Techniques for Optimizing Calibration Intervals[1]

AUTHORS

Howard Castrup, Ph.D.
President, Integrated Sciences Group

Dr. Castrup is a leading authority in the fields of calibration interval analysis, measurement uncertainty analysis and measurement decision risk analysis. Dr. Castrup designed and built the world's first large scale automated calibration interval analysis system and has authored numerous articles and technical reports on the subject of uncertainty growth control and management. In addition, he developed the measurement decision risk methodology of the Navy's Equipment Tolerancing System (ETS) and is the author of the statistical measurement process control (SMPC) method for evaluating measurement error and uncertainty based on system output. Dr. Castrup earned his B.S. and Ph.D. in Engineering from UCLA.

Kirt M. Johnson
Chief Scientist, American Systems Engineering Corporation

Mr. Johnson has almost twenty years experience in Navy test and calibration. He has participated in the evaluation of factory test systems in support of Navy acquisition efforts and has contributed to a number of major ATE systems, covering hardware, system software, and test program sets for NAVAIR, NAVSEA and the Marine Corps. His involvement in the METCAL Program covers more than a decade, where he has focused on equipment tracking and reporting systems, particularly in relation to automated calibration interval analysis. Mr. Johnson has a B.S. in Engineering from UCLA.

ABSTRACT

Periodic calibration comprises a significant cost driver in the life cycle of Navy test and calibration equipment. It also provides a major safeguard in controlling uncertainty growth and reducing the risk of substandard weapon system performance in the Fleet.

Concepts central to calibration interval analysis and management are described. Guidelines and techniques are presented that permit optimizing intervals with respect to both life cycle support costs and costs due to suboptimal weapon system performance. Special focus is given to mathematical reliability modeling methods and to calibration history data management requirements.

INTRODUCTION

Why calibrate?

We calibrate to control measurement errors and uncertainties to "acceptable" levels. For calibration or test equipment, acceptable levels of uncertainty are defined by the tolerance limits of the equipment's calibrated parameters.

MEASUREMENT DECISION RISK

Since measurements have associated with them a degree of uncertainty, there is always a chance that the value of an out-of-tolerance weapon system parameter will be measured by a test system as being within tolerance specifications. The probability that this will happen is called *false accept risk*. False accept risk results in out-of-tolerance items being returned to and used by the Fleet.

Likewise, there is a chance that the value of an in-tolerance weapon system parameter will be measured as being outside tolerance specifications. The probability for this is called *false reject risk*. False reject risk causes unnecessary rework, maintenance or repair.

False accept and false reject risks, taken together are referred to as *measurement decision risk*.

There are a number of ways to control measurement decision risk [1-3]. In brief, suffice it to say that the smaller the measurement uncertainty relative to the tolerance limits of a parameter being measured, the lower the measurement decision risk (other factors being equal).

CONTROLLING DECISION RISK

Weapon system Testing Quality

A major component of measurement uncertainty is *measurement bias* uncertainty. In the context of this paper, the measurement bias of a parameter is the systematic difference between the actual value of the parameter and its stated or declared value. The stated or declared value is commonly referred to as the *nominal* value.

It can be shown [4,5] that the greater the decision risk in a calibration measurement the wider the spread of accepted test system parameter biases around the nominal value. What this is equivalent to saying is the greater the calibration measurement decision risk, the more likely it is to find out-of-tolerance test system parameters slipping through the calibration process.

Similarly, the greater the measurement decision risk during testing, the more likely it is to find out-of-tolerance weapon systems slipping through the testing process. This assertion has been corroborated by the Navy's Equipment Tolerancing System (ETS) [4,5].

The end result of accepting out-of-tolerance items is a potential loss of weapon system or other capability in the Fleet due to individual parameters functioning outside their tolerance limits. This potential loss of capability is particularly dangerous, since out-of-tolerances are not readily perceived by equipment users.

[1] Presented at the ASNE Test & Calibration Symposium, Arlington, December 1994.

Equipment will *seem* to be functioning properly, when in fact it is not.

