
3 : 1	10	3.33	9.43	8.89
2 : 1	10	5.00	8.66	7.50
1.9 : 1	10	5.26	8.50	7.23
1.8 : 1	10	5.56	8.31	6.91
1.7 : 1	10	5.88	8.09	6.54
1.6 : 1	10	6.25	7.81	6.09
1.5 : 1	10	6.67	7.45	5.56
1.4 : 1	10	7.14	7.00	4.90
1.3 : 1	10	7.69	6.39	4.08

Eq. (3) provides a fairly constant chance of under 0.7% of false acceptance for TURs from 4:1 to 1.5:1 although the chance of incorrect rejection is 2% at 4:1 rising to 8.2% at 2:1. Even if it is suggested when both the uncertainty and the specification are stated at the same coverage probability, this is also applicable when specification limits have been treated as absolute, with rectangular probability distribution.

Eq. (4) maintains a very small and severe False-Accept Risk but on the contrary it has a worse False-Reject Risk.

NCSL Recommended Practice 10

The method described in the NCSL-RP10[13] is expressed by the formula:

$$TL = 1.25 - \frac{1}{TUR} \times SL \qquad \text{Eq. (5)}$$

With this method the acceptance test limits change in function of the TUR as per table below.

TUR	*SL*	*U*	*TL*
4 : 1	10	2.50	10.00
3 : 1	10	3.33	9.17
2 : 1	10	5.00	7.50
1.9 : 1	10	5.26	7.24
1.8 : 1	10	5.56	6.94
1.7 : 1	10	5.88	6.62
1.6 : 1	10	6.25	6.25
1.5 : 1	10	6.67	5.83
1.4 : 1	10	7.14	5.36
1.3 : 1	10	7.69	4.81

This practice has an aggressive False-Accept Risk but high False-Reject Risk.

MIL-STD-45662A & ANSI Z540.1 Requirements

The old US military standard MIL-STD-45662A[14] and the old ANSI Z540.1[15] required that measuring and test equipment be calibrated with a system that is at least four times more accurate. That permitted limit can be equated to a quantifiable acceptable risk of reporting out-of-tolerance equipment as being in-tolerance. When system limitations prohibit this 4:1 ratio, guard band test limits can be computed and employed to present a risk no greater than that which would result from using a 4:1 ratio.

The acceptance test limit can be set reducing the specification limit by a portion of the measurement uncertainty so as to establish a False-Accept Risk equivalent of a 4:1 ratio.
This was analyzed in more detail in [2] by Bill Hutchinson.

The following table provides the multiplier *M* to be used to re-establish a similar False-Accept Risk as a TUR of 4:1.

TUR	*SL*	*U*	*M*	*TL*	*False-Accept Risk*
4 : 1	10	2.50	1.000	10.00	0.7893 %
3 : 1	10	3.33	0.975	9.75	0.7666 %
2 : 1	10	5.00	0.935	9.35	0.7755 %
1.9 : 1	10	5.26	0.930	9.30	0.7767 %
1.8 : 1	10	5.56	0.920	9.20	0.7685 %
1.7 : 1	10	5.88	0.915	9.15	0.7787 %
1.6 : 1	10	6.25	0.905	9.05	0.7777 %
1.5 : 1	10	6.67	0.895	8.95	0.7726 %
1.4 : 1	10	7.14	0.885	8.85	0.7761 %
1.3 : 1	10	7.69	0.875	8.75	0.7784 %

As mentioned before, the False-Accept Risk for a 4:1 TUR is 0.79 %. With the multiplier of the table above for TUR less than 4:1 the False-Accept Risk is maintained less than this value and also the False-Reject Risk is maintained under acceptable and reasonable limit.

