NASA
Reference
Publication
1342

1994

Metrology — Calibration and Measurement Processes Guidelines

NASA Metrology and Calibration Working Group

NASA
Reference
Publication
1342

June 1994

Metrology — Calibration and Measurement Processes Guidelines

H. T. Castrup
W. G. Eicke
J. L. Hayes
A. Mark
R. E. Martin
J. L. Taylor

JPL
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

NASA
National Aeronautics and
Space Administration

Office of Management

Scientific and Technical
Information Program

Foreword

This Publication is intended to assist in meeting the metrology requirements of National Aeronautics and Space Administration (**NASA**) Quality Assurance (**QA**) handbooks by system contractors. The Publication is oriented to mission-imposed requirements generated by long-term space operations. However, it is equally valid for use in all NASA program

Acknowledgements

The principal authors of this Publication are:

HOWARD T. CASTRUP — is the principal of Integrated Sciences Group (ISG), a company involved in designing and developing computer-hosted scientific analysis and advanced decision support systems. Castrup is a leader in calibration interval analysis and his pioneering research in test/calibration decision analysis. He is the author of a statistical process control methodology which permits the determination of in-tolerance probabilities for precision measuring and test equipment without the use of higher-level intercomparison standards. He earned his BS and PhD in Engineering from UCLA with concentration in solid state electronics.

WOODWARD G. EICKE — is a consulting metrologist in the field of electrical measurements, standards and instrumentation. His experience includes over thirty-five years with the National Bureau of Standards (now NIST) in the areas of precise electrical measurements, electrical standards, instrumentation, automation, measurement assurance, and other related fields. Eicke is author of more than twenty published papers for scientific and technical journals, and has served on numerous professional society and NBS committees and participated in standards writing. He attended George Washington University where he earned his BS and MS in Engineering.

JERRY L. HAYES — is the principal of Hayes Technology, an engineering consultant firm. He has provided metrology and calibration program consulting services for several aerospace companies and the Department of Defense. He served as Technical Director of the Navy Metrology Engineering Center and established policies and objectives for Navy-wide programs. He has authored numerous papers regarding calibration and measurement controls for quality in calibration programs and tests. Hayes has been the recipient of many awards and honors from peer groups. He earned a BS in Mechanical Engineering from the University of California at Berkeley.

JAMES L. TAYLOR — has over twenty years experience in design, analysis, and management of computer-based data acquisition system projects. He has been responsible for conducting studies and developing conceptual designs, as well as systems design for industry and the Department of Defense. Taylor has published texts on design techniques for computer-based data acquisition systems, and fundamentals of measurement error, and has taught measurement techniques and system design courses for numerous aerospace and industrial firms. He earned a BS in Applied Math and Physics and a MS in Applied Math and Engineering.

We especially thank Dr. Robert B. Abernethy for the use of personal reference material and gratefully appreciate the constructive contributions and critical review from the following:

NASA Metrology and Calibration Working Group

Robert Burdine—NASA HQ (Code Q)/MSFC
Troy J. Estes—WSTF
Mark A. Hutchinson—LaRC
Fred Kern—LaRC
Kristen Riley—KSC
Herman Watts(Tom Weiss)—Stennis(Sverdrup)

Other Organizations

Norman B. Belecki—NIST
Randy Humphries Jr.—Boeing SSF
Ralph T. Johnson—Sandia National Laboratories
C. Ted Lynch—NASA HQ (Vitro Corporation)

Jet Propulsion Laboratory (JPL)

Gordon Dillinger
Daniel C. Griffin
Frank T. Hartley
Daniel LoGiurato
Lothar Kirk
John R. Radbill
James A. Roberts
Robert H. Steinbacher
James H. Wise

Robert E. Martin, Manager of JPL's Instrumentation Section, is the conceptual creator and prime motivator of this document. Without his vision, active support and energetic participation, this document would not have been started and completed.

Alexander Mark, a Member of Technical Staff in JPL's Instrumentation Section, was responsible for technical editing and production. Without the excellent input and critical review from all those who participated, the task could not have been accomplished. As editor, I humbly apologize to the expert authors and reviewers whose opinions may have differed somewhat from the outcome. In particular, one only has to look at the diversity of the principal authors to understand that we were not always in agreement on the presentation of the entire subject material. This publication contains a wide-range view of the world of metrology and goes well beyond calibration into the entire end-to-end measurement process. It is not perfect and we recognize that it may contain flaws both in quality and content. But, it is our feeling that it is better to accept some error than not to disseminate the material at all.

Most important, this Publication would not exist without the recognition of a priority need by the NASA Metrology and Calibration Working Group and provision for funding support through various Code Q organizations at NASA Headquarters.

The original document was produced entirely on an Apple Macintosh IIci using the following software:
Microsoft Word, Microsoft PowerPoint, Claris MacDraw Pro, Aldus SuperPaint, and Design Science MathType.

In early 2008, the original document file was converted to a PC-based Microsoft Word document using Design Science MathType as the equation editor. This converted document contains a few typographical corrections and minor modifications to Appendix F for technical clarification.

Contents

Figures

Tables

Acronyms

ACE	Automated Calibration Equipment
ANSI	American National Standards Institute
AOP	Average-Over-Period
AR	Accuracy Ratio
ATE	Automated Test Equipment
A/D	Analog to Digital
BIPM	International Bureau of Weights and Measures *(Bureau International des Poids et Mesures)*
BOP	Beginning-Of-Period
CGPM	General Conference on Weights and Measures *(Conference General des Poids et Mesures)*
CIPM	International Conference of Weights and Measures *(Conference Internationale des Poids et Mesures)*
CMRR	Common Mode Rejection Ratio
CMV	Common Mode Voltage
CRM	Certified Reference Material
DID	Data Item Description
DVM	Digital Voltmeter
EOP	End-of-Period
ESS	Pure Error Sum of Squares
FS	Full Scale
ISO	International Organization for Standardization *(Organisation Internationale de Normalisation)*
IT	In-Tolerance
IV	Indicated Value
LCL	Lower Confidence Limit
LSS	Lack of Fit Sum of Squares
MAP	Measurement Assurance Program
MLE	Maximum-Likelihood-Estimate
MSA	Manufacturer's Specified Accuracy
MTBF	Mean-Time-Between-Failure
MTBOOT	Mean-Time-Between-Out-of-Tolerance
NASA	National Aeronautics and Space Administration
NBS	National Bureau of Standards (*now NIST*)
NHB	NASA Handbook
NIST	National Institute of Standards and Technology (*was NBS*)
NSRDP	National Standard Reference Data Program

OOTR	Out-Of-Tolerance-Rate
pdf	Probability Density Function
PRT	Platinum Resistance Thermometer
QA	Quality Assurance
RM	Reference Material
RMAP	Regional Measurement Assurance Program
RMS	Root-Mean-Square
RSS	Root-Sum-Square
RTI	Relative To Input
RTO	Relative To Output
SI	International System of Units *(Système International d'Unités)*
SMPC	Statistical Measurement Process Control
SPC	Statistical Process Control
SRM	Standard Reference Materials
S/N	Signal to Noise
TAR	Test Accuracy Ratio
TME	Test and Measurement Equipment
UCL	Upper Confidence Limit
UUT	Unit under test
VIM	International Vocabulary of Basic and General Terms in Metrology *(Vocabulaire International des Termes Fondamentaux et Généraux de Métrologie)*

1. INTRODUCTION

1.1 Purpose

Methodologies and techniques acceptable in fulfilling metrology, calibration, and measurement process quality requirements for NASA programs are outlined in this publication. The intention of this publication is to aid NASA engineers and systems contractors in the design, implementation, and operation of metrology, calibration, and measurement systems. It is also intended as a resource to guide NASA personnel in the uniform evaluation of such systems supplied or operated by contractors.

1.2 Applicability

This publication references NASA Handbooks, and is consistent with them. The measurement quality recommendations are at a high level and technical information is generic. It is recommended that each project determine functional requirements, performance specifications, and related requirements for the measurement activity. Suppliers may use this document as a resource to prepare documentation for doing tasks described in this document.

1.3 Scope

A broad framework of concepts and practices to use with other established procedures of NASA is provided. The publication addresses the entire measurement process, where the term “process” includes activities from definition of measurement requirements through operations that provide data for decisions. NASA’s programs cover a broad range from short-term ground-based research through long-term flight science investigations. Common to all programs are data used for decisions (accept a system, launch a spacecraft) and data used for scientific investigations (composition of a planet’s atmosphere, global warming) to establish scientific facts.

> Measurement systems include hardware and software put in place to measure physical phenomena. In their simplest form, measurement systems can be considered to be a logical arrangement of equipment from one or more fabricators, possibly coupled with application software, integrated within a process so physical phenomena such as pressure, temperature, force, etc., can be measured, quantified, and presented.

Specifically, this publication is not limited to test equipment calibration and measurement standards activities. To provide a realistic assessment of data quality, the total process should be considered. The measurement process is covered from a high level through more detailed discussions of key elements within the process. Emphasis is given to the flowdown of project requirements to measurement system requirements, then through the activities that will provide measurements with known quality that will meet these requirements.

For many years, metrologists, calibration and repair specialists, measurement system designers, and instrumentation specialists have utilized widely known techniques which are conceptually simple and straightforward. With the proliferation of computing technology and philosophical changes occurring in quality management, the field of metrology is undergoing evolutionary and revolutionary change. Methodology for determining measurement uncertainty is becoming extremely complex in terms of system and component error analysis and manipulation of equations that require a good foundation in mathematics.

Total Quality Management (TQM) is becoming the way of doing business. The new environment is characterized by increased competition, scarcer resources, and a need to deliver high-quality products and services on schedule, with as little risk and at the lowest cost possible. Emphasis is on doing the right thing the right way with continuous improvement. This forces increased understanding of what a measurement implies and the decisions based on the measurement. This document is intended as a resource to help both management and technical personnel gain the tools and knowledge necessary to achieve acceptable quality in measurement processes.

Several changes from "business as usual" in the metrology community are reflected in the efforts underway to implement adaptations of the ISO 9000 series as replacements to the NHB 5300.4 series documents. In addition, NASA is working toward compliance with The U.S. National Standard (ANSI/NCSL Z540-1/ISO Guide 25) as it affects general requirements for calibration laboratories and measuring and test equipment. The ISO/TAG4/WG3 *Guide to the Expression of Uncertainty in Measurement* and the interpretation provided in NIST Technical Note 1297 are likewise being considered as changes from "business as usual."

> The complete implementation of the above philosophies has not yet taken place at the time of publishing this document. The developing strategies are imminent, but present a "moving target" for the authors. Therefore, the core of this publication concentrates on the presentation of traditional measurement methodologies with enhanced reinforcement of good engineering practices. As the practices of the measurement community evolve, the techniques presented within will be valuable to all who are responsible for the quality of the measurement.

Readers will vary from managers to personnel concerned with detailed activities. To help the reader, the following sections are suggested for different interests:

- Section 2 (*Quality Recommendations*) defines quality recommendations in high-level terms. The total measurement process is emphasized. This section is intended for all personnel.
- Section 3 (*Measurement Requirements*) describes the derivation of measurement requirements and includes the entire measurement process. Managers who depend on measurements should scan this section, especially the ten stages of Section 3.2.1 and the example in Section 3.2.7. Software is becoming increasingly important in measurement processes, and is addressed in Section 3.5. Personnel responsible for defining measurement requirements should read this section in detail. Other measurement persons should be familiar with this section.

Sections 4 through 6 detail the key elements of the measurement process. Examples of measurement systems are included. These sections are intended for members of the measurement community who will design, implement, and operate the measurement process.

- Section 4 (*Measurement System Design*) presents a systematic design approach for measurement systems, identifies the elemental errors associated with a measurement process, reviews methods for combining errors, and provides the specific steps needed to develop and evaluate a measurement process.
- Section 5 (*Measurement Traceability*) provides the foundation necessary for establishing traceability to measurement standards. Included are methods and techniques to assist in the traceable transfer of known values to final data.
- Section 6 (*Calibration Intervals*) discusses concepts, principles, and methods for the establishment and adjustment of intervals between calibrations for test and measurement equipment.
- Section 7 (*Operational Requirements*) deals with the operations phase of the measurement process at a higher level than that of Sections 3 through 6. This section is primarily intended for operational personnel who must provide data with known quality. Managers should scan Section 7.1, which discusses quality.
- Section 8 (*Recommendations for Waiver/Deviation Requests*) should be read by managers and measurement personnel.

The appendices primarily delve into state-of-the-art innovations and techniques for error analysis, development of statistical measurement process control, optimization of calibration recall systems, and evaluation of measurement uncertainty. The techniques presented in these appendices will likewise be valuable to the establishment of quality measurements.

- Appendix A (*Definitions*) contains the terms used in this publication since it is recognized there are different definitions, connotations, and preferences for specific terms used in the aerospace and metrology communities.
- Appendix B (*Mathematical Methods for Optimal Recall Systems*) provides the mathematical and detailed algorithmic methodology needed to implement optimal calibration interval analysis systems as described in Section 6. This appendix should be read by technical specialists responsible for calibration interval system design and development.
- Appendix C (*Test and Calibration Hierarchy Modeling*) provides mathematical methods and techniques to link each level of the test and calibration support hierarchy in an integrated model. These methods enable analysis of costs and benefits for both summary and detailed visibility at each level of the hierarchy. This appendix should be read by technical specialists responsible for calibration interval system design and development.
- Appendix D (*Statistical Measurement Process Control* ***(SMPC)*** *Methodology Development*) describes statistical measurement process control methodology in generalized mathematical terms. The SMPC methodology overcomes traditional SPC methods which are difficult to implement in remote environments. This appendix is not intended for the casual reader, but should be read by technical specialists responsible for developing information regarding the accuracy of the monitoring process. The methodology is especially useful in cases where astronomical or terrestrial standards are employed as monitoring references, and for reducing dependence on external calibration in remote environments.

- Appendix E (*Error Analysis Methods*) provides the measurement system designer with mathematically invigorating tools to develop measurement system error models and analyze measurement system errors.
- Appendix F (*Practical Method for Analysis of Uncertainty Propagation*) describes an evolutionary nontraditional uncertainty analysis methodology that yields unambiguous results. The term "practical" suggests that the methodology is usable or relevant to user objectives, such as equipment tolerancing or decision risk management. In using this methodology, rigorous construction of statistical distributions for each measurement component is required to assess measurement uncertainty. Application software is presently being developed for user-interactive computer workstations.
- Appendix G (*Determining Uncertainty of an Example Digital Temperature Measurement System*) is founded on an example temperature measurement system given in Section 4. It is very detailed in the identification and analysis of error sources to determine the measurement uncertainty and should be read by technical specialists responsible for the design of measurement systems. The methodologies presented parallel those provided in NIST Technical Note 1297 and the ISO/TAG4/WG3 *Guide to the Expression of Uncertainty in Measurement*..
- Appendix H (*The International System of Units **[SI]***) contains traditional information on the metric system. It is contained in this publication for the convenience of all readers.
- Acronyms are defined at the beginning of this document. A reference section is at the end.

> Throughout this publication, references are made to "space-based" activities. For the purpose of definition, "space-based" includes all activities that are not Earth-based, i.e. satellites, humanly-occupied on-orbit platforms, robotic deep-space probes, space- and planet-based apparatus, etc.—all are included in the term "space-based" as used in this document.

2.0 QUALITY RECOMMENDATIONS

2.1 Introduction

Measurement quality can be described in terms of our knowledge of the factors that contribute to the differences between the measurement and the measurand, and the extent of our efforts to describe and/or correct those differences.

Two attributes of a measurement provide the quality necessary for decisions:

(1) The measurement must be traceable to the National Institute of Standards and Technology (**NIST**), an intrinsic standard, or to a consensus standard accepted by contractual or similar documents.

(2) Measurement uncertainty must be realistically estimated and controlled throughout the measurement process.

Measurement quality assures that actions taken based on measurement data are only negligibly affected by measurement errors. The *complete* measurement process should be included in the objective definition of measurement quality. The following issues should be considered when making a measurement.

- The measurement process quality should be consistent with the decision's need for measurement data. The measurement process should be consistent with economic factors in providing adequate quality and avoid an over-specified, expensive process.
- Measurement system reliability design requirements should be defined and specified so that design objectives are clearly understood.
- Uncertainty is a parameter of the complete measurement process, not a parameter limited to instruments used in the process.
- Control of uncertainty of limited parts of the process, such as calibration of electronic instruments, is a necessary condition for objective definition of uncertainty, but clearly is not a sufficient condition.
- Uncertainty of a chain or sequence of measurements grows progressively through the sequence.
- Uncertainty in the accuracy ascribed by calibration to a measuring attribute grows with time passed since calibration.

2.2 Measurement Functions

Measurement quality requirements are applicable to the measurement processes associated with the following functions:

(1) Activities where test equipment accuracy is essential for the safety of personnel or equipment.

(2) Qualification, calibration, inspection, and maintenance of flight hardware.

(3) Acceptance testing of new instrumentation.

(4) Research and development activities, testing, or special applications where the specification/end products of the activities are accuracy sensitive.

(5) Telecommunication, transmission, and test equipment where exact signal interfaces and circuit confirmations are essential.

Measurement processes used for purposes other than those specified above are considered to have uncontrolled uncertainty and should be limited to

(1) Applications where substantiated measurement accuracy is not required.

(2) "Indication only" purposes of nonhazardous and noncritical applications.

2.3 Measurement Quality Recommendations

2.3.1 Requirement Definition

The measurement quality requirement should be objectively defined early in the activity and drive the measurement process design.

Early definition of the measurement uncertainty and confidence level should be done so the measurement process is responsive to its objective. The measurement process cannot be defined by organizations in the measurement disciplines until the measurement quality requirement, traceable to the decision, is known.

2.3.2 Requirement Traceability

The measurement quality requirement should be traceable to the decision need that will use the data from the measurement.

The requirement should be responsive to the user of the data, and should not be defined only by organizations in the measurement or metrology disciplines.

2.3.3 Implementation Cost

The measurement quality implementation should be cost-effective in providing the needed quality, but not an over-specified quality.

The implementation should define the decision risk to provide adequate quality at the least cost. Some measurements may have a broad uncertainty range, so quality can be implemented economically. Critical decisions with high risk may need measurement uncertainties that are difficult to achieve, with corresponding higher costs.

2.3.4 Uncertainty Identification

The measurement should be treated as a process, with all contributors to bias and precision errors (from the sensor, through data reduction) identified. Uncertainties should reflect a realistic representation of the process so the process uncertainty, and prediction for growth, is meaningful.

Uncertainties must be a realistic representation of the actual physical measurement process. Sensors may disturb the measurand. Thus, they may not provide an accurate representation of the measurand, and so may not provide the correct data needed for a good decision. In such a case, uncertainties from both the sensor intrusion effects, and the relationship of the sensor output to the data reduction equations, are necessary to correctly define the complete uncertainty. The effect of software must be included. Operator characteristics or environmental changes are important sources of uncertainty and so must be included. From the planning viewpoint, consideration of all uncertainties early in the activity is essential to allow the total uncertainty budget to be allocated to the measurement process elements.

Since uncertainties grow with time since test or calibration, measurement decision risk also increases with time since calibration. This is the underlying motivation for recalibrating and retesting regularly. When uncertainty grows beyond predicted limits, insidious “soft” failures occur in the measurement system. “Soft” failures cause a measurement device to generate data beyond stated uncertainty limits. Usually these failures go undetected by the user and/or operator.

2.3.5 Design Documentation

The measurement process design should be documented in written form with an auditable content so that it may be used during the operations phase.

Usually, design documentation will be used by persons in the operation and data reduction phases who did not design or develop the measurement process. The documentation will help operation personnel to monitor uncertainties throughout the period of the measurement, so any uncertainty growth with time can be better defined. Characteristics of the operation phase, which may be under time pressure to correct failures, should be considered. The design documentation also should be auditable. Extensive documentation is not necessarily needed. For instance, a short-term research activity might be documented as a single-page memo that summarized the measurement process, its uncertainties, and included measurement quality traceability. A long-term spaceflight activity will need extensive formal documentation and should take into consideration use of alternate personnel during the flight duration.

2.3.6 Design Review

The measurement process design should pass a review before implementation of the measurement process with representation from technically qualified persons and from the data user organization.

A review should be held before the implementation of the measurement process. The purpose of the review is to ensure that all design requirements have been addressed. The review members should include persons technically competent in relevant disciplines (metrology, sensors, software, etc.), and persons from the user organization. This review could be a half-hour informal meeting to a formal preliminary design review, depending on the scope of the measurement and the phase of the activity. Despite the level of formality, every measurement process should be subjected to some review before implementation. This recommendation is intended to assure both technical competence and satisfaction of the decision organization.

2.3.7 Quality Control

The measurement quality should be monitored and evaluated throughout the data acquisition activity of the operations phase. This should be done to establish that the uncertainty is realistically estimated and controlled within the specified range, and that out-of-control exceptions are objectively identified.

Objective definition of data quality is needed to support the decision process. Rigorous monitoring is necessary to provide the objective definition.

2.3.8 Quality Documentation

The total measurement process should be documented so that decisions based on measurement results can be objectively evaluated.

The measurement process should be documented to the extent necessary to enable an objective estimate of risks associated with decisions based on measurement results.

2.4 Relevant Quality Provisions

Quality provisions relevant to the above measurement quality recommendations are found in the following NASA Handbooks:

- NHB 5300.4(1B), "Quality Program Provisions for Aeronautical and Space System Contractors"
- NHB 5300.4(1C), "Inspection System Provisions for Aeronautical and Space System Materials, Parts, Components and Services"
- NHB 5300.4(1D-2), "Safety, Reliability, Maintainability and Quality Provisions for the Space Shuttle Program"
- NHB 4200.1, "NASA Equipment Management Manual"

3. MEASUREMENT REQUIREMENTS

3.1 Objectives of the Measurement Process

To assure adequate space system performance, it is essential that technical requirements be developed , defined, and documented carefully. Clearly defined measurement requirements lead to the high reliability and quality needed to assure successful system performance and mission achievement. They assure that decisions (including establishing scientific fact from measurements) are based on valid information and that only acceptable end-items proceed from suppliers into flight hardware and support systems. Many of these items are the sensors, detectors, meters, sources, generators, loads, amplifiers, filters, etc., integrated to form the measurement system of a space-based system. The definition and understanding of measurement processes and their requirements raise such questions as:

- What is a measurement? What characterizes it?
- Why is the measurement being made?
- What decisions will be made from the measurement?
- What performance requirements do the measurements seek to validate?
- What measurement and calibration system design requirements will support the performance requirements?
- What level of confidence is needed to assure that measurements yield reliable data and that the risks of using inadequate data are under control?

> **MEASUREMENT** — The set of operations having the object of determining the value of a quantity.

Measurements are subject to varying degrees of *uncertainty*. The uncertainties need to be estimated. From the estimate, the validity of the measurement can be assessed; the risks associated with decisions based on these measurements can be quantified; and corrective actions can be taken to control growth in the measurement uncertainty.

Measurements provide data from which decisions are made:

- To continue or stop a process
- To accept or reject a product
- To rework or complete a design
- To take corrective action or withhold it
- To establish scientific fact.

> The more critical the decision, the more critical the data. The more critical the data, the more critical the measurement.

Hardware attribute measurements should be made during development to evaluate expected system performance capabilities and the tolerance limits within which satisfactory performance is assured. Other measurements, made during the development stage, confirm performance capabilities and tolerances after production and before product delivery. Later, measurements are made by the end user during acceptance tests, before launch or deployment, during deployment exercises, and following mission completion. These tests and measurements, in one way or another, involve decisions made to confirm compliance of the hardware with documented performance specifications.

> Measurements made during development create performance requirements (specifications) from which other (production, acceptance, deployment and post-mission) measurement requirements.

All valid measurement processes call for specificity of:

- Measurement parameters
- Parameter ranges
- Allocation and control of uncertainties
- Time limits to which the requirements apply
- Environments in which they will operate.

These characteristics are used to establish the measurement control limits and design requirements for both measurement and calibration systems.

Determination and control of measurement process uncertainty and it's relation to hardware attribute tolerances is a way to define and control risks taken during decision processes.

> The objective of the design and control of measurement processes is to manage the risks taken in making decisions based on measurement data.

The objective of the measurement process for space systems is to monitor the integrity of the performance parameters of space hardware, instrumentation, and ground support equipment, and to allow sound decisions for taking actions. The objective of calibration is to determine initial bias errors, correct for these, and then to monitor and control the growth of measurement uncertainty. This assures that decisions being made about the hardware from the measurement data are made within acceptable risk limits. Two principles of the measurement activity should be considered:

> **PRINCIPLE 1** — Measurements only estimate the value of the quantity being measured. There is always some uncertainty between the value of the measurand and the data representing the measured quantity. The uncertainty may be very small, such as the case of the measurement of a one-volt standard by a higher-level standard, but the uncertainty always exists. The uncertainty must be estimated and controlled to provide a measurement with known quality.

> **PRINCIPLE 2** — Measurements are made to support decisions or establish facts. If measurement data are not used in a decisions[1], the measurement is unnecessary.

A decision must be based on data with known quality so measurement data errors will have only a negligible effect on the decision. Measurement quality has two attributes: (1) the measurement must be traceable and (2) the measurement must have a realistic estimate of its uncertainty. The "realistic estimate of uncertainty" attribute leads to a third principle:

> **PRINCIPLE 3** — Every element of the measurement process that contributes to the uncertainty must be included.

3.2 Defining Measurement Requirements

3.2.1 Measurement Requirements Definition Sequence

Determining measurement process requirements can be viewed as a ten-stage sequence that flows down as follows:

> **STAGE 1 —** ***MISSION PROFILE***
>
> Define the objectives of the mission. What is to be accomplished? What reliability is needed and what confidence levels are sought for decisions to be made from the measurement data?

> **STAGE 2 —** ***SYSTEM PERFORMANCE PROFILE***
>
> Define the needed system capability and performance envelopes needed to accomplish the Mission Profile. Reliability targets and confidence levels must be defined.

> **STAGE 3 —** ***SYSTEM PERFORMANCE ATTRIBUTES***
>
> Define the functions and features of the system that describe the System's Performance Profile. Performance requirements must be stated in terms of acceptable system hardware attribute values and operational reliability.

[1] The use of the term "decisions" to include scientific data, as another use of measurement data, is shown here.

STAGE 4 — *COMPONENT PERFORMANCE ATTRIBUTES*

Define the functions and features of each component of the system that combine to describe the System's Performance Attributes. Performance requirements must be stated in terms of acceptable component attribute values and operational reliability.

STAGE 5 — *MEASUREMENT PARAMETERS*

Define the measurable characteristics that describe component and/or system performance attributes. Measurement parameter tolerances and measurement risks (confidence levels) must be defined to match system and/or component tolerances and operational reliability.

STAGE 6 — *MEASUREMENT PROCESS REQUIREMENTS*

Define the measurement parameter values, ranges and tolerances, uncertainty limits, confidence levels, and time between measurement limits (test intervals) that match mission, system, and component performance profiles (Stages 2, 3, and 4) and the measurement parameter requirements (Stage 5).

STAGE 7 — *MEASUREMENT SYSTEM DESIGNS*

Define the engineering activities to integrate hardware and software components into measurement systems that meet the Measurement Process Requirements. Definition must include design of measurement techniques and processes to assure data integrity.

STAGE 8 — *CALIBRATION PROCESS REQUIREMENTS*

Define the calibration measurement parameter values, ranges, uncertainty limits, confidence levels, and recalibration time limits (calibration intervals) that match measurement system performance requirements to detect and correct for systematic errors and/or to control uncertainty growth.

STAGE 9 — *CALIBRATION SYSTEM DESIGNS*

Define the integration of sensors, transducers, detectors, meters, sources, generators, loads, amplifiers, levers, attenuators, restrictors, filters, switches, valves, etc., into calibration systems that meet the Calibration Process Requirements. Definition must include design of calibration techniques and processes to assure data integrity.

STAGE 10 — *MEASUREMENT TRACEABILITY REQUIREMENTS*

Define the progressive chain of calibration process requirements and designs that provide continuous reference to national and international systems of measurement from which internationally harmonized systems measurement process control is assured.

Stages 1 through 4 describe the performance requirements of the complete system and each of its parts. These are the system and component capabilities converted to written specifications essential to successful mission achievement. Stages 5 and 6 apply the measurement parameters derived during development that characterize the attributes of the hardware. Because of NASA and contractor technical and management objectives, Stages 5 and 6 are the critical efforts that establish the technical objectives and requirements that the measurement process designs shall meet.

The output of Stage 6, Measurement Process Requirements describes

- *Measurement parameters* — (voltage, pressure, vacuum, temperature, etc.)
- *Values and range* — (3–10 volts, 130–280 pascal, 0 to –235 degrees celsius, etc.)
- *Frequency/spectra range* — (18–20 KHz, 10–120 nanometers, 18–26 GHz, etc.)
- *Uncertainty limit* — (±0.1% full scale, ±0.005 °C, etc.)
- *Confidence level* — (3 standard deviations, 99.73% confidence limits, 2 , etc.)
- *Time limit* — (one flight, six months, five cycles, etc.) for which the uncertainties are not to be exceeded at the confidence levels given.

Stage 7, Measurement System Design, is part of a larger system design activity that focuses on the measurement process. Engineering analysis of the measurement process is done to allocate performance to the system components. Section 4 describes detailed techniques used during design. Also, in Stage 7, provisions for testing and calibration are included in the measurement process.

Stages 8 through 10 are directed at establishing the calibration and measurement traceability capabilities needed to support the operational measurement system and are discussed in Section 5. Fundamental to calibration and measurement traceability is the control of measurement uncertainty, which in turn is controlled by design (Stages 7 and 9) and the establishment and adjustment of calibration intervals. Section 6 deals with this subject.

In the ten-stage flowdown of determining measurement process requirements, two aspects are indigenous to the process. They are: the underlying operational requirements and the special circumstances of state-of-the-art limits and practicality where a waiver or deviation from standard requirements is prudent. These matters are covered in Sections 7 and 8, respectively.

3.2.2 System Characteristics and Measurement Parameters

To get measurement process requirements at Stage 6 of the definition sequence Stages 1 through 4 need to be examined to determine the characteristics (values, tolerances, etc.) of the materials, articles, processes, and experiments.

Often characteristic studies are done. These studies

- Determine theoretical performance capabilities
- Estimate performance degradation over time
- Establish test attributes
- Allocate tolerances at specific measurement sites
- Establish measurement conditions
- Identify where measurements will be made
- Show the confidence levels needed for measurement decisions.

These characteristics are often in system parameter documents or their equivalent. These are the characteristics that affect system functions, features, interchangeability, coordination, reliability, quality, and safety. Characteristics must be described in enough objective detail to include the performance tolerance limits within which the wanted performance lies, or beyond which unsafe or inadequate performance lies. From these article or process characteristics, Stage 5 defines measurement parameters that translate the defined characteristics into measurable terms. These are often the same phenomena, such as temperature or voltage, but they also include characteristics that are only representations of the hardware feature.

For those articles that form a system assembly process, candidate measurement parameters that represent performance characteristics include the following:

- Power inputs needed for article operation
- Signal inputs to emulate interactive hardware operations
- Output signals from the article (especially those parameters that measure nonlinear outputs near specification limits, those outputs sensitive to other component parameters, and those outputs sensitive to two or more inputs that may interact)
- Measurements to control or monitor the process or progress of the article through a series of tests.

More information than just characteristics, values, and tolerances is needed to define measurement requirements. The environment in which the measurements will be done needs to be identified in detail. Is it hostile to the measuring systems? What are the pressures, temperatures, humidity, radiation levels, sound intensity, etc., at which measurements will be done? It will be impossible to do uncertainty analyses without this knowledge. Also, information is needed regarding the intended sites where the measurements will happen and whether they are remote, accessible to human contact, etc.

3.2.3 Establishing Measurement Classifications

Another facet of defining measurement requirements calls for consideration of the relative importance of all measurement processes involved in a given program or mission. Indicators of importance are useful in identifying confidence level requirements on measurement uncertainties in a program or mission.

> The greater the importance of the decision, the higher the confidence the decision makers need in their measurement data. Therefore, important measurement data must be obtained at high confidence levels.

The importance of measurements can be classified, first, to the importance of their application (mission, experiment, fabrication process, inspection, fault analysis, etc.) A second classification, complementary to the first, would involve the degree of difficulty in the measurement process, especially as it relates to the measurement uncertainties and sensitivities needed versus the capability, or state of the art, of the measurement systems.

3.2.3.1 Criticality of Application

NASA Handbook 5300.4(lD-2), Appendix A, defines criticality categories throughout NASA. These represent priority requirements that could apply to all aspects of NASA programs including measurement processes. The categories of criticality are paraphrased here as follows:

Category 1 — *Measurements that affect loss of life or vehicle.*

Category 2 — *Measurements that affect loss of mission.*

Category 3 — *Measurements that affect performance other than Category 1 and Category 2.*

Category 3 is unspecific about subordinate categories. The criticality of measurements should perhaps be classified in terms of the confidence to be expected in making decisions from measurement data. (These subcategories may not be in precise order of importance, since they are influenced by circumstances).

Subcategory 3.1 — *Measurements monitoring mission tasks and sensing changes to steady-state mission conditions.*

Subcategory 3.2 — *Measurements of components and systems under development that generate design specifications. Measurements of fabrication processes that produce goods to design specifications.*

Subcategory 3.3 — *Measurements made to test and confirm that products meet design specifications. Measurements made to test and confirm that measurement equipment meets performance specifications. Measurements made to test and confirm that uncertainties (errors) have been determined and corrected and controlled.*

Subcategory 3.4 — *Measurement of components and systems to determine their maintenance status. Measurement or monitoring environments within which end-items and test systems operate.*

3.2.3.2 Difficulty of the Measurement

The degree of difficulty of each measurement may have a direct effect on it's cost and quality. Measurements deserving the most attention can be rated in terms of degrees of difficulty in meeting measurement requirements, where that difficulty may lead to hardware with lowered performance capability. The following classifications are suggested:

Difficulty Degree A *MOST DIFFICULT OR IMPOSSIBLE MEASUREMENTS*

Al *Measurements of selected parameters that cannot be made because of lack of available measuring devices and methods.*

A2 *Measurements that can be made, but to meet program requirements, require methods that are extremely expensive, or time-consuming.*

A3 *Measurements of space-based calibration processes that cannot be supported readily by simple on-vehicle or astronomical or terrestrial measurement references.*

(Difficulty degrees Al, A2 and A3 usually force use of alternative performance parameters that may only slightly characterize system performance, but can, at least, be measured at reasonable difficulty levels.)

Difficulty Degree B *MEASUREMENTS THAT CANNOT MEET THE NHB 5300.4(1B) MEASUREMENT UNCERTAINTY REQUIREMENTS*

Bl *That uncertainties in any article or material measurement process shall be less than 10 percent (1/10) of the measured parameter tolerance limits.*

B2 *That uncertainties of calibration processes be less than 25 percent (1/4) of the measured parameter tolerance limits.*

Difficulty Degree C *MEASUREMENTS MADE IN ENVIRONMENTS HOSTILE TO OPTIMUM MEASURING SYSTEM PERFORMANCE.*

3.2.4 Establishing Confidence level Requirements

> A method is needed to express the degree of confidence that is wanted for each measurement process.

Confidence levels are related to the quality and reliability goals of the experiment, the hardware or the process. These provide the designer of the measurement process with goals that determine control of uncertainty. in the measurement process. Otherwise, the measurement process designer must guess at the quality and reliability goals of the experiment, hardware, or process. Therefore, the characteristic studies must also show the confidence levels at which the characteristic tolerances will be controlled. From these, measurement uncertainty analyses can be done, decisions regarding tests can be made, and where and how often to test can be established.

Confidence levels have a direct effect on cost, schedule, and data reliability for the measurement system design, its production, its calibration, and its maintenance. Finding a way to assign proper confidence levels is needed to help planner and designer alike and is addressed in the next section.

CONFIDENCE LEVEL (α), that is, the probability that the measurand value lies within the uncertainty interval of the measurement, is expressed in this publication in terms of standard deviation, sigma or σ.

For a high-confidence measurement requirement, the system planner or designer needs guidance about the confidence levels to require for system uncertainty estimates. The correlation of critical applications and difficult measurements suggest that a matrix of these two elements can be formed. This can present a decision base for assignment of proper confidence levels and a sense of priority for the planning and costs for development and designs of the measurement processes. Table 3.1 presents a suggested approach to choosing confidence level requirements to accompany measurement uncertainty requirements.

TABLE 3.1 Measurement Uncertainty Confidence Level Assignments for Measurement Applications and Degrees of Difficulty

TABLE 3.1

Measurement Uncertainty Confidence Level Assignments for Measurement Applications and Degrees of Difficulty

DEGREES of DIFFICULTY	CATEGORIES of CRITICALITY of APPLICATIONS 1	2	3.1	3.2	3.3	3.4	OTHER
A1	*	*	*	*	*	*	*
A2	1	1	2	3	4	5	6
A3	1	1	2	3	4	5	6
B1	1	2	2	3	4	5	6
B2	2	3	3	4	5	6	6
C	2	3	4	5	6	6	7
OTHER	2	3	4	5	6	7	7

Legend:

Matrix Intersection Number	Confidence Level	No. of Standard Deviations (Sigma)
1	99.99994	5.0
2	99.994	4.0
3	99.73	3.0
4	95.45	2.0
5	91.37	1.8
6	86.64	1.5
7	68.27	1.0

* Measurement cannot be performed. Alternative parameter must be selected.

3.2.5 Establishing Measurement System Reliability Requirements

The previous section provided guidance on confidence level assignments for measurement uncertainty requirements. Still, some way is needed to describe the period over which the uncertainty estimate can be depended upon and how to translate that time into a useful design target. Two elements are involved in the description. First, the time within which the uncertainty can be "guaranteed"—this element is equivalent to the calibration interval. Second, the population (percentage) of measurement data that can be expected to be within the uncertainty limits at the end of the "guaranteed" time. This is the end-of-period (**EOP**) in-tolerance probability or the measurement reliability requirement.

For practical purposes, the measurement reliability requirements and the confidence level requirements coincide.

The specified measurement uncertainty is to be contained within the measurement reliability/ confidence level requirements over the course of the calibration interval. For example, the first element could be a 6-month calibration interval; the second element would be a 95.45% EOP measurement reliability, corresponding to a 2–standard deviation confidence level.

With the uncertainty, both the interval and the measurement reliability must be specified to fully convey the design requirements for the measurement system.

This is necessary to assure that rapid uncertainty growth during the calibration interval does not add unreasonable uncertainties to the measurement process when the measurement is being performed. Unfortunately, neither the confidence level or the calibration interval are useful to the planner unless they are translated into terms, or a single term, that designers can use. Calibration interval mathematical models use a term that appears to fulfill this need. It is similar to the term mean-time-between-failure (**MTBF**) used as a reliability target in hardware and system design specifications.

MEAN-TIME-BETWEEN-OUT-OF-TOLERANCE (MTBOOT) reflects the mean time between "soft" failures for measuring instruments and systems. For this purpose, "soft" failures are defined as those that cause a measurement device to generate data beyond stated uncertainty limits. These soft failures usually go undetected by the user and/or operator.

By contrast, MTBF failures are "hard" ones, resulting from extreme component degradation or failure and subsequent inability to reach performance limits (ranges or frequencies) and usually, are *readily detectable to the user and/or operator*. The exponential calibration interval mathematical model (see Appendix B) uses MTBOOT values to establish calibration intervals to match desired percentage in-tolerance goals for program applications. For example, typical general-purpose military test, measurement, and diagnostic equipment have percent in-tolerance probability targets of from 72 to 85% EOP.

For a specified calibration interval, percent in-tolerance (measurement reliability) goals create specific MTBOOT requirements. For example, a one-year calibration interval on an instrument that behaves according to the exponential model, whose recalibration percent in-tolerance (**IT**) is to be greater than 95% IT EOP, results in an MTBOOT requirement of 40,500 hours. This would mean that the instrument designer would have to target his or her design for an MTBOOT equal to or greater than 40,500 hours if the one-year interval is to be achieved. (Under normal circumstances, most MTBFs would be at least equal to or greater than a specified MTBOOT.) A four-month interval with measurement reliability targets of 95% IT EOP would lead to an MTBOOT of 13,500 hours. For the same four-month interval, if ≥99% IT EOP were a requirement, the MTBOOT would increase to 68,700 hours. Were these values of MTBOOT unachievable in the design, the interval would have to be shortened, the allowable out-of-tolerance percentage increased (that could lead to an increased risk of wrong decisions being made from the measurement process through lowered measurement reliability), or the mission objectives re-evaluated to adapt to the lowered measurement reliability.

Table 3.2. reflects example measurement reliability requirements versus MTBOOT for a one-year, six-month and three-month calibration interval assuming a 40-hour work-week usage, and for systems whose uncertainties grow exponentially with time. (MTBOOTs for shorter or longer intervals/usage would vary linearly with time.) The figures in the table are based on the following mathematical relationship:

$$\text{MTBOOT} = -\text{Usage Hours per Year}/\ \ln R$$

Where R = confidence level or measurement reliability.