In the modern arena of technologically intensive warfare, fielding systems with out-of-tolerance parameters can seriously jeopardize the success of Fleet missions or even the survival of Fleet components. Obviously, it is not in the best interest of National security to minimize the importance of controlling and managing measurement uncertainty.

Controlling Measurement Uncertainty

In the majority of cases that we have analyzed over the last seventeen years, measurement bias has appeared as the major contributor to measurement uncertainty, dominating random error and other measurement process errors. Accordingly, maintaining a tight control on measurement bias goes a long way to controlling total measurement uncertainty.

A measure of the control of a parameter's measurement bias is its *in-tolerance probability*. In the parlance of measurement decision risk analysis, this is referred to as *measurement reliability*. The higher the measurement reliability, the lower the measurement bias uncertainty and the lower the measurement decision risk.

Reliability Targets

It can be readily appreciated that measurement reliability is a useful variable to track in controlling measurement bias. As with other technical parameters, we first establish control of measurement reliability by setting an objective or *target*.

Table 1. Weapon system Test Decision Risk vs. Test Parameter Measurement Reliability. Test Accuracy Ratio = 10:1.

End Item EOP Reliability = 95%

Test System Measurement Reliability (%)		Measurement Decision Risk (%)	
EOP	**AOP**	**False Accept**	**False Reject**
50	70.52	0.70	1.10
60	77.24	0.62	0.93
70	83.42	0.56	0.79
80	89.20	0.49	0.66
85	91.97	0.46	0.60
90	94.68	0.42	0.54
95	97.34	0.37	0.46
99	99.46	0.30	0.36

End Item EOP Reliability = 70%

Test System Measurement Reliability (%)		Measurement Decision Risk (%)	
EOP	**AOP**	**False Accept**	**False Reject**
50	70.52	1.72	1.96
60	77.24	1.51	1.69
70	83.42	1.32	1.46
80	89.20	1.15	1.25
85	91.97	1.06	1.14
90	94.68	0.96	1.03
95	97.34	0.84	0.90
99	99.46	0.68	0.71

Tables 1 and 2 show false accept risks for weapon system testing as a function of measurement reliability target. These tables show false accept and false reject risks for cases where weapon systems are received for testing with end-of-period (EOP) in-tolerance probabilities of 95% and 70%. Test system in-tolerance probabilities are keyed to average-over-period (AOP) levels corresponding to various EOP targets. As expected, the lower the test system parameter measurement reliability, the higher the false accept and false reject risks.[2]

Table 2. Weapon system Test Decision Risk vs. Test Parameter Measurement Reliability. Test Accuracy Ratio = 2:1.

End Item EOP Reliability = 95%

Test System Measurement Reliability (%)		Measurement Decision Risk (%)	
EOP	AOP	False Accept	False Reject
50	70.52	1.77	12.00
60	77.24	1.69	9.50
70	83.42	1.60	7.55
80	89.20	1.49	5.92
85	91.97	1.43	5.16
90	94.68	1.36	4.41
95	97.34	1.26	3.57
99	99.46	1.10	2.55

End Item EOP Reliability = 70%

Test System Measurement Reliability (%)		Measurement Decision Risk (%)	
EOP	AOP	False Accept	False Reject
50	70.52	6.55	11.84
60	77.24	5.95	10.14
70	83.42	5.38	8.55
80	89.20	4.81	7.20
85	91.97	4.50	6.54
90	94.68	4.15	5.83
95	97.34	3.71	5.00
99	99.46	3.06	3.88

SETTING RELIABILITY TARGETS

Reliability targets can be set on the basis of acceptable levels of measurement decision risk. The question remains, what constitutes an acceptable level? To answer this question, we need to consider how measurement decision risk impacts weapon system utility, how much it costs when desired levels of utility are not maintained, and what it costs in terms of test and calibration support to achieve these levels.