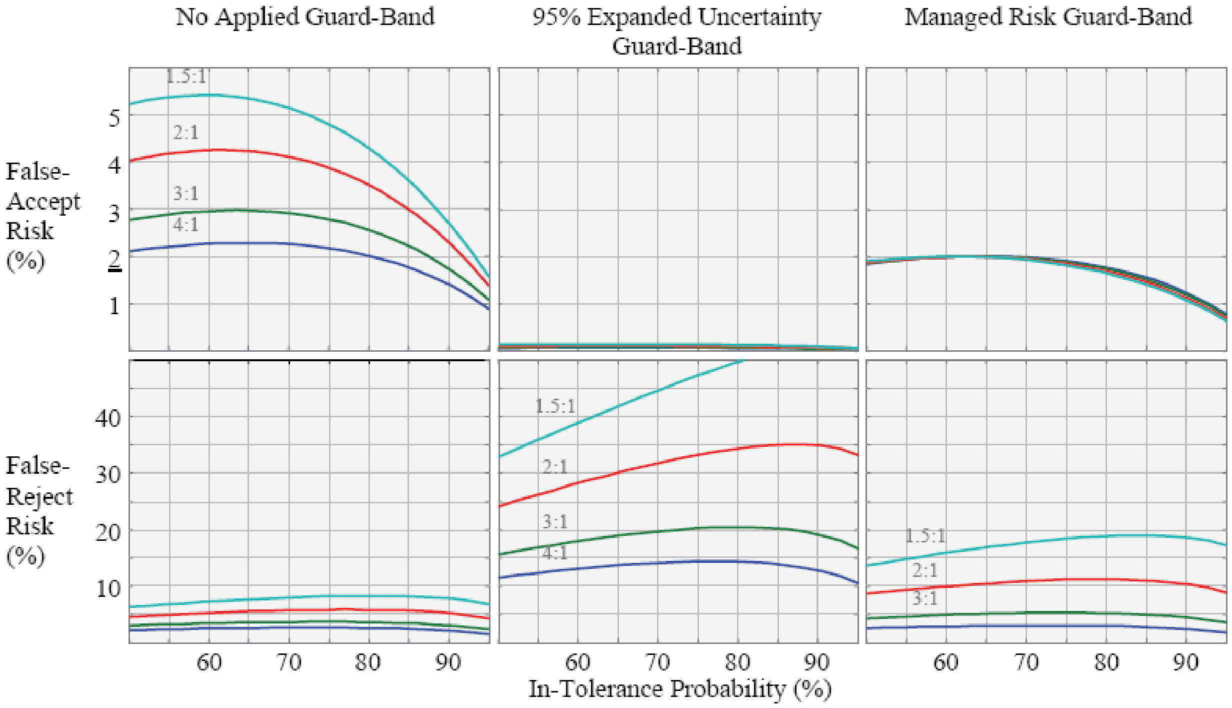
This is valid when both, the uncertainty U and the specification SL, are stated at the 95% coverage probability.

Managed risk

The new ANSI Z540.3 (2006)[16], which replaced the older Z540.1 standard, requires that a process be used that assures a False-Accept Risk of 2% or less.

Michael Dobbert of Agilent Technologies has recently presented two papers[6.7] where he analyzes in more detail the Z540.3 requirement. Using Monte Carlo simulation and statistical analysis he proposes one method of achieving it, which maintains the False-Accept Risk of 2% or less with only limited knowledge of the a priori probability that a device is in-tolerance and at the same time minimizes the False-Reject Risk.

As per following figure, with the guard band set to the 95% expanded uncertainty, the false-reject risk can be disproportionately high. With the managed risk method, the impact on False-Reject Risk is significantly less.



The conformity test limits to maintain the False-Accept Risk of 2% or less are established by the formula:

$$TL_{PFA2\%} = SL - U_{95\%} \times \left[1.04 - e^{(0.38 \cdot \log(TUR) - 0.54)}\right]$$ [8] Eq. (6)

With this method the conformity test limits change in function of the TUR as per table below.

TUR	*SL*	*U*	*TL*
4 : 1	10	2.50	9.87
3 : 1	10	3.33	9.48
2 : 1	10	5.00	8.59
1.9 : 1	10	5.26	8.44
1.8 : 1	10	5.56	8.27
1.7 : 1	10	5.88	8.08
1.6 : 1	10	6.25	7.85
1.5 : 1	10	6.67	7.60
1.4 : 1	10	7.14	7.30
1.3 : 1	10	7.69	6.95

---

[8] The log( ) function is the natural logarithmic function.

Conclusion

There are various methods for defining a guard band. All of them could be technically and metrologically correct in the proper context. The False-Accept Risk is reasonably insensitive to TUR while False-Reject Risk is sensitive and can be significantly different, lower or higher, depending of the guard banding technique chosen.

It is important to mention the difference between "unconditional" False-Accept Risk versus "conditional" False-Accept Risk. Unconditional False-Accept Risk, which is the one mentioned in this paper, is average False-Accept Risk for a population of calibrated devices. It is an appropriate statistic to consider while managing more than a single instrument. However, a conditional False-Accept Risk is appropriate when attention is given to a specific instrument and it is usually a statement about the False-Accept Risk given a measured value. The larger guard bands tend to be based on the conditional False-Accept Risk perspective, while smaller guard bands may assume populations of devices.

Taking above into account it would says the 95% expanded uncertainty guard band (ILAC-G8 approach), M3003-M2, RSS and RP-10 methods tend to be based on the conditional False-Accept Risk perspective while Managed Risk, M3003-M3 and Hutchinson's methods may assume populations of devices.

The 95% expanded uncertainty guard band is the most conservative for the False-Accept Risk but completely ignores the False-Reject Risk.