TABLE 3.2 Mean Time Between Out-of-Tolerance (MTBOOT) Design Values for Confidence Level/Measurement Reliability Goals for Equipment Following the Exponential Reliability Model

MEASUREMENT PROCESS CONFIDENCE LEVELS		*MEASUREMENT SYSTEM MTBOOT (Khrs)* [*]		
SIGMA	**RELIABILITY GOAL**	**FOR 1 YR**	**FOR 6 MO.**	**FOR 3 MO.**
5.0	99.9999%	3,467,000	1,733,000	867,000
4.0	99.994	34,667	17,333	8,667
3.3	99.9	2,059	1,030	515
3.0	99.73	743	372	186
2.6	99	206	103	51.5
2.0	95.45	44.7	22.4	11.2
1.96	95	40.5	20.3	10.1
1.8	91.37	23.0	15.5	7.75
1.65	90	19.7	9.85	4.93
1.5	86.64	14.5	7.25	3.65
1.44	85	12.8	6.4	3.2
1.08	72	6.33	3.17	1.58
1.0	68.27	5.45	2.73	1.36
0.84	60	4.07	2.04	1.02
0.67	50	3.0	1.5	0.75

[*] *(2,080 usage hrs/yr @ 40 hrs/wk)*

Specific values of MTBOOT and implied values of MTBF can be used for definition of system reliability design requirements. They can be used by program planner and system designer alike.

3.2.6 Finalizing Measurement Requirements

Once the measurement parameters, measurement values, applications, environment, and tolerances (including confidence/reliability limits) have been defined, the final definition of measurement requirements is nearly complete.

> If the measurement process supports an experiment, article, or fabrication process, NHB 5300.4(1B) requires that the measurement uncertainty be less than ten percent (1/10) of the tolerances called out for the parameter. If the measurement relates to a calibration measurement process, NHB 5300.4(1B) requires that combined uncertainties of the calibration measurement system will be less than 25 percent (1/4) of the tolerances called out for the parameter.[2]

Finally, the ten-stage definition process generates a measurement requirement that includes:

- The parameter to be measured, including the range and specific values of the parameter, and its location and point of measurement
- The process characteristics, such as static or dynamic, bandwidth/frequency spectrum, etc.
- The measurement modes, such as absolute, gage or differential pressure, volumetric or mass flow, temperature conduction, convection, radiation, etc.
- The environment (pressure, temperature, moisture, electromagnetic interference, etc.) in which the measurement is to be done, including measurement sites and operators
- The data to be acquired throughout the measurement process, including data rates and data bandwidths
- The measurement uncertainty requirements associated with each value of the parameter
- An expression of the confidence limits within which the uncertainties must be contained. These limits would be determined by considering the criticality of the application and the difficulty of the measurement
- The time limits between measurements or tests to assure control of hardware performance spread and a definition of the percent of items or measurement data to be found operating within performance and uncertainty limits.

Equipped with these clearly defined measurement requirements, the designer of the measurement process can continue in an orderly manner to develop specifications to meet a specific design goal and to complete a successful measurement system design.

---

2 These "rules of thumb" ratios of 1/10 and 1/4 are simplified methods of assuring that test or calibration process measurement uncertainties do not negatively affect decisions made from the measurement data. When these rules cannot be met, far more complicated alternatives are available to determine measurement uncertainty requirements. These include individualized measurement uncertainty analyses and measurement statistical process control techniques discussed elsewhere in this document.

3.2.7 Example—Measurement Requirement Definition of a Solar Experiment

An example is presented below to illustrate the ten-stage measurement requirements definition process. The example starts with space science mission requirements and, through the first six stages, develops the Solar Experiment instrument system requirements. In Stage 6, the example switches to the development of requirements of the ground test system needed to support the flight system. Examples covering the operational measurement system design are provided in Section 4.

STAGE 1 — Mission Profile

A space mission named the Solar Experiment is planned that includes, as one of several tasks, an experiment to determine the variability of solar ultraviolet (**UV**) irradiance over a year's cycle. Extreme fluctuations in irradiance are expected to be found based on rough measurements (±30% of indicated value) taken on Earth-based instruments whose uncertainty was increased by atmospheric interference. For the mission, measurement data uncertainty of less than ±10% of indicated value (**IV**) is wanted with 24-hour-per-day, ten-second increment data transmission capability. Mission reliability is targeted at 99.73% (3σ). The Solar Experiment's mission application has been designated by management as a Criticality Category 3.1.

STAGE 2 — Measurement System Performance Profile

The phenomena to be detected are UV intensity and spectra. The measurable characteristics are determined to be units of power (watts/square meter—W/m^2) and spectra (wavelengths of 120 to 400 nanometers.) Measurement difficulty is high and has been assigned Degree A3.

To avoid compromising the mission reliability goal, the reliability goal of each mission component (experiment) must have a reliability goal significantly higher than that of the mission reliability goal. Confidence levels for the Solar Experiment's goals must be significantly higher than the mission's reliability goal of three sigma.

Using Table 3.1, the critical application and difficulty confidence level matrix, a target of 4σ (99.994% confidence level) appears proper for the Solar Experiment's part of the mission.

STAGES 3 and 4 — Measurement System and Component Performance Attributes

The fluctuation in ultraviolet radiation can be measured in several ways: by differential techniques, by absolute techniques, and by a combination of the two. An absolute technique is chosen as the objective. Calibration and testing of the experiment's instrumentation system will be done in the environment of the launch site's test laboratory. Measurement value ranges are set at 1 to 100 milliwatts per square centimeter with a spectrum of 120 to 400 nanometers. The measurement uncertainty requirement is ±10% IV to meet the data accuracy requirement at a confidence level of 4σ. The performance interval over which the uncertainty is to be maintained is 1 year. To provide the design criteria for system and/or component reliability, an MTBOOT corresponding to a 4σ 1-year test interval is assigned. (After one year the system is to be transmitting measurement data, 99.994% of which is within uncertainty limits of ±10% IV.) A 24-hour day, full-time data transmission operational requirement generates 8760 hours per year of usage time. Presuming the instrumentation system's uncertainty will degrade exponentially, an MTBOOT requirement of about 146,000,000 hours is assigned. Shown earlier, MTBOOT is calculated from the equation:

> MTBOOT = –Usage Hours per Year/ ln R
> Where R = confidence level or measurement reliability.

An MTBOOT (or even an MTBF) of 146,000,000 hours is an extremely high requirement that the designers may find impossible to meet. It may call for the extraordinary design features discussed earlier. It may also need a request for waiver of the 99.994% (4σ) confidence level requirement to something closer to 3σ. However, even a 3.29σ requirement translates to 99.9% levels which, for a one year interval would establish an approximate 8,756,000-hour MTBOOT. Obviously, the final design for the Solar Experiment instrumentation system will be difficult. While prototypes have been said to be available with "accuracies of ±5% of indicated value," the confidence levels of the uncertainty estimates were determined to be no better than 3σ, with no account taken for uncertainty growth over a full year, although long-term photodiode sensor and optical element stabilities were said to be excellent. An attentive reevaluation of the capability of the prototype will be needed to confirm that uncertainties at higher confidence levels over the year's interval will match the ±10% requirement. If all efforts fail, it may become necessary for the planners to rethink the need for a 10% data accuracy requirement for the Solar Experiment, or a 3σ mission reliability target. They also could consider changing the data sampling rate to reduce the 24-hour per day operational requirement to, say, 8 hours per day. This would reduce the MTBOOT by 2/3.

STAGES 5 and 6 — Measurement Parameters and Measurement Process Requirements

The sequence now calls for an assessment of the Solar Experiment instrumentation system to determine how and to what requirements its first calibration and retesting after one year will be done. Since the instrument can detect power and spectra, its own first calibration and retesting will need a source or stimulus and comparator with proper characteristics to emulate the UV solar irradiance phenomenon. This requirement calls for a source and comparator testing system that can generate and detect 1 to 100 milliwatts/square centimeter across the 120–400 nanometer spectra. As prescribed by NHB 5300.4(1B), the uncertainty of this test system is to be 10% of that of the Solar Experiment's goal, or ±1% IV. It has a Category 3.1 application assignment. The degree of difficulty is Bl in expectation of the inability to meet the ±1% IV requirement. From the Table 3.1 application and difficulty matrix, a 4σ (99.994%) confidence level requirement is assigned. The calibration interval for the test system can be short, except its calibration is expected to be expensive and time-consuming. Six months is considered an acceptable target. Calibration of the test system will be done in a calibration and/or standards laboratory environment. Test system usage is planned to be 40 hours per week. Presuming that the test system's uncertainty will degrade exponentially, the MTBOOT requirement is 17,333,000 hours, corresponding to 99.994% measurement reliability and a 6-month calibration interval with 40 hour per week usage.

STAGE 7 — Measurement Systems Designs

The test system that will be designed to meet the Measurement Process Requirements stages is a series of three calibrated standard Deuterium lamps operating in an ambient air medium. These serve as 1–100 milliwatt/cm^2 power sources operating across the full power range at spot wavelengths in the spectrum of 120–400 nm with proper shielding and focusing hardware to assure that random uncertainty sources are minimized. Three lamps are used to meet the MTBOOT requirements, to allow process-controlled statistical intercomparisons of the three to increase measurement uncertainty confidence levels, and to compensate for the gaps in the wavelength spectrum. Also, measurement techniques will be devised so that the largest bias errors of the experiment's instrumentation system are corrected for in its embedded computer software, as are wavelength extrapolations. While an uncertainty of ±1% IV to 4σ for 6 months for the new design is not achievable, ±1% IV at 3σ for 4 months is. The 3σ at 4 months requirement results in an MTBOOT of 248,000 hours. By comparison, if the new system were to achieve a 4σ confidence level it would create a short calibration interval of only 65 calendar hours, or less than every three days. Conversely, if the original ±1.0% IV tolerances could be relaxed to ±1.33% IV, the 4σ at 4 months requirement could be met. However, this ±1.33% IV, "4x4" system needs an MTBOOT of 11,556,000. This would be the equivalent of saying that a ±1%, "3x4" system with a 248,000-hour MTBOOT is equal to a ±1.33% "4x4" system with a 11,556,000-hour MTBOOT. If an MTBOOT were too high to meet, designing to a lowered confidence level, a shorter interval, and a somewhat wider tolerance would allow a much lower MTBOOT and provide some design relief. The design will trade off use of expensive high-reliability components, parallel and redundant circuits, etc., for spending effort on a better understanding of the uncertainty estimation and improvement process. In the case at hand, a 25% tightening of tolerances from ±1.33% to ±1.0% netted a 4500% reduction in MTBOOT. This dramatic change is the result of the drop from an extremely high confidence level—4σ/99.994%—to a more moderate one—3σ/99.73%. Section 6 will shed more light on these intriguing trade-off possibilities.

STAGES 8 AND 9 — Calibration Process Requirements and Calibration System Designs

Requirements for the calibration system to support the test system are defined in terms of the need to calibrate the standard lamps and the related optical elements. Intercomparison devices and reference standard lamps will be needed in the calibration/standards laboratory to characterize and to determine the bias and precision errors of the lamps if they haven't been determined before. In any event, the bias errors must be determined periodically and either corrected out, or a certificate issued to tell the test system user the correction factors to apply when testing the instrumentation system. The same power and spectra requirements exist—1 to 100 milliwatts/square centimeter and 120 to 400 nanometers wavelengths. Per NHB 5300.4(1B), the calibration system uncertainty is to be 25% or less of the uncertainty of the test system. This results in a preliminary uncertainty requirement of ±0.25% for the calibration system. While a one-year calibration interval is desirable, due to the difficulty of sending the reference standard lamps to NIST for standardization, a six-month interval is chosen to reduce expected MTBOOT requirements, reduce the bias errors in the calibration process, and reduce calibration uncertainties. While the criticality of application is still Category 3.1, the difficulty of measurement is below Degree C, labeled OTHER on the matrix. This results in a confidence level and measurement reliability requirement of 2σ, or 95.45%. The usage of the calibration system is expected to be less than 1,500 hours per year because of its specialized application. The calibration system MTBOOT is 16,154 hours for a 95.45% measurement reliability, 6-month calibration interval, and 1,500 hours per year usage rate. From these requirements a new calibration system emerges that has an optical comparator, is environmentally controlled and vibration isolated, and uses a bank of three standard reference lamps and statistical analyses techniques for enhanced uncertainty determinations and control.

STAGE 10 — Measurement Traceability Requirements

To assure that the measurement processes are nationally and internationally correlated, the calibration system's reference standards need recalibration (standardization) at NIST or an equivalent facility whose measurement processes meet the NHB requirements and which are themselves internationally standardized.[3] The standard lamps used as references in the calibration system will be periodically rotated to NIST for calibration so fresh lamps, within their 6-month intervals are always in use. To maintain the high confidence levels called for, the bank of reference lamps in the calibration laboratory is intercompared with the freshly calibrated lamp from NIST to confirm that all are within uncertainty limits. NIST is requested to provide an estimate or realization of the absolute values of the power and spectra, or to provide corrections for bias differences discovered during the NIST standardization process. NIST is also requested to furnish correction factors for operation in vacuum, versus ambient air in the laboratory. For traceability to continue to the international level, NIST will send their national reference standard lamp or suitable transfer standard lamps to the Bureau International des Poids et Mesures (**BIPM**), and other nations' laboratories noted for lamp calibration competence (NPL in the UK, for example), and to confirm vacuum to air correction coefficients. This will assure that international standardization is controlled and that measurement uncertainty estimates are valid.

In this and other similar cases, each nation, including the U.S., has established reference standards for a particular quantity. They do not rely on a single international standard. Instead, they conduct periodic intercomparisons and measure the difference between the as-maintained standards representing a particular unit (here, the unit of irradiance—watt/meter^2).

During the intercomparison process, it is important to note that NIST should be requested to provide the uncertainty estimate for their measurement process and the confidence levels that accompany the estimates (so that adjustments to required program confidence levels can be made, if needed.) NIST should be requested to confirm that their measurement uncertainty estimates account for the degradation over time of their systems, so that when standardization values are "certified" by them, they warrant that the values are within the specified uncertainty limits to the confidence stated *at the time of their measurements.* This assurance is often unclear in NIST reports. (The calibration laboratory should also realize that its own standards' uncertainty will degrade with time.) Using the 25% (1/4) NHB ratio requirement, the uncertainty limit for NIST for the standard lamps is ±0.25%/4, or, ±0.06% IV at 2σ. This would be equivalent to ±0.09% IV at 3σ. If the NIST certificate showed an uncertainty estimate of less than 0.09% IV at 3σ, the uncertainties could be ignored as having a minor contribution to the calibration laboratory calibration chain. If the uncertainty is greater than the equivalent of 0.06% IV at 2σ, the uncertainty of the NIST value should be combined with the calibration laboratory uncertainty estimates for comparison with the program measurement requirements. It is desirable that the measurement uncertainties of the Solar Experiment instrumentation system should have been derived from the stack of uncertainties spilling down from international standards laboratories, through NIST, through the calibration laboratory, through the test laboratory to the solar instrumentation system. Performing these hierarchical calculations can be onerous, iterative tasks. The use of the NHB uncertainty ratios (1/10 and 1/4) between the layers of the measurement process chain simplifies this uncertainty assessment process. It allows independent relationships among laboratories as long as the uncertainty estimates of each can be trusted and fully stated and that the uncertainties are sufficiently small to meet the NHB ratio requirements. The problem is that uncertainty statements are rarely stated fully and adequately to execute sound planning and

---

3 Where measurements are being made with state of the art techniques, activities at all levels should be carefully coordinated with NIST to ensure traceability at the desired (or near) level.

requirement definition. Further, it is often impossible to meet ratio requirements because of limits in the state of the art of the measurement process. This topic will be explored further in Section 6.

> By pursuing the ten stages described here and establishing rigor throughout the measurement chain, adequate uncertainty definition is assured, weak spots are made visible, and compensation or corrections are applied to assure measurement process control.

3.2.8 Compensating for Difficult Requirements

It often seems that the most critical and difficult of measurements are the high-priority ones, yet they are the most apt to produce measurement requirements nearly impossible to satisfy. Often, a lack of capability is a result of state of the art limits: i.e., present technology has yet to produce the needed equipment or techniques of measurement, especially for long-term space-based situations. While technological development efforts should be pursued to resolve the fundamental uncertainty limit problem , especially on the higher priority measurements, parallel efforts to compensate for limits can be taken by any of the following actions:

- Measuring alternative, more easily measured parameters
- Making more independent measurements
- Retesting the end-item hardware at more frequent intervals, especially before deployment
- Relaxing end-item tolerances where no criticality category is endangered or when end-item quality is not degraded excessively
- Applying alternative measurement schemes of higher net accuracy
- Using embedded, intrinsic, or astronomical reference standards to improve long-term stabilities
- Using multiple sensors and measurement paths
- Applying computer enhancements with statistical process control methods.

These and other innovative compensation methods may be needed to meet severe measurement requirements for long intervals, high confidence, and low uncertainties.

3.3 Calibration Considerations

Measurement processes are accompanied by errors and uncertainties that cannot be eliminated. However, they can be quantified and limited or controlled to “acceptable” levels. Calibration is done for this purpose.

> Calibration compares the declared value of an attribute or parameter of a calibrating artifact, such as a reference standard, against the declared value[4] of an attribute of a unit under test (**UUT**).

When the UUT is a test instrument or another calibrating instrument, the result of calibration is usually a decision whether the calibrated attribute is within stated tolerances. Following calibration, the attribute may or may not be adjusted or corrected to within tolerance. When the UUT is used as a standard, its declared value is usually corrected and uncertainties involved in performing the calibration are reported. When the UUT is a parameter of a design prototype undergoing initial standardization, the calibrating artifact provides a reference against which parameter declared values are set. Uncertainties in the calibration are quantified and used to establish the parameter's specified tolerances

All measurements involve a stimulus and a response. Figures 3.1 through 3.3 illustrate the principal basic configurations.

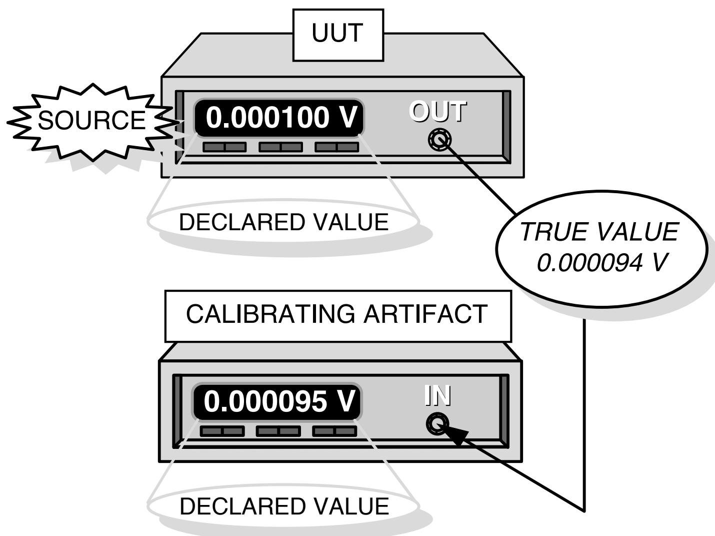


FIGURE 3.1 — CALIBRATION CONFIGURATION—UUT AS SOURCE.
In this configuration, a property of the UUT provides the stimulus. The UUT's declared attribute value is its nominal value or an indicated output. The calibrating artifact provides the sensor. The calibrating artifact's declared attribute value is displayed or otherwise shown.

From this, it can be seen that the question "why calibrate?" has been transformed into two questions: (1) Why quantify measurement error and uncertainty and control them to acceptable levels? and (2) What *are* acceptable levels of measurement error and uncertainty? To answer the first question, it will be useful to examine what is calibrated and why. As discussed in later sections, calibration comprises part of a measurement support infrastructure called the *test and calibration hierarchy*. In this hierarchy, fundamental standards are used to calibrate reference (interlab) standards that are, in turn, used to calibrate transfer standards that then are used to calibrate measurement devices.

---

4 Chapter 5 distinguishes between a "reference standard" and a "direct reading apparatus." The declared value of a reference standard is usually a documented quantity obtained through calibration with a higher-level artifact. The declared value of a direct reading instrument is usually a digital readout, a meter reading, or equivalent. In the simplest cases, the declared value is a nominal rating. Thus, the declared value of a 5-cm gage block, for example, is 5 centimeters. The concept of a declared value can be extended to components. For ex-ample, the declared value of a 100Ω resistor is 100 ohms.

The goal of calibration is the transfer of accuracy from a calibrating standard to an artifact that comprises an end-item or one that will be used to calibrate or test other artifacts. In this usage, the accuracy of the standard and the uncertainties in the transfer process are factors in establishing the subject parameter's tolerances. Following the test and calibration traceability down the vertical chain (see Figure 5.1), it becomes apparent that inaccurate reference standards beget inaccurate transfer standards, which beget inaccurate working standards, which beget inaccurate test systems, which beget inaccurate end-items and/or erroneous end-item test results.

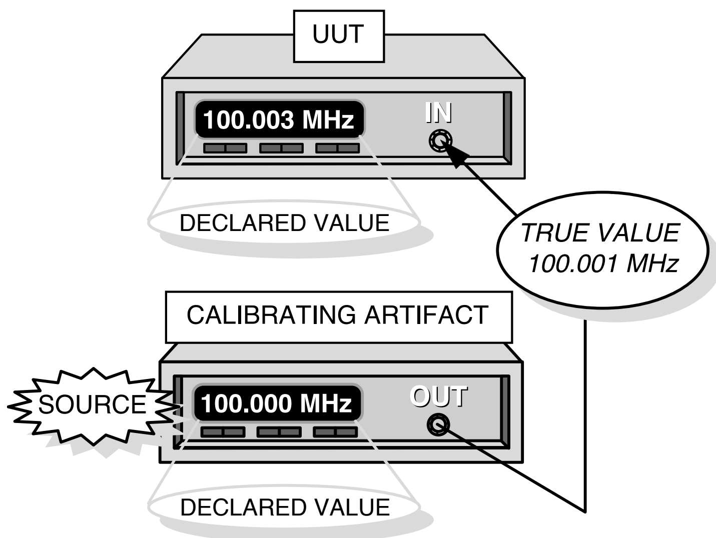


FIGURE 3.2 — CALIBRATION CONFIGURATION—CALIBRATING ARTIFACT AS SOURCE.

In this configuration, the calibrating artifact provides the stimulus. The calibrating artifact's declared value is its nominal or indicated value. The UUT provides the sensor. The sensor responds to the stimulus and drives a display. The displayed reading is the UUT's declared attribute value.

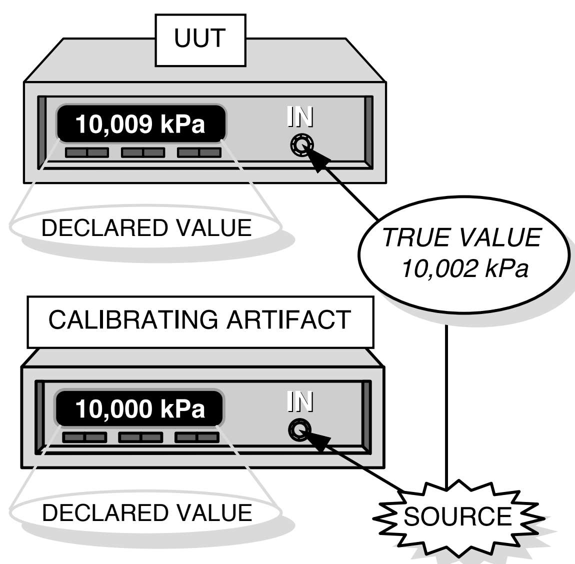


FIGURE 3.3 — CALIBRATION CONFIGURATION—EXTERNAL SOURCE.

In this configuration, the stimulus is supplied by a source external to both the calibrating artifact and the UUT. Each artifact responds to the stimulus and drives a display. The displayed readings are the calibrating and UUT's declared attribute values.

With these considerations in mind, the ultimate purpose of controlling measurement error and uncertainty within a test and calibration hierarchy (i.e., the ultimate purpose of calibration) is either the accurate standardization of end-item parameters, in design and development applications, or the control of erroneous end-item testing in product production and equipment management applications or scientific measurements.

Answering the question of what constitutes acceptable levels of error or uncertainty within test and calibration traceability calls for an analysis of the accuracy to which end-items must be standardized or tested. This accuracy should be established based on end-item performance objectives. For example, a level of uncertainty that is acceptable in day-to-day measurement applications, such as checking automobile tire pressure, may not be acceptable in highly critical applications, such as monitoring nuclear reactor core temperatures, or in state-of-the-art applications. Working backward from end-item accuracy requirements enables the quantification of accuracies needed for test system calibration. Working backward from these accuracies enables the determination of accuracies needed for calibration of calibrating systems, and so on. The method for doing an analysis of this kind is discussed in Section 4 and is presented in detail in Appendix C.

3.4 Space-based Considerations

3.4.1 Space-based Measurement System Implications

The designers of measurement processes and equipment intended for long-duration space operations should consider providing functional and physical metrology architecture designed to fit techniques and methodologies that will permit calibration and/or evaluation. The architecture should use self-calibration, self-test, self-monitoring, and stable reference standards technologies to minimize and facilitate space-based metrology control. The following should be considered:

- Design sound strategies for on-board calibration calling for minimum skill, a minimum of reference standards, and minimum interference with ongoing operations
- Institute a policy to ensure that on-board standards, including critical test equipment, are regularly calibrated in terms of national standards for measurement traceability
- Implement measurement quality assurance policies to ensure long-term measurement integrity
- Establish tolerances of measurable attributes commensurate with equipment performance objectives
- Verify that available test process[5] accuracies and stabilities are adequate for testing and monitoring end-item attributes
- Verify that available calibration process accuracies and stabilities are adequate for ensuring proper test process accuracies
- Verify that attribute stabilities are such that attribute values will stay within tolerance limits over the period of intended use with a specified level of confidence.

---

[5] In the context used, the terms “test process,” “measurement process,” and “TME” (Test and Measurement Equipment) are used interchangeably throughout this document and can be considered to be equivalent for practical purposes.

Calibration requirements created by long-term space-based missions pose special problems. Ease of calibration and minor repair or adjustment is frequently a low-priority item in the design of instrumentation. For example, unlike most other space-oriented hardware, equipment in a humanly-occupied space-based platform will need regular calibration access and adjustment over the platform lifetime. To meet this objective, lifetime calibration and maintenance requirements should be addressed during the earliest design phase.

A requirement for long calibration intervals means that high MTBOOT design targets will result. These will be difficult to meet unless the designs are very simple, minimize components used, and use redundant circuitry in critical measurement paths. Humanly-executed space-based calibrations are discouraged for several reasons, such as time, space, weight and priority considerations. For those measurement systems whose calibration intervals are estimated to be shorter than the mission duration requirement, special in-place calibration or interval extension schemes should be tried. The following should be considered:

- Provide internal instrument reference standards having a long-term accuracy commensurate with expected mission profiles
- Use built-in measurement standard references at selected points in the operating range
- Use carefully characterized astronomical artifacts as intrinsic-type measurement references, such as thermal, radiation, intensity, and noise references
- Use Earth-to-space-to-Earth comparison signals
- Replace unstable measurement system components with easily installed, small, modular, freshly calibrated units—use modular design to ease calibration, maintenance, and replacement
- Use higher accuracy (>10:1) measurement processes to compensate for increasing uncertainty over time so that the calibration interval matches the time where uncertainty growth has reached a point equal to a 10:1 process before recalibration is due
- Build in redundant and compensating measurement circuitry to improve reliability
- Provide physical adjustment points that are readily accessible without major disassembly of the equipment—all easily accessible adjustments should be sealed after calibration
- Use alternative or multiple measurement sensors with comparison devices
- Standardize easily accessible interfaces to instrumentation to simplify calibration
- Tighten end-item hardware tolerance requirements to create more conforming hardware that can tolerate the lowered confidence levels generated by the increasing uncertainty over time of the measurement process
- Provide access for sensor calibration and the capability of being calibrated in position or in place
- Design instrumentation and racking to allow complete calibration in place
- Make corrections and adjustments via software

- Measure end-items more frequently to assure higher confidence that parameter growth beyond performance limits is detected earlier and that a higher population of end-items are operating well within tolerances when deployed
- Use measurement statistical process control schemes to improve uncertainty.

These, and any other schemes that can be devised, should be considered to implement space-based calibration support. However, it should be cautioned that all measurement systems need complete calibration at some point to assure adequate continued performance.

> So-called self-calibration or self-test systems are useful, but are rarely substitutes for complete periodic calibrations—they serve mainly as interval expanders or limited range stopgap devices. Also, note that statistical measurement process control (**SMPC**) is a tool to analyze results and permit better decisions to be made. Ultimately, to ensure that any standard or instrument is "in calibration" calls for comparison to a known representation of the same unit.

Evaluating the adequacy of test and calibration process accuracies is done through measurement decision risk analysis. Further information on measurement decision risk analysis will be found in Section 4.

3.4.2 SMPC for Space-based Hardware

Measurement assurance support is usually viewed as a process in which the accuracy of a measuring instrument or system is maintained over its life cycle through either periodic calibration or testing. For items remotely operated and monitored, such as those deployed in space-based environments, periodic calibration or testing is more difficult than with terrestrial applications. In certain applications, such as deep-space probes, periodic calibration is nearly impossible. Exceptions are cases where terrestrial or astronomical references can be used. In such cases, the use of SMPC methods may be advisable.

SMPC methods enable the estimation of measurement parameter biases and in-tolerance probabilities through statistical intercomparisons of measurements made using closed sets of independent measuring attributes. A measuring attribute is regarded here as anything which provides a declared value, as interpreted in Section 3.3. In this sense, a measuring attribute may provide a measurement, a value comparison, or a quantified stimulus. Attributes in a set may be as few as two or as many as can be imagined. The set may include both calibrating units and units under test in either one-to-many or many-to-one configurations.

In traditional calibration and testing, the calibrators are ordinarily required to be intrinsically more accurate than the units under test. Therefore, measurements made by calibrators are held in higher regard than measurements made by units under test. If a calibrator measurement shows a unit under test to be out-of-tolerance, the unit under test is considered at fault. In making statistical intercomparisons, the SMPC methods do not distinguish between calibrators and units under test. Measurement intercomparisons provide bias and in-tolerance probability estimates for units under test and calibrators alike. Consequently, the SMPC methods can be used to evaluate the status of check standards as well as Test and Measurement Equipment (**TME**) workload items.

Check standard and TME recalibrations may be done on an attribute set without recourse to external references, if SMPC methods are applied under the following conditions:

(1) The measuring attributes in the set are statistically independent.

(2) The attributes in the set exhibit enough variety to ensure that changes in attribute values are uncorrelated (i.e., tend to cancel out) over the long term.

(3) Drift or other uncertainty growth characteristics of the attributes in the set that have been defined before deployment.

(4) The attributes in the set have been calibrated or tested before deployment.

If these conditions are met, application of the SMPC methods can serve to make payload measuring systems somewhat self-contained. This subject is covered in detail in Section 6.4 and Appendix D.

3.5 Software Considerations

Major measurement systems typically are computer-based. They contain software that can affect measurement quality. As the cost of computer hardware decreases, software will be contained in the smallest measurement systems. It is certain that the importance of software to measurement quality will increase during the life of this publication. Software development, and its effect on operations, is important to NASA's measurement processes.

3.5.1 Software Requirements

Software requirements for measurement systems should follow the requirements flowdown defined in the ten-stage sequence of Section 3.2.1. Also, two factors will make software use in NASA measurement systems particularly important:

(1) NASA measurements are often associated with spaceflight tests, where stringent time pressure because of launch commitment is typical.

(2) Software control of measurements for long-term spaceflight operations will often be more practical than hardware changes.

The potential need to change measurement system software quickly during testing and operations, makes it necessary to consider special software requirements.

(1) Software modularity, which will minimize effects of changes made under the duress of test conditions, should be stressed.

(2) Test cases that help objective definition of measurement uncertainty during the operations phase should be required.

(3) Software maintenance during the operations phase of long-term spaceflight missions should be given great emphasis.

(4) All requirements connected to the operations phase should be reviewed critically to make certain they are testable under the expected operations environment.

(5) Provision for regression testing targeted to the operations environment should be required, particularly for long-term spaceflight missions.

3.5.2 Software Development

Software development must follow a structured, accepted development method, such as NASA's Software Acquisition Life Cycle, to assure software quality. Besides normal software development methods, measurement software should consider:

(1) Verifying modularity by detailed inspections or walk-throughs that consider software changes made in the operations environment. These activities can start in the software architecture phase, then continue throughout the software development.

(2) Specifying exact hardware configurations for software test cases. Tests done during operations can then reproduce results obtained in acceptance tests, or provide objective explanations of the effect(s) of hardware changes. Measurement uncertainty monitoring during operations must also be based on a known hardware configuration.

(3) Documenting acceptance test results related to measurement quality in a form directly usable during operations.

3.6 Considerations for Waiver of Requirements

The effective implementation of the requirements normally results in a level of performance and risks acceptable to the project. Any deviation from these requirements usually requires a formally approved written waiver. The waiver should identify the risk resulting from the deviations and identify the original requirement(s), reason/justification for the request, and show what effect the waiver/deviation will have on performance, safety, quality, and reliability. The measurement classifications earlier discussed in Section 3.2.2 can aid in the preparation of a waiver request. The recommended standards for waiver or deviation requests are discussed in Section 8.

While it is intended that flight equipment be designed to perform within specification throughout the flight environmental design and test ranges, it must be recognized that sometimes out-of-specification performance at extreme flight environment limits may be justified and approved by waiver. For instance, an instrument or an engineering assembly may need complex sophisticated temperature compensation circuitry to provide in-specification operation throughout the required flight temperature range. Instead of incurring great cost, mass, and perhaps reliability penalties, an alternative approach would allow out-of-specification performance at temperatures near the extreme flight temperature range. This would be prudent for consideration when the following qualifying conditions exist:

(1) The out-of-specification performance is predictable and repeatable.

(2) The performance will be within specification when the flight equipment temperature range is within the allowable flight temperature boundaries.

(3) The out-of-specification performance will produce no permanent degradation in the flight equipment.

(4) The allowable flight temperature range will include all temperature prediction uncertainties and reflects not-to-be-exceeded limits in flight.

(5) The flight equipment development engineering organization can prove by analysis or test that the above four conditions hold true for the flight equipment being addressed.

Flight equipment components that have been characterized with proven temperature sensitivities incompatible with the product assurance environmental temperature ranges might be assigned tailored design and test temperature limits with an approved waiver.

4. MEASUREMENT SYSTEM DESIGN

4.1 Measurement System Design Approach

The previous section described the derivation of measurement requirements. This section provides the approach for design of measurement process hardware to achieve the required performance attributes established in Section 3. It identifies the various errors associated with the measurement process chain, reviews methods of combining errors, reviews the measurement system specifications established in Section 3, and presents a systematic design approach for measurement systems.

It is critical that the system designer provide visibility into the process of going from requirements to specifications to physical systems. A structured process enables timely and significant design reviews at critical points.

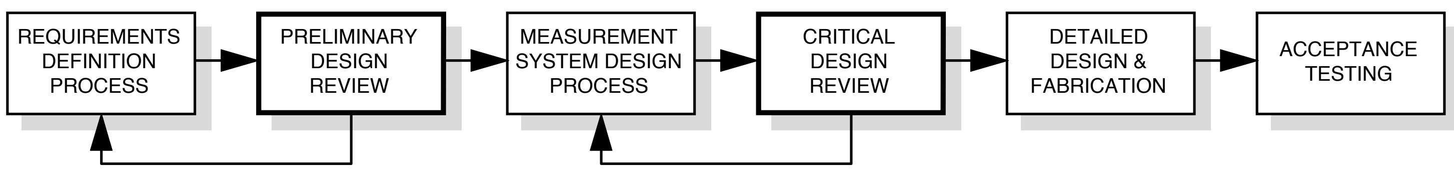


FIGURE 4.1 — OVERVIEW OF THE MEASUREMENT SYSTEM DESIGN PROCESS.

Figure 4.1 is an overview of the design process that features two essential reviews. One review is at the finish of the requirements definition phase and one is at the completion of design. Other reviews may also be incorporated to review progress on specific elements. Since the design focuses on supplying a system to satisfy the requirements, it is important that the *Preliminary Design Review* critique the requirements to establish completeness. For a measurement system, the requirements describe types of measurements (e.g., temperature, pressure, etc.), measurement range (e.g., ±100 KPa for a pressure measurement), required accuracy (e.g., ±0.1% full scale within 3 standard deviations for 1 year), bandwidth (e.g., 10 Hz), etc.

Once approved, the requirements document is usually placed under configuration control. The second major review is termed *Critical Design Review* and is a review of the system specifications and associated drawings. During this review, it is the responsibility of the designer to prove that each requirement has been satisfied by relating system specifications and attributes to requirements. The calibration methods necessary to achieve the required measurement system are presented and the measurement system specifications are established at this review.

An example of a measurement system design process of translating requirements into system specifications is illustrated in Figure 4.2. The process shown is for a digital measurement system (i.e., a system with analog inputs converted into corresponding digital format.) There are two key aspects of a digital system used in developing specifications—measurement uncertainty and bandwidth. First, regarding measurement uncertainty, error propagation techniques are used to decompose parametric measurement requirements into individual measurement requirements. Error budgeting and prediction methods are used with candidate equipment performance specifications to establish performance specifications for the various components of the measurement chain.

Second, bandwidth is a critical requirement that is decomposed and used to establish system specifications including anti-alias filter characteristics, sampling rates, and throughput.

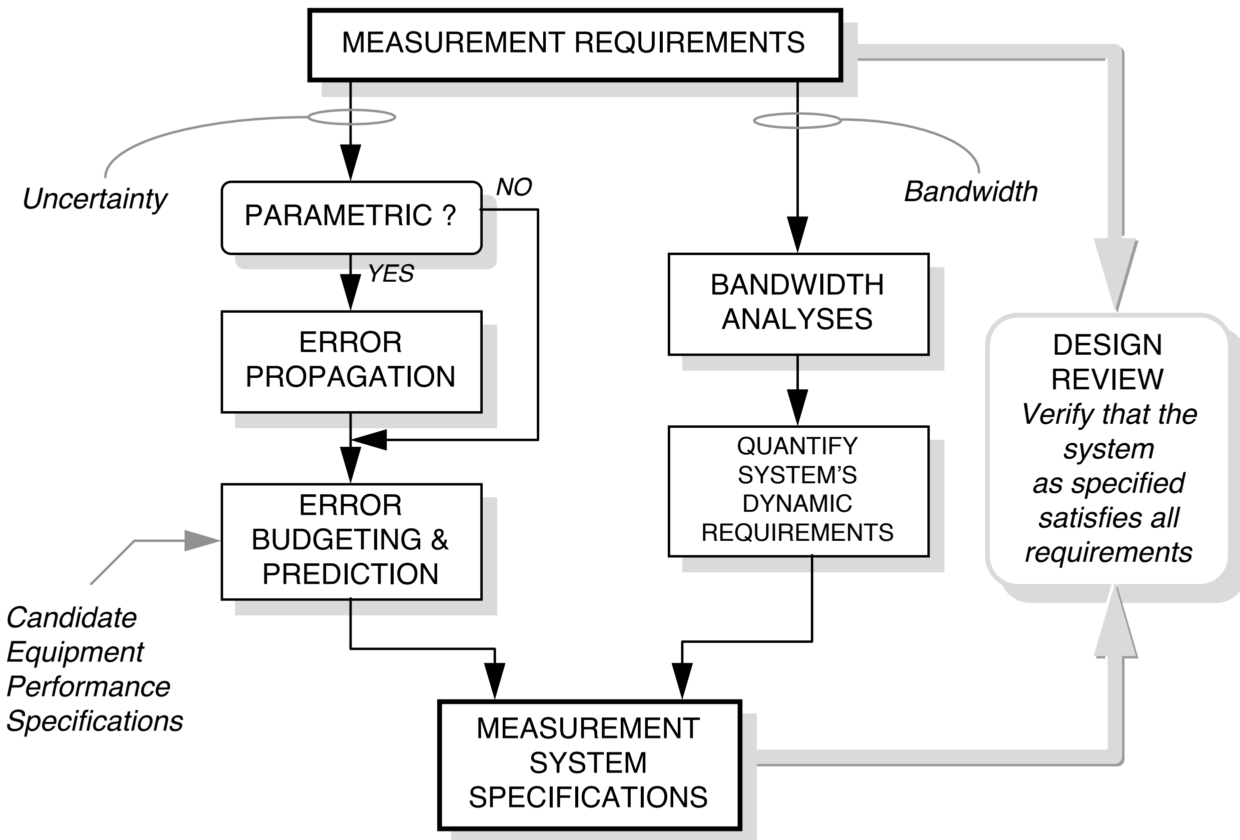


FIGURE 4.2 — EXAMPLE OF A MEASUREMENT SYSTEM DESIGN PROCESS.

It is assumed the measurement requirements have been analyzed to establish measurement system specifications and the measurement requirements have been formalized (Section 3).

> Once the specifications have been established, it is the designer's responsibility to prove that the system when built will comply with the requirements.

The specific steps associated with designing a measurement process are

(1) Identify physical phenomena to be measured and specific detailed requirements.

(2) Select candidate measurement equipment and interpret their specifications.

(3) Construct an error model of the process and predict measurement system performance, including MTBF/MTBOOT that match confidence levels and time limits.

(4) Identify calibration requirements.

(5) Evaluate the effects of changing environment on the measurement process.

(6) Manage the measurement decision risk.

4.2 Identifying Physical Phenomena to be Measured

At the least, the following information should be established where applicable for each measurement.