This issue is dealt with by the Navy's ETS and elsewhere [1,4,5]. In particular, the model incorporated in ETS permits a determination of weapon system utility and costs associated with weapon system performance in scenario-specific contexts. In the model, "performance" costs are balanced against test and calibration support costs in such a way that total cost can be minimized. Performance costs include such considerations as cost of hardware, cost of mission loss, liability incurred in hitting the wrong targets, etc. Support costs include equipment cost, labor cost, downtime cost, parts cost, etc.

---

[2]Tables 1 and 2 were developed using ETS.

A byproduct of the cost minimization process is the determination of optimal reliability targets.

UNCERTAINTY GROWTH

As stated earlier, reliability targets provide objectives for managing measurement bias uncertainty. This uncertainty is not a static quantity. It begins to grow from the time of test or calibration and increases throughout the test or calibration interval.

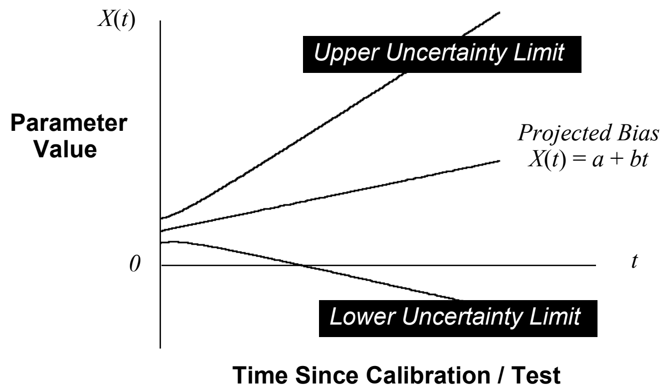

Figure 1. Bias Uncertainty Growth. The case shown is for a parameter whose value drifts linearly with time. Because of irregularities in usage, environment and other random variables, the uncertainty in the projected bias also changes with time, as shown by the upper and lower uncertainty limits.

Controlling Uncertainty Growth

When the bias uncertainty of a parameter has grown to the point where the in-tolerance percentage is equal to the reliability target, it is time to recall the parameter for re-testing or re-calibration.

Stated mathematically, if we represent the measurement reliability for a parameter by $R(t)$ and the reliability target by R^*, the parameter should be re-tested or re-calibrated when

$$R(t) = R^* . \qquad (1)$$

The occurrence of this event will not be detectable to the user of the item. The time of occurrence of the out-of-tolerance event can, however, often be predicted by inferring the interval of time required for Eq. (1) to hold.

This is accomplished by choosing a mathematical model to represent the function $R(t)$ and fitting this model to observed test or calibration history data [6-8]. Once the fit is achieved, the variable t can be solved for.

RELIABILITY MODELING AND ANALYSIS

Mathematical Modeling

Several models have been found that are useful for modeling uncertainty growth over time [1,7,8]. Each model is described in terms of a mathematical function, characterized by a set of coefficients. For example, the exponential model is characterized by the parameters R_0 and λ:

$$R(t) = R_0 e^{-\lambda t}; \qquad (2)$$

while the Weibull model is characterized by the coefficients R_0, λ and β:

$$R(t) = R_0 e^{-(\lambda t)^\beta}. \qquad (3)$$

References 1 and 7 describe eight reliability models that have been found useful for calibration interval analysis. These are shown in Table 3.

Table 3. Applicable Reliability Models

MODEL	APPLICABILITY
Exponential	Parameters whose out-of-tolerance rates are constant with time.
Weibull	Parameters whose out-of-tolerance rates increase or decrease with time, due to wear out or burn in.
Mixed Exponential	Items or parameter composed of several components, each of which has a constant out-of-tolerance rate.
Random Walk	Parameters whose out-of-tolerances occur as a result of random fluctuations in parameter value.
Restricted Random Walk	Parameters whose out-of-tolerances occur as a result of random fluctuations that are constrained within definable limits.
Modified Gamma	Parameters whose out-of-tolerances occur as a result of stresses accumulated over time.
Mortality Drift	Parameters whose out-of-tolerance rates increase or decrease monotonically with time.
Warranty	Parameters that tend to go out-of-tolerance within a narrow time interval.