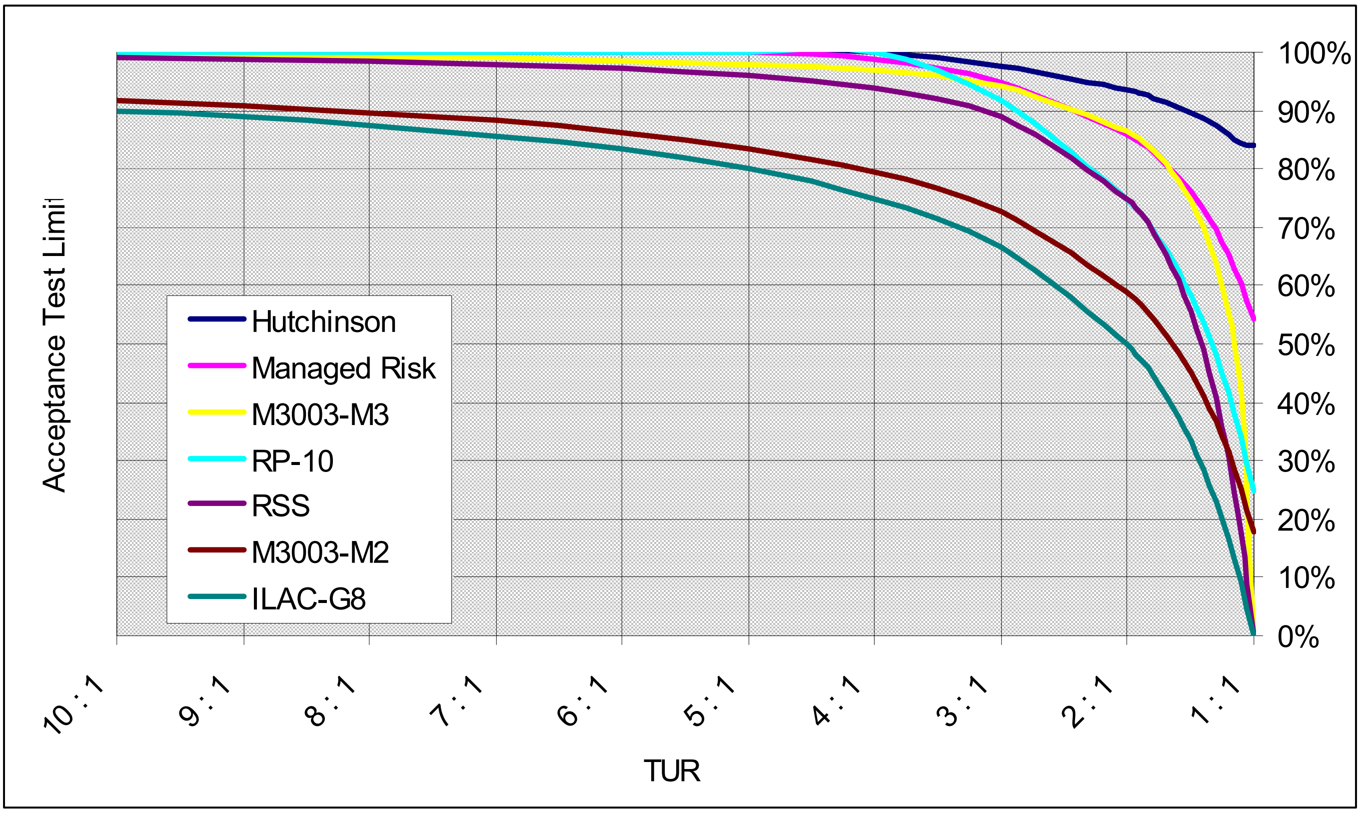
M3003-M2, RSS and RP-10 methods provide a better False-Reject Risk than 95% expanded uncertainty guard band but still high if compared with the other methods.

Managed Risk and M3003-M3 methods provide similar results with a controlled False-Accept Risk and a False-Reject Risk significantly lower than 95% expanded uncertainty guard band. Both can be applied either with specification quoted as absolute limit or when specification are stated with normal distribution.

Hutchinson's method provides the best False-Reject Risk but the worst False-Accept Risk even if it respects the Z540.3 2% risk requirement. It could be indicated when the specification are quoted with normal distribution and at the same 95% confidence level of the uncertainty.

Whichever of the presented approaches is used should, of course, be done with the agreement and understanding of the customer/user depending of the criticality and reliability of the application and of the financial consequences of uncontrolled False-Reject Risk.