4.2.1 Process Characteristics

Establish the process characteristics and use this information in the selection of the sensors. The rate at which changes occur in the parameters being measured and the systematic or repetitive nature of occurrence are of special significance in determining how the measurement should be made. The two general classes of process phenomena are static and dynamic. Dynamic processes can be further divided into transient, periodic, and random. Time relationships are not as important in the measurement of static processes as in the dynamic process measurements.

4.2.2 Measurement Mode

Establish the required measurement mode. For example, determine if the measurements are direct, absolute, relative, differential, or inferential measurements. Direct measurement is feasible only in those cases where the measurand can directly actuate the sensor. There are many physical quantities for which direct detection is not possible: for example, mass flow, Mach number, or altitude. In such cases, one must rely on some functional relationship between the quantity one wishes to measure but cannot, and other related quantities that can be measured. For fluid flow measurements, determine whether the desired quantity is volumetric or mass flow.

4.2.3 Method of Transduction or Energy Transfer

The physical process that provides a usable output in response to the specific measurand should be identified. For example, when measuring temperature, establish the primary mode of heat transfer (conduction, convection, or radiation).

4.2.4 Measurement Location

Measurements are generally made at a point. As such, errors can result if there is a spatial gradient in the process. Also, the sensor installation may cause a process or system disturbance, such as the weight of an accelerometer on a light structure or the flow disturbance of a Pitot probe in a duct.

4.2.5 Measurement Range

Quantify the range of measured values. The setting of the parameter range should provide for the uncertainty in the actual range of the measurand. This measurement range is later used for establishing the “full scale” of the designed instrumentation system.

4.2.6 Measurement Uncertainty

Establish the acceptable measurement uncertainty over the required range and the required confidence levels and time limits.

4.2.7 Measurement Bandwidth

Quantify the frequency content of physical phenomena to allow establishment of filter bandwidths to pass the desired signal while suppressing noise and/or set digital sampling rates.

4.3 Selecting Candidate Equipment and Interpreting Specifications

For each measurement, select candidate equipment whose characteristics and performance are consistent. Since there are no industry standards regarding error definitions or performance specifications, one must use caution when interpreting manufacturer's performance specifications. Specification completeness and specification interpretation must be addressed.

4.3.1 Specification Completeness

The designer should review performance specifications for similar equipment from different manufacturers to determine whether the manufacturer has listed all relevant performance specifications for the candidate equipment. Note all omissions, and be attentive to specifications that differ significantly from manufacturer to manufacturer. Since each item specified can affect the measurement process depending on configuration and application, it is the designer's responsibility to determine which specifications are important for the specific application.

4.3.2 Specification Interpretation

Performance specifications for measurement equipment are quantified and published to describe a specific equipment's measurement attributes. There may be differences in the specifications among different manufacturers for similar items due to differences in the manufacturing and testing process. If the manufacturer integrates several subsystems together to form a product, the specifications will generally apply to the integrated system and not the individual subsystems. Thus, published specifications are assumed to reflect the manufacturer's testing process.

> *Beware* — Occasionally, manufacturer's specifications may be generated by the manufacturer's marketing department and may have only a casual relationship to the expected performance of measurement attributes. Establishing this relationship ordinarily falls to the user.

For measurement equipment, performance specifications can be categorized as either application-related performance specifications or intrinsic errors. For a data acquisition system, application-related performance specifications include source current, input impedance, input capacitance, common mode rejection, temperature coefficients, and crosstalk. The magnitude of errors resulting from these depends on the specific application. In contrast, intrinsic errors are those errors inherent to the system. Typical intrinsic errors include offset, gain accuracy, nonlinearity, hysteresis, repeatability, and noise.

Except for repeatability, drift, and noise, the intrinsic errors can generally be called bias errors. The manufacturer's specifications are interpreted to be absolute limits or windows for each error source. A gain error specification of ±0.1% full scale (**FS**) is interpreted to mean the gain error should be less than ±0.1% FS (within stated confidence levels and time limits). Manufacturer specs are statements of performance. If the manufacturer's specs will be used as references for estimating uncertainties, the instrument user needs to do the necessary calibration to ascertain these claims. Should an experiment be done which shows that the gain error exceeds ±0.1% FS, it can be concluded the equipment's gain performance is out of specification.

Intrinsic errors, such as repeatability and noise, are classified as precision errors. As such, they are normally distributed. The specifications for these errors must state either the statistically determined standard deviation (e.g., $\pm 3\sigma$) or the bounds. There is significant variation among manufacturers in reporting such precision errors as noise. Typical units specified include $\pm 3\sigma$, peak-to-peak, etc. Since noise depends on gain and bandwidth, the specification is incomplete unless both these parameters are given.

The requirement in NHB 5300.4(1B), *Quality Program Provisions for Aeronautical and Space System Contractors*, Section 9, *Article or Material Measurement Processes* establishes a tight requirement for the measurement system designer. It states that "random and systematic errors in any article or material measurement process shall not exceed ten percent of the tolerance of the parameter being measured." This 10% requirement (known to many as the 10:1 requirement) places much emphasis on the proper interpretation of the specifications furnished by the manufacturer of the measuring devices and accessories that will comprise the measurement system.

First, the accuracy or uncertainty specification needs close examination to assure that all the needed information is included for use in the system uncertainty computation equations. Usually, this information isn't available in the written specification. In addition to a statement of the measurement uncertainty of each parameter that the instrument measures, also needed is the time span (one month, 6 months, 3 years) that the uncertainty covers and standard deviations or σ confidence limits (one, two, or three) within which the stated uncertainty is contained. If this information is not available from specification sheets, the designer must go directly to the instrument manufacturer's engineers to determine those values.

Next, the environmental limit of the instrument must be determined to identify those contributors to other uncertainties that can and cannot be corrected or compensated for. These include thermal responses, vibration sensitivity, moisture effects, radiation effects, etc.

Finally, the "fine print" of the specifications must be examined to be sure there are no caveats regarding performance limits, such as loading effects, frequency response, interface impedances, data flow rates, line power fluctuations (regulation), distortion effects, etc.

4.4 Evaluating Measurement System Errors

> Understanding, identifying and quantifying the various error sources is a prerequisite for determining design adequacy and establishing calibration requirements.

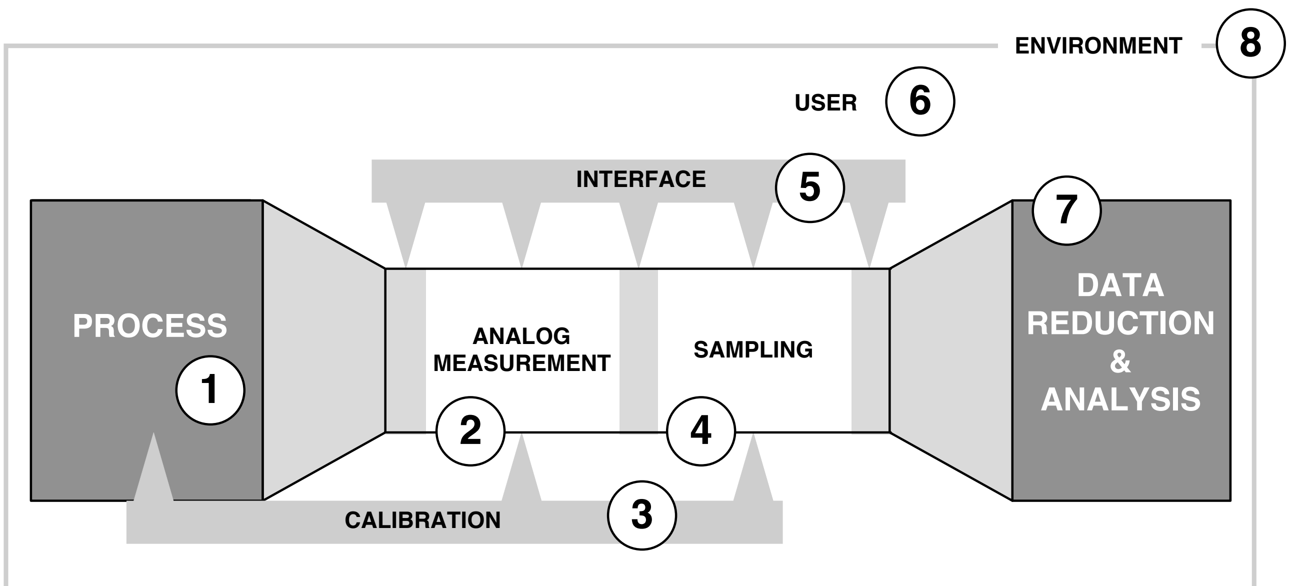
It is preferable to err on the side of providing too much information rather than too little. One should

- Clearly describe the methods used to calculate the measurement result and its uncertainty
- List all uncertainty components and document how they were evaluated
- Present the data analysis in such a way that each of its important steps can be readily followed and the calculation of the reported result can be independently repeated if necessary

- Give all correction factors and constants used in the analysis and their sources.

One should ask "Have I provided enough information in a sufficiently clear manner that my result can be updated in the future if new data become available?"

The individual measurement uncertainties established because of error propagation relate to the uncertainty of the complete measurement process and include many error sources, as illustrated in Figure 4.3. Knowledge of these errors is important in both establishing the estimate of uncertainty and in establishing the calibration requirements.



1. SENSOR, SPATIAL, INTERACTION
2. INTRINSIC EQUIPMENT
3. CALIBRATION & INTRINSIC EQUIPMENT/METHOD
4. QUANTIZATION, APERTURE, DISTORTION, STABILIZATION
5. CABLING, LOADING, CMV, EMI
6. HUMAN
7. TRUNCATION, ROUNDOFF, SOFTWARE, PROCESSING EQUATIONS
8. PRESSURE, TEMPERATURE, HUMIDITY INDUCED

FIGURE 4.3 — SOURCES OF ERROR WITHIN A MEASUREMENT CHAIN.

4.4.1 Sensing Errors

Measuring physical phenomena with sensors, which in themselves may influence the measurand's value, can introduce errors to the measurement process. Typical examples are: pressure measurements add volume; temperature measurements add thermal mass; and acceleration measurements add mass. Typical error sources in this category include spatial errors, interaction errors, and sensor errors. These are owed to disturbances caused by insertion of a probe in a moving fluid.

Sensing errors are generally omitted from uncertainty estimates because of the difficulty in quantifying this class of errors. However, this practice will nearly always lead to a significant underestimate of the total measurement process uncertainty. Figure 4.4 shows an example of sensing errors. Two thermocouples are inserted in a stream of flowing gas to measure the temperature rise of the gas. Heat is added to the gas immediately downstream of T_1. The temperature of T_2, the downstream thermocouple, is significantly higher than that of T_1 and the wall. The value of the bulk gas temperature rise at the two planes will be used in the data reduction equation:

$$Q = M\,Cp\,(T_1 - T_2)$$

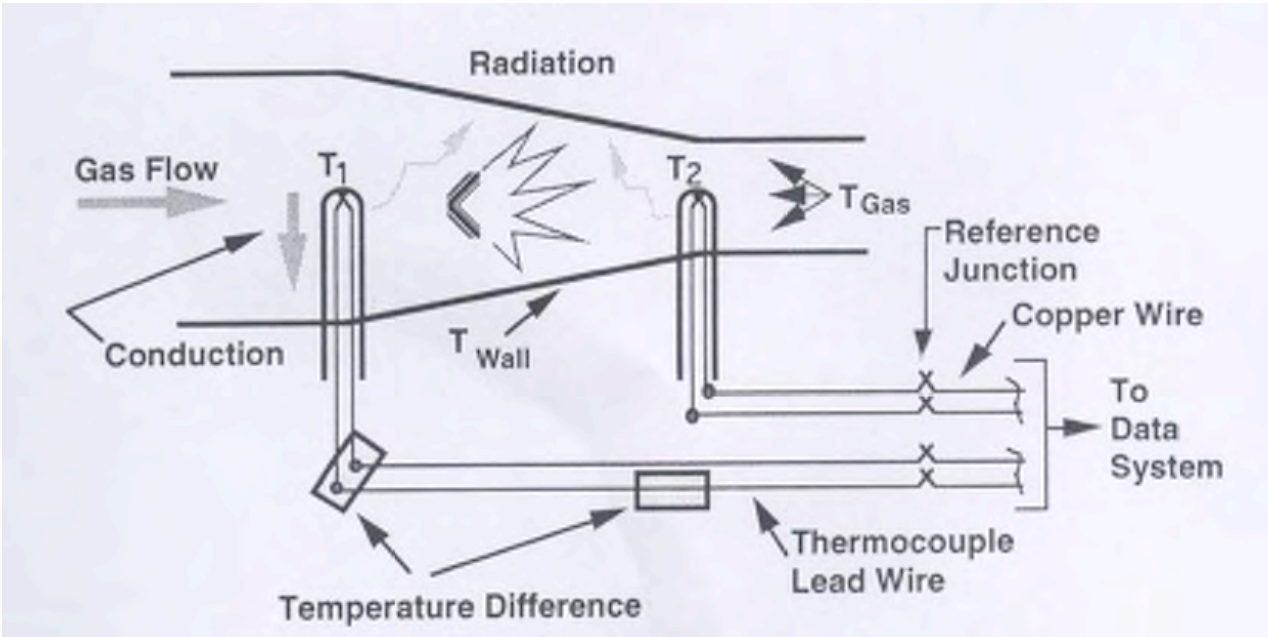


FIGURE 4.4 — EXAMPLE OF POTENTIAL SENSING ERRORS.

The following errors owed to the sensors can happen in this example:

- The gas will have a temperature gradient unless the wall temperature is equal to the gas temperature, which is not a realistic case. Each thermocouple measures the gas temperature at a single point, which will not represent the bulk gas temperatures.
- The velocity of the fluid flowing around the probe sets up a boundary layer complicating heat transfer from the fluid to the probe.
- The thermocouple probe conduction to the cold wall will lower the measured temperature from the measurand. Parallel conduction paths exist; the protecting sheath, the two thermocouple wires, and the insulating material. If T_2 is at a different temperature relative to the wall than T_1, the conduction errors will be different.
- Radiation from the thermocouple probe to the wall will lower the measured temperature from its value. The temperature will also be dependent on the respective surface conditions (i.e., emissivity or absorption) of the probe and wall.
- Thermocouple wire of the same type will have calibration differences resulting from slightly different composition.
- Temperature differences along the thermocouple wire may create errors because of in-homogeneity of the thermocouple wire and local work hardening of the wire.
- The increased resistance of the thermocouple wire, and resistive imbalance between the two different thermocouple materials, will increase the common mode voltage (**CMV**) error over that of copper wire.

- The response time of the thermocouple wire/probe will create a time-lag error in the measured value, depending on the dynamics of the measurand. The thermal mass of the thermocouple will influence the response time.

These, and other errors will cause the measured value to be different from the value needed for the data reduction equation—the temperature difference of the bulk gas. Analysis of these potential errors is necessary to disclose all uncertainties in the total sensing uncertainty.

4.4.2 Intrinsic Errors

The equipment that comprise a measurement chain, such as sensors, signal conditioners, amplifiers, etc., contribute to the measurement's error because of error sources inherent to the measurement and conversion system. This category includes such error sources as gain inaccuracy, nonlinearity, drift, hysteresis, offset, and noise.

If the magnitude and direction of the intrinsic error of a measuring attribute are known, the error can be factored out of measurements made on the attribute. Usually, the magnitude and direction of intrinsic errors are unknown. Yet, they can be accounted for statistically if their distributions are known. Often, information about the statistical distributions of intrinsic bias errors can be inferred from calibration history, as discussed in Section D.3 of Appendix D.

4.4.3 Sampling Errors

Representing a continuous phenomenon with a set of discrete samples introduces measurement errors. Typical error sources resulting from sampling are aliasing, aperture and resolution. These errors are generally minimized during the design process through analyses and later specification of filter characteristics, sampling rates, etc.

Converting continuous phenomena into a set of equally spaced discrete values introduces an error called aliasing by which high-frequency energy (either information or noise frequencies) manifests at lower or alias frequencies. The classic example used to show aliasing is the stagecoach wheel movement in a Western movie. The camera is operating at a fixed frame rate converting the continuous wheel movement into discrete values. What appears to be a reversal of the wheel movement is a result of aliasing. For a digital measurement system, aliasing can distort the measured value by introducing errors at various frequencies within the bandwidth of interest. System designers account for this by (l) filtering the analog signal to eliminate frequencies outside the band of interest and by (2) choosing sampling frequencies based on frequency and dynamic distortion considerations.

4.4.3.1 Overview of Aliasing

Aliasing is the process whereby two or more frequencies that are integral multiples of each other cannot be distinguished from each other when sampled in an analog to digital (**A/D**) converter. A folding frequency identifies the frequencies about which aliased data are folded down to the frequency range of interest.

> **NYQUIST FREQUENCY** — the frequency at which data are sampled at twice the upper data bandwidth limit. Also known as a folding frequency.

When data are sampled by an A/D converter, data from frequencies higher than the Nyquist frequency will fold like an accordion pleat down to frequencies ranging from one-half the Nyquist frequency down to the low-frequency limit of the system.

If the sampling rate of an A/D converter is less than the frequency components above the Nyquist frequency (f_N), the data will appear in the sampled data below f_N. This phenomenon is known as "aliasing." Data frequencies in the original data above f_N will be aliased and added to the data in the range $0 \le f \le f_N$ and defined relative to f_N by $f_{alias} = (2nf_N \pm f)$ where $n = 1, 2, 3, \ldots$

> Aliased data cannot be distinguished by a computer, nor can aliased data be eliminated after it has been sampled. Once A/D conversion is completed, there is no way to know from the sampled data whether aliasing has occurred. Even if it were possible to know, there is no way to correct the digital data for alias-induced errors.

Because aliasing can introduce errors into digital data, aliasing must be made negligible by assuring that the sampled analog signal has no significant components above f_N. This is accomplished by using analog low-pass filters at the input to the A/D converter. Under no circumstances should an analog-to-digital conversion be attempted without the use of analog low-pass anti-aliasing filters. It is very desirable that anti-aliasing filters have a flat frequency response over the widest possible range below the cutoff frequency (f_c). To provide a margin of safety, the upper value of f_c of the anti-aliasing filter should be set below f_N. The value of f_c relative to f_N depends on the anti-aliasing filter roll-off, the sampling frequency, the type of analysis to be performed, and the signal above f_N.

A/D conversion systems are being used that employ over-sampling. A relatively unsophisticated analog low-pass filter is used prior to the A/D converter to suppress aliasing in the original signal and the A/D converter operates at a much higher rate than is required for the data upper frequency limit. The over-sampled data are digitally filtered and decimated. The characteristics of the analog low-pass filter are not critical to the resulting data and the digital filter characteristics are much easier to control and are less costly.

Most low-pass filters produce frequency dependent phase shifts within f_c and may introduce errors that distort the data signal. In some analyses, the phase errors are unimportant (e.g., autospectrum analyses). However, amplitude domain analyses, such as probability density and distribution, as well as frequency domain analyses, such as shock response spectra and cross spectra, can be adversely affected. In addition, frequency response functions and time domain analyses, such as cross correlation, can also be adversely affected.

4.4.3.2 Description and Mechanism of Aliased Data

Figure 4.5 illustrates three sine waves, each one simultaneously sampled by the A/D converter. If the plots were laid over one another, the sampled points (indicated by the symbol X) would all lie on top of one another. A computer would reconstruct them into the same sine wave as the middle plot. The middle plot could be real data or could be aliases of the other two, or aliases of a theoretically infinite number of sine waves.

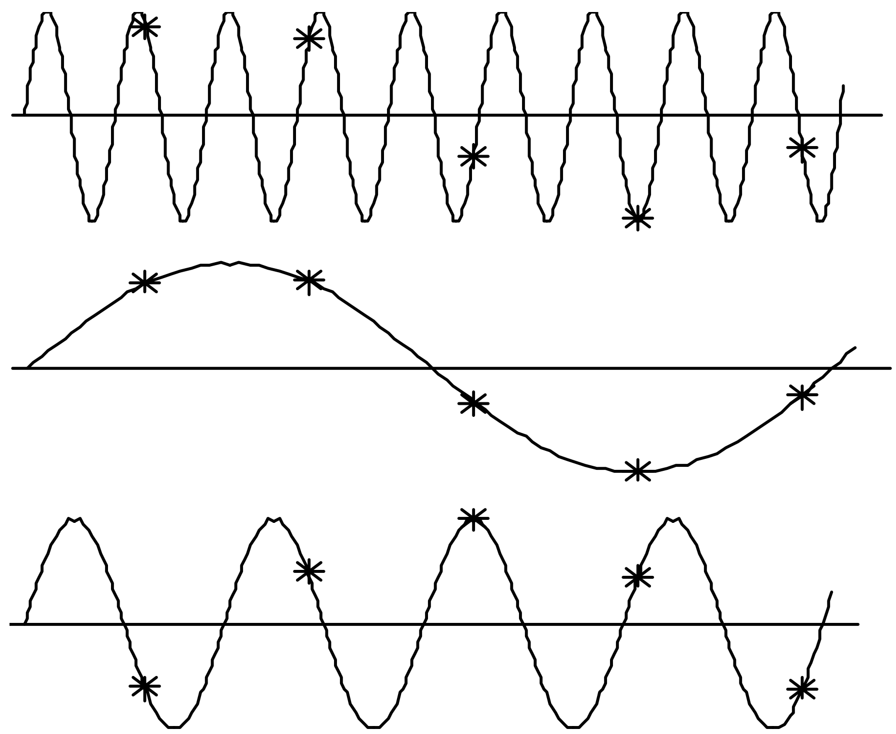


FIGURE 4.5 — SIMULTANEOUS SAMPLING OF THREE SINE WAVES.

The frequency of the top sine wave is nine times that of the middle sine wave, while the lower one is four times that of the middle one. Once the data are sampled, the computer has no way of distinguishing between the aliased data and the real data. The computer will reconstruct the data to the lowest frequency to fit the data points.

Figure 4.6 below shows how aliased data would appear in a continuous power spectral density (PSD) plot where data from higher frequencies are aliased down to the frequency range of interest.

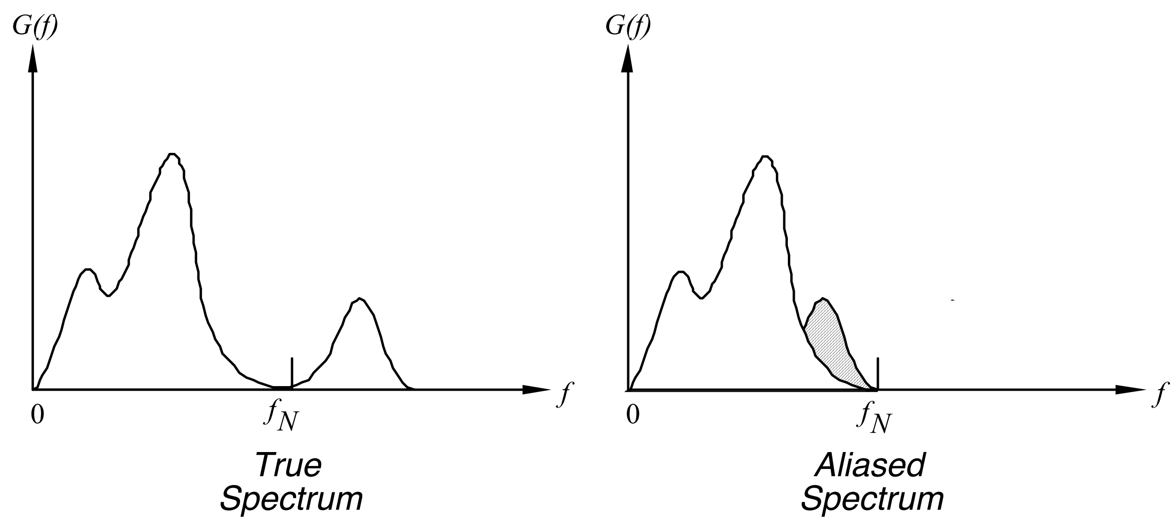


FIGURE 4.6 — POWER SPECTRAL DENSITY ALIASING.
The left-hand plot shows the true spectrum, while the right-hand plot shows the aliased spectrum as a result of folding.

Frequency folding from data above the Nyquist frequency occurs in an accordion-pleated pattern, as shown in Figure 4.7. Data sampled at integral multiples of data between 0 and the Nyquist frequency will appear in the frequency range of interest, as shown. If, for example, the Nyquist frequency is 100 Hz, data at 30 Hz would be aliased with data at 170, 230, 370, 430 Hz, etc. The dashed line crossings represent these frequencies.

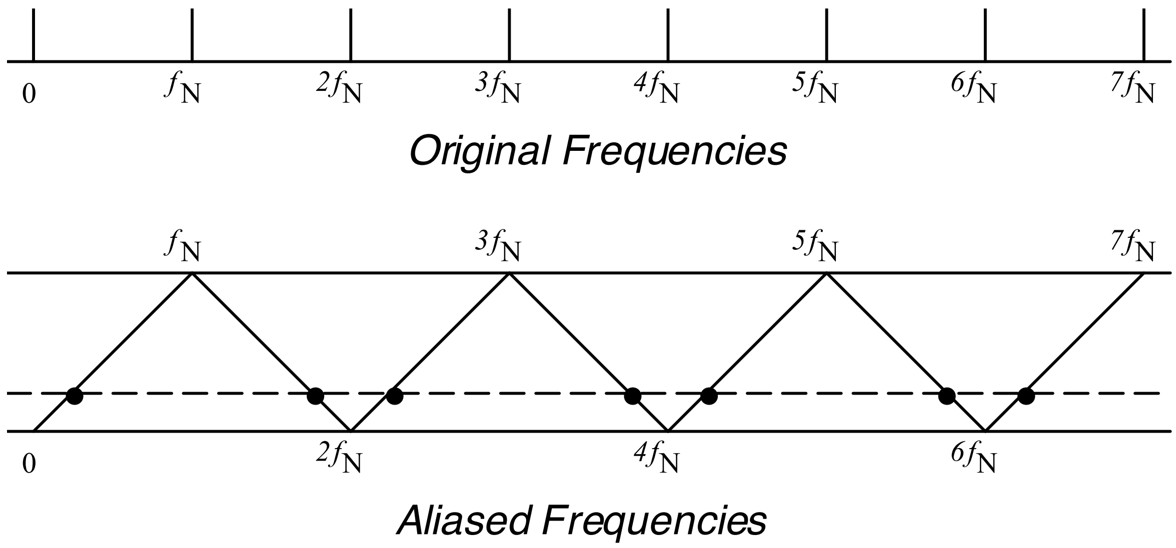


FIGURE 4.7 — DATA FOLDING RELATIVE TO THE NYQUIST FREQUENCY.

Data which can be aliased must be removed prior to sampling. There are two methods which can eliminate aliased data:

(1) The use of high-quality anti-aliasing filters.

(2) Higher sampling rates than all data frequencies, on the order of *at least 5 to 10 times* the highest significant frequency.

The advantages and disadvantages of these two methods are discussed below.

4.4.3.3 Methods for Avoiding Aliased Data

There are two methods that can be used to eliminate aliased data. The first method utilizes high-quality, low-pass anti-aliasing filters. When properly chosen and applied they eliminate the possibility of aliased data. In the second method, an unsophisticated low-pass filter with a high cutoff frequency f_c is used and the data are sampled at a higher rate so that no data can exist above the Nyquist frequency (over-sampled), and then digitally filtered and decimated. While both methods provide valid data, the first is preferred whereby the presence of unknown high-frequency signals can be aliased into the real data. If the existence of high frequencies are not a problem, then the second method is preferred. Analog anti-aliasing filters are more expensive than digital anti-aliasing filters, and the control of digital filter parameters is far superior.

Anti-Aliasing Filters — Analog filters are used prior to data sampling because once sampled, aliased data cannot be separated from true data. Digital filters alone will not eliminate aliased data because the data must be sampled prior to digital filtering. Two general types of filters are available for anti-aliasing: (1) constant amplitude filters, and (2) linear phase filters.

Constant amplitude filters, e.g., brickwall (elliptic) and Butterworth, have the advantage of a relatively flat frequency response within the passband. However, if not chosen properly they can exhibit large phase errors in the region of cutoff and have greater overshoot and ripple in response to a step function input.

Phase response of Butterworth filters is linear to approximately half of the cutoff frequency, but overshoot and ripple cannot be eliminated. If possible, Butterworth filters should be restricted to half the cutoff frequency in those cases where intrachannel phase response is a factor.

Properly designed brickwall filters can be obtained which have the best compromise between roll-off, intra- and interchannel phase response, overshoot and ripple. Intrachannel phase preservation is important in processing transients, e.g., shock response spectra. For cases in which interchannel phase is important, phase response between channels must be closely matched.

Linear phase filters, e.g., Bessel, exhibit very good phase response even beyond f_c, but the amplitude response starts to fall at approximately half f_c. Overshoot and ripple response to a step function is minimal over the frequency band. The rate of filter attenuation beyond f_c is less than the constant amplitude filters, requiring higher sampling rates to achieve the same anti-alias rejection as constant amplitude filters.

Anti-Alias Filter Selection Methodology — There are three variables to be considered in the selection of anti-aliasing filters: the rate of filter roll-off, the dynamic range of the system, and the sampling rate. The selection of one affects the others, so all must be considered together. Figure 4.8 illustrates filter selection with ideal constant amplitude filters. The method and result is the same for linear phase filters, except that filter roll-off beyond f_c is not as great as in the case of constant amplitude filters. The selection of filter type should be based on data acquisition system parameters, data processing, and analysis requirements in each case.

The filter must be chosen to provide sufficient roll-off to attenuate aliased data below the noise floor of the system where aliased data fold back within the data bandwidth frequency range. The noise floor is usually fixed in the system, so the filter characteristics are chosen to accommodate the signal to noise ratio (**S/N**). In addition to data foldover, the filter response is effectively folded over also.

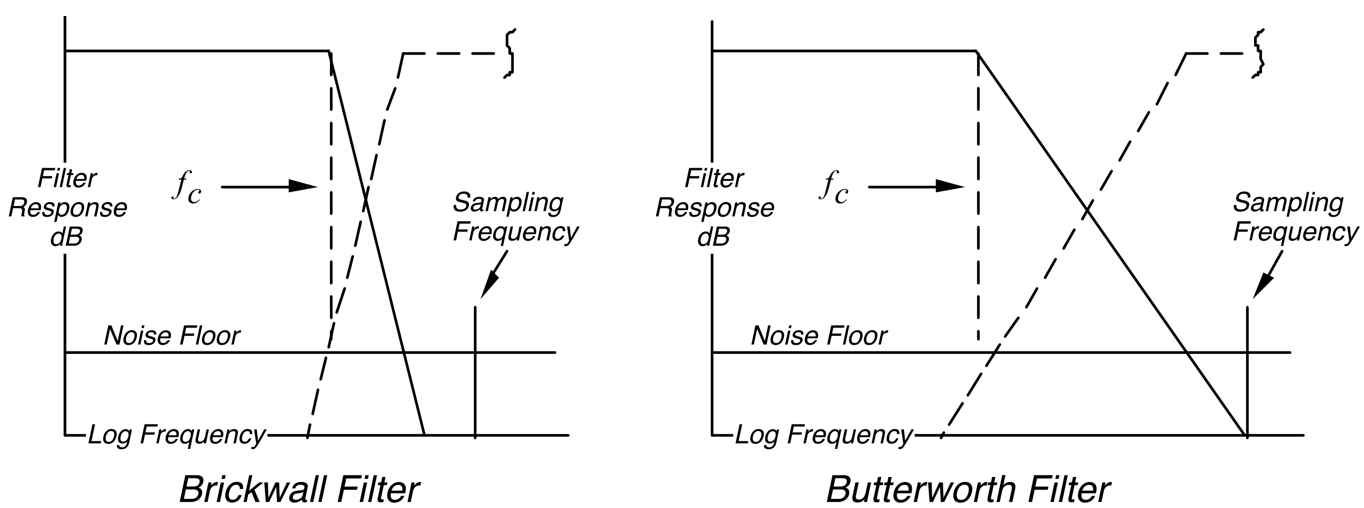


FIGURE 4.8 — ANTI-ALIASING FILTER SELECTION EXAMPLES.

The dashed lines represent the typical roll-off for the "folded" filters. The filter roll-off rate is compared to the system S/N at the frequency where the anti-aliasing filter response crosses the system noise floor.

The minimum sampling rate is set to at least twice the roll-off/noise floor crossing frequency. Even for the sharpest roll-off filters, the sampling rate should not be less than 2.5 times the data f_c.

If a white noise distribution is assumed, the S/N within the narrow resolution bandwidth of the analyzer can be considerably less than the data system bandwidth, because the energy in a narrow filter is less than that in a wide filter for noise of the same spectral density. The spectral analysis amplitude noise floor can be lower than the total system noise floor. This S/N is a function of frequency. In addition, the analog front-end and anti-aliasing filter may not have as much S/N as the data acquisition system. This can occur when the data acquisition system is designed to make use of the S/N available for large digital word lengths. For example, a sixteen-bit word length provides at least 90 dB of S/N.

Alias Elimination by High Sampling Rates — Data can be sampled at frequencies higher than the highest frequencies in the data sample. This presupposes a knowledge of the frequency distribution of the data sample. Current data systems are of high quality, but they may suffer from spurious inputs from such unintentional manufacturer design flaws as intermodulation distortion. Intermodulation can occur between telemetry bands, crosstalk between data channels, and crosstalk between heads on an analog recorder, etc. A high-frequency spectral analysis may be required to determine whether spurious signals can be aliased down to the data frequency band from higher data frequencies than expected. While this is a valid method to eliminate aliases, the uncertainty of the data content above the sampling rate poses some risk.

After the data are sampled, digital filters and decimation are used to limit the data to the desired frequency range. Control of digital filter parameters is far superior to that of analog filters. For that reason, the method is preferred by some data-processing experts.

4.4.3.4 Phase Distortion

Phase distortion is the deviation from a straight line of the phase in a frequency versus phase plot. Phase distortion of a complex waveform translates into amplitude distortion. In computing the power spectral density of a time history, the relative phase of each of its components does not change the value of the data. Yet, the amplitude distortion can cause an error in the computation of shock response spectrum. All filters in the data acquisition and analysis systems will affect phase distortion, and therefore, the shock response spectrum. These errors will be a function of the relative amplitudes of the spectral components, the frequencies of the spectral components, and the phase in different transients. Because of the random distribution of the amplitudes, frequencies, and phase in different transients, each time history will exhibit errors that will result in different errors for each. If a given time history is repeatedly analyzed (and no other errors exist) then the data will consistently have the same errors and the same shock response spectrum will be computed each time. This will instill a false sense of confidence in the user.

4.4.4 Interface Errors

The equipment and cabling of a measurement chain is characterized by such electrical properties as resistance, capacitance, etc. These input/output properties may change as either a result of connecting equipment or the environment. Typical error sources in this category include loading, CMV, noise, cabling, and crosstalk. Many of these errors, caused by loading, CMV, etc., are addressed during design and analyses used to establish specifications, such as common mode rejection ratio (**CMRR**), crosstalk specifications, input/output impedances, etc.

4.4.5 Environment Induced Errors

Variations in temperature may affect the measurement system by introducing such error sources as offset and gain. These errors are generally minimized during the design process through analyses and subsequent specification of temperature coefficients, and specifications for environmental conditioning of temperature sensitive equipment. Also, the designer must include analyses of other environmental factors, such as humidity and altitude (pressure), depending on the specified end use of the system.

4.4.6 Calibration Induced Errors

Calibration equipment and procedures are usually incorporated into a system during design to provide a way to quantify and eliminate bias errors. While there are errors associated with the calibration process, these may generally be considered negligible if the ratio of permissible uncertainty (tolerance) of the calibration to calibrating equipment uncertainty is about four or more.

The rationale behind this assumption is as follows. Let ε_1 represent the permissible uncertainty of the calibration and ε_2 represent the calibrator uncertainty that is given as $\varepsilon_2 \leq 0.25\ \varepsilon_1$. Assuming that the errors are statistically independent, they root-sum-square (**RSS**) as follows: where ε_T is the observed error in the calibration process:

$$\varepsilon_T = \sqrt{\varepsilon_1^2 + (0.25\varepsilon_1)^2}$$

$$\varepsilon_T = 1.03\ \varepsilon_1\ .$$

The error induced using a calibrator that is about four times as good as the system being calibrated is about 3 percent of the system error.

If the accuracy ratio of the calibration standard is not sufficiently high, then the uncertainty associated with the standard is included as an error source in the determination of bias uncertainty. A more complete discussion is given in Sections 5.1 and 5.7.

4.4.7 Data Reduction and Analysis Errors

Correlation of data reduction methods and the characteristics of the measurand must be an important part of the design activity. The application of software must be well understood to prevent errors from such sources as misapplied algorithms, truncation, and roundoff. The potential for software induced errors during data reduction cannot be ignored. The software issues discussed in Section 5.9 should be given full consideration.

4.4.8 Operator Errors

Human errors, especially in the operational phase of the work, may be a significant error source. This is particularly true if manual data acquisition methods are used. Human errors may cause gross mistakes that will show good data points as outliers, which might be removed erroneously.

4.4.9 Error Propagation

Often, multiple measurements are needed to establish a parameter. For example, consider the parameter *Specific Impulse* that is computed based on measurements of thrust and propellant flow.

Since there is an uncertainty associated with each of these measurements, there is an uncertainty associated with the parameter *Specific Impulse*. The Taylor Series Expansion is a numerical technique that is often used to describe the relationships between individual measurement uncertainties and parameter uncertainty at an operating point.

Consider a parameter F that depends on several measurements denoted M_i.

$$F = f\left(M_1, M_2, ..., M_n\right) \tag{4.1}$$

To a first-order approximation, the change in the function F, denoted δF, is related to the changes in the measurements M_i, denoted ΔM_i as follows:

$$\delta F = \left[\frac{\partial F}{\partial M_1}\right]\Delta M_1 + \left[\frac{\partial F}{\partial M_2}\right]\Delta M_2 + ... + \left[\frac{\partial F}{\partial M_n}\right]\Delta M_n \tag{4.2}$$

where the partial derivatives $\left[\frac{\partial F}{\partial M_i}\right]$ will be evaluated at an operating point. This is a simplification of the Taylor Series Expansion. It is assumed the partial derivatives exist at the point and that the remainder term is zero. Since measurement uncertainty can, for practical purposes, be considered a randomly distributed variable, it has been a common practice to change Eq. (4.2) as follows:

$$\delta F = \sqrt{\left(\frac{\partial F}{\partial M_1}\Delta M_1\right)^2 + \left(\frac{\partial F}{\partial M_2}\Delta M_2\right)^2 + ... + \left(\frac{\partial F}{\partial M_n}\Delta M_n\right)^2} \tag{4.3}$$

where the ΔM_i are interpreted to be measurement uncertainties.

From a design viewpoint, the parameter uncertainty, δF, is a stated requirement along with the parametric relationship F. The unknowns are the allowable individual measurement un-certainties, ΔM_i.

Since Eq. (4.3) has n unknowns, a unique solution does not exist. Equation (4.3) gives the designer a mechanism for budgeting uncertainties to each of the n measurements. The examples in Sections 4.5.6 and 4.5.7 are prepared to illustrate the technique.

4.5 Combining Errors

Once we have determined the sources of the various measurement system errors, we need to have a method for quantifying them and combining them into a single estimated uncertainty value.

4.5.1 Error Classifications

The various error sources of a measurement process can be categorized as either *bias errors* (fixed or systematic errors) or *precision errors* (random errors.) The bias error is the difference between the mean of the measured values and the measurand value shown in Figure 4.9. The magnitude of this error is important if the absolute accuracy is required. If repeated observations of the measurement are made, the observed values will appear to be randomly distributed about the mean value. The repeatability of the measurement depends on the precision errors. If, at a specific value of the measurement, the bias and precision errors are known, they can be combined to establish an estimate of the uncertainty associated with the measurement.

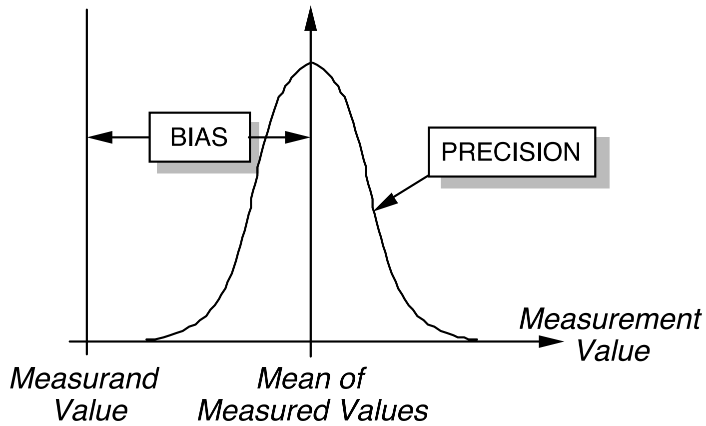


FIGURE 4.9 — COMPONENTS OF MEASUREMENT UNCERTAINTY.

Unlike experimental approaches that can be used to quantify a specific measurement system's error, the designer's task is to

- Estimate the uncertainties of a proposed measurement chain by analyzing the measurement process
- Quantify the error sources using manufacturer's specifications, analysis, and/ or engineering judgment
- Combine the error source uncertainties to establish an estimate of measurement uncertainty.