Analyzing Calibration/Test History Data

In modeling measurement reliability vs. time, a method that has proved to be especially powerful is the method of *maximum likelihood estimation*. With this method, the coefficients of a selected model are solved for in such a way that the probability of observing the particular history obtained for the parameter of interest is maximized.

This involves assembling test or calibration history data in a time series of the form

Time Since Measurement	Number Tested or Calibrated	Number In-Tolerance
t_1	n_1	g_1
t_2	n_2	g_2
.	.	.
.	.	.
.	.	.
t_k	n_k	g_k

The tabulated data are then portrayed as a plot of % observed in-tolerance vs. time since test or calibration. Next, the mathematical reliability model is fit to the plotted points using the maximum likelihood method. An example of such a fit is shown in Figure 2.

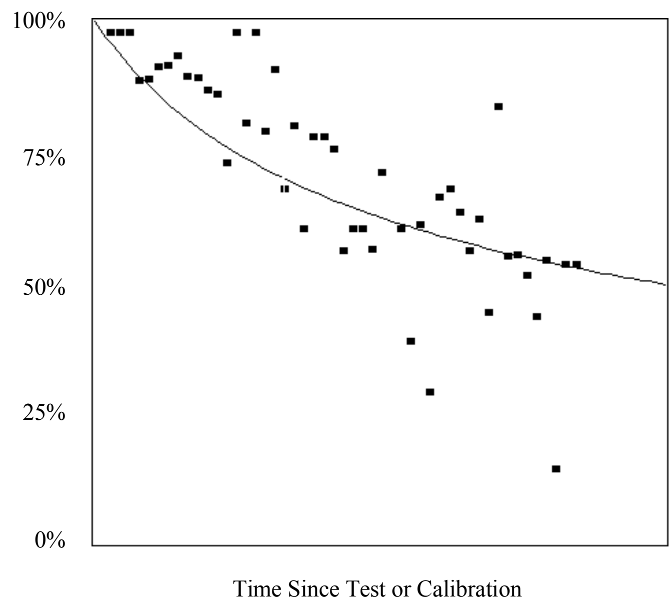

Figure 2. A Measurement Reliability Time Series. The observed time series is represented by the filled squares. The reliability model is represented by the curve.

It should be mentioned that, although fitting a reliability model to data, such as is done in Figure 2, appears graphically simple, it actually involves a complex mathematical process that has been developed over years of research [9-12]. This research is ongoing and promises to yield new methods that will be increasingly adaptable to desktop computing workstations.

USING THE CORRECT MODEL

Two major steps in arriving at an optimal test or calibration interval are (1) selecting an appropriate reliability target and (2) using maximum likelihood methods to fit reliability models to observed time series. A third major step involves selecting the best available reliability model.

The importance of this third step cannot be minimized. Because inventories of test equipment represent a variety of measurement disciplines, design approaches and fabrication technologies, no single reliability model is appropriate for all types of equipment [1,7,8].

To see that this is so, consider Figure 3. Figure 3 shows the output of an automated interval analysis program [14]. The inner curve is a fit of the exponential model to the data. The middle curve is a fit of a model referred to as the modified gamma model and the outer curve is a fit of the well known Weibull model. For the case shown, the Weibull model is the correct one.