TUR	*SL*	*U*	*TL*						
			ILAC-G8	*M3003 M2*	*M3003 M3*	*RSS*	*RP-10*	*Hutchinson's Method*	*Managed Risk*
			Eq. (1)	*Eq. (2)*	*Eq. (3)*	*Eq. (4)*	*Eq. (5)*		*Eq. (6)*
10 : 1	10	1.00	9.00	9.18	9.95	9.90	10.00	10.00	10.00
5 : 1	10	2.00	8.00	8.36	9.80	9.60	10.00	10.00	10.00
4 : 1	10	2.50	7.50	7.94	9.68	9.38	10.00	10.00	9.87
3 : 1	10	3.33	6.67	7.26	9.43	8.89	9.17	9.75	9.48
2 : 1	10	5.00	5.00	5.89	8.66	7.50	7.50	9.35	8.59
1.9 : 1	10	5.26	4.74	5.67	8.50	7.23	7.24	9.30	8.44
1.8 : 1	10	5.56	4.44	5.43	8.31	6.91	6.94	9.20	8.27
1.7 : 1	10	5.88	4.12	5.16	8.09	6.54	6.62	9.15	8.08
1.6 : 1	10	6.25	3.75	4.86	7.81	6.09	6.25	9.05	7.85
1.5 : 1	10	6.67	3.33	4.52	7.45	5.56	5.83	8.95	7.60
1.4 : 1	10	7.14	2.86	4.13	7.00	4.90	5.36	8.85	7.30
1.3 : 1	10	7.69	2.31	3.67	6.39	4.08	4.81	8.75	6.95



References

[1] ISO17025:2005, *General requirements for the competence of testing and calibration laboratories,* International Organization for Standardization, 2005

[2] Hutchinson, Bill, *"Setting Guardband Test Limits to Satisfy MIL-STD-45662A Requirements"*, 1991 NCSL International Workshop & Symposium

[3] Deaver, David K., *"How to Maintain Your Confidence (In a World of Declining Test Uncertainty Ratios)"*, 1993 NCSL International Workshop & Symposium http://assets.fluke.com/appnotes/Calibration/ddncsl93.pdf

[4] Deaver, David K., *"Managing Calibration Confidence in the Real World"*, 1995 NCSL International Workshop & Symposium http://assets.fluke.com/appnotes/Calibration/ddncsl95.pdf

[5] Deaver, David K., *"Guardbanding and the World of ISO Guide 25. Is There Only One Way?"*, 1998 NCSL International Workshop & Symposium http://assets.fluke.com/appnotes/Calibration/ddncsl98.pdf

[6] Dobbert, Michael,*"Understanding Measurement Risk"*, 2007 NCSL International Workshop & Symposium http://www.agilent.com/metrology/ncsli2007_dobbert.shtml

[7] Dobbert, Michael, *"A Guard-Band Strategy for Managing False-Accept Risk"*, 2008 NCSL International Workshop & Symposium http://www.agilent.com/metrology/ncsli2008_dobbert.shtml

[8] ISO14253-1:1998, *Geometrical Product Specifications (GPS) -- Inspection by measurement of workpieces and measuring equipment -- Part 1: Decision rules for proving conformance or non-conformance with specifications,* International Organization for Standardization, 1998

[9] ILAC-G8:1996, *Guidelines on Assessment and Reporting of Compliance with Specification*, International Laboratory Accreditation Cooperation, 1996 http://www.ilac.org/documents/g8_1996.pdf

[10] SIT/Tec–015/07, *Strutture di riferibilità e documentazione tecnica nei laboratori di taratura elettrici*, Servizio di Taratura in Italia, 2007, pp 24-31 http://www.sit-italia.it/SIT/Documenti/Documenti_Tecnici/SIT_Tec_015_07.pdf

[11] M3003:2007, *The Expression of Uncertainty and Confidence in Measurement*, United Kingdom Accreditation Service, 2007, pp 69-73 http://www.ukas.com/Library/downloads/publications/M3003.pdf

[12] *Guide to the Expression of Uncertainty in Measurement* (GUM), BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML – , International Organization for Standardization, 1995

[13] NCSL RP-10, *Establishment and Operation of an Electrical Utility Metrology Laboratory*, National Conference of Standard Laboratories International, 1991.

[14] *Military Standard 45662A - Calibration System Requirements*, United States of America Department of Defense, 1988

[15] ANSI/NCSL Z540.1-1994, *Calibration Laboratories and Measuring and Test Equipment - General Requirements*, National Conference of Standard Laboratories International, 1994

[16] ANSI/NCSL Z540.3-2006, *Requirements for the Calibration of Measuring and Test Equipment*, National Conference of Standard Laboratories International, 2006.

[17] Uncertainty & Confidence in Measurements http://metrologyforum.tm.agilent.com/uncert.shtml