Estimates of standard deviation confidence limits usually are difficult to obtain from manufacturer's literature, as are performance time limits. It is recommended that the manufacturer's engineering staff be contacted directly for this information.

To aid the designer, Table 4.1 is provided as a guide for interpreting and establishing estimates of uncertainties for the various error sources.

TABLE 4.1 Error Source Classifications

TABLE 4.1
Error Source Classifications

ELEMENTAL ERROR	ERROR CLASSIFICATION	ESTIMATION METHOD
SENSING ERRORS		
Spatial	*Bias*	*Engineering Judgement*
Interaction	*Bias*	*Engineering Judgement*
Probe	*Bias*	*Engineering Judgement*
INTRINSIC ERRORS		
Offset	*Bias*	*Manufacturer's Specs*
Gain	*Bias*	*Manufacturer's Specs*
Nonlinearity	*Bias*	*Manufacturer's Specs*
Hysteresis	*Bias*	*Manufacturer's Specs*
Repeatability	*Precision*	*Manufacturer's Specs*
Drift	*Precision or Bias*	*Manufacturer's Specs*
Noise	*Precision*	*Manufacturer's Specs*
Source Current	*Bias*	*Manufacturer's Specs*
SAMPLING ERRORS		
Aliasing	*Bias*	*Application Analysis*
Aperture	*Bias*	*Application Analysis*
Resolution	*Bias*	*Manufacturer's Specs*
INTERFACE ERRORS		
CMV	*Bias or Precision*	*Application Analysis*
Noise	*Precision*	*Application Analysis*
Cabling	*Bias or Precision*	*Application Analysis*
Crosstalk	*Bias or Precision*	*Application Analysis*
ENVIRONMENT INDUCED ERRORS		
Offset	*Bias or Precision*	*Application Analysis*
Gain	*Bias or Precision*	*Application Analysis*

4.5.2 Common Units and Confidence levels

Different units, such as % Full Scale, % Reading, μV RTI, mV RTO, etc., are used by manufacturers to specify equipment performance. Therefore, it is necessary to pick a common unit and to convert all error source uncertainty. For a specific application with candidate equipment, this will call for establishing such operating conditions as signal levels, gain, and bandwidth parameters. Once selected, all error source uncertainty should be converted into the same units.

> The uncertainty value should be of the same confidence level. Manufacturer's specs can be 1, 2, or 3σ, and typically, engineering judgment is a 2σ estimate. To achieve a meaningful combining of error sources, they must be converted to common units and confidence levels.

4.5.3 Establishing the Total Bias Estimate

At a specific measurement value, the various biases listed in Table 4.1 are established and combined to provide the measurement's total bias, B_T. At a different measurement value, these elemental biases are in themselves variables with unknown distributions.

From a design viewpoint, the error sources reported by manufacturers as specifications represent ranges (e.g., nonlinearity of ±0.1% FS.) The uncertainty for any error source can be interpreted to be the specified value with a confidence level depending on the standard practice of the manufacturer. The confidence level of the uncertainty must be determined for each bias source.

There are various ways of establishing estimates of how these bias error sources, such as non-linearity, hysteresis, offset, etc., combine to form total bias. These include summing the absolute values of all bias error sources to form total bias and applying the Root-Sum-Square (**RSS**) method. For example, the RSS can be used to establish an estimate of total bias, as follows:

$$\text{Bias, } B_T = \sqrt{b_1^2 + b_2^2 + \ldots + b_n^2} \tag{4.4}$$

While there is no mathematical basis for using the RSS method to establish B_T unless all terms are statistically independent, the rationale behind using this method is that it does provide for some error cancellation. It is unreasonable to assume that all the biases are cumulative. In practical measurement chains, there will be a canceling effect because some errors are positive and some are negative.

In combining nonsymmetrical bias limits, apply the RSS method to the upper limits to determine the combined upper limit. The lower limits should be treated likewise. The result will be nonsymmetrical bias limits.

Using the above methods of combining biases to establish an estimate of total bias is considered conservative, but the effects of calibration methods have yet to be considered. It is here in the design process that calibration and the frequency of calibration are established based on a consideration of the biases and their magnitudes. The estimate of total bias would then be adjusted accordingly.

The concept of the total bias is relevant to the above discussion. The total bias is the difference between the measurand's value and the mean of the measured value. A calculated total bias uncertainty is derived during design activities from the manufacturer's data of such bias error sources as shown in Table 4.1. The calculated total bias is dependent on sources that include unknowns. Further, the measurand's value is not known, so there is usually no rigorous equation that defines the bias error. The calculated bias, calibrations, verified manufacturer's data, and comparisons with other measurements by independent methods will help the effort to estimate the total bias. *But, generally the estimate of total bias error must be based on engineering judgment.*

4.5.4 Establishing the Total Precision Estimate

A review of the error classifications in Table 4.1 shows that the errors generally classified as precision errors are repeatability and noise. Of these, noise is generally the dominant uncertainty.

Within a measurement system, the primary noise sources include noise generated by thermal processes within conductors and semiconductors, white noise generated by thermal processes within resistors, and systematic noise such as that caused by line frequency, power supply ripple, electromagnetic interference, digital logic, etc. Active system elements, such as amplifiers, are principal sources of noise. Since the magnitude of noise depends on both gain and bandwidth, the manufacturer's specifications should include a measure of the magnitude of the noise and the corresponding gain and bandwidth.

The RSS technique is also the method commonly used to establish an estimate of total precision. The mathematical basis assumes that these elemental precision uncertainties are randomly distributed and statistically independent. Thus,

$$\text{Precision,} \quad B_T = \sqrt{s_1^2 + s_2^2 + \ldots + s_n^2} \tag{4.5}$$

This is also called the precision index. Note that since these are random variables, the magnitude of each precision uncertainty is generally expressed in terms of standard deviation (i.e., ±1 σ represents 68.3%, ±2σ represents 95.5%, ±3σ represents 99.7%, etc.) Thus, precision errors must be adjusted to the same sigma level before they are combined.

4.5.5 Establishing the Total Uncertainty Estimate

Measurement uncertainty, U, is a function of bias and precision. To combine the two separately estimated uncertainties, two methods are currently accepted: U_{ADD} and U_{RSS} .

$$U_{ADD} = \pm\left(B_T + t_\alpha S_T\right) \tag{4.6a}$$

$$U_{RSS} = \pm\sqrt{\left(B_T\right)^2 + \left(t_\alpha S_T\right)^2} \tag{4.6b}$$

where t denotes the Student T statistic and α is the confidence interval.

If the bias and precision error estimates are propagated separately to the end test result and the equation used to combine them into uncertainty is stated, either U_{ADD} or U_{RSS} can be used. Monte Carlo simulations were used in studies to compare the additive (U_{ADD}) and root-sum-squared (U_{RSS}) values. The results of the studies comparing the two intervals are:

- U_{ADD} averages 99.1% coverage, while U_{RSS} provides 95% coverage based on bias limits assumed to be 95% (2σ for normally distributed biases and 1.65σ for rectangularly distributed biases.
- U_{ADD} averages 99.7% coverage, while U_{RSS} provides 97.5% coverage based on bias limits assumed to be 99.7% (3σ for normally distributed biases and 1.73σ for rectangularly distributed biases).
- Because of these coverages, U_{ADD} is sometimes called U_{99} and U_{RSS} is called U_{95}.
- If the bias error is negligible, both intervals provide 95% confidence.
- If the precision error is negligible, both intervals provide 95% to 99.7% confidence depending on the assumed bias limit size.
- When the interval coverages are compared, U_{ADD} provides a more precise estimate of the interval size (98% to 100%) than 93% to 100% for U_{RSS}.

The "Student T" value is a function of the degrees of freedom (ν). The degree of freedom ν is the number of observations in the sample (the sample size) minus the number k of population parameters that must be estimated from these sample observations. For large samples (i.e., $N > 30$), t_α is set equal to 2. It is acceptable practice for t_α to be taken as 2 during the design process. This corresponds to a 2σ (95.45%) confidence level.

The key procedure in establishing total uncertainty estimates is as follows:

(1) Study the measurement system and data algorithm to figure out which elements must be considered in the uncertainty analyses.

(2) For each measurement, make a list of every possible error source and estimate its uncertainty interval based on its correspondence to a set confidence level.

(3) Classify the uncertainties according to the categories of bias and precision.

(4) Propagate the total bias and precision index to the end-measurement results, as described earlier.

(5) Calculate total uncertainty by one or both methods as shown above.

(6) Document the bias, precision and total estimates and the uncertainty formulas used.

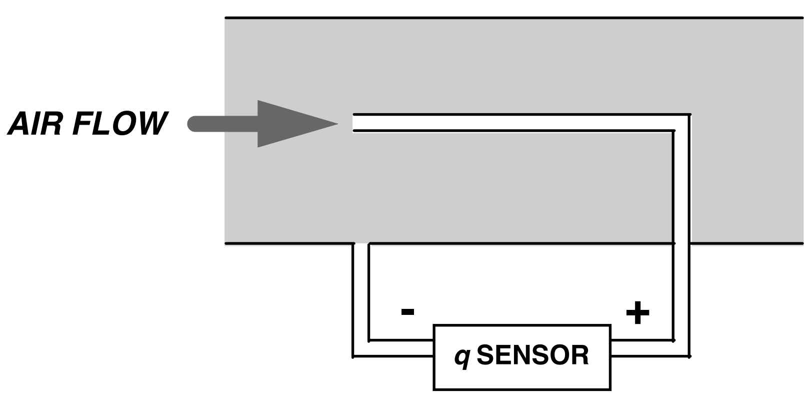
Documentation of the methodology used is as important as the choice of methodology.

4.5.6 Example—Budgeting Measurement Uncertainty in the Design Process

Consider the requirement to develop a measurement system to measure the velocity of air in a low-speed duct with a Pitot static probe (see sketch below). Using the Bernoulli equation for incompressible fluids, the velocity, V, is related to the difference between the Pitot pressure and the stream static pressure, which here is q, and to fluid density, ρ, as follows:

$$V = \sqrt{2q / \rho}$$

where q is in units of pascals (N/m^2), ρ is in units of kg/m^3, and V is in units of m/sec.



The requirement is that the uncertainty in velocity must be less than ±1% V at 3σ when q equals 2400 Pa. For this example, assume that fluid density, ρ, is given as 1.000 kg/m^3. How accurate must the q measurement be to achieve ±1% V?

Approach

Using error propagation, the expression for the uncertainty in V, (δV)

$$\delta V = \sqrt{\left(\frac{\partial V}{\partial q}\Delta q\right)^2 + \left(\frac{\partial V}{\partial p}\Delta p\right)^2}\ .$$

Since we have one variable, the above simplifies to

$$\delta V = \pm\left[\frac{dV}{dq}\middle|\Delta q\right]$$

The derivative is

$$\frac{dV}{dq} = \frac{1}{2}\frac{1}{\sqrt{\frac{2q}{\rho}}}\frac{2}{\rho} = \frac{1}{\sqrt{2q\rho}}\ .$$

At this dymamic pressure, $\frac{\partial V}{\partial q}$ = 0.0144, V = 69.3 m/sec, and ∂V = ± 1% = ± 0.693 m/sec.

Thus, the maximum allowable error in the q measurement is

$$\Delta q = 0.693 / 0.0144 \approx 48 \text{ Pa or } \pm 2\% \text{ or Reading, at 3 sigma.}$$

An alternate method of determining the design requirement measurement of q is as follows:
From $\delta V = \pm\left[\frac{dV}{dq}\middle|\Delta q\right]$, multiply by dq and divide by V

$$\frac{\partial V}{V} = \frac{1}{2}\frac{1}{q}dq\,,$$

and, therefore, the measurement requirement for q is 2% for a 1% measurement of V.

Interpretation of Solution

The computed uncertainty in q of ±2% reading at 3σ is the specification for errors in the measurement including sensor, data acquisition, etc., and applies only when q = 2400 Pa.

> ***Note*** — While calculus was used to establish the derivative (the sensitivity of V to changes in q), this could have alternatively been established numerically, as follows:

$$\frac{dV}{dq} = \text{Change in } V \,/\, \text{Change in } q$$

Let q change from its base value of 2400 by 1%. Thus,

$$V_0 = \sqrt{2q_0/\rho} = \sqrt{2\frac{(2400)}{1}} = 69.28 \text{ m/sec}$$

$$q_1 = 0.99 \quad q_0 = 0.99\ (2400) = 2376 \text{ Pa}$$

$$V_1 = \sqrt{2q_1/\rho} = \sqrt{2\frac{(2376)}{1}} = 68.93 \text{ m/sec}$$

$$\frac{dV}{dq} \approx \frac{V_0 - V_1}{q_0 - q_1} = \frac{69.28 - 68.93}{2400 - 2376} = 0.0145 \, .$$

4.5.7 Example—Establishing Maximum Allowable Errors

In this example, we specify that fluid density, ρ, equals 1.000 Kg/m^3. Typically, fluid density is given by

$$\rho = \frac{P}{RT}$$

where P is fluid pressure in pascals, T is fluid temperature in kelvins, and R is the gas constant. For air, R=287 J/kgK. Using error propagation, establish the maximum allowable errors in the three measurements (q, P, and T) when P equals 96,000 Pa and T equals 334.5 K to achieve ±1% (3σ) in fluid velocity, V.

Approach

Apply Eq. (4.3) to establish the relationship as follows:

$$\delta V = \sqrt{\left(\frac{\partial V}{\partial q}\Delta q\right)^2 + \left(\frac{\partial V}{\partial P}\Delta P\right)^2 + \left(\frac{\partial V}{\partial T}\Delta T\right)^2}$$

where

$$\frac{\partial V}{\partial q} = \frac{V}{2q} = 0.0144 \quad \text{(from example 4.1)}$$

$$\frac{\partial V}{\partial P} = \frac{-V}{2P} = -0.00036$$

$$\frac{\partial V}{\partial P} = \frac{V}{2T} = 0.104 \, .$$

Thus,

$$\pm 0.693 = \sqrt{(0.0144\Delta q)^2 + (-0.00036\Delta P)^2 + (0.104\Delta T)^2} \, .$$

Since there are three unknowns, a unique solution does not exist. Still, maximum error limits can be established for each measurement by specifying two variables to be zero and solving for the third.

Therefore, the maximum allowable errors at 3 sigma are

$$\Delta q = \pm\ 48 \text{ Pa}$$
$$\Delta P = \pm\ 1925 \text{ Pa}$$
$$\Delta T = \pm\ 6.7 \text{ K} .$$

Interpretation of Solution

These are maximum allowable errors for each measurement if the errors in the other two are zero and include sensor, data system, etc. In practice, the designer would establish error budgets for measurements less than these maximums and use the above equation to ensure compliance with the $\pm 1\%$ V at 3σ specification. The designer would also take into account the time requirements over which the maximum allowable errors must not be exceeded. This, then, would generate the MTBOOT/MTBF target which the design is to meet.

4.6 Constructing Error Models

When we measure a physical attribute by any means (e.g., eyeballing, using off-the-shelf instruments, employing precise standards, etc.), we are making an estimate of the value of the quantity being measured. Two features of such estimates are *measurement error* and *measurement uncertainty*. The terms error and uncertainty are often interchanged, but there is a subtle distinction between the two. For example, the result of a measurement after correction can unknowingly be very close to the unknown value of the measurand, and thus have negligible error, even though it may have a large uncertainty.

4.6.1 Measurement Uncertainty

Measurement errors are never known exactly. In some instances they may be estimated and tolerated or corrected for. In others, they may be simply acknowledged as being present. Whether an error is estimated or acknowledged, its existence introduces a certain amount of measurement uncertainty.

> **UNCERTAINTY** — a parameter, associated with the result of a measurement, which characterizes the dispersion of the values that could reasonably be attributed to the measurand.

The assessment of uncertainty requires critical thinking, intellectual honesty, and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one — it depends on one's detailed knowledge of the nature of the measurand and of the measurement methods and procedures used. The utility of the uncertainty quoted depends on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value.

Some sources of uncertainty — not necessarily independent — are:

- Incomplete definition of the measurand and imperfect realization of the definition of the measurand
- Sampling — the sample measured may not represent the defined measurand
- Instrument resolution or truncation
- Values assigned to measurement standards and reference materials
- Values of constants and other parameters obtained from external sources and used in the data algorithms
- Approximations and assumptions incorporated in the measurement methods and procedures
- Variations in repeated observations of the measurand under apparently identical conditions
- Inadequate knowledge of the effects of environmental conditions on the measurement procedure, or imperfect measurement of environmental conditions, or unknown uncertainties of the measurement equipment used to determine the environmental conditions.

Mistakes in recording or analyzing data can introduce significant unknown error in the result of a measurement. Large mistakes can usually be identified by proper data review — small ones could be masked by or even appear as random variations.

In instances where the value of an error is estimated, the uncertainty in the estimate can be used to indicate a range of values surrounding the estimate. In instances where the error is not estimated but simply acknowledged, an uncertainty estimate serves to define a range of values that is ordinarily expected to contain the error, whatever its value might be. In both cases, the uncertainty estimate is made to establish regions of values that bound the error with some level of probability or "confidence." The limits of such regions are referred to as *confidence limits*. The term "expanded uncertainty" is also used.

4.6.2 Measurement Error

The difference between the measurand value[6] and the measurement estimate of this value is referred to as *measurement error*.

> **ERROR —** the difference between the result of a measurement and the value of the measurand.

Measurement error for a given measuring artifact and measurand may be bias (systematic) or precision (random). Bias errors are classified as those whose sign and magnitude remain fixed over

[6] In accordance with the ISO/TAG4/WG3 *Guide to the Expression of Uncertainty in Measurement*, a measurand is defined as "a specific quantity subject to measurement." As defined, a measurand is a *specific* quantity and as such, is definite, certain, unique, or particular. The definition implies that the value of a measurand is the "truth." To add the term "true" to "value of a measurand" is redundant. Therefore, the term "true value of a measurand" (often abbreviated as "true value") is generally not used in this publication. Where used, the terms "value of a measurand" (or of a quantity), "true value of a measurand" (or of a quantity), or simply "true value" are viewed as equivalent.

a specified period of time or whose values change in a predictable way under specified conditions. Precision errors are those whose sign and/or magnitude may change randomly over a specified period of time or whose values are unpredictable, given randomly changing conditions.

Typically, error estimates are attempted only for bias errors. This does not mean that all bias errors can be estimated. It may not be possible to estimate the value if

(1) the sign and magnitude are either not measured or not communicated

(2) the sign and magnitude vary in an unknown way over periods of time between measurement or

(3) both (1) and (2).

An example of an unknown bias error is the bias of a measuring attribute of an instrument drawn randomly from a pool of like instruments where its sign and magnitude are unknown. In such a case, all that can be done is to form a distribution of values, weighted by probability of occurrence, that attribute biases may assume. Estimates of these probabilities may be based on prior calibration or test history data taken on like instruments or may derive from heuristic or engineering estimates based on stability and other considerations.

The designer's objective is to configure and specify the individual system components so the integrated performance satisfies the overall requirements, including the targeted measurement accuracy. A mechanism is needed that will help the analytical evaluation of the candidate system's performance. This is traditionally done using error models.

Error models are simple schematic illustrations of a measurement process used to

- Identify the error sources associated with the measurement equipment (i.e., the published intrinsic errors, such as nonlinearity, gain error, hysteresis, etc.)
- Identify and quantify installation-related errors, such as those owed to the environment, CMV, electrical loading, and cabling, in addition to spatial and disturbance errors
- Identify and quantify application-related errors, such as those caused by improper sampling, improper data collection and reduction.

The specific steps used in constructing an error model follow:

1. Draw a simple schematic diagram of the process showing major hardware and software components.
2. Establish signal levels.
3. Identify and quantify intrinsic equipment errors and confidence estimates.
4. Choose consistent units and confidence levels.
5. Identify and quantify installation-related errors and application-related errors.
6. Combine errors to establish estimate of uncertainty

4.7 Example—Developing a Temperature Measurement System

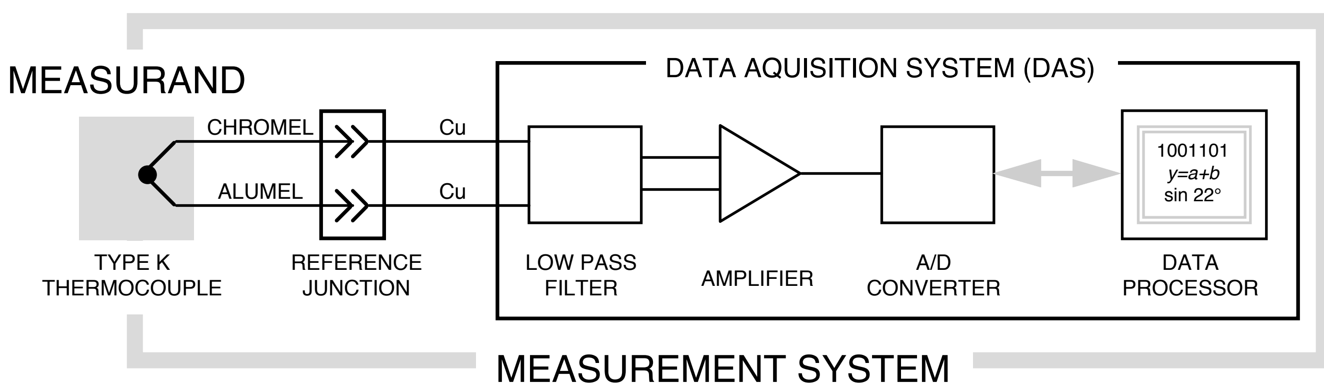
Consider the problem of measuring the temperature of a moving fluid which nominally is in the range of 30–70 °C. Past experience has shown that a Chromel–Alumel thermocouple is useful for measurements in the range from 0 to 1260 °C. Therefore, it has been decided that an ISA Type K Chromel/Alumel thermocouple configured in a grounded sheathed probe through a bulkhead into the fluid stream will be used.

The following specifications have been established for this measurement:

- *Range of Temperature to be Measured*: 20–100 °C
- *Bandwidth*: 0–10 Hz
- *Uncertainty*: ±3 °C, 3σ at 60 °C, for one year
- *Principal Mode of Heat Transfer*: Natural convection from fluid to probe, conduction from probe to thermocouple
- *Measurement Sensor*: ISA Type K Chromel–Alumel thermocouple.

4.7.1 Temperature Measurement System Equipment Selection and Specification Interpretation

The basic elements comprising the example temperature measurement system are shown in the following sketch:



Since thermocouples are differential measurement devices, the voltage input to the measurement system depends on the voltage generated by the thermocouple and the subsequent voltage generated at the reference junction. For this example, the equipment items needed are the thermocouple, the reference junction, a system to measure voltage, and a method of correlating measured voltage to temperature.

Thermocouple

The accuracy of a thermocouple depends on the type and homogeneity of wire material, the grade of the wire, and the temperature range in which it will be used. Most thermocouples are nonlinear from the low to high limits of their nominal working range, however most have good linearity when used in a reasonably narrow portion of the thermocouple material's total range.

For best results, thermocouples should be calibrated before using. They should be calibrated at the temperature range of interest to lessen and quantify errors due to variations in materials and manufacturing. Calibration will allow for careful selection of thermocouples, which may significantly reduce the measurement's uncertainty.

The thermocouple's indicated versus measurand temperature can be influenced by installation techniques. Complicated heat transfer effects produced by the measurand, protective housing, measurand vessel, environment, and measurand dynamics can have a profound impact on the measurement accuracy. If the measurand is a moving gas, several temperatures may exist simultaneously making it necessary to decide what is being measured. It is not good practice to correct a poor installation by the use of computed correction factors. For proper temperature measurement, one should make a thorough analysis of each installation.

A Type K (Chromel–Alumel) thermocouple is useful for measuring temperatures from 0 to 1260 °C. The manufacturer's published Limits-of-Error for a Type K thermocouple over the temperature range 20–100 °C is ±2.2 °C. Because of material impurities and variability in the manufacturing process, the actual emf versus temperature characteristics may differ from the published characteristics for the manufacturer's reference Type K thermocouple. This is interpreted as bias error. The manufacturer does not provide any information on the confidence level associated with the stated uncertainty interval. From many calibrations of wire samples, the user has established that the confidence level of the uncertainty is 3σ.

Often, the measurement uncertainty requirement is impossible to meet. If the requirement had been, for example, ±1 °C, 3σ, and given the manufacturer's published data of ±2.2 °C at 20–100 °C for a reference thermocouple, segments of a roll of thermocouple wire would have to be individually calibrated to find lengths that would reduce the Limits-of-Error to less than ±1 °C. If this cannot be accomplished, the measurement uncertainty specification must be relaxed.

Reference Junction

It is critical that the reference junction temperature be known exactly. The typical specifications for reference junctions include an accuracy statement for the junction temperature, and for multiple thermocouple junctions, a statement of temperature uniformity. Typical uncertainties published are ±0.25 °C for junction temperature and ±0.1 °C for uniformity. Usually, the manufacturer is silent on the uncertainty confidence level. Experience has shown the confidence level to be between 2 and 3σ. The uncertainties are interpreted as bias errors.

Data Acquisition System

Using a nominal sensitivity for Type K thermocouples of 40 μV/°C, the voltage range corresponding to a temperature range of 20–100 °C is 0.8 to 4.0 mV. The data acquisition system must be capable of measuring time-varying phenomena of these magnitudes at frequencies from zero to 10 Hz. The following specifications are considered to be representative for a quality multichannel data acquisition system. Here the manufacturer specifies 99% (~3σ) confidence level for uncertainty values.

- *Gain Accuracy*: ±0.05% FS ±0.003% /°C
- *Nonlinearity*: ±0.02% FS
- *Time Zero Stability*: ±5 μV relative to input (**RTI**) ±1.25 mV relative to output (**RTO**)
- *Temperature Zero Stability*: ±0.5 μV/°C RTI ±0.1 mV/°C RTO
- *Zero Offset*: ±10 μV, Channel-to-Channel
- *Noise*: ±8.5 μV RTI ±0.75 mV RTO, ±3σ with a 10-Hz filter installed
- *Resolution*: ±0.003% FS
- *Common Mode Rejection Ratio*: 120 dB
- *Static Crosstalk*: 120 dB

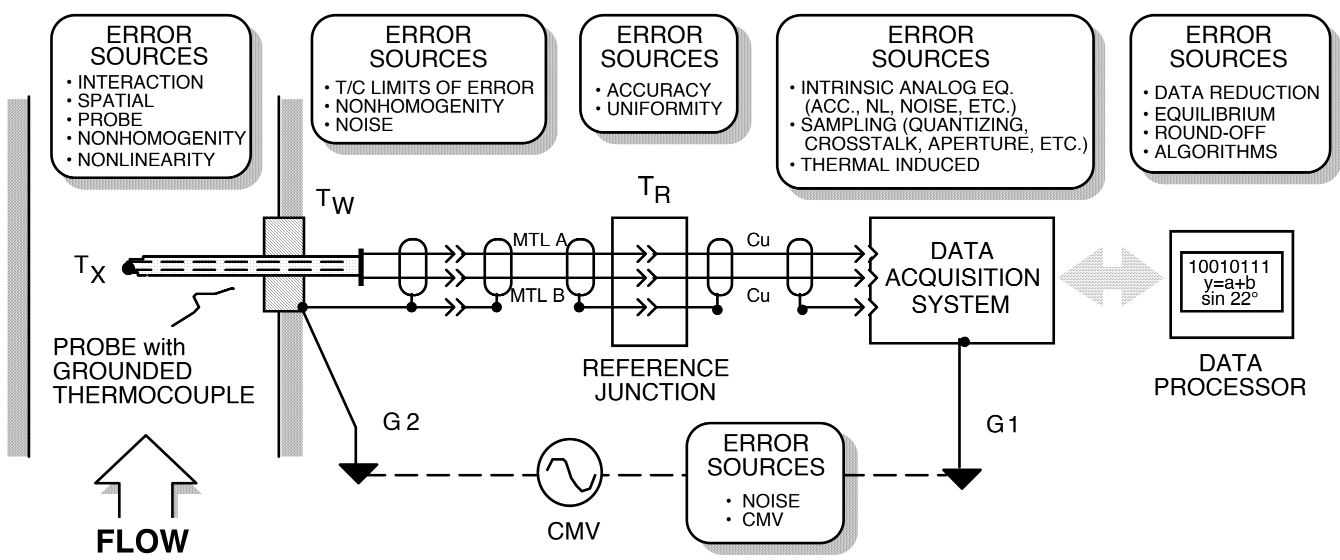
Interpretation of these errors is provided below.

4.7.2 Example Temperature Measurement System Error Model

This example illustrates the traditional process of developing an error model for the temperature measurement system and establishing an estimate of uncertainty.

> **NOTE** — The example is repeated in detail in Appendix G. There, the reader will find some techniques differing from the traditional approach taken below, a more detailed treatment of the identification of error sources, and development of mathematical expressions for establishing the estimate of uncertainty.

STEP 1. Draw a simple schematic diagram of the process.



STEP 2. Establish signal levels.

Because of a nominal sensitivity for a Type K thermocouple of 40 μV/°C, the voltage corresponding to 20–100 °C is 0.8–4.0 mV.

An amplifier gain of 1000 is chosen for the measurement system. This provides an input voltage to the analog-to-digital converter of 0.8–4.0 V. The selected converter has a full-scale input of ±10 V.

STEP 3. Identify and quantify intrinsic equipment errors and confidence levels.

Gain Accuracy: ±0.003% /°C, ±0.05% FS *[Given]*

Nonlinearity: ±0.02% FS *[Given]*

Time Zero Stability: ±5 μV RTI ±1.25 mV RTO *[Given]*

Using a gain of 1000, the time zero stability error is converted to % FS by multiplying the RTI component by 1000 and summing this with the RTO component.

∴ Time Zero Stability: ±0.0625% FS

Temperature Zero Stability: ±0.5 μV/°C RTI ±0.1 mV/°C RTO *[Given]*

This error can be restated in term of % FS as:

∴ Temperature Zero Stability: ±0.006% FS/°C

Zero Offset: ±10 μV, Channel-to-Channel *[Given]*

This error can be restated in terms of % FS as:

∴ Zero Offset: ±0.1% FS

Noise: ±8.5 μV RTI ±0.75 mV RTO *[Given]*

This error can be stated in % FS by RSSing the components where the RTI component is adjusted by gain.

∴ Noise: ±0.085% FS

Resolution: ±0.003% FS *[Given]*

The confidence level for uncertainties is 3σ, based on conservative engineering estimates and experimental measurement data analysis.

Step 4. Choose consistent units and confidence levels.
For this example, it is desirable to use °C to represent all errors. Since the thermocouple and reference junction are already in °C, it is only necessary to convert the measurement system errors

into °C. Since the system gain has been picked to be 1000, the maximum input voltage can be ±10 mV (computed by dividing the converter's full-scale input of ±10 V by the gain of 1000). Given a nominal sensitivity of 40 μV/°C, the full-scale input of 10 mV corresponds to about 250 °C. The above specifications can be restated as follows:

- Gain Accuracy: ±0.125 °C , ±0.0075 °C/°C
- Nonlinearity: ±0.05 °C
- Time Zero Stability: ±0.15 °C
- Temperature Zero Stability: ±0.015 °C/°C
- Zero Offset: ±0.25° C, Channel-to-Channel
- Noise: ±0.2125 °C
- Resolution: ±0.0075 °C

All the above error sources have been estimated to a 3σ confidence level or adjusted to 3σ where higher or lower confidence levels were used.

Step 5. Identify and quantify installation- and application-related errors.

- *Common Mode Voltage (CMV)*

  The error, $\bar{e}_{cmv}$ resulting from a common mode voltage of e_{cmv} can be computed using the CMRR (common mode rejection ratio) specifications as follows:

$$\bar{e}_{cmv} = \frac{G \cdot e_{cmv}}{\log^{-1}\left(CMRR/20\right)} .$$

  For a CMRR of 120 dB *[Given]* and an estimate of CMV of 10 V, the error is $\bar{e}_{cmv} = 0.01\,V$ which is 0.1% FS or 0.25 °C.

- *Static Crosstalk*

  This computation is similar to CMV, where an estimate of maximum voltage between channels is used. Assuming 10 V maximum, the error is the same as CMV.

- *Temperature-Induced Errors*

  The effects of temperature on both gain and zero offset can be computed using the temperature coefficients stated in Step 4 and an estimate of maximum temperature change. Assuming a maximum temperature change of 10 °C, gain and offset errors are:

- *Thermal Gain Accuracy: ±0.08 °C*
- *Thermal Zero Stability: ±0.15 °C*

Step 6. Combine errors to establish uncertainty estimate.

- *Bias estimate*
  - Thermocouple: ±2.2 °C
  - Reference Junction Accuracy: ±0.25 °C
  - Reference Junction Uniformity: ±0.1 °C
  - Gain Accuracy: ±0.125 °C
  - Nonlinearity: ±0.05 °C
  - Zero Offset: ±0.25 °C
  - Resolution: ±0.008 °C
  - CMV: ±0.25 °C
  - Static Crosstalk: ±0.25 °C

  ∴ Total bias estimate based on RSS of above: ±2.26 °C at 3 σ

- *Precision estimate*
  - Zero Stability: ±0.15 °C
  - Noise: ±0.21 °C
  - Thermal Gain Accuracy: ±0.08 °C
  - Thermal Zero Stability: ±0.15 °C

  ∴ Total precision estimate based on RSS of above: ±0.31 °C at 3 σ

Since the bias limits were determined to be 3σ with normal distribution (99.7%), the uncertainty estimate is:

$$U_{ADD} = \pm\left(2.26\,°\mathrm{C} + 2\times 0.31\,°\mathrm{C}\right) = \pm\ 2.88\ °\mathrm{C} \text{ with a confidence level of } 99.7\%$$

$$U_{RSS} = \pm\sqrt{\left(2.26\,°\mathrm{C}\right)^2 + \left(2\times 0.31\,°\mathrm{C}\right)^2} = \pm\ 2.34\ °\mathrm{C} \text{ with a confidence level of } 97.5\%\ .$$

Therefore, either of these uncertainty estimates *may* meet the ±3 °C uncertainty requirement of the measurement as specified.

The word "may" is used here because the uncertainty specification was established to be ±3 °C, 3σ at 60 °C, *for one year*. Yet, as one can observe, none of the manufacturer's data specified confidence levels for uncertainty values in terms of a time element. At this point, critical engineering judgment and uncertainty growth analyses are required to support whether or not the uncertainty estimates will meet the one-year requirement.

> The measurement system designer *must* consider the *time duration* of the specification and be aware that the calibration certification is only applicable at the instant of calibration. In addition, most manufacturer's data does not specify confidence levels for uncertainty values in terms of a time duration. The designer must not overlook this very important aspect when estimating uncertainty, especially for systems design of remote long-term applications.

The designer should pay particular attention to the material covered in Section 3.2.5 and Table 3.2 regarding the establishment of measurement system reliability requirements as they apply to mean-time-between-out-of-tolerance (MTBOOT).

4.8 Consideration of Calibration Techniques to Reduce Predicted Bias Errors

Generally, a measurement system's predicted bias errors, as established from interpreting manufacturer's specifications and other supporting analyses, dominate the uncertainty calculation. This is a consequence of using worst-case limits to quantify error sources. Bias errors are fixed by definition, so many of these can be effectively reduced through calibration. The designer's task is to review the predicted bias error terms and incorporate calibration techniques within the measurement system so that these can be effectively reduced, if needed. Methods commonly used include:

- Inserting known stimuli at sensor input (in situ calibration)
- Inserting known stimuli at measurement system input
- Simulating known inputs (e.g., creating imbalance with Wheatstone bridge configurations and substituting known resistors for potentiometric measurements, such as resistance temperature devices, or viewing deep-space radiation using a blackbody at a known temperature)
- Calibration of individual measuring system components
- Calibration by use of a reference material having the same general composition as the material being measured—for example, using a gas mixture of known composition to calibrate a gas analyzer
- Calibrating range by viewing two or more known points (triple point of water and melting point of zinc).

Where there is more than one measuring system for a given parameter, relative performance can be found by interchanging measuring systems relative to the sensors and by applying SMPC methods.

4.9 Consideration of Uncertainty Growth in the Measurement System Design Process

Immediately following test or calibration, the uncertainty in the recorded value of a measurement parameter begins to grow in response to several factors. These factors include environmental stress, usage stress, storage and handling stress, stray emf, vibration and mechanical shock, and so on. Uncertainty growth reflects shifts in parameter value described by a variety of mechanisms, including:

- Linear drift
- Random fluctuations
- Periodic oscillations
- Spontaneous quantum jumps
- Response to discrete events.

> The specific manner in which uncertainty growth is accounted for depends on the mechanism.

Suppose that parameter values shift because of linear drift. Linear drift is described according to

$$Y(t) = Y(0) + \kappa t \ , \tag{4.7}$$

where $Y(t)$ represents the parameter value after a time t has passed since test or calibration, and κ is the parameter drift rate. In practice, the coefficient κ is an estimated drift rate, based on engineering or other data that are themselves characterized by an uncertainty $\sigma_\kappa(t)$ that grows with time (and other stresses) since test or calibration. Given this, estimates of the parameter value are obtained from

$$\hat{Y}(t) = Y(0) + \kappa t \pm z_\alpha \sigma_y(t) \ , \tag{4.8}$$

where

$$\sigma_y^2(t) = \sigma_y^2(0) + \sigma_\kappa^2(t) \ , \tag{4.9}$$

and where z_α is the two-sided normal deviate, obtained from a standard normal or Gaussian distribution table, for a $(1-\alpha) \times 100\%$ confidence level. The quantity $\sigma_y^2(0)$ is the variance in the parameter value immediately after test or calibration.

A straightforward method for getting the coefficient κ is to fit $\hat{Y}(t)$ in Eq. (4.8) to observed values for $Y(t)$ using regression analysis. In this approach, measured values $Y_1, Y_2, \ldots, Y_n$ are sampled at various times $t_1, t_2, \ldots, t_n$. . Using linear regression methods gives

$$\kappa = \frac{\sum_{i=1}^{n}(t_i - \bar{t})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(t_i - \bar{t})^2} , \tag{4.10}$$

where

$$Y_i \equiv Y(t_i),$$

$$\bar{t} = (1/n)\sum_{i=1}^{n} t_i ,$$

and

$$\bar{Y} = (1/n)\sum_{i=1}^{n} Y_i .$$

Similarly, the variance $\sigma_\kappa^2(t)$ is obtained from

$$\sigma_\kappa^2(t) = \left[\frac{1}{n} + \frac{(t - \bar{t})^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}\right] s^2 , \tag{4.11}$$

where

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left[Y_i - \hat{Y}(t_i)\right]^2 .$$

Measurement parameter uncertainty growth for the linear drift model is depicted in Figure 6.2 of Section 6. Other uncertainty growth mechanisms and associated models are described in Appendix B.

> A word of caution about uncertainty growth is due. If, for example, drift is a concern, then the established value for the measurement uncertainty is only valid at the time of calibration.

If drift can be characterized as discussed above, it may be possible to correct for this or to change the estimated uncertainty to include this based on engineering judgment. A more practical method would be to incorporate a mechanism within the measurement system that allows drift to be measured and compensated for.

4.10 Consideration of Decision Risk in the Measurement System Design Process

Because of measurement uncertainties, incorrect decisions may result from information obtained from measurements.

The probability of making an incorrect decision based on a measurement result is called *measurement decision risk*. Since uncertainties grow with time since test or calibration, measurement decision risk also increases with time since calibration. This is the underlying motivation for doing recalibrations or retests regularly.

Measurement decision risk may take several forms—the most common are *false accept risk* and *false reject risk*. A false accept is an event in which an unacceptable item or parameter is wrongly perceived as acceptable during testing or calibration. Acceptance criteria are ordinarily specified in terms of parameter tolerance limits. An acceptable parameter is one that is in-tolerance. An unacceptable parameter is one that is out-of-tolerance.

Therefore, false accept risk is usually defined as the probability that an out-of-tolerance parameter will be accepted by testing or calibration. This definition is relevant from the viewpoint of the testing or calibrating organization. An alternative definition is sometimes used which is relevant to the receiving organization. From this viewpoint, false accept risk is the probability that an out-of-tolerance item or parameter will be drawn at random from a given lot of accepted items or parameters.

False reject risk is similarly defined as the probability that an in-tolerance item or parameter will be rejected by testing or calibration. False accept and false reject criteria can be used to establish parameter tolerances, among other things. False accept and false reject risks are described mathematically in Appendix C.

4.10.1 False Accepts

Certain negative consequences may arise because of false accepts. *Test process false accepts* can lead to reduced end-item capacity or capability, mission loss or compromise, loss of life, damaged corporate reputation, warranty expenses, shipping and associated costs for returned items, loss of future sales, punitive damages, legal fees, etc.

Calibration process false accepts lead to test system populations characterized by parameters being out-of-tolerance at the beginning of their usage periods. In Appendix B it is shown that the higher the beginning-of-period (**BOP**) out-of-tolerance probability, the higher the average-over-period (**AOP**) out-of-tolerance probability. High AOP out-of-tolerance probabilities lead to higher measurement decision risks encountered during test system calibration. These higher risks, in turn, make test systems more prone to measurement decision risk during end-item testing.

4.10.2 False Rejects

Both test process false rejects and calibration process false rejects lead to unnecessary rework and handling. Since higher rejection rates imply poorer production controls, *test process false rejects* also create an excessively pessimistic view of the quality of the end-item production process. This view may lead to more frequent disassembly and repair of production tools, machinery, molds and templates than is necessary.