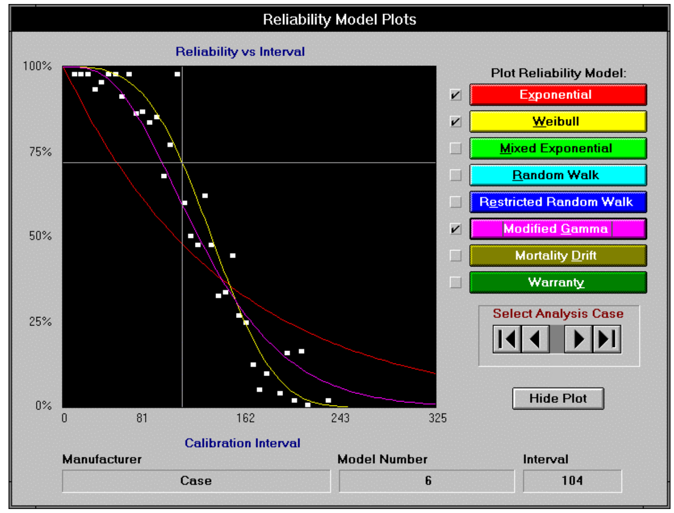

Figure 3. Maximum Likelihood Reliability Model Fits. The inner curve is the exponential model, the middle curve is the modified gamma model, and the outer curve is the Weibull model. The horizontal line is the reliability target, and the vertical line is the correct interval. Even though state-of-the-art curve fit methods were used for all three models, only one yielded the correct interval.

Notice that, although all models were fit to the same data using state-of-the-art maximum likelihood methods, the intervals recommended by each are different:

Computed Intervals (Reliability Target = 85%)

Exponential Model	Modified Gamma Model	Weibull Model
102 days	312 days	373 days

If we had complacently assumed that the exponential model, for example, was applicable to the case in question, we would have been off by more than a factor of three. This means that we would have set an interval that caused us to calibrate the item in question three times more often than necessary, with three times the support cost and one third the availability.

SELECTING THE CORRECT MODEL

A method that has been found to be effective in selecting the correct reliability model (or one close to it) is the method documented in references 1 and 7. With this method, each model of a set of reliability models is fit to test or calibration history data taken on a given item. The "quality" of the fit is then evaluated, with the model with the highest quality of fit being selected as the correct one.

As is discussed in References 1 and 7, the quality of fit is in part a function of the statistical "goodness of fit," tempered by a decision algorithm that evaluates "representativeness of fit" and factors in economic considerations.

A modification of this approach has recently been implemented [14]. The modified method has been successfully tested against simulated calibration history featuring a variety of uncertainty growth mechanisms.

MANAGING CALIBRATION DATA

As was discussed earlier, the determination of a calibration interval involves the analysis of data arranged as observed percent in-tolerance vs. time since calibration or test (see Figure 2). Such data are assembled from recorded results of calibration or testing, organized into "calibration histories." A calibration history consists of an unbroken sequence of calibration or testing results accompanied by the date of service for each service action.

Since measurement decision risk is encountered at the equipment parameter level, an ideal calibration history would be one that is maintained for each parameter of interest. Until automated or real-time desktop calibration procedures become more widely proliferated, however, it is not economically feasible to maintain histories by parameter. At present, the best that can be expected is the maintenance of calibration histories at the equipment serial number level.

With the maximum likelihood method of analysis, histories for individual serial numbered items are pooled into homogeneous groupings, usually at the manufacturer/model level. The pooled data are then organized into successive windows of time. In each time window, the number observed in-tolerance is divided by the number calibrated to arrive at the observed percent in-tolerance.

The observed percents in-tolerance for the time windows are arranged chronologically in a time series. Maximum likelihood methods are then applied to the time series to select the appropriate reliability model and to calculate the optimal calibration interval for the homogeneous grouping.

Certain data management requirements must be met to ensure that this process produces a correct interval. The minimum requirements are summarized in Table 4. Additional requirements that pertain to maximizing the potential of interval analysis for life cycle cost management are described in References 7 and [13].

In Table 4, the term "condition received" refers to whether the item was in-tolerance when received for calibration. Since calibration interval analysis attempts to set optimal recall schedules to prevent excessive use of out-of-tolerance items, it is important to be able to distinguish whether or not an out-of-tolerance condition is user detectable.

Table 4. Test/Calibration Interval Analysis Minimum Data Requirements

REQUIREMENT	DESCRIPTION
Continuity	Calibration histories should be free of missing service actions. If missing service actions are present, they should be detectable.
Completeness	Each record should provide all information necessary for analysis. This information includes as a minimum (1) identification of the item serviced, (2) any special usage classification or designation, (3) date of service, (4) condition received prior to adjustment or other corrective service, (5) service action taken, (6) condition released.
Consistency	Each record in a serial numbered item's calibration history should reflect uniformity with respect to parameters calibrated, tolerances used, procedure used, etc. If this is not the case, then the observed time series is contaminated and the resulting interval will be suboptimal.

Note that the term "condition released" refers to whether the item was in-tolerance when returned to service. Apart from the effects of false accept risk, it is usually assumed that items are returned to service in an in-tolerance condition.

CONCLUSION

We have seen that the ingredients of an optimal interval are the following

Table 5. Ingredients for Optimal Intervals

INGREDIENT	REQUIREMENT
Test/calibration data quality	As a minimum, data must meet the requirements of Table 4. Experience shows that investments in data quality are returned many times over in terms of reduced support and performance costs.
Appropriate reliability targets	Targets should be adjusted to minimize total cost. Analysis of targets takes into account both support costs and costs due to false accept risk.
Maximum likelihood analysis	Data should be analyzed using maximum likelihood or equivalent methods to best fit reliability models to test/calibration history time series.
Appropriate reliability models	Reliability models should be used that represent actual uncertainty growth behavior. No single model is appropriate for modeling all cases.

It cannot be over stressed that, if any one of these ingredients is missing, the resulting interval will be suboptimal. Suboptimal intervals correspond to those that fail to meet reliability targets. The consequences of suboptimal intervals are shown in Table 6.

Table 6. Consequence of Suboptimal Intervals

PROBLEM	CONSEQUENCE
Short Interval	High calibration cost, cost of downtime, logistics problems.
Long Interval	High measurement decision risk, reduced probability of acceptable weapon system performance, potential high performance cost.

It should be remarked that, until recently, it has been virtually impossible to quantify the cost of a long interval. Since the advent of ETS [4] and other tools [2], however, we can now evaluate such costs, placing periodic calibration on a return on investment footing. When the investment is made to optimize calibration intervals, the return is the assurance that Fleet readiness will not be compromised by poor measurement reliability.

REFERENCES

[1] NASA Reference Publication 1342, *Metrology — Calibration and Measurement Processes Guidelines*, Jet Propulsion Laboratory, June 1994.

[2] *AccuracyRatio* 1.1, Integrated Sciences Group, Copyright 1992.

[3] Deaver, D., "How to Maintain Your Confidence," *Proc. NCSL 1993 Workshop and Symposium*, Albuquerque, July, 1993.

[4] Navy Metrology Research & Development Program Technical Report, *ETS Methodology*, Dept. of the Navy, Metrology Engineering Center, NWS, Seal Beach, March 1989.

[5] Castrup, H., "Calibration Requirements Analysis System," *Proc. NCSL 1989 Workshop and Symposium*, Denver, July 1989.

[6] Glassman, J., "Intervals by Exception," *Proc. NCSL 1970 Workshop and Symposium*, Gaithersburg, October, 1970.

[7] National Conference of Standards Laboratories Recommended Practice RP-1, *Establishment and Adjustment of Calibration Intervals*, November, 1989.

[8] Larsen, J., "Ageless Periodic Recall Models," NAVAIR 17-35AMN93-1, February 1993.

[9] Jennrich, R. and Sampson, P., "Application of Stepwise Regression to Non-Linear Estimation," *Technometrics*, Vol. 10, No. 1, 1968.

[10] TRW Equipment Recall Optimization System (EROS), TRW DSSG, Redondo Beach, 1978.

[11] Jackson, D. and Castrup, H., "Reliability Analysis Methods for Calibration Intervals: Analysis of Type III Censored Data," *Proc. NCSL 1987 Workshop and Symposium*, Denver, July 1987.

[12] Morris, M., "A Sequential Experimental Design for Estimating a Scale parameter from Quantal Life Testing Data," Technometrics, Vol. 29, No. 2, May 1987.

[13] Wyatt, D. and Castrup, H., "Managing Calibration Intervals," Proc. NCSL 1991 Workshop and Symposium, Albuquerque, August 1991.

[14] IntervalMAX, Copyright 1994, Integrated Software Associates.