Calibration process false rejects create an excessively pessimistic view of the EOP in-tolerance percentage of test systems. Since test system calibration intervals are adjusted because of this percentage, calibration process false rejects lead to unnecessarily shortened test system calibration intervals. This results in unnecessary operating expenses and increased downtime.

5. MEASUREMENT TRACEABILITY

5.1 General

Common measurement references are critical to the worldwide exchange of goods, products, information, and technology. Transferring these common references in a controlled manner to thousands of individual measurements made every day is the goal of traceability. NASA measurement traceability extends from the ground-based operations to measurements made aboard space-based platforms and planetary probes. Decisions based on measurements will affect the day-to-day well-being of the crew, the performance of the on-board and ground-based systems and the ongoing scientific experiments.

Measurement traceability is a sequential process in which each measurement in a chain of measurements, starting with accepted reference standards, depends on its predecessor as shown in Figure 5.1.

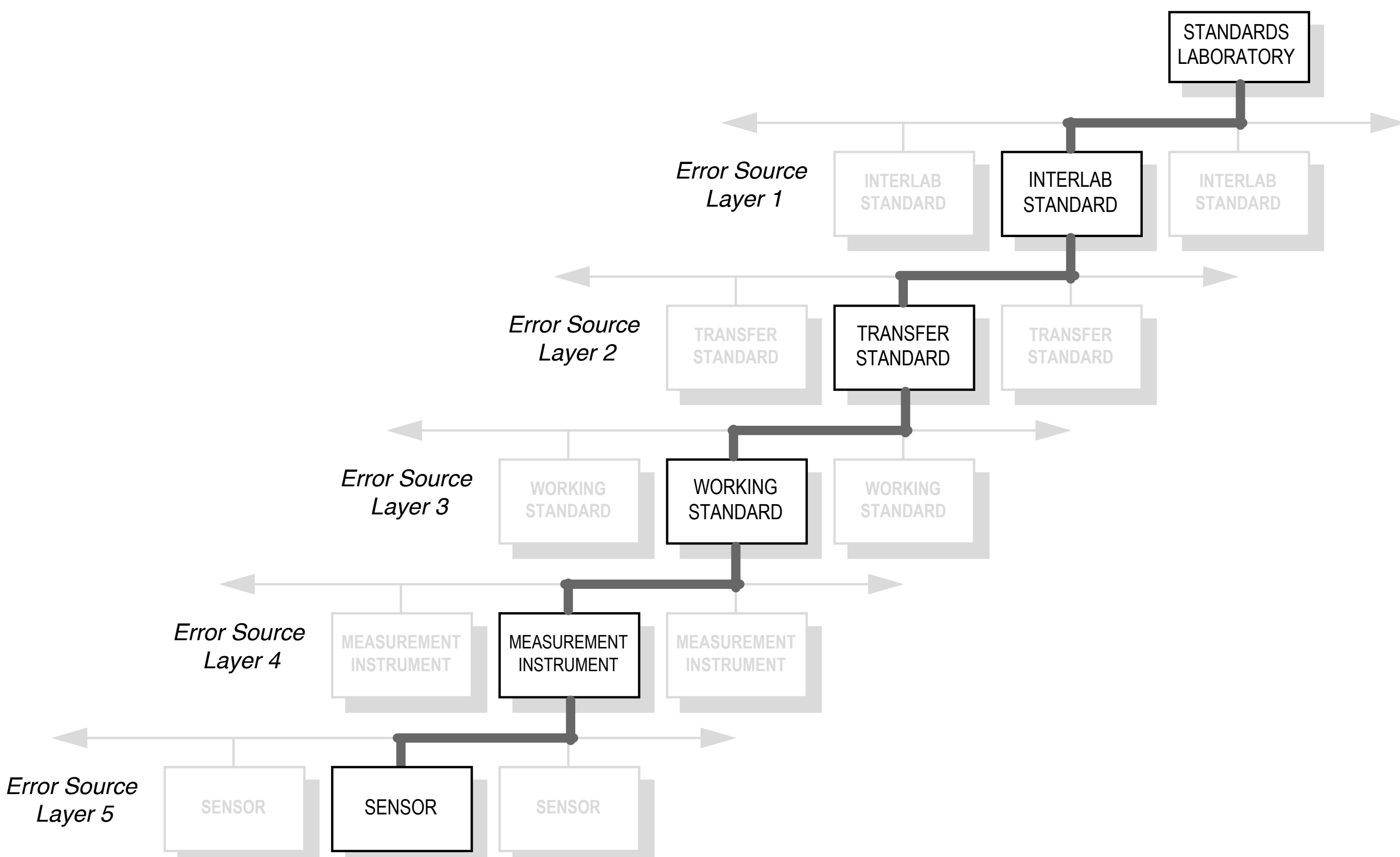


FIGURE 5.1 — VERTICAL CHAIN EFFECTS ON UNCERTAINTY IN MEASUREMENTS.
The top of the chain (Standards Laboratory) is assumed to be the accepted authority. Therefore, the resultant data can pass through at least five layers, each with multiple sources of error.

The chain may only be one link or it may involve many links and several reference standards. For example, temperature measurements using a thermocouple rely on the temperature scale and the unit of voltage.

One of several definitions of traceability is:

TRACEABILITY — The property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons.

Traceability is a hierarchical process. There are other definitions of traceability and many of these are discussed by Belanger. In the United States, traceability begins at NIST and ends with an operational measurement, i.e., a rocket motor temperature. It is a measurement chain that is no better than its weakest link. At each link or stage of the traceability chain, errors are introduced that must be quantified, and their effects combined, to yield a realistic estimate of the uncertainty with respect to the accepted standards (usually NIST). At each level, a standard will calibrate an unknown. Both may be a single-valued or a standard artifact standard, or an instrument. The chain may have only one link or it may involve many links and several reference standards.

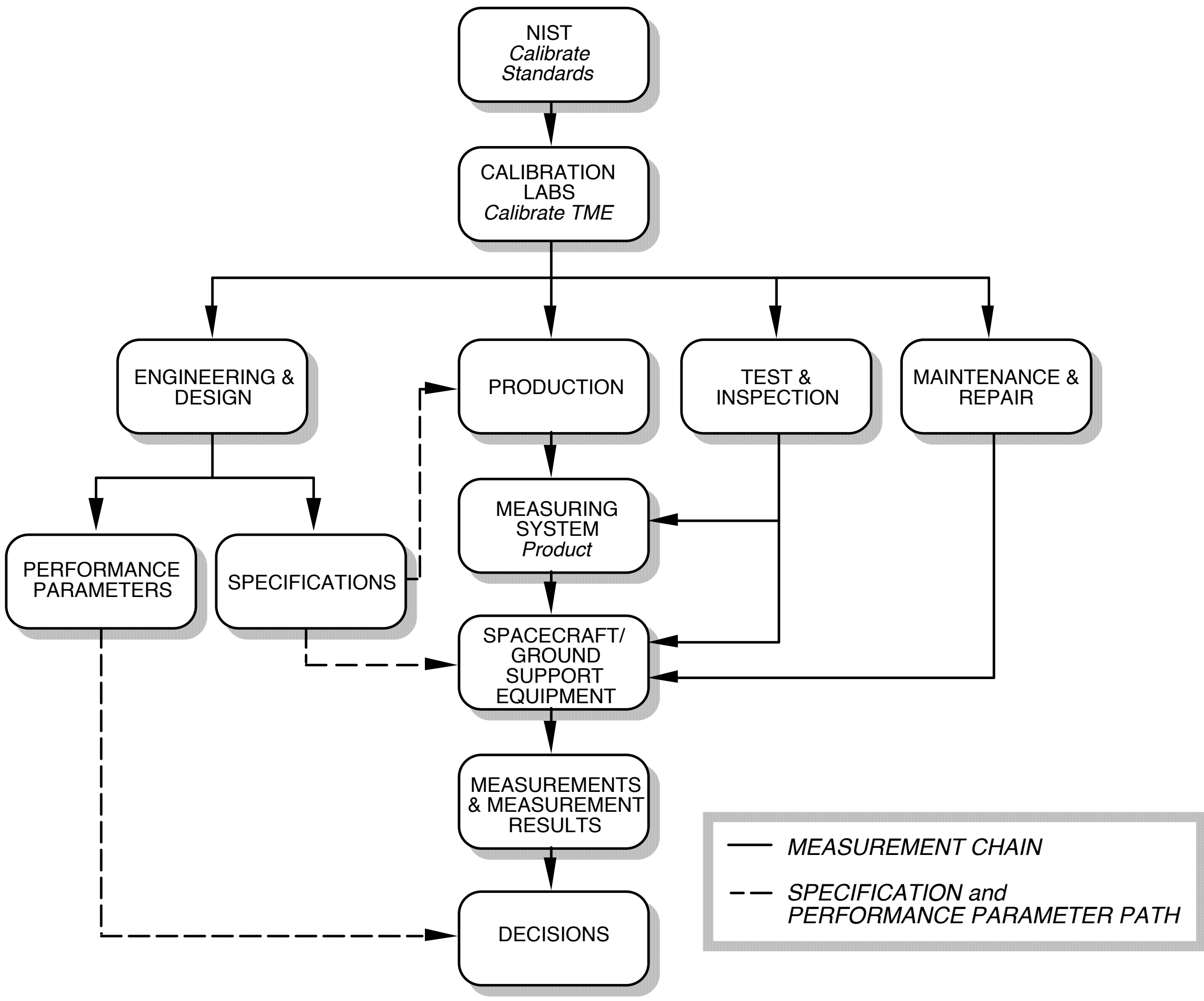


FIGURE 5.2 — HIERARCHICAL NATURE OF TRACEABILITY.
Solid lines represent the measurement paths, with each line representing one or more measurements of one or more quantities. The dashed lines are established specifications based on previously made measurements.

Figure 5.2 is a simplified illustration of the hierarchical nature of traceability. It begins with national standards and ends when the measurement result will be used to make a decision. The quality of the decision depends on the quality of the traceability paths. The box labeled "Calibration Labs" represents many laboratories of varying capabilities and may be multilayered.

They may go directly to NIST or to another calibration laboratory. At each stage, there are error sources producing measurement uncertainties propagated to the next level. Also, the paths in most cases are usually parallel, coming together when the product (measuring process) is placed in operation (in a spacecraft or ground support equipment). The result of this complicated process is a measurement result used to make a decision. To get a single measurement, the result may involve a similar path for each measurement quantity involved in the final measurement. Consider the measurement of temperature using a thermocouple. In the field, it involves (1) a calibrated thermocouple, (2) a calibrated reference junction, and (3) a calibrated voltage measuring instrument.

Traceability is the melding together of measurement standards, measurement techniques, periodic calibration, data analysis, statistical process control, and sound decision making for each link of the measurement chain. This information, necessary to reconstruct the measurement, must be documented and preserved to ensure the integrity of the traceability. For each link, documentation should contain the assigned values of the final item, a stated uncertainty of the result, an uncertainty budget, the standards used in the calibration, and the specification of the environmental conditions under which the measurements were made. The allowable degradation in accuracy (increase in uncertainty) is often specified for each link in the chain as an accuracy ratio.

5.1.1 Components of a Measurement

Every measurement M_{obs} of a quantity is an estimate of the magnitude ($\{N\}$). This estimate is a pure number that represents the value of the measurand of the quantity expressed in terms of the unit of measure (θ) used to make the measurement. Furthermore, M_{obs} has an error (ε) that is unique to that measurement. Mathematically it can be represented by the following relationship.

$$M_{obs} = \{N\} \bullet \theta + \varepsilon \quad . \tag{5.1}$$

For differing units representing a quantity, different values for $\{N\}$ will result. This can be seen by considering the measurement of an invariant quantity using two different units. Since the quantity is invariant, the following relationship results:

$$\{N\}_A \cdot \theta_A = \{N\}_B \cdot \theta_B \quad , \tag{5.2}$$

where the subscripts A and B represent measurements in terms of different units. If two slightly different representations of the same unit are used to make measurements, there will be small differences in $\{N\}$. The difference is quantified by Eq. (5.3.)

$$-\frac{\delta_{\{N\}}}{\{N\}} = \frac{\delta_\theta}{\theta} \quad . \tag{5.3}$$

The function of calibration is to reduce δ_θ to an acceptable magnitude. To achieve measurement uniformity and assure traceability for a given quantity:

- There must be only a single unit of measure for each quantity
- The uncertainty of the unit with respect to its definition must be known

- The uncertainty of the measurement process must be known.

5.1.2 Definition of Tolerance, Uncertainty, and Accuracy Ratio

Following are the definitions of tolerance, uncertainty, and accuracy ratio:

Tolerance — Tolerance is a *condition* imposed on a measurement by the designer or other agency. Tolerance is defined as *"the total permissible variation of a quantity from a designated value."*

Uncertainty — Uncertainty is *"a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand."* Measurement uncertainty is a *property* of the measuring system and all prior measurement chain errors. Obviously, the measurement uncertainty must never exceed the tolerance.

Accuracy Ratio — Accuracy ratio (**AR**) or test accuracy ratio (**TAR**) are terms used to describe the relationship between specified tolerances and measurement uncertainty. AR or TAR is the ratio of the tolerance of the instruments being tested to the uncertainty of the standard.

ACCURACY RATIO (AR) is: $AR = \frac{tolerance}{uncertainty}$

The realization of accuracy ratios is sometimes impossible because of requirements for hardware, materials, measuring processes, and the state of the art. The calibration of an 8-1/2-digit digital voltmeter (DVM) is an example of instrumentation approaching the quality of the standard. Most calibration laboratories maintain the volt over an extended period to about ±1 ppm but are called on to calibrate DVMs having a performance in the 1-ppm region.

5.1.3 The Metric System

A coherent, universally accepted system of units of measure is critical to measurement uniformity and traceability. Over the years, various systems of units have been adopted, but each has been less than universal until the adoption of the International System of Units (SI) by the 11th General Conference on Weights and Measures (CGPM) in 1960. The SI is frequently called, simply, the metric system. It is proper to refer to the SI as the *modernized* metric system. There have been efforts to adopt the modernized metric system in the United States, particularly the Metric Conversion Act of 1975. There has been little or no movement to metrication until recently. Now, by law, United States Government activities must metricate in a reasonable time.

Section 5146 of Public Law 100-418 , the Omnibus Trade and Competitiveness Act of 1988, amends Public Law 94-168, the Metric Conversion Act of 1975. Specifically, Section 3 of the latter act is amended to read as follows:

> It is therefore the declared policy of the United States
>
> (1) to designate the metric system of measurement as the preferred system of weights and measures for United States trade and commerce:
>
> (2) to require that each Federal agency, by the date certain and to the extent feasible by the end of the fiscal year 1992, use the metric system of measurement in its

procurement, grants, and other business-related activities, except to the extent that such use is impractical or likely to cause significant inefficiencies or loss of markets to United States firms, such as when foreign competitors are producing products in non-metric units;

(3) to seek ways to increase understanding of the metric system of measurement through educational information and guidance and in Government publications; and

(4) to permit the continued use of traditional systems of weights and measures in nonbusiness activities.

The notice published in the Federal Register states:

> Under both this act and the Metric Conversion Act of 1975, the "metric system of measurement" is defined as the International System of Units [SI] as established in 1960 by the General Conference of Weights and Measures and interpreted or modified by the Secretary of Commerce. (Sec. 4(4), Pub. L. 94-168; Sec. 403(1)(3), Pub. L 93-380.)

Although universal, there are a few very small variations among nations regarding names, symbols, and other matters. An overview of the SI is given in Appendix H. All material is the SI as interpreted for use in the United States. Also, the SI is dynamic and is continually undergoing revision. Though the material in Appendix H is stable, it is important to verify it has not changed.

5.2 Measurement Standards

Units of measure must be realized experimentally besides, as well as conceptually defined. Such work is scientifically demanding, requiring years of research, and is usually restricted to national laboratories, universities, and other scientific institutions. To serve their own needs, nations have established legal standards of measure and often, by law, have decreed that all measurements must be traceable to their national standards. Because of errors in realizing the unit, small but significant differences between as-maintained units may exist among nations.

The measurement standard is the primary tool for traceability. A measurement standard may be a physical object (artifact) adopted by national or international convention, a physical phenomenon or a constant of nature (an intrinsic standard), a standard reference material (**SRM**), or in some situations a consensus physical standard. An example is the Rockwell Hardness Tester, which is generally accepted to measure the hardness of steel. The purpose of an SRM is to provide a common reference point whereto a specific species of measurements is referred to ensure measurement compatibility with time and place.

Traditionally, standards have been thought of as devices specifically designed for that purpose. In the context of NASA, the concept of standards must be extended to cover all instruments and apparatuses used to calibrate or verify proper operation of the operational equipment aboard a space-based platform and on the ground. This includes all equipment traditionally thought of as "test" equipment. When a DVM will calibrate or verify a panel meter, the DVM is the "standard." (A standard is a reference device for a calibration process.

5.2.1 Intrinsic Standards

An intrinsic standard is based on one or more physical phenomena of high reproducibility, or constants of nature. Originally, these standards were primarily confined to national laboratories but are finding their way to other metrology laboratories. Examples are; the triple point of water and other temperature-fixed points to define the temperature scale, the ac Josephson effect to define the representation of the SI volt, and cesium beam clocks for time and frequency. *Intrinsic standards can be realized anywhere (if an appropriate level of competence exists and the system embodying the intrinsic standard can be well-characterized), eliminating the need for calibration at a higher echelon such as NIST*. (A Josephson volt can be readily realized in a Dewar at cryogenic temperatures. However, the process of using it to measure a source at room temperature is fraught with difficulties. The process may be idiot-proof at ±5 ppm, but to achieve 0.05 ppm requires expertise and good procedures.) For international consistency, the phenomenon is fully described and the values of the constants are assigned by international agreement. The procedure by which measurements are made with intrinsic standards must be fully documented and agreed upon to prevent procedural variations.

5.2.2 Artifact Standards

An artifact standard uses one or more physical properties to represent the desired unit or quantity. For example, the thermal expansion of mercury is used to measure temperature changes. Artifact standards are the most common and *all* must be calibrated periodically in terms of a higher order (echelon) standard. Examples of artifact standards are quartz oscillators, standard resistors, gauge blocks, etc.

5.2.3 Ratio Standards

Ratio standards are dimensionless standards used to scale various quantities and can, in principle, be derived locally. For example, the calibration of a precision voltage divider can be done without reference to an external standard. Sometimes, calibration services are available for certain types of ratio apparatus. Ratio measurements are vital tools for scaling units.

5.2.4 Reference Materials

In certain situations, the accepted reference standard is a reference material (**RM**), certified reference material (**CRM**) as defined by the International Standards Organization (**ISO**) Guide 30-1981(E), or a material that has been carefully characterized by NIST and sold as an SRM. Through its use, traceability to the accepted national standards is achieved. For example, mixtures of gases of known composition are used to calibrate systems designed to measure the composition of an unknown gas mixture. When properly used, these materials usually calibrate the entire measurement system and provide traceability.

5.2.5 Other Standards

There are circumstances where there are no national standards. For example, NIST does not maintain a standard for hardness testing. To ensure uniformity, one or more agreed upon standards have been recognized. *Where more than one standard exists, they may not give the same measurement results*. To avoid ambiguity, the particular standard used must be clearly specified. They may or might not be recognized internationally or even nationally.

5.3 United States Standards

In the United States, NIST, formerly the National Bureau of Standards (**NBS**), has, by law, the responsibility to establish, maintain, and disseminate the physical units for the nation. To meet this responsibility NIST provides a wide range of calibration services, develops and distributes SRMs, operates a standard reference data program, and provides measurement expertise for a wide range of disciplines. Besides fulfilling its role of disseminating standards, NIST is very active in developing new measurement techniques where none exist or where major improvements are needed. Measurement service activities at NIST are coordinated by

The Office of Measurement Services
National Institute of Standards and Technology
Gaithersburg MD 20899

5.3.1 NIST Physical Measurement Services Program

The physical measurement services of NIST are designed to help those engaged in precision measurements achieve the highest possible levels of measurement quality. There are hundreds of services available and each class is described in *NIST Calibration Services Users Guide* (NIST SP250). The general areas are dimensional measurements, mechanical measurements, thermodynamic quantities, optical radiation measurements, ionizing radiation measurements, electromagnetic measurements, and time and frequency measurements. They are the highest order of calibration service available in the U.S. by providing a direct link between clients and the national measurement standards. NIST will only calibrate standards or specific instrumentation that meets certain high performance standards. For general information about services contact

Calibration Program
National Institute of Standards and Technology
Gaithersburg MD 20899

NIST urges direct contact with the staff member responsible for the particular calibration area for specific questions or problems.

5.3.2 NIST SRM Program

NIST has an extensive reference material program covering a wide range of materials sold throughout the world. These materials are primarily SRMs certified for their chemical composition, chemical property, or physical property, but include other reference materials. They serve three main purposes:

(1) To help develop accurate methods of analysis;

(2) to calibrate measurement systems; and

(3) to assure the long-term adequacy and integrity of measurement quality assurance policies.

It is probable that SRMs will find use in certain life-support systems aboard future humanly-occupied space-based platforms. Two examples are the use of one or more SRMs to monitor the composition of a habitation atmosphere and to monitor composition of recycled water.

NIST publishes the SRM Catalog (SP260) of available materials every two years. The current catalog lists over 1000 materials. For further information contact

Standard Reference Materials Program
National Institute of Standards and Technology
Gaithersburg MD 20899

As part of the SRM program, many special publications are available from NIST. One in particular is applicable to traceability (Handbook for SRM Users, NIST SP260-100, 1985).

5.3.3 NIST National Standard Reference Data Program (NSRDP)

NSRDP is a nationwide program established to compile and critically evaluate quantitative physical science data and assure their availability to the technical community. For information contact

Standard Reference Data Program
National Institute of Standards and Technology
Gaithersburg MD 20899

5.4 International Compatibility

Representatives of most nations have established systems of legal units based on the SI units that may result in small differences in certain national as-maintained units. Although the differences are small, such differences may be important to NASA's space program, particularly in the exchange of technology among the participating nations. The differences range from negligible for most quantities, to significant for others. Significant differences generally occur for derived quantities and such evolving measurement areas as millimeterwave standards. The U.S. and other nations are constantly seeking to effect better international agreement among national standards using a wide range of tools to ensure compatibility.

5.4.1 Reciprocal Recognition of National Standards

NIST has established a program to recognize the equivalency of standards between NIST and the national standards organizations of selected other countries. For each quantity, through experiments or careful evaluation of a participating nation's capability, participants establish the equivalency for their national standards. These equivalency accords are nonbinding but do provide evidence that the national standards are equivalent. (They *do not* assure equivalency at lower levels however.) In the United States, the Department of Defense accepts the accords on equivalency while the Nuclear Regulatory Commission does not. Several agreements exist and more are being negotiated between NIST and the national laboratories of Japan, Canada, Italy, Germany, and other countries. The NIST Calibration Program is cataloging such agreements and should be consulted for details.

5.4.2 BIPM Calibrations

The BIPM was established under the Treaty of the Meter as the international metrology laboratory. One of its missions is to provide calibration services to signatories of the treaty. Many nations with small central metrology laboratories use BIPM. Although these nations use BIPM, the accuracy and precision of their measurement systems place limits on the level of agreement to be found between the standards of such nations and those of major industrial nations.

5.4.3 International Comparisons

Bilateral and multilateral international comparisons of national standards directly measure differences among the participating laboratories. The BIPM is taking a very active role in organizing and managing such comparisons. International comparisons are usually important to reciprocal agreements. Many nations, including the U.S. do many comparisons with no regard to reciprocal agreements.

5.4.4 NIST Calibrations

NIST provides direct calibration services to some nations to ensure measurement compatibility. Calibration at BIPM does not necessarily provide NIST traceability. Calibration at NIST provides traceability to the U.S. units, but does not guarantee the results of each measurement made in the customer's laboratory.

5.5 Calibration Transfer Techniques

The heart of traceability is the ability to transfer units, derived quantities and other agreed-on reference standards, with the least degradation in accuracy. Calibrations fall into two broad classes:

(1) Devices such as calibrated standards and specific values determined in terms of national standards and

(2) Instruments or standards measured to determine if they are within assigned specified limits of error relative to national standards.

The difference is in the way the results are reported. In the first case, a specific value is reported and in the second, it is reported as either in or out of tolerance (specification). The minimum information that must be supplied is illustrated by the content of a typical NIST report. Note that a NIST report of test generally has nothing to do with calibrations. A NIST Report of Calibration gives (1) the value of the item calibrated (2) the uncertainty of the calibration for the accepted reference standard, and details about the overall uncertainty (3) the conditions under which the measurements were carried out, and (4) any special information regarding the calibration. It *does not* include uncertainties for effects of transport to and from the calibrating laboratory, drifts with time, effects of environmental conditions (i.e., temperature, humidity, barometric pressure, etc.). Sometimes these errors may be greater than the reported uncertainty of the calibration. Generally, calibration transfer techniques are one of the following types.

5.5.1 Traditional Calibration

Traditionally, instruments and standards are transported to and from the calibration laboratory, by hand or common carrier. *This method is the simplest and most straight forward, but it suffers from the weakness that the calibration is guaranteed valid only at the time and place it was carried out.* It is the user's responsibility to assess other factors that can introduce errors into the traceability chain. Despite the possible shortcomings, it is the easiest and still the most widely used calibration transfer technique. Some guidelines to aid in getting the best possible calibration at the local level are listed below:

(1) Pay close attention to the total transportation process, including packing, mode of transport, time in transit, and the carrier. Manufacturers and the calibration laboratories can frequently help to minimize transport effects.

(2) Always calibrate standards to be sent to the calibrating laboratory with the remaining (at home) standards before and after transport. A significant change shows potential problems; a small or no change shows that the transport process has not affected the item.

(3) Understand the effect of environment on the item and evaluate any effects if the local environment differs significantly from the one in which the item was calibrated. The environment is that of the physical location of the item, and not the room. A digital voltmeter may be housed in a confined space and be at a temperature significantly different from the general environment. A thorough understanding of the equipment and standards is critical to minimizing environmentally induced errors.

(4) Artifact-based instruments and standards are not absolutely stable with time and, therefore, must be recalibrated periodically by the strategies discussed in Section 6.

5.5.2 Measurement Assurance Program (MAP) Transfers

The concept of the MAP was developed by NIST in the 1970s. In its simplest form, a MAP is a calibration technique in which the calibrating laboratory calibrates its client's measurement process instead of the client's standard.

A MAP is to metrology what quality control or assurance is to manufacturing. Sound measurement assurance programs at all levels in the calibration chain are essential to traceability. A MAP does two things

(1) Ties a single measurement to a reference base, and

(2) Establishes the uncertainty of a measured value relative to this reference base.

Well-designed and implemented MAPs are critical for ensuring long-term, high-level performance of on-board and ground-based space application systems.

> Most MAPs are carried out at the calibration laboratory level, but some, including critical day-to-day operational measurements, could be adapted for use throughout the total system.

Much has been written about MAPs, but the reader should become familiar with two publications: one by Belanger and the other by Croarkin. The first is an overview of MAP programs for calibration laboratories, and the second is an excellent tutorial on MAP methodology. Much of the material in both is applicable to MAPs at all levels.
All MAPs have two distinct parts:

(1) Transfer of the unit or quantity to a given laboratory or system. This is the calibration process, and it sets the lowest limit of the uncertainty for the process.

(2) The day-to-day measurement process used to monitor the local process including standards and instruments between external calibrations. Note that when an artifact is externally

calibrated, the user assumes its value is constant (or predictable), unless there is evidence to the contrary. Therefore, the internal actions taken between calibrations to monitor the local process and provide evidence are as important as the calibration itself.

The first, the calibration model, describes the relationship among reference standards, unknowns, instrumentation, and the operating environment. For each calibration process, there is a unique model. The second is the statistical model used for error estimation and uncertainty determination. When this model is used in conjunction with the calibration model, various error sources can be identified and quantified. Operationally, MAPs rely on the use of a check standard to monitor the process continuously. By repeated measurements on the same object (a check standard), process errors are quantified. The statistical analysis of the data leads to the estimate of the measurement process bias uncertainty. Croarkin discusses several possible check standards.

In a MAP, the entire system used to perform a calibration, and to provide traceability from the standards of the supporting standards laboratory, is viewed as a process. The system includes standards, instruments, environment, procedures, personnel, and such activities as moving standards and evaluating errors. The supporting standards laboratory and its components are also taken into account. Two techniques are used to evaluate the process: a “blind” test of the process output and statistical process control techniques. The former is used on a periodic basis (perhaps yearly), while the latter is used continuously to ensure the integrity on a day-to-day basis.

The “blind” test is typically carried out using a well-characterized transport standard or precision instrument (artifact) whose calibrated values are unknown to the process output. The artifact is selected so that its parameters and their proven performance levels are adequate to sample the type of measurement critical to the objectives or purpose of the measurement process. The artifact is treated as a normal workload item by the process output, except that it may be measured repeatedly, or used in a special pattern of measurements designed to determine the process precision and improve the accuracy of measurement of the process offset(s). The artifact is characterized before and after this sampling process by the supporting laboratory. All data from both laboratories are used to determine the errors (offsets) of the process output and their characteristic statistical properties. This approach has been used as (1) a quality control check on a measurement process (2) a tool to identify and correct measurement problems, and (3) a way to achieve traceability where very low uncertainties or very high confidence levels are required of the process.

When used alone, this technique suffers from the same weakness as that found in periodic instrument calibration; i.e., it cannot determine exactly when, between samples, a measurement process has gone out of control (when the measurement errors exceed the process requirement). However, when it is complemented with the application of statistical process control techniques, a full measurement (quality) assurance policy results and nearly eliminates any likelihood that a poor product (bad measurements) can get out.

Typically, the way a measurement assurance policy is carried out is through the use of a “check” standard. This is an instrument or device similar to and, if possible, of higher quality than, the items being measured by the process. The measurements made on the check standard do not need to be as complete as those made on the process output, but the same measurements must be made repeatedly. The frequency is determined by the stability of the system, the statistical characteristics of the data, and the process requirements on a statistical basis.

NIST offers a number of MAP services (see NIST SP250) that serve as "blind" sampling for calibration processes. NIST requires that participants in NIST MAPs demonstrate that their measuring process is in a state of statistical control between transfers.

5.5.3 Regional Measurement Assurance Program (RMAP) Transfers

RMAPs or group MAPs are an outgrowth of the NIST MAP program. Instead of one laboratory interacting unilaterally with NIST, several establish a program in which one or more transport standards are circulated among participants to measure differences among laboratories. During the interchange period, NIST provides a MAP service with one of the participants. From this set of measurements, the measurement processes of all laboratories are evaluated and traceability is achieved. For a well-planned RMAP, the extra step adds a very small increment to the overall uncertainty. RMAPs can be used to ensure close agreement among any group of facilities.

5.5.4 Round Robins

Round robins are an audit tool to identify systematic differences and estimate measurement capability among the participants. Well-devised round robins provide *realistic traceability* by directly assessing the capability of a number of laboratories. Most round robins are based on a technique developed by Youden.

For example, one laboratory may serve as the pivot by circulating well-characterized artifacts among the participants and analyzing the round robin results. (Usually two artifacts are used. With one, the analysis is more difficult and not as much information is obtained.) Each artifact is measured by each participant, and all results are then analyzed. The two artifacts do not need to be identical, but they must evaluate the same measuring process. The round robin done by the Kennedy Space Center for voltage, at the 10 V level, illustrates the idea.

EXAMPLE—10 V ROUND ROBIN

Two 10 volt solid-state references were circulated among the participating laboratories. They were measured by each participant with the participant's as-maintained unit of voltage and measuring processes. For each participant, the measured value of one standard was plotted as a function of the other, as shown in Figure 5.3. Interpretation is straightforward. If the points had been distributed in the four quadrants in a random or a shotgun-like pattern, the experimental errors would have been random and much greater than the systematic errors. Here, the points are along a straight line showing systematic differences among laboratories. Furthermore, because of the closeness of each point to the line, the bias uncertainty for each set of measurements is small. From these data, one concludes that there are systematic biases in the measuring processes among the participants. NIST disseminates the unit at the 10 V level to better than 1×10^{-6}. It is possible to maintain the local unit to an uncertainty of about 1×10^{-6} using MAP techniques (the circle in the center). If one laboratory were known to be correct, then the offset of the others could easily be estimated. Here, the pivot laboratory was known to be in very close agreement with NIST, and the three points at 0,0 are for that laboratory since it served as the reference.

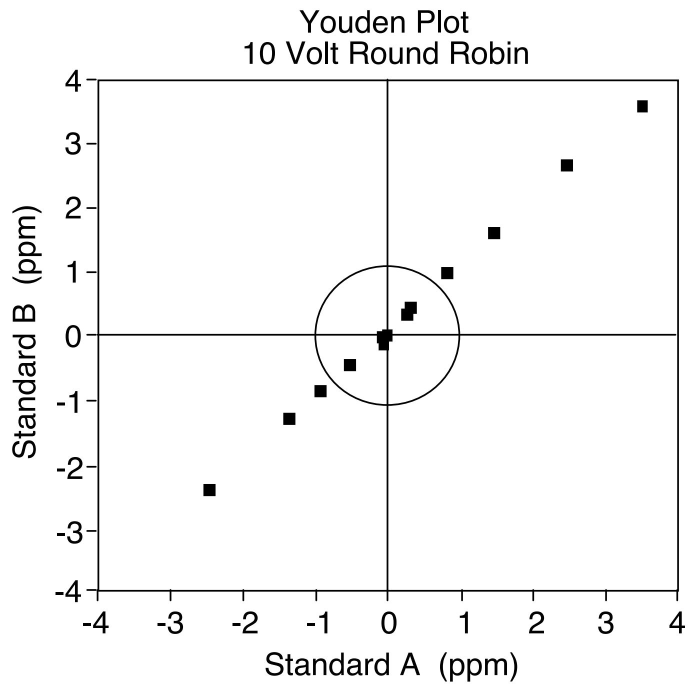


FIGURE 5.3 — A YOUDEN PLOT FOR A 10 V ROUND ROBIN.

A total of 11 laboratories participated with one serving as the pivot, or control. The points indicate the difference between the pivot laboratory (3 points near 0) and the participating laboratory. The circle has a radius of 1×10^{-6} that indicates the potential capability of the laboratories. Note that only three laboratories fall within the circle (Pivot lab excluded).

The degree of closeness to the line is an indicator of individual internal precision, while scatter along the line indicates systematic effects between laboratories.

5.5.5 Intrinsic Standards

An intrinsic standard is a calibration transfer standard because it reproduces a unit locally without recourse to NIST. It is, however, important that the methodology used in the use of such a standard be fully evaluated and verified by comparison with NIST or a similar laboratory. For example, though the temperature scale can be realized by fixed points and a platinum resistance thermometer, the methodology should be independently verified.

5.5.6 SMPC Methods Transfers

If calibrations are done on a diverse workload base whose measurable attributes derive their values from independently traceable sources, then transfer of accuracy can take place from the workload to the calibrator. This “consensus traceability” is possible with statistical process control methods described in Section 6 and Appendix D. Moreover, if the measured quantities include known terrestrial or astronomical references, the SMPC methods enable a transference of accuracy from these references to orbital or space-based platforms.

5.6 Calibration Methods and Techniques

The methodology for making measurements is crucial to traceability and the decision making process. It calls for the integrated understanding and application of the following major elements:

- The physical laws and concepts underlying the total measuring process
- Reference standards

- Instrumentation
- Control and understanding of environmental effects (including operators or technicians) on the measurement process
- Data reduction and analysis
- Error estimation and analysis.

Calibration techniques vary depending on the category of equipment being calibrated. All measurements are the comparison of an unknown to a known and calibrations are no exception. Categories are

- Reference standards
- Test and measurement equipment (TME)
- Systems.

5.6.1 Calibration of Reference Standards

Most reference standards are fixed. They are usually an artifact that is the representation of a unit at a single point. Examples are gauge blocks, standard lamps, and standard resistors. Although chiefly used at the highest accuracy levels, reference standards are among the easiest to calibrate. Often for a specific quantity, there are several standards covering a wide range of values. Standards are usually calibrated by comparing them to one or more known standards of the same approximate magnitude. These comparisons or calibrations are made by either measuring differences (Δ) between the standard(s) and the unknowns (X)

$$\Delta = X - S \tag{5.4}$$

or ratios (K)

$$K = \frac{X}{S} \; . \tag{5.5}$$

In either case, the value of the standard must be independently determined, or known, to calculate X. Since the two objects differ only slightly, the instrumentation need only cover the range of the maximum expected difference (ratio). For example, a 10 V solid-state voltage standard is calibrated by measuring the difference to 1 μV (0.1 ppm) between the standard and the unknown using a DVM. If the largest measured difference is 100 ppm, then the range of the DVM need only be ±1000 μV and the resolution only ±1 μV. The accuracy required of the DVM is only 1 part in 1000 or 0.1 percent, well within the capability of today's high-accuracy DVMs.

The product of most standards calibrations is a correction figure or a value. Standards are rarely adjusted to be within fixed bounds. Generally, corrections are made to the nominal value of the standard for its calibrated value, temperature, pressure, and perhaps other influence factors, to obtain a value to be used with the standard to perform calibrations.

5.6.2 Calibration of TME

TME is the link between the world of calibration and the end-user; it is the major workload of the calibration laboratory. TME can be as simple as a hand-held meter or as complex as an automated test stand that measures many parameters. Although many calibration techniques used are similar to those used for standards calibration, there are significant differences, as described below:

- TME is generally calibrated to a specified accuracy, usually the manufacturer's specified accuracy over its operating range or ranges. For many newer microprocessor-based instruments, it is possible to store corrections that are applied automatically to individual readings. More and more instruments take advantage of software corrections to enhance instrument performance.
- The instrument is calibrated on each range at a sufficient number of points (including zero) to determine the required performance parameters.
- Corrections are seldom supplied unless requested by the user.
- Minor adjustments may be made to bring indicated reading of the instrument into better agreement with the correct or "true" value. Major out-of-tolerance conditions usually need repair by a competent repair facility.
- Good practice requires that the calibrating facility maintain records and report to the customer the as-found and as-left conditions of instruments.

5.6.3 Calibration of Systems

Equipment used to make operational measurements is the reference standard for that measurement process. The measurements are used to make decisions based on the indication of the instrument (not the "true" value). For TME, the equipment is calibrated to the manufacturer's specifications. Broadly speaking, a single piece of measuring equipment might consist of a sensor and a data acquisition system, as illustrated in Figure 5.4.

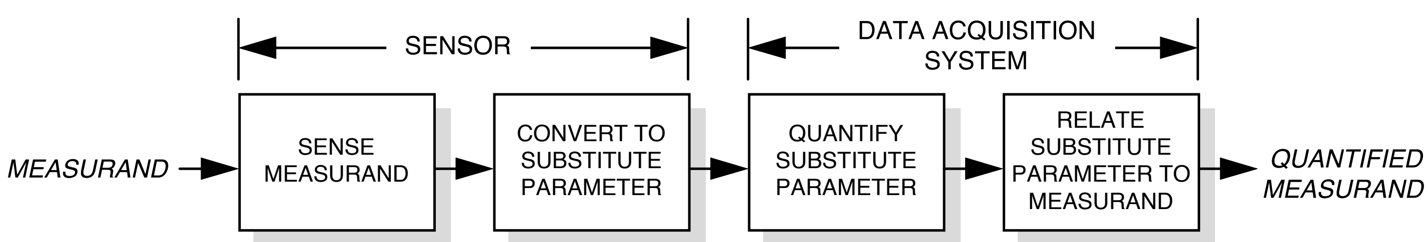


FIGURE 5.4 — INDIRECT MEASUREMENT OF PHYSICAL VARIABLE.

The sensor senses the quantity to be measured (the measurand) and converts it to a substitute parameter (usually electrical). The substitute parameter is then transmitted to the data acquisition system where it is quantified and related to the original parameter being measured. Also, there are several subelements, such as signal conditioners, transmission lines, connectors, etc. There is no single strategy to calibrate such a system. Two strategies, neither of which is well suited to every case, are

(1) Calibration of each operating entity individually; a process that may mean partial disassembly of the system. This method may overlook certain sources of error that might adversely affect the overall system calibration (for example, interaction between subsystems).

(2) Calibration as a system using suitable standards. While in some ways this is the simplest approach, it does not necessarily identify the source of any out-of-tolerance subsystems. For cases where the measurand is a physical quantity that has no reasonable substitute measurand (a flowing gas at a known temperature, for example), system calibration is not practical.

To further compound the problem, complex test systems and measurement systems designed to measure many parameters often provide control based on the results of some function thereof. There is a real possibility that there will be interactions among the various elements in the system. To calibrate such a system totally may be nearly impossible because of the interactions. For example, some high-accuracy digital voltmeters measuring a dc voltage may be affected by ac signals coupled to the dc path where ac signals are a part of the total measuring system. The size of the resulting error depends on the instrument, the magnitude of the coupled ac current, and its frequency. (Usually, the effect on the measured dc voltage is proportional to the square of the ac current.) A nonexhaustive list of the major categories of error sources includes

- Measurand-sensor interface errors
- Sensor conversion errors
- Signal conditioning errors
- Transmission from sensor to data acquisition system errors
- Data acquisition system errors
- Algorithm errors (both sensor and data acquisition system)
- Software errors
- Operator and operational/procedural errors.

The most effective action to ensure the long-term calibration of any system is to address the calibration and maintenance problems early in the design phase. One approach is to integrate reference standards and associated calibration means into the system with sound calibration techniques. Such a system only requires that the internal standards be routinely calibrated.

5.6.4 Calibration Using SRMs

Reference materials are used to calibrate complete measuring systems that are used to measure the concentration of particular substances in a mixture—particularly in the fields of chemistry and medicine. These materials are applied to the input of the measuring system and the output observed. The result is the direct measurement of any instrumental offset that can be used as a correction to routine measurements of the quantity of interest. This direct calibration method and may have only a limited range, thereby requiring reference materials containing various amounts of the substance of interest. For example, pH standards (Sums) are used to calibrate or verify a pH meter.

5.6.5 Scaling

Real-world measurements of a quantity may be made over many decades, and the measurements should be traceable to national standards. National laboratories, including NIST, cannot provide calibration services for all possible multiples and submultiples. However, suitable standards and methodology for realizing submultiples and multiples of most units can be readily available at the local level. The two principal methods for scaling are the additive and ratio techniques.

5.6.5.1 Additive Scaling

As the name implies, additive scaling is the process of calibrating multiples or submultiples of the reference standard using only the mathematical operations of addition and subtraction. Additive scaling requires that the sum of the parts be equal to the whole. Not all standards are truly additive. For example, two 10.00000 Ω resistance standards connected in series are not equal to 20.00000 Ω because of lead and contact resistances. Mass calibrations, on the other hand, are an example of an additive scaling process. Starting with the kilogram, larger and smaller mass standards are calibrated by comparing multiple mass standards (weights) with single standards of equivalent mass using sound experimental designs and a suitable 1:1 comparator (a balance). Another important example of using additivity is that of the dead weight gauge used to calibrate pressure transducers. Different pressures are developed in the system by changing the weights.

5.6.5.2 Ratio Scaling

Multiplication and division are used to scale by ratio. The precise mechanism used depends on the particular measurement discipline. Ratio is a dimensionless quantity that can be independently established to a high degree of accuracy—it finds wide use in many disciplines, particularly in electrical measurements. Resistance measurements are made by using a bridge as the ratio scaling device. To avoid the effect of lead resistance, resistors are scaled with precisely known resistance ratios in such a way that no current flows by defining leads and contacts. The resistance ratios are embodied in special circuits that may be calibrated using additive techniques.

5.6.5.3 Scaling Using Defined Fixed Points

The temperature scale is defined with (1) certain intrinsic standards known as defined fixed points (2) interpolating devices (transducers), and (3) the defined mathematical relationship relating the property measured to the thermodynamic temperature. Several interpolating devices are needed to cover the complete range of temperatures, but for space applications, the platinum resistance thermometer (**PRT**) is the most important. By measuring the resistance at selected fixed points and using the defined mathematical relationship between resistance and thermodynamic temperature, the temperature scale from about –259 to 960 °C is realized. The PRT can then be used to measure temperature or calibrate other temperature transducers by direct compensation.

5.7 Calibration Traceability vs. Error Propagation

Measurement errors happen at every link in a chain of measurements, from the realization of a measurement unit to the final measurement result. Also, standards and instruments are subject to errors arising from transportation, drift with time, use and abuse, subtle component changes, environmental effects, and other sources. At each link, the errors must be estimated, combined, and unambiguously communicated to the next link (level). The parameter used to disseminate information about measurement errors is the measurement uncertainty. This section addresses the issue of errors, their estimation, combination, and propagation in the TME calibration chain. More discussion from the instrument designer's perspective is given in Section 4. The effect of uncertainty on calibration interval is discussed in Section 6. The quality of the measurement uncertainty estimate plays major roles in both traceability and calibration intervals.

5.7.1 Evaluation of the Process Uncertainty

At each calibration level, the steps necessary for the reliable evaluation of the process uncertainty are discussed below. A stable measurement process is a prerequisite to estimating the measurement uncertainty.

STEP 1 — All sources of error must be identified and classified according to type (bias or precision) .

Identification is done by attentive and thorough analysis of the physical principles and concepts underlying the particular measurement and augmented by auxillary experiments and data. In addition to the basic methodology, one must consider secondary effects that can affect the measurement. For example, low-level voltage measurements are sensitive to thermally generated emfs caused by temperature differences within the measuring circuit.

STEP 2 — Individual or groups of errors must be quantified.

Bias and precision errors are estimated differently but must be expressed so that, they can be combined to convey the total uncertainty and its composition in a meaningful way to the user. The errors must be stated at the same confidence levels.[7]

Bias (systematic) uncertainties cannot be directly estimated. Instead, they are estimated using sound engineering judgment and ancillary experiments. The bounds of each bias error is estimated through an understanding of the physical laws and concepts underlying the measurement and an analysis of the *actual* process. They are usually combined using Eq. (4.4), which is based upon the underlying assumptions expressed in Section 4.4.3, to get the total bias uncertainty (B_T). Estimating each error is a judgment call. A conservative practice is to estimate bias error as the “maximum” possible bias. The problem is that “maximum” is subjective. What does “maximum” mean? Present-day thinking is that bias uncertainties are expressed at either the 99.994% (4σ) or 99.73% (3σ) confidence level. That is, the chance that the estimated bias uncertainty will exceed that stated is 6 in 100,000 for the first and 270 in 100,000 for the latter. The confidence level may be arbitrarily chosen, but in any error analysis the chosen level must be stated.

Precision (random) uncertainties are estimated by replication of measurements and ancillary experiments. They can be estimated individually and combined through Eq. (4.5), or by the application of SMPC to yield an overall estimate of σ_r. The SMPC method is preferred for several reasons:

(l) It directly estimates σ_r from operational data from the measuring process.

---

[7] To be consistent with Section 4, σ will be used throughout. All references to σ can be replaced with s for small or medium-sized data sets. Since this section deals primarily with the calibration chain, which usually has extensive data at each link, σ is more applicable.

(2) Because it is operational and ongoing, σ_r provides *continuous* information about the process.

(3) It can provide information on the day-to-day and long-term performance of the process (detect process changes).

(4) The day-to-day process variations that would otherwise be systematic are randomized.

> STEP 3 — Bias and precision uncertainties are combined to estimate the process sigma (σ_t. Calculate the total uncertainty using a suitable multiplier:
>
> $$U = K\sigma_t$$

There are several methods that can be used to combine precision and bias errors, one of which is given in Eq. (4.6a), that is

$$U = \pm\left(B_T + t_\alpha S_T\right)$$

which is a special case of the equation given in "Step 3" above. Here, the multiplier K is t_α, the Student T statistic at the confidence interval α. Equation (4.6a) also assumes that the bias errors are estimated at the same probability level. For a well-characterized measurement process with a large data base, the statistic simply becomes that gotten from the normal distribution. This is usually the case for most calibration processes. Typical multipliers in metrology are 2 and 3, which correspond to α equal to the 95.45% and 99.73% confidence levels for a large number of degrees of freedom. Within the metrology community, both nationally and internationally, there are efforts proceeding to develop methods for expressing uncertainty.

> STEP 4 — The measurement process and uncertainty estimates must be documented and unambiguously communicated to the user.

At the least, the documentation must include the following:

(1) A statement of the combined uncertainty of the measurement.

(2) The confidence level to which the uncertainty is estimated.

(3) The interval over which the uncertainty and confidence level apply.

5.7.2 Propagation of Uncertainty in the Calibration Chain

Errors made at higher levels are propagated to the next level. Since the true error cannot be measured directly, the uncertainty is the tool by which error estimates are transferred down the chain.

> All uncertainties propagated from a higher level are taken as bias at the current level.

This is true for both precision and bias errors. Therefore, it is essential that the estimate of the uncertainty be a valid reflection of the measurement process.

Note that this is not true in the BIPM recommendations. A different approach is recommended by the *Comité International des Poids et Measures* (Recommendation 1 [CI-1981], Metrologia 18 [1982], page 44). The expression of the uncertainty of measurement in calibrations does *not* contain bias (systematic) errors. Uncertainty values are calculated after corrections have been made for all known bias errors. Thus, calibration certificates which are in accordance with BIPM procedures state only precision (random) uncertainty values.

5.8 Calibration Adjustment Strategies

Calibration assumes the object being calibrated, and hence, the quantity that it represents, changes. A well-designed process will choose the calibration interval and methodology so that changes will have only a negligible effect on operational measurements. When an adjustment is needed, depending on the object, three possible actions can be taken. First, a known correction can be applied to the results of all observations. Second, the object can be physically adjusted to bring its values to within certain specified limits. Last, many microprocessor-based instruments can store software corrections in nonvolatile memory and automatically apply them to each measurement.

5.8.1 Reference Standards

Reference standards are usually fixed. The calibration process yields the current value that is used with corrections for influence factors to calibrate other items. Predictions of the sign and magnitude of the drift with time should be obtained based on the calibration history of the reference standard and used to predict the present value. Adjustments are rarely made to reference standards, thus the adjustment strategy is: “Do not adjust, but monitor drift.”

5.8.2 Direct Reading Apparatus

TME and most other instruments are designed for direct reading. That is, the indicated value is assumed to be correct to within a specified tolerance. When a calibration shows the value to be out of tolerance, one of the following actions must be taken:

(1) The instrument or system can be adjusted to bring it into specification either locally or by a qualified service center. When adjusting an instrument to bring it into specifications, it is important to make certain that the adjustment is within the operating adjustment band specified for the instrument.

(2) Many instruments can store corrections in nonvolatile memory. In use, the instrument logic handles proper application of the correction to display the correct value. Procedures for using such features must be unambiguous. Several measurements should be taken after calibration to ensure that corrections were properly installed.

(3) For systems having computing capability, the corrections can be applied during the data-processing phase.

(4) The calibrating laboratory must notify the user when a calibration shows a value to be out of tolerance as found.

Adjustments can be harmful if a software correction is too large. In such a case, the instrument may be out of its design envelope. All software-applied corrections must include limits to ensure that the correction is within design limits.

Three strategies for adjustment of indicated reading to the center of the tolerance band are currently being used in calibration laboratories:

(1) Adjust at each calibration to the center of the tolerance band.

(2) Adjust to the center of the tolerance band only when the indicated reading exceeds a specified percentage of the tolerance limit, such as 70% of tolerance limit.

(3) Adjust to the center of the tolerance band only when the indicated reading exceeds the tolerance limit.

> The policy for adjusting TME during the calibration process and the adjustment action taken must be documented and available for analysis of calibration interval.

5.9 Software Issues

No other technogical artifact is changed as often as software. When some new functionality is needed, one perceives that software can easily be changed to fit this need, but anyone who has written and debugged software realizes that interactions can be extremely complex.

Software-influenced elements of the measurement chain act as black boxes, greatly simplifying design and use, *and misuse,* of measurement systems. With some effort, one can ascertain measurement quality for each link of the measurement chain through analyses of the standards and techniques used, data results, and decision-making processes. Often, one neglects the application of these analyses to the software "black box." The software-driven computational and control power present in contemporary data acquisition systems implicitly claims achievement of superior accuracy when it might be only apparent precision. There is a tendency to be lulled by this tempting and superficial simplification of the measurement process. An understanding of the software is a vital element of the measurement traceability process.

Metrology software guidelines are primarily formulated to improve the reliability of metrology operations and secondarily to reduce the cost of maintaining metrology systems. As helpful as these guidelines are, managers, engineers, and technicians involved with metrology operations should be persuaded to use them. Acceptance is an evolutionary process achieved by education at all levels. Therefore, the first set of guidelines should be minimal with plans to continue to more extensive guidelines over time.

5.9.1 Software Documentation

The minimal set of documentation for metrology software has the following sections:

- *Software Requirements* — Description of what the software is supposed to do.
- *Software Architecture Design* — Gives a high-level picture of how the system is put together and serves as a "road map" for the source code.

- *Software Version Description* — Contains commented source code and is the real detailed description of how the software works.
- *Software Testing* — Provides a set of test cases and procedures to prove that the system satisfies the requirements and continues to satisfy the requirements when changes are made.
- *Software User's Guide* — Tells the new or unskilled user how to run the system and describes error indications and recovery procedures.

For a small system, these sections will easily fit into a single binder, although to simplify revision, the sections may be considered separate documents.

5.9.2 Software Configuration Management

Configuration management is a critical but often neglected function in small installations and projects.

> When a change is made to metrology-related software and the new version executes the set of controlled test cases in an acceptable manner, and is formally approved, a version package should be placed in a secure controlled environment and obsolete versions removed from service. Secure copies of the obsolete version should be retained until they are of no known value. This is essential to maintain measurement traceability.

The version package should include the following as a minimum: source code, object code, and test results. If requirements have been changed, or the user interface has changed, revisions to the requirements document and user's manual should be included.
A reliability performance goal can be set to determine when changes should be allowed and how large a change should be permitted. For instance, a freeze on all changes not related to debugging can be imposed when the failure intensity rises above the performance

5.9.3 Software Standards

The development and maintenance of metrology software are a special case of software development and maintenance. Therefore, standards for metrology software should be selected and tailored from the general NASA software standards to take advantage of the expertise and effort that have gone into these standards. In particular, the Data Item Descriptions (**DID**s) supporting NASA "Information System Life-Cycle and Documentation Standards" should be tailored to provide appropriate guidelines for documents and procedures. The DIDs for this standard are prepared in a tree structure so that sections in higher level DIDs are expanded by lower level DIDs for use by larger, more complicated projects. For metrology software, only the top one or two levels of DIDs need to be considered, and these should be tailored to provide proper guidelines. The following list of DIDs is suggested as a basis for tailoring:

SMAP-DID-P200-SW	Software Requirements
SMAP-DID-P310-SW	Software Architectural Design
SMAP-DID-A200	Testing

SMAP-DID-P400	Version Description
SMAP-DID-P500	User's Guide
SMAP-DID-M920	Configuration Management

Although the proposed package of DIDs looks imposing, it probably would only total about fifteen pages if it were reformatted into a single document and edited to exclude deleted and redundant material.

6. CALIBRATION INTERVALS

6.1 General

6.1.1 Purpose

Concepts, principles and methods for the establishment and adjustment of intervals between calibration for TME and standards are discussed in this section. The material presented has a twofold purpose. For ground testing or measuring applications, the material is intended to guide NASA agencies and contractors in selecting or designing calibration recall systems. For space-based testing or measuring applications, the material is intended to provide alternatives to periodic TME recalibration and to indicate factors to be considered in designing systems for extended periods of use without recalibration or testing.

6.1.2 Scope

General information for establishing and adjusting calibration intervals is presented in this section. Section 6.2 is devoted to management considerations, and Section 6.3 discusses technical details of the calibration interval problem. The SMPC methodology as an alternative or supplement to periodic TME recalibration is discussed in Section 6.4. Concepts relevant to the technical management of calibration SMPC system design and development projects are also given in Section 6.4. Technical specialists should read Appendices B and D.

6.1.3 Background

The establishment and adjustment of calibration intervals are activities that often drive test and calibration support infrastructure managers to distraction. For most organizations, personnel are not conversant with this highly specialized and often misunderstood subject. Nevertheless, the task of developing calibration recall systems ordinarily falls to individuals with minimal background. This usually means “starting from square one,” only to discover after extensive effort that the ensuing systems fail to achieve desired objectives and/or are unacceptable to auditors from customer organizations.

The reasons for this are varied. First, the problem is complicated by the fact that calibration is concerned with so many different types of equipment, e.g., electrical, electronic, microwave, physical, radiometric, etc. Second, each organization requiring calibration of TME and standards is confronted with its own unique minimum reliability requirements, failure definitions, cost constraints and testing procedures, as determined by the product to be provided and by the individual customer’s application requirements. Third, it is often difficult to ascertain precisely what the goals of a calibration interval establishment and adjustment methodology should be. This is due in part to seemingly conflicting objectives that typically accompany product quality assurance. Generally, these objectives are

- The customer’s requirement for accurate, high-performance, high-quality products
- The producer’s requirement for a high probability of product acceptance
- The requirement for minimizing test and calibration costs, a requirement usually associated with the producer, but often of concern to both producer and customer.

Although satisfying all three requirements is often difficult, methods and techniques have emerged for establishing and adjusting calibration intervals that promote meeting both product assurance and cost control objectives.

6.1.4 Basic Concepts

To appreciate the need for maintaining calibration intervals and motivate the methodologies necessary for their determination and adjustment, it is worthwhile to review several basic ideas. First, it is important to keep in mind that test and calibration infrastructures are established to ensure that end-items, such as communication equipment, navigation systems, attitude control systems, etc., perform as intended. Performance of such systems can be related to the various measurable attributes that characterize them. For example, the ability of a microwave communication system to receive a weak signal is a function of its antenna gain (as well as other parameters). Hence, antenna gain is a measurable attribute by which communication system performance can be quantified. In this section, it is assumed that end-items will not perform as intended unless the values of their various measurable attributes are maintained within definable limits. Providing assurance that these limits are maintained is the primary motivation for testing and calibration.

The extent to which the value of a parameter of a given item of TME can be known at calibration is determined by a number of variables. These include the uncertainty of the calibrating equipment, the precision with which measurements can be made, the stability of the measurement process, the skill of the person performing the calibration, etc. Immediately following calibration, knowledge of a parameter's value is constrained to a range of values that can be fairly well-specified. After a time, however, this range becomes less well defined. Because of inherent random processes and the diversity of usage and environmental stresses, parameter values tend to vary randomly. This random variation spreads the distribution of parameter values from their "starting" values at time of calibration (defined as BOP in Section 5). As time passes, the spread of parameter values increases. Thus the uncertainty surrounding the value of each calibrated parameter grows as time elapses since calibration.

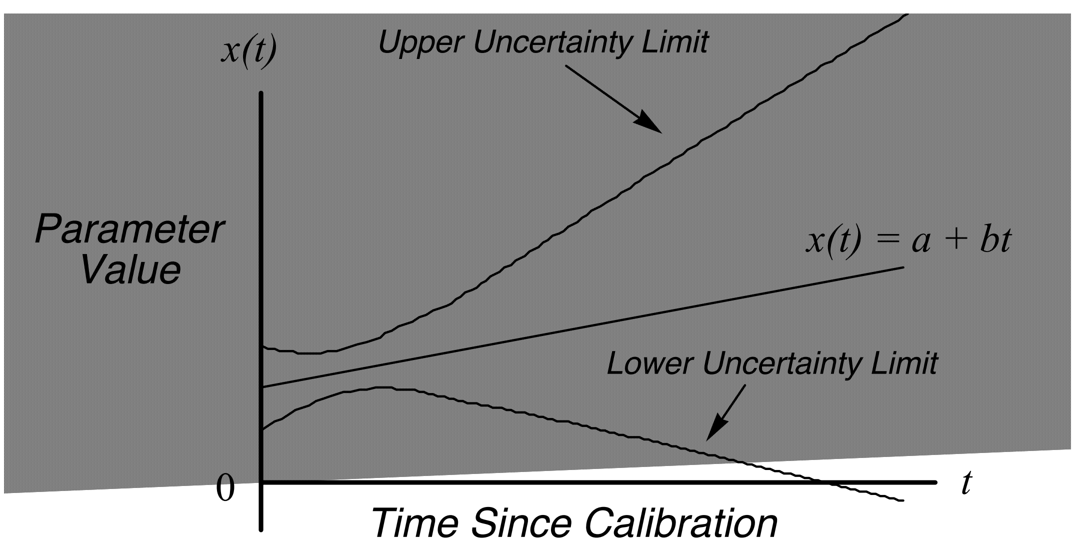


FIGURE 6.1 — PARAMETER UNCERTAINTY GROWTH.

Knowledge of the value of a calibrated parameter becomes less certain as time elapses since calibration. The case shown depicts a parameter whose value is known to drift linearly with time. The increased spreading of the upper and lower uncertainty curves is typical for this kind of time dependence.

> TME and standards are calibrated at periodic intervals to limit the growth of measurement uncertainty to acceptable limits. The calibration interval is determined from considerations of whether the expected level of uncertainty growth has exceeded these limits.

It should be noted that in many organizations acceptable uncertainty limits are subjectively arrived at. In organizations concerned primarily with ensuring measurement integrity, such as high-level standards laboratories, such subjective determinations tend to be conservative: i.e., they tend to lead to intervals between calibrations that are often shorter than may be economically justifiable. Conversely, in organizations that are concerned primarily with economics rather than with measurement integrity, intervals between calibrations often tend to be longer than that which is justifiable for prudent measurement uncertainty control.

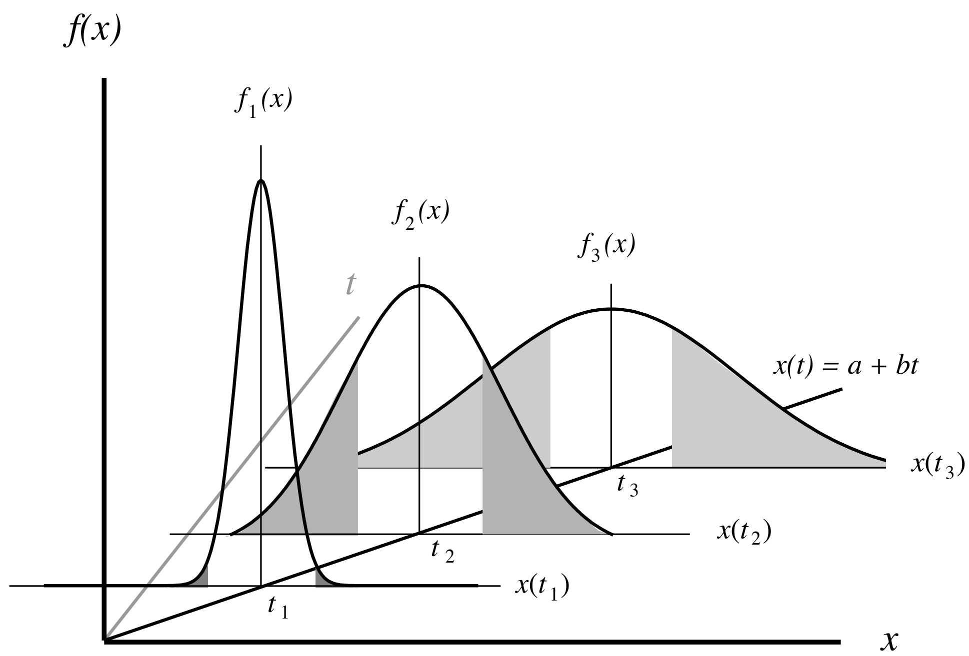


FIGURE 6.2 — MEASUREMENT UNCERTAINTY GROWTH.
Growth in uncertainty is shown for the parameter of Figure 6.1. The confidence in our knowledge of the parameter's value diminishes as time since calibration elapses. This confidence is indicated by the bell-shaped distribution curves for times $t_1 > t_2 > t_3$. The wider the spread of the curve, the greater the uncertainty in the parameter value. The shaded areas represent the probability for parameter out-of-tolerance. This probability increases as time elapsed since calibration increases.

This section describes approaches for determining intervals between calibrations that are commensurate with both cost constraints and measurement integrity requirements.

6.2 Management Considerations

Certain management concepts relevant to the implementation and operation of TME calibration recall systems are discussed here. The concepts presented relate to designing, developing and maintaining a capability to establish optimal intervals between TME calibrations.

6.2.1 Establishing the Need for Calibration Interval Analysis Systems

TME employed to verify the uncertainty of measurement processes require calibration to ensure that their verifying attributes are performing within appropriate accuracy specifications. Since the uncertainties in the values of such attributes tend to grow with time since last calibrated, such TME require periodic recalibration. For cost-effective operation, intervals between recalibrations should be optimized to achieve a balance between operational support costs and the TME accuracy requirements.

Different TME designs exhibit different rates of uncertainty growth. In addition, uncertainty growth rates are influenced by different conditions of usage and environment. Consequently, not all optimal TME recalibration intervals are alike. If recalibration is to be optimized, therefore, a unique interval is needed for each TME model employed under each specified set of usage and environmental conditions. Establishing such intervals requires the application of advanced calibration interval analysis methods.

6.2.2 Measurement Reliability Targets

TME are calibrated at periodic intervals to hold the growth of measurement uncertainty to within acceptable limits. In so doing, the prolonged use of out-of-tolerance TME is prevented and the validity of TME calibrations, tests, or other verifications are enhanced.

As Figure 6.2 shows, as the uncertainty in the value of a TME parameter grows, the probability that the parameter will be found in-tolerance decreases. Controlling uncertainty growth to within an acceptable maximum is, therefore, equivalent to controlling in-tolerance probability to an acceptable minimum. This acceptable minimum is referred to as the *measurement reliability* (or percent in-tolerance).

What constitutes an appropriate measurement reliability target is determined by the requirements for calibration accuracy. Measurement reliability targets are usually referenced to the end of the calibration interval (EOP targets) or to a value averaged over the duration of the calibration interval (AOP targets).

6.2.3 Calibration Interval Objectives

> The immediate objective of calibration interval analysis systems is the establishment of calibration intervals which ensure that appropriate measurement reliability targets are met.

A goal of any calibration interval analysis system should be that the cost per interval is held to a minimum. This requirement, when coupled with the requirement for meeting measurement reliability targets, leads to the following objectives of effective calibration interval analysis systems:

- Establishment of appropriate measurement reliability targets

- Establishment or adjustment of intervals to meet reliability targets

- Employment of algorithms and methods that arrive at the correct intervals in the shortest possible time

- Calibration intervals determined with a minimum of human intervention and manual labor.

Since the early 1960s experience with alternative approaches has shown that these objectives can be accomplished by employing the statistical calibration interval analysis methodologies described in this Section and in Appendix B.

In addition to these objectives, calibration interval analysis systems should permit easy and expedient implementation of analysis results. The results should be comprehensive, informative and unambiguous. Mechanisms should be in place to either couple the analysis results directly to an associated equipment control system or to transfer information to the equipment control system with a minimum of restatement or translation.

6.2.4 Potential Spin-offs

Because of the nature of the data being processed and the kinds of analyses being performed, calibration interval analysis systems are inherently capable of providing “spin-offs.”

One potential spin-off is the identification of TME with exceptionally high or low uncertainty growth rates (“dogs” or “gems,” respectively). As will be discussed in Section 6.3, dogs and gems can be identified by TME serial-number and by manufacturer and model. Identifying serial number dogs helps weed out poor performers; identifying serial-number gems helps in selecting items to be used as check standards. Model-number dog and gem identification can assist in making procurement decisions.

Other potential spin-offs include the potential to:

- Provide visibility of trends in uncertainty growth rate or calibration interval

- Identify users associated with exceptionally high incidence of out-of-tolerance or repair

- Project test and calibration workload changes to be anticipated as a result of calibration interval changes

- Identify calibration or test technicians who generate unusual data patterns.

> Calibration interval analysis systems also offer some unique possibilities as potential testbeds for evaluating alternative reliability targets, adjustment policies, and equipment tolerance limits in terms of their impact on calibration workloads.

6.2.5 Calibration Interval Elements

Implementing the capability for calibration interval analysis within an organization can have an impact on facilities, equipment, procedures, and personnel. To assist in evaluating this impact, several of the more predominant elements related to calibration interval analysis system design, development, and maintenance are described below. These elements include

- Data collection and storage
- Reliability modeling
- Statistical analysis of calibration results
- Engineering analysis
- Logistics analysis
- Cost/benefits
- Personnel requirements
- Training and communications.

6.2.5.1 Data Collection and Storage

Calibration history data are required to infer the time dependence of TME uncertainty growth processes. These data must be complete, homogeneous, comprehensive, and accurate.

Completeness — Data are complete when no calibration actions are missing. Completeness is assured by recording and storing all calibration results.

Homogeneity — Data are homogeneous when all calibrations on a homogeneous equipment grouping (e.g., manufacturer/model) are performed to the same tolerances using the same procedure.

Comprehensiveness — Data are comprehensive when "condition received" (condition as received for calibration), "action taken" (correction, adjustment, repair, etc., executed during calibration), and "condition released" (condition as deployed following calibration) are unambiguously specified for each calibration. Date calibrated, date released, serial or other individual ID number, model number, and standardized noun nomenclature are also required for comprehensiveness. For detection of facility and technician outliers, the calibrating facility designation and the technician identity should be recorded and stored for each calibration. Finally, if intervals are to be analyzed by parameter, the procedural step identification number is a required data element.

Accuracy — Data are accurate when they reflect the actual perceived condition of equipment as received for calibration, the actual servicing actions executed, and the actual perceived condition of equipment upon return from calibration. Data accuracy depends on calibrating personnel using data formats properly. Often data accuracy can be enhanced by designing these formats so that provision is made for recording all calibration results noted and all service actions taken. Instances have been encountered where deficiencies not provided for on data input formats tend to make their presence known in unrelated data fields. For example, stabilizing adjustments made on in-tolerance parameters are sometimes wrongly (but intentionally) recorded as out-of-tolerances.

6.2.5.2 Reliability Modeling

Uncertainty growth processes are described in terms of mathematical reliability models. Use of these models greatly facilitates the determination of optimal calibration intervals and the realization of spin-offs already noted. Reliability modeling is described in Section 6.3 and in

Appendix B.

6.2.5.3 Statistical Analysis of Calibration Results

Since equipment parameter drift and other fluctuations are subject to inherently random processes and to random stresses encountered during usage, the analysis of parameter behavior requires the application of statistical methods. Statistical methods are used to fit reliability models to uncertainty growth data and to identify exceptional (outlier) circumstances or equipment. The methods are described in Appendix B.

6.2.5.4 Engineering Analysis

Engineering analyses are performed to establish homogeneous TME groupings (e.g., standardized noun nomenclatures), to provide integrity checks of statistical analysis results, and to develop heuristic interval estimates in cases where calibration data are not sufficient for statistical analysis (e.g., initial intervals).

6.2.5.5 Logistics Analysis

Logistics considerations must be taken into account to synchronize intervals to achievable maintenance schedules. Interval synchronization is also required in setting intervals for TME models, such as mainframes and plug-ins that are used together.

6.2.5.6 Costs and Benefits

Operating Costs — Obviously, higher frequencies of calibration (shorter intervals) result in higher operational support costs. However, because of uncertainty growth, longer intervals lead to higher probabilities of using out-of-tolerance TME for longer periods of time.

Determination of the balance between operational costs and risks associated with the use of out-of-tolerance TME requires the application of methods described in Section 5 and Appendix C. These methods enable optimizing calibration frequency through the determination of appropriate measurement reliability targets.

Development and Maintenance Costs — Cost and benefits trade-offs are also evident in budgeting for the development and maintenance of calibration interval analysis systems. A significant factor is the anticipated system life expectancy. Designing and developing interval analysis systems that employ state-of-the-art methods can be costly. On the other hand, such methods are likely to be more applicable to future TME designs and to future technology management requirements than less sophisticated methods, which translates to greater system longevity and lower life cycle maintenance costs.

Another significant factor is the benefit to be derived from calibration interval analysis system spin-offs. Cost savings and cost avoidance made possible by these supplemental diagnostic and reporting capabilities must be included with operational cost factors in weighing system development and maintenance costs against potential benefits.

6.2.5.7 Personnel Requirements

Highly trained and experienced personnel are required for the design and development of statistical calibration interval analysis systems. Besides advanced training in statistics and probability theory, personnel must be familiar with TME uncertainty growth mechanisms in particular and with measurement science and engineering principles in general. Knowledge of the calibration facility and associated operations is required, as is familiarity with calibration procedures, calibration formats, and calibration history databases. In addition, both scientific and business programming knowledge are invaluable for system development.

6.2.5.8 Training and Communications

Training and communications are required to apprise managers, engineers and technicians about what the interval analysis system is designed to do and what is required to ensure its successful operation. Agreement between system designers and calibrating technicians on terminology, interpretation of data formats, and administrative procedures is needed to ensure that system results match real-world TME behavior. In addition, an understanding of the principles of uncertainty growth and an appreciation for how calibration data are used in establishing and adjusting intervals are required to promote data accuracy.

Comprehensive user and system also required to ensure successful system operation and longevity.

> Unfortunately, calibration interval systems are not immune to "improvements" made by personnel unfamiliar with system theory and operation.

A prime example of this is found in a Southern California company whose calibration interval system was designed and developed in 1978. Because it employs advanced methodologies and is fully automated, the system is considered technologically viable by today's standards. Regrettably, its data integrity has been seriously compromised by personnel unfamiliar with its design principles. These individuals mistakenly decided that certain important data elements were superfluous and could be eliminated.

6.2.6 Extended Deployment Considerations

For some applications, TME cannot be calibrated according to recommended or established calibration schedules. In these instances, alternatives or supplements to calibration are advisable. One alternative involves the use of high-accuracy ratios between TME parameters and end-item attributes. In cases where this is not feasible, a statistical process control supplement is recommended.

6.2.6.1 Calibration Alternative—Using High Accuracy Ratios

Experimentation with a prototype decision support system has shown that TME parameters that are inherently and significantly more accurate than the attributes they support seldom require periodic calibration. Roughly speaking, TME parameters with significantly tighter tolerances than the attribute tolerances they support can forego calibration for extended periods. This is because the values accessible to a parameter are usually physically constrained by design to prevent the parameter from attaining values at extreme divergence from the stated tolerance limits. This means that the range of values accessible to a TME parameter will remain well within the tolerance limit of the end-item attribute it supports in cases where the relative attribute-to-TME parameter tolerance ratio is large. This ratio is traditionally referred to as the TME-to-end-item "accuracy ratio."

A high accuracy ratio between a TME parameter and an end-item attribute implies that the relative uncertainty between the measurement process and the attribute is low. From the discussion in Section 4, it can be seen that this corresponds to a situation in which the end-item average utility is insensitive to test process uncertainty.

What constitutes a “high” accuracy ratio is determined by case-by-case analyses. Such analyses extrapolate parameter uncertainty growth to extended periods. This is done to determine whether maximum expected TME parameter uncertainties lead to inadequate testing of the attribute(s) to be supported.

6.2.6.2 Calibration Alternative—Implementing SMPC Methods

SMPC methods have been developed in recent years to supplement periodic calibration of test and calibration systems. These methods can be incorporated in automated test equipment (**ATE**), automated calibration equipment (**ACE**) and end-items to provide on-line indicators of in- or out-of-tolerance probability at the attribute or parameter level.

The methods employ Bayesian identities that permit role-swapping between calibrating or testing systems and units under test or calibration. By role-swapping manipulation, recorded measurements can be used to assess the in-tolerance probability of the testing or calibrating parameter. The process is supplemented by knowledge of time elapsed since calibration of the testing or calibrating parameter and of the unit under test or calibration. The methods have been extended to provide not only an in-tolerance probability for the testing or calibrating parameter but also an estimate of the parameter’s error or bias.

Using these methods permits on-line statistical process control of the accuracies of TME parameters. The methods can be incorporated by embedding them in measurement controllers.

The SMPC methods work best with a repository of intercomparison results to draw from. This is an important point in selecting or specifying ATE or ACE memory sizes. If the new methods are to be implemented, adequate controller or other memory should be planned for storing intercomparison histories for parameters of interest.

6.3 Technical Considerations

Several ideas are key to the development of optimal calibration recall systems. These ideas are central to defining the calibration interval problem as one that addresses the control of TME measurement uncertainty. The link between the calibration interval problem and measurement uncertainty control is established through transitioning of TME parameters from in-tolerance to out-of-tolerance states.

6.3.1 The Calibration Interval Problem

To summarize the material presented so far, the calibration interval problem consists of the following:

> Determine intervals between TME calibrations that limit or control TME measurement uncertainties to acceptable levels.

TME measurement uncertainties are controlled to limit end-item test-decision risk. Test-decision risk is, in turn, limited to control end-item measurement uncertainties. Finally, end-item measurement uncertainties are controlled to ensure acceptable end-item utility or performance. In this way, calibration intervals impact end-item performance. In keeping with the primary objective of test and calibration support infrastructures, i.e., the support of end-items, calibration intervals should be managed so that their impact on end-item performance is beneficial.

For TME and calibration standards installed onboard satellites or deep-space probes not accessible for periodic recalibration, the principles of calibration interval analysis can still be used to evaluate whether these devices can hold their respective tolerances over the duration of the mission they support.

6.3.2 Measurement Reliability

End-item utility is related to the uncertainty of the process surrounding verification of end-item compliance with specifications. In Section 4 it was pointed out that a major component of test process uncertainty is the uncertainty in the measuring parameters of the associated TME. As implied by Figure 6.2, parameter uncertainty can be expressed in terms of parameter in-tolerance probability.

For a given population of TME, the in-tolerance probability for a parameter of interest can be measured in terms of the percentage of observations on this parameter that correspond to in-tolerance conditions. In Appendix B, it is shown that the fraction of observations on a given TME parameter that are classified as in-tolerance at calibration is a *maximum-likelihood-estimate* (**MLE**) of the in-tolerance probability for the parameter. Thus, since in-tolerance probability is a measure of test process uncertainty, the percentage of calibrations that yield in-tolerance observations provides an indication of this uncertainty. This leads to using “percent observed in-tolerance” as the variable by which test process uncertainty is monitored.

The percent observed in-tolerance is referred to as *measurement reliability*, which is defined as

> MEASUREMENT RELIABILITY – The probability that a measurement attribute (parameter) of an item of equipment is in conformance with performance specifications.

An effective way to impose a limit on measurement process uncertainty involves the application of a minimum acceptable measurement reliability criterion or *measurement reliability target*. A primary objective of optimal calibration interval analysis is, accordingly,

> Establish measurement reliability targets commensurate with end-item utility objectives, and test and calibration support cost constraints.

The connection between end-item utility and TME measurement reliability has been described. Cost considerations are another matter. Since costs involve not only obvious factors, such as cost of calibration and repair, but also include indirect costs associated with false accepts/rejects (system downtime, product liability lawsuits, warranty expenses, etc.), finding the balance between attaining a desired level of measurement reliability and what it costs to attain it is a multifaceted and difficult process. The process is described in Appendix C.

In practice, many organizations have found it expedient to manage measurement reliability at the instrument rather than the parameter level. In these cases, an item of TME is considered out-of-tolerance if one or more of its parameters is found to be out-of-tolerance. Variations on this theme are possible.

6.3.3 Calibration Interval System Objectives

The effectiveness of a system designed to control test process uncertainty is measured in terms of how well actual TME in-tolerance percentages match established measurement reliability targets. A primary objective of any system created to determine and adjust TME calibration intervals is

> Estimate calibration intervals that yield the desired measurement reliability target(s), i.e., determine "optimal" intervals.

Since measurement uncertainty grows with time since calibration (see Figures 6.1 and 6.2), measurement reliability decreases with time since calibration. The particular time since calibration that corresponds to the established measurement reliability target is the optimal calibration interval. In some applications, periodic TME recalibrations are not possible (as with TME on-board deep-space probes) or are not economically feasible (as with TME on-board orbiting satellites). In these cases, TME measurement uncertainty is controlled by designing the TME and ancillary equipment or software to maintain a measurement reliability level which will not fall below the minimum acceptable reliability target for the duration of the mission.

A second objective of calibration interval analysis systems is

> Determine optimal intervals in the shortest possible time at minimum expense and minimum negative impact on resources.

In practice, the relationship between time since calibration and measurement reliability is sought in a number of ways. Not all approaches work. Some work in principle, but fail to do so within the lifetime of the TME of interest.

In many instances, the connection between the out-of-tolerance process and calibration interval is not well understood. This leads to intervals that are suboptimal with respect to the above objectives. It is worthwhile to consider the consequences of such suboptimal systems. Appendix B describes these consequences in detail and provides guidelines for establishing optimal systems.

6.3.4 The Out-of-Tolerance Process

TME are subjected to stresses that occur randomly during use and/or storage. For many electrical and electronic TME parameters, these stresses cause shifts in value that occur randomly with respect to magnitude and direction. Although the parameters of certain mechanical and dimensional TME may shift or drift in ways that are fairly predictable, they too are subject to stresses that cause random changes in value. Besides sensitivity to externally applied stresses, high-precision TME also exhibit shifts in parameter values arising from inherent random processes.

Just as gases of randomly moving molecules expand to fill containers, random TME parameter variations tend to spread across the spectrum of all accessible values. This is the principle behind uncertainty growth. The rate at which parameter values spread is the uncertainty growth rate. Since uncertainty growth arises from random processes, out-of-tolerances occur as random events. Out-of-tolerance events can be used to infer information about underlying uncertainty growth processes.

The uncertainty growth process can be determined by constructing "experiments" in which samples of TME are calibrated at various times elapsed since calibration. (In practice, experiments of this kind are not carried out. Instead, samples are taken from calibration history data.) Measurement reliability estimates are obtained for each sample by dividing the number observed in-tolerance by the number calibrated in the sample. These estimates are arranged chronologically to form a *time series* (see Appendix B). The uncertainty growth process is inferred from the time series through measurement reliability modeling. The calibration interval determination process is summarized in Table 6.1.

TABLE 6.1 Calibration Interval Key Ideas

Measurement Reliability
- *Probability that a TME parameter is in-tolerance*

Measurement Reliability Targets
- *Percent in-tolerance objectives for TME parameters*

Goals of Optimal Calibration Intervals
- *Establish recalibration schedules that ensure that measurement reliability targets are maintained*
- *Determine intervals in the shortest possible time at minimum expense and minimum negative impact on resources*

The Out-of-Tolerance Process
- *Out-of-tolerances occur as random events*
- *The uncertainty growth process governs the rate of these occurrences*
- *The uncertainty growth process can be described as a time series*
- *The out-of-tolerance process is modeled using time series analysis*

Measurement Reliability Modeling
- *Represent the time series with mathematical reliability models*
- *Construct the likelihood functions*
- *Obtain maximum likelihood estimates of reliability model coefficients (analyze the time series to infer the uncertainty growth process)*
- *Select the appropriate reliability model*

Calibration Interval Estimation
- *Set the reliability model equal to the reliability target and solve for the interval*

6.3.5 Measurement Reliability Modeling

A number of uncertainty growth processes are possible. Each process corresponds to a particular mathematical description or model. Each model consists of a mathematical form characterized by statistical parameters. Models are used to represent the observed measurement reliability time series described in the previous section.

> A model is considered as a possible representative of an uncertainty growth process when its statistical parameters have been adjusted to achieve the closest agreement possible between the model and the observed time series.

The method employed for achieving this agreement is referred to as MLE. The MLE method is described in Appendix B. By submitting each model to a statistical and engineering selection procedure, the model that best represents the uncertainty growth process can be identified.

The selected model is used to compute measurement reliability as a function of time. The desired calibration interval is determined by setting the computed measurement reliability equal to the measurement reliability target established for the TME under study. The procedure is described in Appendix B.

6.3.6 Calibration Interval Assignment and Adjustment

Calibration data must be reviewed periodically to refine or modify existing calibration intervals. This is motivated by three considerations. First, the “accuracy” with which reliability modeling represents the out-of-tolerance process is generally influenced by the amount of calibration data used to estimate the reliability model coefficients and to select the appropriate model. Other factors being equal, the more data, the better the results. Second, as TME populations age, their characteristic uncertainty growth rates may accelerate. By reviewing updated calibration data periodically, uncertainty growth rate changes can be detected and adjusted to. Third, periodic review is required to respond to changes in calibration procedures. A calibration procedure change may produce changes in recorded out-of-tolerance rates and require discarding of calibration history before the date of the change.

An interval adjustment may either shorten or lengthen an interval. In the discussion that follows, both adjustments are treated as being equal, with no distinction made between the QA approval requirements for, or advisability of, each. The discussion merely assumes that any interval adjustment (longer or shorter) is based on supporting data and that the adjustment is made in such a way as to strive toward meeting specified reliability targets. There are three major levels at which calibration interval adjustments are implemented:

(1) Adjustment by serial number.

(2) Adjustment by model number family.

(3) Adjustment by instrument class.

6.3.6.1 Adjustment by Serial Number

Serial Number Analysis — Even though serial numbered items of a given manufacturer/model group are inherently similar, they are not necessarily identical. Also, the nature and frequency of usage of individual items and their respective in-use environmental conditions may vary. Thus, some may perform better and others may perform worse than the average. For this reason, some organizations analyze calibration intervals at the individual serial-number level. The various methods used base these analyses on the calibration history of each item and give simple-to-complicated rules or look-up procedures for interval adjustment. Most of these methods assume that the "correct" calibration interval for an individual instrument is subject to change over its life span, and that, therefore, only data taken from recent calibrations are relevant for establishing its interval.

It has been shown that the relevant data required ordinarily cannot be accumulated at the single serial-number level to establish a "correct" interval for an individual item. Even if the restriction of using only recent data could be lifted, it would normally take somewhere between fifteen and sixty years (often longer than the instrument's useful life) to accumulate sufficient data for an accurate analysis.

> These considerations argue that calibration interval adjustment for a given serial-numbered item cannot ordinarily be justified solely on the basis of an analysis of calibration data taken on the serial number.

Serial-Number Assignment and Adjustment — Although calibration interval analysis at the serial-number level may not be feasible in most applications, calibration interval adjustment may be feasible at this level if such adjustment is made with the cognizance that sufficient data must be accumulated to justify the action. Appropriate serial-number interval adjustment approaches involve calibration interval analysis at the model-number level or at some other grouping level, with interval *adjustment* performed at the serial-number level.

These adjustments take into account whether calibration data taken on the serial-numbered item in question are homogeneous with calibration data taken on the grouping. The decision whether to adjust would be influenced by statistical tests of this homogeneity to evaluate the appropriateness of calibrating the serial-numbered item at the frequency established by the calibration interval for the group.

Special measurement reliability target requirements may pertain to the serial-numbered item. If a given serial-numbered item requires a higher measurement reliability than is normally assigned for routine applications, the computed interval (see Appendix B) for the grouping, based on this higher target, can be assigned to the individual item.

Parameter Within Serial-Number Analysis — If calibration data are recorded and analyzed by instrument parameter, further serial-number calibration interval fine-tuning is possible. This involves accumulating and analyzing data on specific parameters for each manufacturer/model level grouping of interest. The recommended analytical methods are the same as those used for analysis at the manufacturer/model level, with reliability targets imposed by parameter instead of by manufacturer/model. This results in calibration intervals being established by *parameter*. Calibration intervals can be assigned at the serial-number level by selecting the shortest applicable parameter interval. In this approach, known as Ferling's method, only those parameters used for each serial-numbered item are involved in the selection process. Further refinement is possible if individual measurement reliability targets are exercised at the parameter level.

6.3.6.2 Adjustment by Model-Number Family

Model-Number Analysis — Each serial-numbered item of a given model-number family is typically built to a uniform set of design and component specifications. Moreover, even though design and/or production changes may occur, items of the same model number are generally expected to meet a uniform set of published performance specifications. For these reasons, most serial-numbered items of a given model number should be expected to exhibit homogeneous measurement reliability behavior over time, unless demonstrated otherwise.

The model-number identification is unique and hence makes possible a systematic accumulation of homogeneous calibration history. In some cases, enough model-number data for a valid statistical analysis can be accumulated in less than a year, where there are large inventories of a model number and short intervals.

The following conditions are necessary to ensure the accuracy and utility of adjustments based on these analyses:

(1) Calibration history data are complete and comprehensive; a good rule is to require data to be maintained by serial number, with all calibrations recorded or accounted for.

(2) Calibration history data are homogeneous. To ensure the validity of the calibration interval "experiment," data must be homogeneous with respect to the level (parameter, serial number, model number, instrument class) at which the interval analysis will be performed and with respect to the calibration procedure and parameter tolerances used.

(3) Calibration history data are reviewed and analyzed, and calibration intervals are adjusted in accordance with the guidelines given in (6) below.

(4) Mathematical failure models are used to model measurement reliability behavior and the model or models used must be appropriate; i.e., they model the process by which equipment transitions from an in-tolerance to an out-of-tolerance state. Mathematical models that have been found useful for this purpose are described in Appendix B. Other models can be found in the reliability analysis and statistics literature or can be specially constructed to represent the specific in-tolerance to out-of-tolerance transition mechanisms of interest.

(5) Analysis techniques for fitting reliability models to calibration history data are based on statistically valid methods. Such methods include the method of moments, maximum-likelihood-estimation, least-squares analysis, or Bayesian estimation. The method advocated in this publication is maximum-likelihood estimation, which is described in Appendix B.

(6) Interval adjustments should be made in a way that does not compromise reliability requirements. Interval extensions that reduce calibration costs are encouraged, provided that reliability targets are adhered to.

Some amplification is needed to determine when review and analysis of calibration history data are appropriate. Review is appropriate when any of the following applies:

(1) Sufficient data have been accumulated to justify a reanalysis.

(2) Some relevant procedural or policy modification (changes in calibration procedure, reliability target, equipment application or usage, etc.) has been implemented since the previous interval assignment or adjustment.

(3) Equipment is known to have a definable performance trend, and enough time has elapsed for the trend to require an interval change.

Notwithstanding these criteria, a quarterly to annual review and analysis should be sufficient for all but "problem" equipment, critical application equipment, etc.

Dog/Gem Identification — The requirements for valid calibration intervals, based on analysis of data sufficient for accurate measurement reliability modeling, and the need for responsiveness to instrument idiosyncrasies both can be accommodated by incorporating a means of statistically identifying exceptional items within a homogeneous grouping. In such schemes, calibration data are indexed by item for the grouping. Items with significantly higher and lower out-of-tolerance frequencies than are characteristic of the group may be flagged by a unique item identifier (e.g., serial number, procedure step number, etc.). Statistical outliers identified in this way are commonly referred to as "dogs" (high out-of-tolerance rate) and "gems" (low out-of-tolerance rate). In particular, the presence of dogs unduly shortens the calibration interval for other items in the grouping. Removing these outliers provides greater assurance that the assigned interval is representative. Finally, flagging outliers ensures responsiveness to individual behavior.

Dog/Gem Management — Various methods may be devised for identifying such outliers. The preferred methods are statistical. Once outliers are identified, considerable latitude is possible regarding their disposition. For example, dogs may require shortened intervals, complete overhaul, removal from service, certification for limited use only, etc. On the other hand, gems may qualify for lengthened intervals, designation as critical support items, or upgrade to higher level standards.

6.3.6.3 Adjustment by Instrument Class

In some cases, sufficient data for calibration interval analysis may not be available at the model-number level. One method of compensating for insufficient model-number data involves the creation of larger, approximately homogeneous groupings of equipment that may contain several model numbers. Such groupings are referred to as instrument classes. Pooling the calibration histories from model numbers within a class often yields sufficient data for analysis. The results of these analyses may be applied to model numbers within the class for which data are sparse or unavailable. Once a class has been defined, statistical homogeneity tests should be performed whenever possible to verify the validity of the equipment grouping.

Several criteria are used to define a class: commonality of function, application, accuracy, inherent stability, complexity, design, and technology. One class definition scheme that has proved useful consists of subgrouping within standardized noun nomenclature categories according to accuracy, stability, complexity, and date of issue.

Calibration interval analysis at the class level is performed in the same way as analysis at the model number family level, with data grouped according to class for interval analysis and by model number for dog and gem analysis. That is, dogs and gems are identified at the manufacturer and/or model number level.

6.3.6.4 Initial Intervals

At the commencement of an equipment's life cycle, its calibration recall process is inaugurated with an initial interval. Because the equipment is new to inventory, calibration history data are usually unavailable. This situation may call for subjective or engineering analysis methods of initial interval assignment. The assignment of initial calibration intervals should utilize all available calibration data and should promote the efficient generation of new data. Numerous methods are currently in use or are projected for future use. These methods are discussed below.

General Intervals — The most expedient way of introducing equipment into the calibration process is to assign an initiating recall cycle that is common for all new items. New items should remain on this interval until their calibration data indicate that an interval adjustment is appropriate. A conservative (i.e., short) interval will accelerate the generation of calibration history, thereby tending to spur the determination of an accurate interval. However, this expedient may set shorter intervals than are necessary, leading to high initial calibration support costs and unnecessary equipment downtime due to frequent recalls for calibration. Fortunately, more accurate initial intervals can be obtained by employing certain refinements, as discussed below.

Engineering Analysis — If the available relevant calibration data are insufficient for analysis, engineering analysis may be needed to establish initial intervals. Initial interval engineering analysis includes establishing similarity between equipment, evaluating manufacturer's recommendations, assigning instrument class designations, or evaluating externally available intervals.

Related Models and/or Similar Equipment — In some cases, a new TME is an updated version of an existing product line. It may be the same as its predecessor except for minor or cosmetic modifications. In such cases, the new TME is assumed to have performance characteristics similar to its parent model. Often, the parent model will already have an assigned calibration interval based on the analysis of calibration history. If so, the new model can be tentatively assigned the recall interval of the parent model.

In like fashion, when no direct family relationship can be used, the calibration interval of similar equipment or equipment of similar complexity and employing similar technologies may be appropriate.

Manufacturer Data/Recommendations — Another source of information is the manufacturer of the equipment. Manufacturers may provide recommended calibration interval information in their published equipment specifications. These recommendations are usually based on analyses of stability at the parametric level. To be valid, the specifications should accommodate three considerations:

(1) The parameter tolerance limits.

(2) The duration over which the parameter values will be contained within these limits.

(3) The percentage of items whose parameters will be contained within these limits over this duration.

Unfortunately, it appears that TME manufacturers are typically cognizant of only one or, at best, two of these points. Accordingly, some care must be taken when employing manufacturer interval recommendations. If the manufacturer recommended intervals *per se* are in question, supporting data and manufacturer expertise may, nevertheless, be helpful in setting accurate initial intervals.

Another option is to require the manufacturer to demonstrate the equipment's capability to meet a prescribed measurement reliability target. The manufacturer should either enter into a product demonstration interval verification test using a random sample of production units or accumulate stability data at the TME parameter level to determine a maximum-likelihood distribution of times to out-of-tolerance. This information can be employed to estimate measurement reliability levels that correspond to times between calibration.

Design Analysis — Another source of information is the TME design. Knowledgeable engineers can provide valuable information by identifying, describing, and evaluating the calibration critical circuits and components of the TME in question. An accurate calibration interval prediction is sometimes possible in lieu of calibration history data when equipment measurement parameter aggregate out-of-tolerance rates (**OOTR**) are determined via circuit analysis and parts performance. (OOTR is the inverse of the mean-time-between-out-of-tolerances [MTBOOT] referred to earlier.) The OOTR can be applied in mathematical reliability models, as if it were obtained from calibration history data, to determine an initial calibration interval estimate.

Instrument Class Assignment — If a new item can be assigned membership in an instrument class, the interval for the class will be applicable as an initial interval. Assignment in a class should be made according to the criteria previously discussed.

External Authority — If engineering analysis is not feasible, calibration intervals determined by an external organization may be used. It is strongly recommended that the external organization be similar to the requiring activity, with respect to reliability targets, calibration procedures, usage, handling, environment, etc. In cases where there are differences in these areas, adjustments must be made in the "borrowed" intervals. The magnitude and direction of these adjustments should be made with the engineering considerations (outlined above) in mind.

Adjustments for reliability targets may sometimes be made mathematically. For example, suppose that a model-number family can be modeled by the negative exponential function $R(t) = \exp(-\lambda t)$, where the parameter λ is the OOTR for the model-number family. Then, if the reliability target and interval for the external authority are R^* and I, respectively, the failure rate parameter λ can be obtained from

$$\lambda = \frac{-\ln R^*}{I} .$$

If the reliability target for the requiring organization is r*, the appropriate interval is calculated as

$$\text{interval} = \frac{-\ln r^*}{\lambda} = I \frac{\ln r^*}{\ln R^*} .$$

Intervals may also be computed from externally generated calibration history data. For example, the Department of Defense shares data among the services. Large equipment reliability databases may also be consulted. As a word of caution, some foreknowledge is needed of the quality and relevance of data obtained externally to ensure compatibility with the needs of the requiring organization.

6.3.7 Multiparameter TME

In discussing the relationship between TME calibration intervals and measurement reliability targets, it is implied that the item for which the calibration interval is being adjusted is the item that has been assigned the measurement reliability target. Previously, measurement reliability targets have been keyed to individual TME parameters. This is a natural consequence of the fact that TME usage requirements are best defined at the parameter level. However, TME are recalled for calibration at the instrument or system level. Thus, the calibration interval applies to the whole TME. This presents a problem if TME comprises more than one parameter, namely, that of establishing a calibration interval for the entire instrument, which is based on measurement reliability considerations for each individual constituent parameter.

6.3.7.1 Multiparameter TME Intervals

Calibration intervals for multiparameter TME can be determined in a number of ways. One of the most effective ways involves describing the TME in terms of a "measurement reliability network" in which each parameter is considered as a component of a functioning entity. The performance of the entity is measured in terms of the contributing performances of each parameter. For many TME applications, not all parameters are considered equal. Some may be highly critical, whereas others may provide only low-level support. Depending upon the application, a definable subset of parameters may have no use at all. These considerations lead to weighting schemes in which parameter criticality is taken into account.

The simplest illustration of such a weighting scheme is a two-parameter TME employed in an application that requires only one of the parameters. In this case, the useful parameter is assigned a weight of 1, whereas the unused parameter is assigned a weight of 0.

If both parameters are of equal criticality, each receives a weight of 0.5. To illustrate how such a weighting scheme relates to calibration interval determination, let $R_1(I)$ and $R_2(I)$ represent the in-tolerance probabilities of parameters 1 and 2, respectively, for a calibration interval I. If each is given an equal criticality weight coefficient, (c), then the "weighted" measurement reliability for the TME is given by

$$R(I) = 0.5R_1(I) + 0.5\,R_2(I).$$

Suppose that an overall measurement reliability target has been determined for the TME. If this target is labeled R^*, then the interval is obtained by setting $R(I) = R^*$ and solving for I. (Note that with criticality weighting schemes, individual parameter measurement reliability targets are implicit in the weighting factors.) If parameter 1 were assigned a weight of, say, $c_1 = 0.7$, then the interval I would be solved from

$$0.7R_1(I) + 0.3\,R_2(I) = R^*.$$

The situation is complicated by the fact that, in addition to unequal criticalities, parameters are not always used at the same frequency or rate, which should somehow be factored into the equation. For example, if parameter 1 is used three times as often as parameter 2, then the in-tolerance status of parameter 1 should have a greater bearing on the TME calibration interval than parameter 2. This is accounted for by the demand weight coefficient, (d). Calculating the demand function is similar to calculating the criticality weights but is slightly more complicated because the sum of products of demand weight values and criticality weights must be normalized to unity. This is facilitated by expressing both the criticality and demand weights as ratios. For example, for a 3:1 demand ratio ($D = 3$) for parameters 1 and 2, combined with the criticality weighting ratio ($C = 0.7/0.3$) of the previous example, the TME measurement reliability for the calibration interval I can be calculated as follows:

$$d_1 w_1 R_1(I) + d_2 w_2 R_2(I) = R^*.$$

where the criticality and demand weighting coefficients are obtained from

$$w_1 = \frac{C}{C+1}$$
$$w_2 = 1 - w_1$$
$$d_1 = \frac{D}{Dw_1 + w_2}$$
$$d_2 = \frac{d_1}{D}$$

Extension to more than two parameters is fairly straightforward. Note that the foregoing assumes that parameters 1 and 2 are independent of one another. If this is not the case, then solving for I becomes considerably more complex and is beyond the scope of this discussion.

The question arises that, since criticalities and demand coefficients are determined at the parameter level, what guides the determination of the measurement reliability target R^* for the TME? The answer lies in the fact that the weights w_1, w_2, ... represent *relative* criticalities of parameters 1, 2, ... The absolute criticalities come about as a result of assigning a criticality to the TME at the instrument level. This criticality is embodied in R^*.

Determination of criticality weighting factors and demand coefficients may be beyond the capability of many TME users. If so, some other technique for solving for I for multiparameter TME is needed that bypasses these determinations. The most promising method reported to date was proposed by Ferling in 1987. In Ferling's method, the interval for the TME is set equal to the shortest individual parameter interval. While this approach may at first appear overly simplified, it works very well from the standpoint of measurement reliability assurance. It offers a moderation of the traditional extreme view that all parameters of a multiparameter TME must be in-tolerance for the TME itself to be considered in-tolerance. By focusing attention on the "least reliable" parameter, Ferling's method does not compromise measurement uncertainty control.

Ferling's method is implemented as follows: If the measurement reliability models (see Appendix B) for the TME parameters are represented by $R_1(t)$, $R_2(t)$, …, $R_k(t)$, and the individual parameter measurement reliability targets by $R_1^*, R_2^*, \ldots, R_k^*$, then TME interval is equal to I_j, where $R_j\left(I_j\right) = R_j^*$.
Note that with Ferling's method parameter criticalities and demand coefficients are incorporated in the individual parameter reliability targets.

6.3.7.2 Stratified Calibration

The use of Ferling's method of setting multiparameter TME calibration intervals suggests a calibration approach that provides maximum support at minimum cost. In this approach, only the shortest interval parameter(s) is calibrated at each TME resubmission. The next shortest interval parameter is calibrated at every other calibration, the third shortest at every third calibration, and so on. Such a calibration schedule is similar to maintenance schedules that have been proven effective for both commercial and military applications. The term applied to a calibration schedule of this type is *stratified calibration*.

In stratified calibration, the shortest parameter interval is compared to intervals for other parameters to develop a scheme in which parameter intervals are whole number multiples of the shortest parameter interval. This ordinarily involves a certain amount of “rounding off” or approximating. For example, suppose that the TME of interest is a three-parameter instrument with parameter intervals of

$$I_1 = 3.3 \text{ months}$$

$$I_2 = 7.6 \text{ months}$$

$$I_3 = 17.1 \text{ months.}$$

A stratification scheme that strictly adheres to measurement reliability requirements would set the parameter intervals at

$$I_1' = 3 \text{ months}$$
$$I_2' = 6 \text{ months}$$
$$I_3' = 12 \text{ months.}$$

From a detailed review of the measurement reliability function, it may turn out that calibration of the third parameter at 18 months does not compromise its measurement reliability to a significant extent. If so, the stratified calibration scheme would be established at

$$I_1' = 3 \text{ months}$$
$$I_2' = 6 \text{ months}$$
$$I_3' = 12 \text{ months.}$$

By focusing calibration on only the subset of parameters that are due for service, stratified calibration schemes can offer significant potential operating cost savings without compromising TME measurement reliability. These schemes allow servicing to be performed without the need for special services out of sync with normal service cycles.

6.3.8 Equipment Adjustment Considerations

During calibration, decisions are made whether to adjust or correct parameters under test. Typically, TME is adjusted to match the values of its calibrating TME. Three categories of adjustment practice are encountered:

1. *Adjust if failed only* — With this practice, parameter values are adjusted only if found out-of-tolerance. This practice has been advocated as beneficial for parameters whose uncertainty growth is best controlled if the values of in-tolerance parameters are not tampered with. It has also been advocated in the past, due to analytical state-of-the-art limitations, that only failed items be adjusted to enable reliability analysis of data. This limitation is no longer applicable.

2. *Adjust always* — This practice advocates optimizing, or adjusting to, “center of tolerance band” all parameters calibrated, regardless of in- or out-of-tolerance status. Analytical

resistance to this practice has softened since the mid ‘1970’s with the development of statistical tools appropriate for the analysis of adjust always data.

3. *Adjust as needed* — The practice of “adjust as needed” employs limits, not necessarily equal to a given parameter’s tolerance limits, which signals a need for adjustment or correction. If parameter values are found to be outside the specified percentage of the tolerance band, they are adjusted to center specification. If parameter values are found to be within the specified percentage of the tolerance band, they are left undisturbed.

Current interval analysis technology can accommodate all three adjustment practices, the only condition being that it must be known whether an adjustment action took place or not. This means that adjustment information must accompany each parameter calibration record.

Certain automated calibration systems adjust parameter values by employing software corrections rather than physical adjustments. Software corrections are not physically intrusive and are, accordingly, usually applied whether parameters are in- or out-of-tolerance. In automated calibration, correction factors are stored internally in the workload TME memory and are applied to all measurements made using the parameter or parameters under consideration.

Over well-behaved portions of a parameter’s operating curve, such corrections are entirely equivalent to physical adjustments. However, if parameter values drift or otherwise transition to unstable portions of their respective operating curves, software corrections alone are not advisable. This is because, in unstable portions of operating curves, parameter values shift at a faster than usual rate. A software correction in an unstable operating region is not as viable over an interval of time as it would be if it were made in a stable region. What this means is that parameters that are functioning in unstable portions of their operating curves and that are adjusted via software corrections would require shorter calibration intervals than if they were operating in stable portions of these curves.

> Software corrections should be limited to stable operating curve regions. Parameters that drift to unstable regions are to be physically adjusted to stable regions as needed.

6.3.9 Establishing Measurement Reliability Targets

Establishing measurement reliability targets involves a consideration of several trade-offs between the desirability of controlling measurement uncertainty growth and the cost associated with maintaining such control. The trade-offs are applicable whether the objective is managing a ground-based calibration interval analysis system or designing TME for spaceflight applications.

Establishment of an appropriate measurement reliability target is a multifaceted process. The major points in establishing a measurement reliability target follow:

- TME measurement reliability is a measure of TME uncertainty
- TME uncertainty is a major contributor to the uncertainty of the end-item’s calibration

- The uncertainty of the end-item's calibration impacts the uncertainty of the measurements made with the end-item
- Measurement uncertainties impact end-item usefulness.

Given that the immediate objective of setting a measurement reliability target is the control of TME measurement uncertainty, the above list provokes three central questions:

(1) How much does TME parameter uncertainty contribute to calibration uncertainty?

(2) How sensitive is end-item uncertainty to calibration uncertainty?

(3) How sensitive is end-item utility to end-item uncertainty?

The impact of TME uncertainty on total test process uncertainty can be established by considering end-item attribute value distributions resulting from testing with TME exhibiting maximum uncertainty (the lowest level of TME measurement reliability achievable in practice) and minimum uncertainty (measurement reliability = 1.0). If the range of end-item attribute values obtained under these extremes is negligible, then TME uncertainty is not a crucial issue, and measurement reliability targets can be set at low levels. In certain cases, it may even be determined that periodic recalibration of TME is not required.

If, however, end-item uncertainty proves to be a sensitive function of TME uncertainty, then TME measurement reliability takes on more significance, and measurement reliability targets must be set at high levels. Establishing optimal TME measurement reliability targets that are commensurate with end-item support requirements involves the use of specialized vertical uncertainty propagation and test decision risk analysis methods. These methods are described in detail in Appendix C. It should be stressed that not all cases are clear-cut with regard to the conditions listed in Table 6.2. Considerable ambiguity and numerous gray areas are likely to be encountered in practice.

For space-based applications, there is often no calibration interval *per se*. TME are operated without recalibration over a period of time that is often equivalent to the mission lifetime. In these applications, designing systems that will perform within required levels of accuracy is equivalent to designing systems that are inherently stable or that can tolerate low measurement reliability targets. From the foregoing, it is apparent that this can be achieved if the TME system is "over-designed" relative to what is required to support end-item tolerances. Such over design may involve the incorporation of highly stable components, built-in redundancy in measurement subsystems, etc. Alternatively, in cases where end-item tolerances are at the envelope of high-level measurement capability, it may be necessary to reduce the scope of the end-item's performance objectives. Another alternative involves the use of supplemental measurement assurance measures, as discussed in Section 6.4 and Appendix D.

TABLE 6.2 Measurement Reliability Target Rough Guidelines

TABLE 6.2
Measurement Reliability Target Rough Guidelines

CONDITION	TME MEASUREMENT RELIABILITY REQUIREMENT (%)
End-item utility is insensitive to attribute value	*<60*
Range of acceptable end-item attribute values is large relative to test process uncertainty	*<60*
Alternative (redundant) independent TME are planned for concurrent use	*< 60 – 90*
End-item application is critical	*>90*
End-item backups are unavailable	*>90*

6.3.10 The Interval Analysis Process

The process of establishing calibration intervals and/or evaluating measurement reliability over extended periods of time is summarized in Table 6.3 and consists of the following steps:

STEP 1. Determine end-item performance requirements in terms of acceptable end-item attribute values.

This involves evaluations of end-item utility versus attribute value for each end-item attribute. Based on these evaluations, meaningful end-item attribute uncertainty limits or *performance tolerance limits* can be established. Testing, measuring, or monitoring end-items to these limits is performed to ensure that end-item attributes will perform as intended.

STEP 2. Determine TME parameter tolerances that correspond to acceptable test process uncertainty.

Controlling end-item attribute uncertainty through testing, measuring, or monitoring requires that test process uncertainty be constrained to appropriate limits. As discussed in Sections 4 and 5, uncertainty in TME parameter values is a major contributor to overall end-item test process uncertainty. TME uncertainty is controlled by calibration to ensure compliance with established TME parameter tolerance limits. In addition, by evaluating false-accept and false-reject risks resulting from relative uncertainties of the test process and the end-item attributes, end-item attribute *test tolerance limits* can be developed that compensate for these risks.

STEP 3. Determine appropriate measurement reliability targets for TME parameters.

Controlling TME uncertainty requires that TME parameters be maintained within tolerance limits at some level of probability commensurate with test process uncertainty constraints. This probability level is the measurement reliability target.

> **STEP 4.** Collect data on TME parameters to provide visibility of TME uncertainty growth processes.

Visibility of the uncertainty growth process for each TME parameter is obtained by sampling the time series that reflects this process. Data can be collected through recording the results of periodic calibrations for TME deployed in ground-based applications or can be accumulated through controlled experimentation during TME design and development. For the latter, care must be exercised to match the experimental conditions with those anticipated in actual usage.

> **STEP 5.** Determine reliability models and coefficients by using maximum-likelihood estimation methods.

For most TME applications, the transition from an in-tolerance to an out-of-tolerance state is essentially a random phenomenon. Transition phenomena can be modeled using mathematical functions characterized by a mathematical form with appropriate coefficients. Sampled uncertainty growth–time series data are used to estimate the values of these coefficients.

> **STEP 6.** Identify the TME parameter uncertainty growth process. Select the appropriate measurement reliability model.

In some cases, the uncertainty growth mechanism and associated uncertainty growth process are known prior to analysis and the appropriate reliability model can be selected a priori. In most cases, however, the uncertainty growth process is revealed through analyzing data employing a set of candidate reliability models. Statistical tests can be applied a posteriori to select the model that provides the best uncertainty growth process representation.

> **STEP 7.** Compute calibration intervals commensurate with appropriate measurement reliability targets.

This involves setting the modeled measurement reliability function equal to the measurement reliability target and solving for the interval. This solution is a *maximum-likelihood interval* estimate. Decisions to adjust existing intervals can be assisted by determination of upper and lower calibration interval confidence limits. If current assigned intervals fall outside these limits, the intervals are adjusted to the maximum-likelihood estimates.

The process of establishing calibration intervals and/or evaluating measurement reliability over extended periods of time is summarized in the following table:

TABLE 6.3 The Calibration Interval Process

TABLE 6.3 ***The Calibration Interval Process***
Determine end-item performance tolerances in terms of acceptable end-item attribute values.
Determine TME parameter tolerances that correspond to acceptable test process uncertainty.
Determine appropriate measurement reliability targets for TME parameters.
Collect data on TME parameters to provide visibility of the uncertainty growth process.
Determine reliability models and coefficients using maximum-likelihood estimation methods.
Identify the TME parameter uncertainty growth process. Select the appropriate measurement reliability model.
Compute calibration intervals commensurate with appropriate measurement reliability targets.

6.3.11 Extended Deployment Considerations

Both TME and end-items are subject to uncertainty growth with time. A TME parameter uncertainty grows with time since calibration. End-item attribute value uncertainty grows with time since last tested.

Previous discussions in this section have focused mainly on calibration recall principles, methods and systems as applied to the problem of controlling uncertainty growth in TME accessible for periodic calibration. In this section, the same principles and methods will be applied to the problem of ensuring measurement uncertainty control for TME and end-items deployed on extended missions. The systems of interest are end-items without on-board testing support and TME without on-board calibration support.

Section 3 discussed metrology requirements for such systems. These requirements relate to designing subsystems to either (1) provide for calibration/test using built-in or on-board references and/or terrestrial and/or astronomical references or (2) to tolerate extended periods without calibration or testing. Using the measurement reliability modeling methods described in Appendix B, designs can be evaluated in terms of how well these objectives will be met. The interval analysis process described in the previous section applies with minor modification. Three cases are of interest and follow below.

6.3.11.1 Case 1—TME With Calibration History

The TME parameter or end-item attribute under consideration has a history of calibration or testing acquired over its operational life. In this case, determining the uncertainty growth process is accomplished as described in Section 6.3.10. For end-items, the procedure is the same as for TME, except that test data, rather than calibration data, are used, and the resulting intervals are test intervals rather than calibration intervals. It should be emphasized that for this procedure to be valid, the operational parameter or attribute tolerances and the conditions of usage must be the same as those planned for the mission of interest. If these conditions are not met, then Case 1 becomes equivalent to Case 2.

6.3.11.2 Case 2—New TME

The TME parameter or end-item attribute under consideration is part of a system that has been developed but that has not been introduced into operation or has not been operational long enough to accumulate a history of calibration or testing. In this case, complete Steps 1 through 3 of Section 6.3.10. When Step 3 is completed, use *empirical uncertainty growth modeling* (see below) to determine the measurement uncertainty growth process for the parameter or attribute.

6.3.11.3 Case 3—TME Under Development

The TME parameter or end-item attribute is in the design phase of its development. In this case, the interval analysis process is summarized in Table 6.4 and detailed here as follows.

STEP 1. *Determining end-item performance requirements in terms of acceptable end-item attribute values.*

This involves evaluations of end-item utility versus attribute value for each end-item attribute. Based on these evaluations, meaningful end-item attribute uncertainty limits or performance tolerance limits can be established. End-items are to be designed to ensure that attributes will perform within these limits over the duration of one or both of the following time intervals:

(1) Established testing intervals. This applies to end-items supported by on-board TME.

(2) The mission of interest or some pre-specified portion thereof. This applies to end-items not supported by on-board TME.

STEP 2. *Determining TME parameter tolerances that correspond to acceptable test process uncertainty.*

For unsupported on-board TME, design and fabrication functions focus on these limits as parameter uncertainty constraints to be maintained over the duration of one or both of the following:

(1) Established calibration intervals. This applies to TME supported by on-board (including built-in) standards.

(2) The mission of interest or some pre-specified portion thereof. This applies to TME not supported by on-board standards.

STEP 3. *Determining appropriate measurement reliability targets for end-item attributes or TME parameters.*

Controlling uncertainty over the course of a mission requires that attributes or parameters be maintained within tolerance limits at some level of probability commensurate with intended applications. This probability level serves as the measurement reliability target. For on-board TME, the intended application is testing of on-board end-items. For end-items, the application is specified according to mission requirements.

For extended deployment applications, measurement reliability targets should constitute design goals for each TME and end-item parameter. Ordinarily, this practice is not followed. For example, a specification for detector sensitivity might read something like the following:

24 hour stability limits:	±0.010 Vdc
90 day stability limits:	±0.020 Vdc
1 year stability limits:	±0.028 Vdc.

Such a specification is incomplete, especially for extended deployment applications. A third qualifier is needed. This third qualifier is the probability that the specified tolerance will be maintained over the intended period. For example, the complete detector specification would look something like the following:

DURATION	TOLERANCE	RELIABILITY TARGET
24 hours	±0.010 Vdc	0.982
90 days	±0.020 Vdc	0.985
1 year	±0.028 Vdc	0.940.

Without the third qualifier, it can be readily perceived that any tolerance can be specified for virtually any duration without reservation. For example, the TME contractor or manufacturer could have claimed a ±0.001 Vdc specification for a 24-hour period. This may be applicable in less than one case out of a thousand, but, if the probability of maintaining this spec for this period is nonzero, the specification can be upheld. It has been claimed by certain TME manufacturers that the probability implicit in parameter specifications is understood to be 1.0, i.e., there is *no* chance for an out-of-tolerance condition at the end of the specified time. There are two reasons why such claims are ill-conceived.

First, cases of 100% measurement reliability have rarely been observed in practice. Instead, out-of-tolerance percentages of 30% or higher have been routinely reported by TME calibrating organizations.

Second, stating tolerance limits in such a way that they carry with them a zero expectation for attribute or parameter out-of-tolerance is suboptimal for measurement uncertainty management. There are three problems related to this concern:

(1) If tolerance limits of ±X are expected to contain *all* values of an attribute parameter of interest, then so do tolerance limits of ±2X or ±3X or ⋯ . The question arises, which should be used?

(2) It might be argued that the tolerance limits ±X are *minimum* limits that will contain *all* values of the attribute or parameter. Some reflection shows that this is impossible unless parameters adhere to distributions with abrupt cutoff points. Such distributions are rarely encountered in practice.

(3) Such all-inclusive tolerances are ordinarily comprised of a curious mix of statistics and engineering fudge factors. While use of such devices may lead to "comfortable" or conservative equipment tolerances, they provide no statistical information on parameter stabilities. This information is essential for effective measurement decision risk management.

Establishing a reliability target for an end-item attribute is equivalent to establishing a maximum end-item attribute uncertainty level corresponding to a minimum acceptable end-item average utility.

STEP 4. *Ascertaining the uncertainty growth process for the end-item attributes of TME parameters of interest.*

In the design/development phase of a system's life cycle, visibility of the uncertainty growth process for each attribute or parameter is obtained in two stages. The first, measurement reliability network modeling and simulation, is applicable to the design phase. The second stage, empirical uncertainty growth modeling, is applicable to the preproduction or prototype phase.

Measurement reliability network modeling and simulation — In this stage, the components that make up the attribute or parameter of interest are integrated in a system configuration model that permits evaluation of measurement accuracy and stability under the range of component values and usage conditions anticipated in practice. These values and conditions are simulated and attribute or parameter responses are noted. Such simulations take into account all factors of usage, operation, storage, shipping, etc., to which the attribute or parameter of interest may be subjected.

Development of an attribute or parameter uncertainty growth model in the design phase requires a detailed specification of component stabilities, circuit topology, operational parameters (ranges of current, frequency, temperature, vibration, etc.), environmental conditions, usage considerations, and any other data that may impact mechanisms whereby the attribute or parameter may transition from a nominal to an out-of-tolerance state.

Model development begins with a mathematical statement of the stability of each component impacting the in-tolerance probability for the attribute or parameter of interest. In this application, the term "stability" refers to a component's rate of uncertainty growth under specified conditions of stress. Components with low-uncertainty growth rates exhibit high stabilities; components with high-uncertainty growth rates exhibit low stabilities. Component stability models are integrated into board-level stability models.

For complex boards, the stability model may consist of an event tree network integrating the stabilities of individual components into a composite description of the entire board. In cases where boards are relatively simple, the model may be a component-like mathematical model (e.g., the mixed exponential model—see Appendix B) that sufficiently represents the aggregate stability modeling of the constituent components.

Empirical uncertainty growth modeling — This stage involves experimentation with preproduction units in which usage conditions are emulated. Such experimentation has as its objective obtaining sampled time series data on system attributes or parameters (see Appendix B) with which to infer the underlying uncertainty growth processes.

To speed up such a process, *functional* reliability preproduction testing normally employs accelerated life techniques to determine anticipated system reliability under conditions of use. Unfortunately, *measurement* reliability experiments to infer the growth processes of interest do not usually benefit from accelerated life testing.

This is because one of the principal "stresses" encountered during usage or storage is time itself. The "response" to this stress of attributes or parameters with precision tolerances consists of such effects as drift as a result of movement toward chemical equilibrium, changes in value caused by thermal molecular motion, etc. These effects cannot always be accelerated in a natural way through the application of intensified controlled stresses.

Moreover, if the set of models chosen as candidates to model an uncertainty growth process has been carefully arrived at, accelerated life testing may not be needed. It may be that data taken over a span of time that is small relative to the intended period of equipment operation will be sufficient to select the appropriate model and estimate its coefficients.

STEP 5. *Determining reliability models and coefficients.*

As stated earlier, measurement reliability can be modeled using mathematical functions characterized by mathematical forms with appropriate statistical parameters. During the preproduction stage of the design/development phase, experimental time series data are sampled (see Empirical Uncertainty Growth Modeling above). Values of measurement reliability model parameters can be obtained through maximum-likelihood analysis as described in Appendix B. The experimental time series data can be used to select the function that best represents the measurement reliability of the attribute or parameter.

During the design stage, measurement reliability model parameters emerge as natural byproducts of the method of measurement reliability network modeling and simulation. Furthermore, since such models are constructed from design considerations, there is no need to select the best model after the fact. Stability models at the package level and measurement reliability models at the attribute or parameter level of integration are based on the same principles as uncertainty growth models at the component and board levels, but their form may be considerably more complex. Package and attribute/parameter level models tend to be constructed using event tree and fault tree approaches in which "what-if" analyses can be applied.

Development of measurement reliability models from component-, board- and package-level models is an area of current research. Several approaches are suggested by methods used in establishing functional reliability predictions for hardware. The following listing provides an overview of some of the more conventional of these methods:

Similar item method — Extrapolation from the known measurement reliability of existing attributes performing similar functions under similar conditions and employing similar design approaches to those intended for the attribute or parameter of interest.

Similar circuit method — Extrapolation from the known measurement reliability of existing circuit configurations and combinations to reliability predictions for circuit configurations and combinations under consideration.

Active element group method — Formation of gross estimates of attribute or parameter measurement reliability based on the number of series active attribute groups required to perform functions. This method provides a feasibility estimate based on design complexity during concept formulation and preliminary design.

Parts count method — Crude estimation of attribute or parameter measurement reliability based on the number of constituent components. This method is strictly applicable to series configurations only. Nonseries elements are handled as "equivalent" series elements. The parts count method ordinarily assumes that times to out-of-tolerance are exponentially distributed with constant failure rates.

Measurement reliability network modeling and simulation develops reliability predictions by simulating attribute- or parameter-value operating curves bounded by statistical confidence limits. Operating curves are simulated on the basis of engineering expectations in response to time and stress. Statistical confidence limits are simulated from attribute or parameter stability models that are, in turn, based on package, board, and component stability models. The time corresponding to a given operating curve crossing a tolerance limit boundary with a predetermined level of confidence constitutes a prediction of the time to out-of-tolerance for the attribute or parameter.

Functional reliability network modeling is covered extensively in the reliability literature. Although many of the same principles apply, measurement reliability modeling is not covered in any known body of established literature. However, such modeling is often performed by TME designers, as indicated by published equipment specifications found in TME catalogs and user manuals.

STEP 6. *Computing calibration intervals commensurate with appropriate measurement reliability targets.*

This involves setting the modeled measurement reliability function equal to the measurement reliability target and solving for the interval. This also involves estimating a lower confidence limit for the computed interval, as discussed in Appendix B.

STEP 7. Evaluating computed test or calibration intervals for suitability for the intended mission.

This involves comparing the estimated calibration-interval lower-confidence limit obtained in Step 6 with the period of extended usage called out in the mission schedule, i.e., with the mission life requirements for the unsupported TME parameter or end-item attribute under consideration. If the lower confidence limit is longer than the mission life, the equipment design is acceptable. If not, then the attribute or parameter is flagged for further work.

STEP 8. Taking corrective action.

In taking corrective action for an attribute or parameter whose estimated calibration interval lower-confidence limit is less than the attribute's mission life, the following alternatives should be considered:

(1) Incorporation of redundant functions. This involves the inclusion of additional attributes or parameters to back up the problem attribute or parameter. To be effective, redundant attributes should be independent and should be used in parallel. Under these conditions, if the problem attribute or parameter and its backups are equivalent with respect to design, fabrication, and maintenance, then the total uncertainty varies as the square root of the number of redundant parallel attributes.

(2) Monitoring of measurement uncertainty using SMPC methods. Incorporate the methods discussed in Appendix D in ATE, ACE, or end-item controllers.

(3) Reevaluation of end-item performance objectives. If the uncertainty growth of a given parameter or attribute cannot be held to a level commensurate with minimum end-item average utility requirements, and compensating design or other measures fail to correct this deficiency, it may be prudent to review the performance objectives of the end-item to determine if these objectives are realistic within the context of available technology. While this practice is less attractive than solving the problem, it may be the only course available.

Establishment of new performance objectives will require revision of relationships between end-item attribute uncertainty and average end-item utility, TME parameter uncertainty and end-item attribute uncertainty, and TME parameter uncertainty and end-item average utility.

TABLE 6.4 Provision for Extended Deployment

TABLE 6.4
Provision for Extended Deployment

Determine end-item performance requirements in terms of acceptable end-item attribute values.

Determine TME parameter tolerances required to ensure acceptable test process uncertainty.

Determine appropriate measurement reliability targets for end-item attributes or TME parameters.

Ascertain the uncertainty growth processes for the end-item attributes or TME parameters of interest.

Determine reliability models and coefficients.

Compute test or calibration intervals commensurate with appropriate measurement reliability targets.

Evaluate computed test or calibration intervals for suitability for the intended mission.

Take corrective action if necessary

- *Incorporate redundant functions*
- *Incorporate SMPC methods*
- *Reevaluate end-item performance objectives.*

6.4 Statistical Measurement Process Control (SMPC) Methods

6.4.1 Basic Concepts

Periodic recall and calibration of TME and standards is not practical for space-based applications. The usual approach for ensuring system measurement integrity in such applications involves the incorporation of redundant capabilities. In applications where this is not time, weight, space, or cost effective, certain compromises may be considered in on-board system performance objectives. Such compromises would allow widening tolerances to limits that could be expected to contain uncertainty growth over the mission life cycle.

Instances may arise, however, where incorporation of redundant functions is not feasible, and/or where on-board system performance objectives and corresponding accuracy requirements are “cast in concrete” and cannot be relaxed. In such instances, measurement assurance can still be supported through the use of SMPC.

SMPC can be employed to monitor the integrity of on-board system calibration through a “bootstrap” approach. In this approach, on-board TME and calibration standards are used to check one another within the context of predetermined uncertainty growth expectations. The process emulates a closed system round robin, conducted periodically, which updates the prevailing knowledge of on-board equipment accuracies.

In traditional SPC applications, the monitoring of testing or calibrating processes is done by using *process control limits*. Process control limits consist of performance specifications expanded to include measurement process uncertainty contributions. These contributions are arrived at by multiplying measurement process uncertainties by statistical confidence multipliers. The multipliers are determined in accordance with the degree of confidence (e.g., 95%) desired in monitoring the process.

Measured values are plotted against these control limits. The resulting plot is called a "control chart." The occurrence of an "out-of-control" value on a control chart is taken to signify an out-of-control process, possibly an out-of-tolerance measuring device. Since the procedure does not rely on external TME or standards, the use of *statistical measurement process control* (SMPC) offers possibilities for reducing dependence on external calibration in remote environments.

It should be noted that identifying the cause of an out-of-control measurement often requires human judgment and analysis. In such an analysis, control charts are studied to detect trends or anomalies that may shed light on whether the measuring device is measuring accurately, whether problems have arisen due to ancillary equipment, or whether the measured values are correct but simply lie outside expected limits. With its reliance on manual control chart monitoring, traditional SPC is difficult to implement in remote environments. If SPC is to be used in these environments, what is needed are more revealing and less ambiguous measures of measurement integrity than out-of-control occurrences.

Such measures are available through the application of methods that will be collectively referred to in this publication as SMPC. SMPC can be applied in cases where TME or standards are used to monitor other TME or standards. In addition to ordinary ground-based testing and calibration applications, these cases include remote applications in which local monitoring is done in an automated or remotely controlled closed system. Also included are cases where astronomical or terrestrial standards are employed as monitoring references.

With SMPC, as with traditional SPC methods, the results of measurements are used to develop information regarding the accuracy of the monitoring process. With SMPC, this information takes the form of in-tolerance probabilities and bias (error or offset) estimates for measuring attributes. In-tolerance probabilities can be used to indicate instances where monitoring devices should be either taken out of service or derated. Bias estimates can be used as error correction factors to be applied to subsequent measurements.

SMPC is described below. Development of this methodology is detailed in Appendix D.

6.4.2 SMPC Methodology

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

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

To illustrate the method, consider the following question, that arose during a proficiency audit:

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

Of course, it is never possible to say with certainty whether a given instrument or another is in- or out-of-tolerance. Instead, the best we can do is try to evaluate out-of-tolerance or in-tolerance probabilities. The application of the method to the proficiency audit example follows.

The measurement configuration is shown in Figure 6.3 and tabulated in column 1 of Table 6.5. For discussion purposes, let instrument 1 act the role of a unit under test (**UUT**) and label its indicated or "declared" value as Y_0 (the "0" subscript labels the UUT). Likewise, let instruments 2 and 3 function as TME, label their declared values as Y_1 and Y_2, respectively, (the "1" and "2" subscripts label TME1 and TME2) and define the variables

$$X_1 = Y_0 - Y_1 = -6$$

and

$$X_2 = Y_0 - Y_2 = -15$$

These quantities can be used to solve for the UUT (instrument 1) in-tolerance probability estimate.

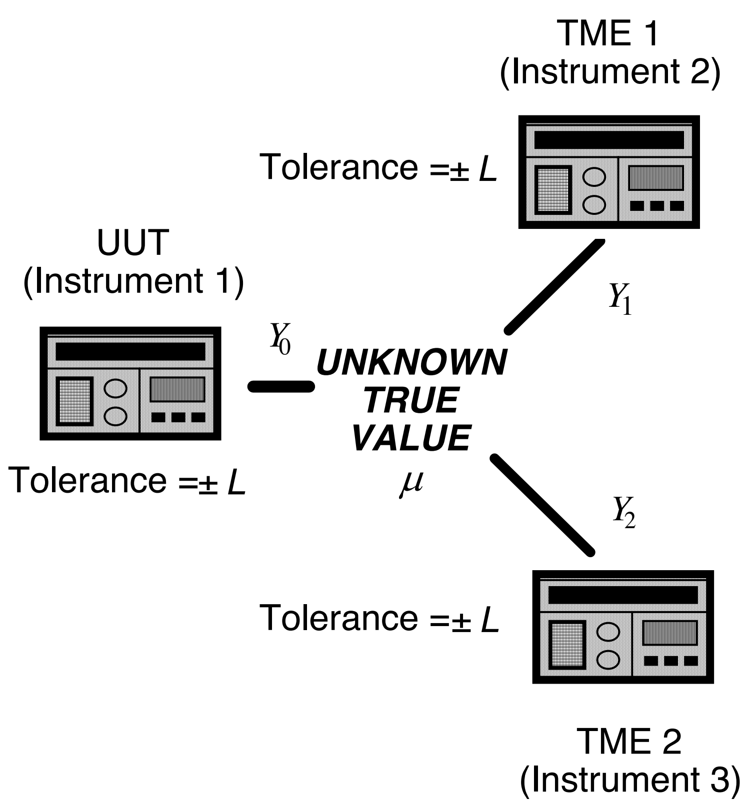


FIGURE 6.3 — PROFICIENCY AUDIT EXAMPLE.

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

TABLE 6.5 Proficiency Audit Results Arranged for SPC Analysis

TABLE 6.5

Proficiency Audit Results Arranged for SPC Analysis

UUT=TME 1	***UUT=TME 2***	***UUT=TME 3***
$L_0 = 10$	$L_0' = 10$	$L_0'' = 10$
$L_1 = 10$	$L_1' = 10$	$L_1'' = 10$
$L_2 = 10$	$L_2' = 10$	$L_2'' = 10$
$Y_0 = 0$	$Y_0' = 6$	$Y_0'' = 15$
$Y_1 = 6$	$Y_1' = 0$	$Y_1'' = 6$
$Y_2 = 15$	$Y_2' = 15$	$Y_2'' = 0$
$X_1 = -6$	$X_1' = 6$	$X_1'' = 9$
$X_2 = -15$	$X_2' = -9$	$X_2'' = 15$

Solving for the In-Tolerance Probability of Instrument 1 — In probability theory, the notation *P(w|x)* is used to denote the probability that an event *w* will occur, given that an event *x* has occurred. For example, *w* may represent the event that a UUT attribute is in-tolerance and *x* may represent the event that we obtained a set of measurements X_1, and X_2, of the attribute's value. In this case, *P(w|x)* is the probability that the UUT attribute is in-tolerance, *given* that we have obtained the measurement results X_1 and X_2.

P(w|x) is a *conditional probability*. We can also form conditional pdfs. For instance, we can form a conditional pdf for a UUT attribute error ε being present given that we have obtained the quantities X_1 and X_2 defined above. We write this pdf $f\left(\varepsilon | X_1, X_2\right) d\varepsilon$. With, $f\left(\varepsilon | X_1, X_2\right) d\varepsilon$ we can estimate an in-tolerance probability for instrument 1 by using it as the pdf in Eq. (6.1).

Following this procedure yields an in-tolerance probability estimate of approximately 77%.

Solving for the In-Tolerance Probabilities of Instruments 2 and 3 — In reviewing the proficiency audit question, it becomes apparent that there is nothing special about instrument 1 that should motivate calling it the UUT. Likewise, there is nothing special about instruments 2 and 3 that should brand them as TME. Alternatively, instrument 2 could have been labeled the UUT and instruments 1 and 3 the TME, as in Figure 6.4 and column 2 of Table 6.5. This rearrangement of labels allows us to calculate the in-tolerance probability for instrument 2 just as we have done for instrument 1. This involves defining the quantities

$$X_1' = Y_0' - Y_1' = +6$$

and

$$X_2^{'} = Y_0^{'} - Y_2^{'} = -9 \, ,$$

and forming the pdf $f\left(\varepsilon \middle| X_1^{'}, X_2^{'}\right) d\varepsilon$. Using this pdf in Eq. (6.1) yields an in-tolerance probability estimate of 99% for instrument 2.

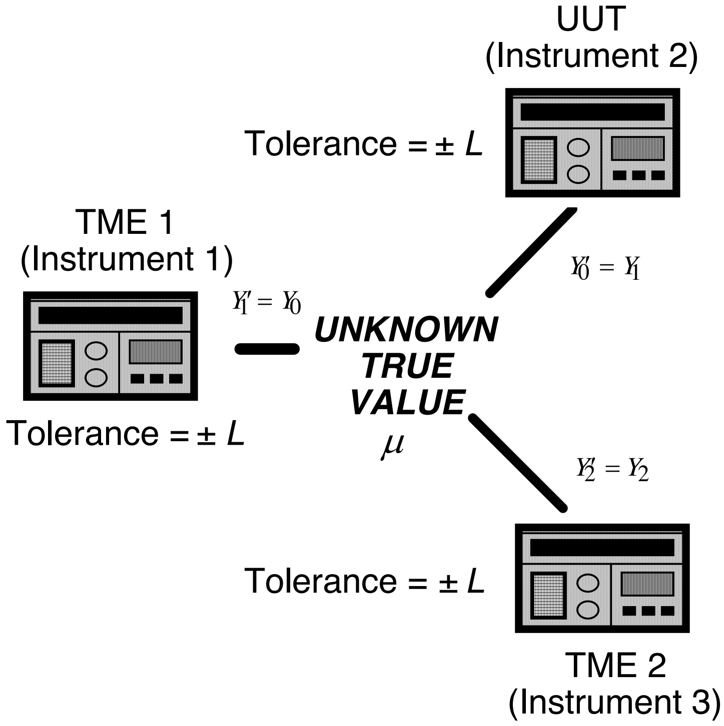


FIGURE 6.4 — EXCHANGING UUT AND TME ROLES.

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

Similarly, if we compute

$$X_1^{"} = Y_0^{"} - Y_1^{"} = +9$$

and

$$X_2^{"} = Y_0^{"} - Y_2^{"} = +15 \, ,$$

construct the pdf, $f\left(\varepsilon \middle| X_1^{"}, X_2^{"}\right) d\varepsilon$ and use this pdf in Eq. (6.1), we get an in-tolerance probability estimate of 69% for instrument 3.

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

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

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

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

Using Eq. (6.2), the biases of instruments 1, 2, and 3 are estimated to be –7, –1, and +8, respectively. As will be discussed later, such bias estimates can be employed as measuring attribute correction factors.

So, as to the proficiency audit question "who's out-of-tolerance?" the answer is that instrument 1 has an estimated bias of –7 and an in-tolerance probability of 77%, instrument 2 has an estimated bias of –1 and an in-tolerance probability of 99%, and instrument 3 has an estimated bias of +8 and an in-tolerance probability of 69%. General-purpose test equipment is usually managed to an end-of-period measurement reliability target of 72%. Accordingly, the results show that instrument 3 should be submitted to a higher level facility for recalibration.

Incidentally, before placing too much stock in the above bias estimates, it is worthwhile to consider that their computed 95% confidence limits are fairly wide:

Instrument 1:	–13.4 to –0.6
Instrument 2:	–7.4 to +5.4
Instrument 3:	+1.6 to +14.4

The wide ranges are due to the wide spread of the measured values and to the fact that all instruments were considered a priori to be of equal accuracy.

Evaluating Attribute In-Tolerance Probabilities — Consider an attribute of an automated TME or standard that monitors or checks n independent subject attributes over a span of time that is short relative to the TME's deployment cycle. This allows us to regard the TME's measuring attribute as fairly stable over the span of time considered. The result of each check is a pair of declared values: the TME attribute declared value and the subject attribute's declared value. Either the pairs of values or their differences are stored and maintained for SMPC analysis.

In the customary view of such checks, the TME is regarded as the automated testing or calibrating system and the subject attributes are regarded as the UUTs. From the SMPC perspective, *any* attribute in the process can be labeled the UUT, with each of the other attributes placed in the role of TME. Thus, the monitoring system's attribute can be considered a UUT and the workload attributes can be imagined to be a set of monitoring TME. Given this scheme, label the monitoring system attribute's declared value as Y_0 and the subject attributes' declared values as Y_i, i= 1, 2, …, n.

In Figures 6.3 and 6.4, UUT and TME comparisons are based on the measurement of an underlying value μ. Ordinarily, monitoring system checks of subject attributes may occur at different times and may involve different values. This is not a problem in applying SMPC methodology to evaluating monitoring-system attributes, however, since the quantities of interest are the *differences* in declared values $X_i \equiv Y_0 - Y_i$, rather than the declared values themselves. These differences do not depend on the precise values pertaining at the time of measurement—only that the same value be measured by both the TME and the UUT.

The combined set of comparisons compiled from test or calibrations of subject attributes yields an in-tolerance probability estimate for the monitoring attribute. This in-tolerance probability estimate can be used in deciding whether to attempt a recalibration of the attribute against an astronomical or terrestrial reference, to derate its accuracy, or discontinue its use.

Computing Attribute Correction Factors — It was shown earlier that using SMPC can provide estimates of the biases of instrument attributes. These estimates can be employed as attribute *error correction factors*.

Suppose, for example, that instrument 1 of the proficiency audit problem is a monitoring system, and instruments 2 and 3 are subject items. Then, following measurements of the attributes of instruments 2 and 3 by the measuring system and application of SMPC, the monitoring system attribute could be assigned a correction factor of β, where β would be calculated using appropriate pdfs as shown in Eq. (6.2). The attribute could be compensated or corrected for "in software" by automatically subtracting the value β from subsequent monitoring-system measurements.

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

By using check standards, not only can the in-tolerance probabilities and biases of the attributes of monitoring systems be more accurately estimated, but in-tolerance probability and bias estimates can also be determined for the check standards. Since check standards are subject to drift and fluctuation, using monitoring systems and associated subject items to check for integrity in this way helps ensure that continuity with the "external universe" is maintained.

Now that we have control data, we can

- Correct for known errors/drifts
- Know when to recalibrate
- Know when the measurement process is out of control—or headed there—and take corrective action.

6.5 Analyzing Measurement Decision Risk

Good measurement system design includes well-defined and documented measurement assurance techniques to verify the adequacy of the measurement process. Conventional procedures for measurement system design, selecting equipment, and interpreting specifications call out nominal ratios of accuracy to be maintained between testing or calibrating systems and units under test or calibration. Use of these nominal ratios while supportable from a measurement assurance standpoint, are not always best from a cost-effectiveness standpoint. Moreover, many instances arise in which nominal ratios cannot be met because of limits in the state of the art. Also, other program control variables are used to avoid setting arbitrary levels, such as in-tolerance percentage targets, to ensure a level of measurement integrity commensurate with program needs.

The following provides guidelines for using new methods that enable rigorous analyses of accuracy ratios, in-tolerance percentage requirements and related parameters. Through use of these methods, test and calibration capabilities can be tailored to meet mission support requirements.

The mathematical procedures and methods that underlie test and calibration optimization are described in Appendix C. This appendix is recommended reading for technical specialists.

6.5.1 Measurement Decision Risk Analysis—General Concepts

All measurement processes are accompanied by measurement error and uncertainty. Since errors and uncertainties can never be eliminated, the potential always exists for making incorrect decisions. Although error and uncertainty cannot be eliminated, they can be limited or controlled to acceptable levels through critical design, testing, and calibration.

Until recently, establishing acceptable levels of error and uncertainty has been a simple process in which nominal standards of high accuracy between verifying and subject units were maintained. Historically, relative accuracies have been such that measurement system uncertainties were required to be ten percent or less of end-item or product tolerances, and calibrating system uncertainties were required to be twenty-five percent or less of the tolerances of subject units. In the marketplace, and in military and aerospace applications, maintenance of these high relative accuracies (or low relative uncertainties) has often proved impossible.

In applications where performance objectives border on state-of-the-art measurement accuracy, the acceptability of the uncertainty ratio between a measurement system and a subject end-item must be evaluated within the context of the application. Also, the acceptability of the uncertainty ratio between a calibrating system and its subject measurement system must be determined within the same context.

Maintaining the accuracy (i.e., controlling the uncertainty) of measurement systems is accomplished through calibration, and maintaining the accuracy of calibrating systems is accomplished through still higher-level calibration. The chain of calibration and test interfaces comprising the foundation of accuracy upon which end-items are tested and evaluated is called measurement traceability. With this in mind, the question “Why calibrate?” in Section 3.3, becomes rephrased as “Why maintain measurement traceability?” The answer to this question is that an accuracy base is needed to ensure that measurement decision risk is acceptable.

Since the accuracy at any given level of the test and calibration hierarchy is affected by the accuracy of supporting levels, the effect of uncertainty at one level on subsequent levels must be accounted for. Moreover, since the primary reason for calibration is the maintenance of an adequate end-item test-and-evaluation accuracy base, accuracy requirements are ultimately determined by end-item performance requirements.

That is, measurement system accuracy requirements are driven by mission performance requirements, calibration system accuracy requirements are driven by measurement system accuracy requirements, calibration standards accuracy requirements are driven by calibration system accuracy requirements, and so on.

To illustrate this concept, let’s examine the process of acceptance or rejection of a manufactured part, a cannonball, for instance, based on its physical measurements.

6.5.2 Measurement Decision Risk Analysis—A Simple Example

The ultimate goal of end-item testing is to ensure that end-items will meet or exceed design objectives. To illustrate how testing and calibration plays a role in attaining this goal, a hypothetical example is considered. In this example, the end-items are taken to be cannonballs and the measurable attribute of interest is the cannonball's diameter. To avoid getting bogged down in extraneous details, the example assumes that the cannonballs will be fired from a frictionless cannon barrel whose bore diameter never varies from precisely 200 millimeters. Moreover, thermal expansion and friction effects are ignored.

For this example, the attribute by which cannonball performance is to be evaluated is the expected range of the idealized cannon. The range of such a cannon is largely governed by the difference between the cross-sectional area of the cannon bore and the cross-sectional area of the cannonballs. This leads to a performance curve that is quadratic with respect to cannonball diameter, as shown in Figure 6.5.

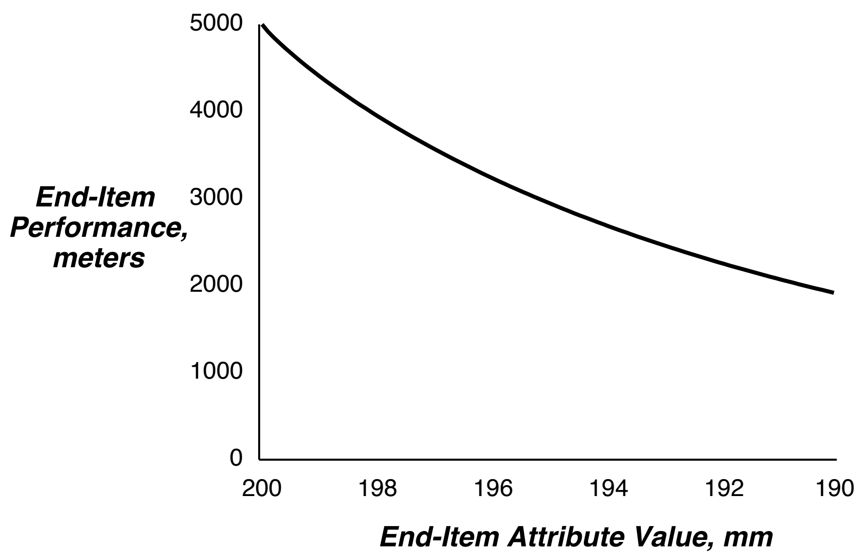


FIGURE 6.5 — END-ITEM PERFORMANCE CURVE.
End-item (cannonball) maximum performance is achieved when the end-item measurable attribute is equal to nominal (200 mm). In the example considered, performance (range) drops off quadratically with deviation from nominal.

The requirements hinge on the "usefulness" or "utility" of the various ranges attainable by the cannonballs. The utility is determined by the cannon's intended application. For example, suppose that the fielded system of interest is intended to achieve cannonball delivery within a specified region not covered by other systems. Cannonballs that fall within this region exhibit maximum utility. Those that fall short of this region exhibit lower utility.

How useful a given end-item (cannonball) will be in a given application is described by its *utility function*. The utility characterizing a given end-item is determined by the extent to which its actual performance matches its performance objectives. For this example, cannonballs that reach or exceed the specified range are characterized by a utility function value of unity. Those that fall short of, but still close to, the specified range are characterized by utility function values less than unity but greater than zero. At some point, the maximum attainable cannonball range becomes "useless." Such a range is assigned a utility function value of zero.

A typical utility function is shown in Figure 6.6 where end-item performance is given in terms of cannonball range. Since range can be directly related to cannonball diameter, according to Figure 6.5, utility can also be specified in terms of end-item attribute value. This is done in Figure 6.7.

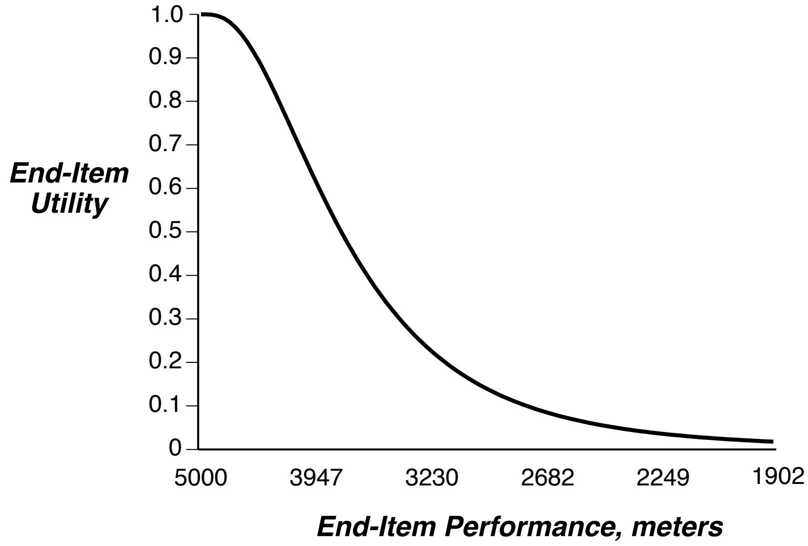


FIGURE 6.6 — END-ITEM UTILITY VERSUS END-ITEM PERFORMANCE.
For the cannonball example, a range of 5000 meters is associated with a utility of 1. As maximum attainable range decreases from 5000 meters, cannonball utility is reduced.

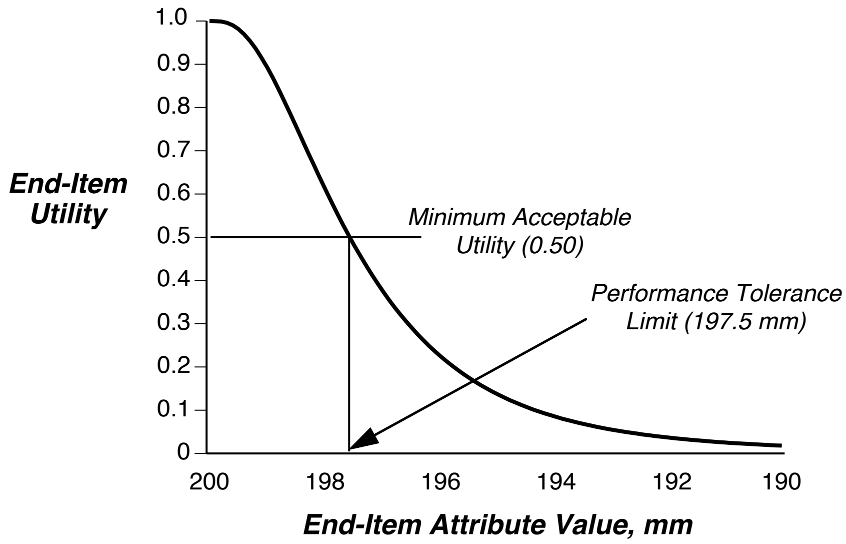


FIGURE 6.7 — END-ITEM UTILITY VERSUS END-ITEM ATTRIBUTE VALUE.
A cannonball diameter (attribute value) of 200 millimeters corresponds to a range (performance) of 5000 meters, which is associated with a utility of 1. As cannonball diameters decrease from 200 millimeters, cannonball utility is reduced.

The relationship between the utility function and end-item attribute values is particularly useful for establishing end-item tolerances. This is done by identifying an attribute value or range of values associated with a minimum acceptable end-item utility. Minimum acceptable end-item utility is determined from mission considerations. For example, suppose that the scope of missions intended for our hypothetical cannon requires that utility shall not fall below 0.50. As Figure 6.7 shows, a utility of 0.50 corresponds to a cannonball diameter of 197.5 millimeters.

Hence, cannonballs with diameters less than 197.5 millimeters are considered to be out-of-tolerance. Also, since the cannon bore diameter is 200 millimeters, cannonballs whose diameters exceed 200 millimeters will not fit in the cannon barrel. This is equivalent to saying that the utility function is equal to zero if the attribute value is greater than 200 millimeters. Accordingly, the cannonball performance tolerance specifications are given as

Upper performance tolerance limit: 200.0 mm

Lower performance tolerance limit: 197.5 mm

Cannonballs produced by a given manufacturer are not all issued with the same diameter. Because of the vagaries of manufacturing, storage, shipping, and handling, cannonballs are produced with diameters that vary relative to their design target value, according to some definable probability distribution. The closeness of the agreement between actual cannonball diameters and the design target value is measured in terms of the spread of this distribution.

Some cannonballs will be larger than the design value and some will be smaller. For purposes of illustration, assume that the production of cannonballs larger than the design value and smaller than the design value are equally likely outcomes. To avoid producing many cannonballs that will be too large to fit in the cannon barrel, the design target would probably be set at some value less than 200 millimeters.

Exactly where to set the design value is an involved process that tries to achieve a viable balance between false-reject risk (the probability that in-tolerance cannonballs will be rejected by testing) and false-accept risk (the probability that out-of-tolerance cannonballs will be accepted by testing.) False reject risk results in unnecessary rework costs suffered by the manufacturer and false-accept risk results in out-of-tolerance products being delivered to customers. Studies have shown that solving the problem involves the analysis of alternative approaches, policies, and input parameters for each specific problem of interest. A methodology is presented in Appendix C.

A useful statistic for evaluating the population of cannonballs delivered by a given manufacturer is the population's *average* utility. Since the utility of an end-item depends on its attribute value, the average utility of a population of end-items depends on the distribution of these values. Thus, a population whose distribution is closely bunched around the end-item design value will have a greater average utility than a population whose distribution is widely spread. Figure 6.8 illustrates this idea. In the figure, the population spread is shown in terms of the population standard deviation, or equivalently, the population *uncertainty*.

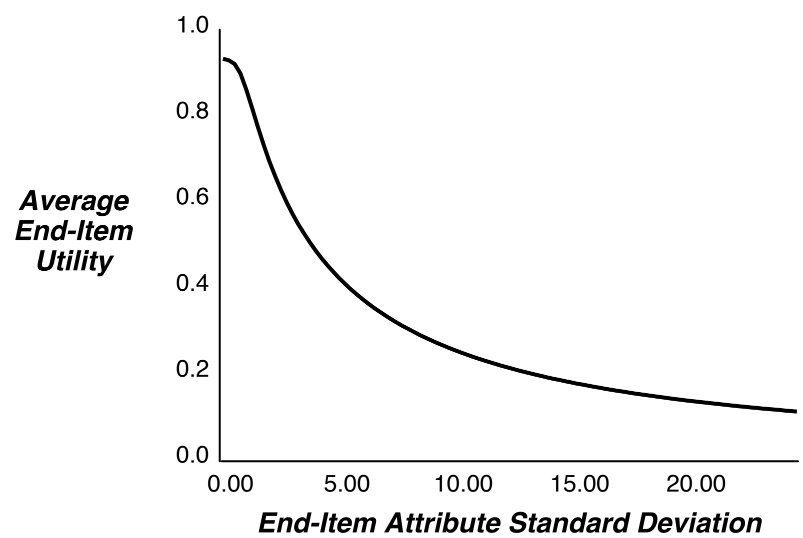


FIGURE 6.8 — AVERAGE END-ITEM UTILITY.

The average utility of a population of end-items delivered by a given manufacturer is related to the spread of the population attribute values. This spread is quantified in terms of population uncertainty or *standard deviation*. As the figure shows, a higher standard deviation corresponds to a lower average end-item utility.

To ensure average end-item utility is at an acceptable level, end-item populations are tested before delivery. Testing is performed using TME to determine whether end-item attribute values correspond to acceptable performance. End-items that "pass" testing are shipped with values spread relative to their design values. The degree of spread reflects the efficacy or accuracy of the testing process.

Because of unavoidable measurement uncertainties in this process, some percentage of delivered end-items will ordinarily be out-of-tolerance. The relationship between end-item population spread and the percentage of end-items out-of-tolerance can be inferred from Figure 6.9. Generally, the greater the spread of the distribution, the higher the out-of-tolerance percentage. As stated earlier, this spread is described by the end-item attribute population standard deviation. For example, Figure 6.9 shows an end-item (cannonball) attribute-value probability distribution characterized by an attribute value standard deviation of 1.0 millimeter.

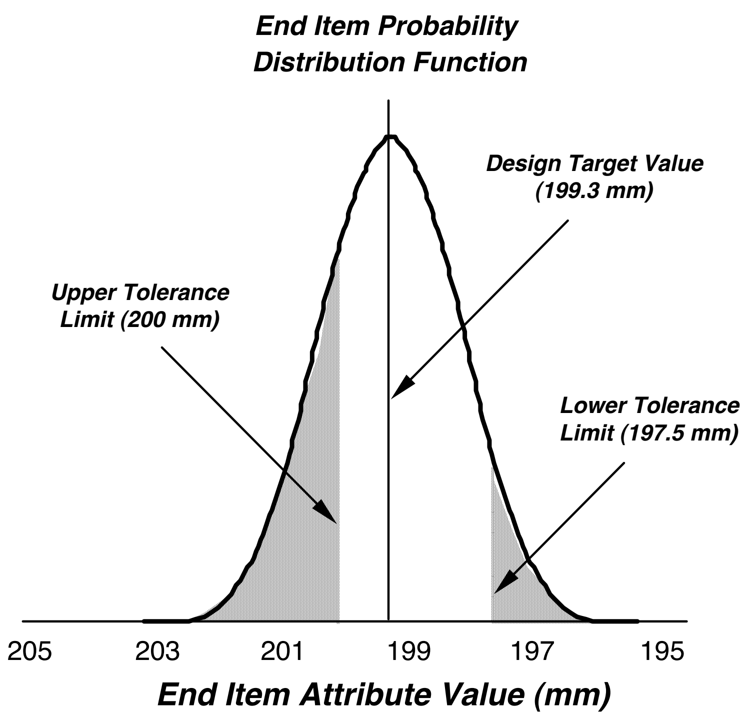


FIGURE 6.9 — END-ITEM PROBABILITY DENSITY FUNCTION.
The probability density function for a cannonball population whose design value is 199.3 millimeters and whose standard deviation is 1.0 millimeters. The height of the curve indicates the probability that a given attribute value will be found in the population. The shaded area represents the fraction of cannonballs made with diameters outside the tolerance limits.

How effective testing is in screening out-of-tolerance end-items depends on the measurement uncertainty that characterizes the test process. A test process characterized by extremely low uncertainty will do a better job of screening out-of-tolerance end-items than will a process characterized by a high uncertainty. This is shown in Figure 6.10. Higher end-item population out-of-tolerance percentages are associated with higher end-item population uncertainties. The more out-of-tolerance end-items that slip through the testing process, the higher will be the uncertainty in the attribute values of items delivered to customers. The logical conclusion is that greater test process uncertainty leads to higher end-item attribute uncertainty.

Since test process uncertainty affects the distribution of end-item attributes, and the distribution of end-item attributes affects average end-item utility, then test process uncertainty affects end-item utility.

> Since test process uncertainty is controlled through calibration, the ultimate benefit of calibration is the assurance of end-item utility.

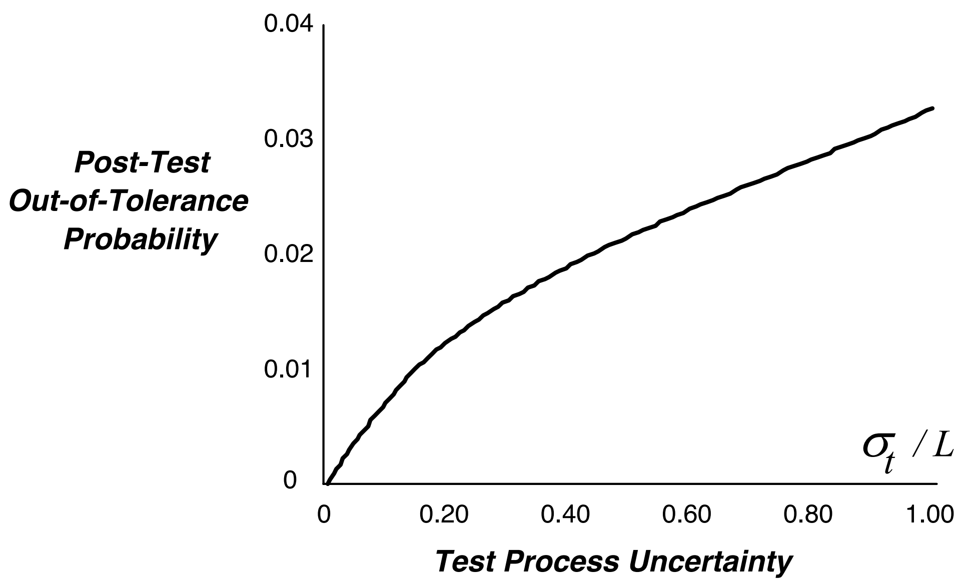


FIGURE 6.10 — END-ITEM UNCERTAINTY VERSUS TEST PROCESS UNCERTAINTY.
The out-of-tolerance probability for a population of end-item attributes that have been screened by testing is governed in part by the uncertainty of the test process. Test process uncertainty is expressed in terms of the ratio of the standard deviation of the test process (σ_t) to the end-item tolerance (L). The figure applies to a pretest population out-of-tolerance probability of 5%.

6.5.3 Measurement Decision Risk Analysis—Methodology

Current methodologies for evaluating measurement decision risks examine these risks in the context of test and calibration infrastructures. This enables the building of integrated models that consider the propagation of uncertainties throughout the infrastructure.

6.5.3.1 The Test and Calibration Support Hierarchy

Test and calibration infrastructures are manifested in test and calibration support hierarchies. These hierarchies consist of support levels whose uncertainties decrease from level to level, from end-items down through to primary reference standards. Figure 6.11 represents a generic test and calibration support hierarchy. As Figure 6.11 shows, each level is separated by an interface through which support requirements are communicated to lower levels and measurement decisions are communicated to higher levels.

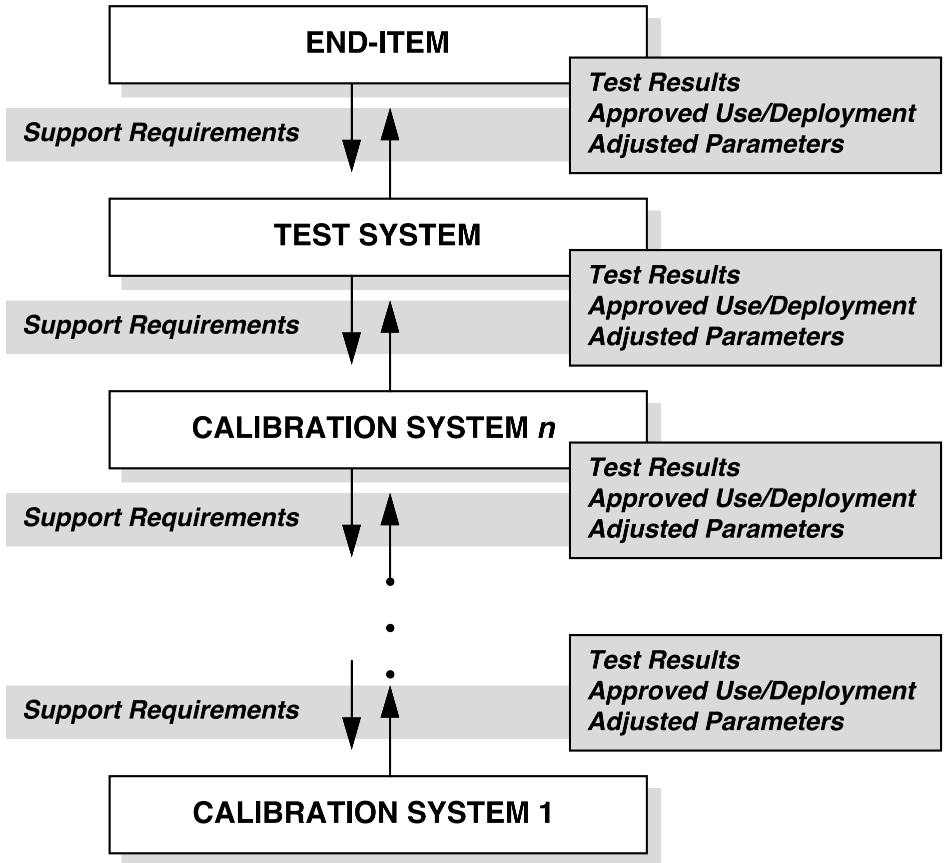


FIGURE 6.11 — THE TEST AND CALIBRATION HIERARCHY.

The hierarchy shows the flow of support requirements from the end-item level to the primary calibration support level. Immediate end-item support requirements are in terms of the measurement process uncertainty that can be tolerated during testing. As can be inferred from Figures 6.8 and 6.10, the utility of an end-item population is affected by this uncertainty. This uncertainty is in turn affected by the measurement process uncertainty accompanying test system calibration. Also, measurement process uncertainty at each calibration level in the hierarchy is affected by measurement process uncertainty at other levels. Because of this, measurement process uncertainties propagate vertically through the hierarchy to affect end-item quality.

6.5.3.2 The Measurement Assurance Cycle

Figure 6.12 represents the overall measurement assurance cycle transacted across each hierarchy interface. The sequence depicted applies to cases where units under test (**UUT**s) are shipped for test or calibration from one hierarchy level to another. In cases where tests or calibrations are done on site, shipping stresses are not a factor (although some similar stress may be induced by routine handling and maintenance).

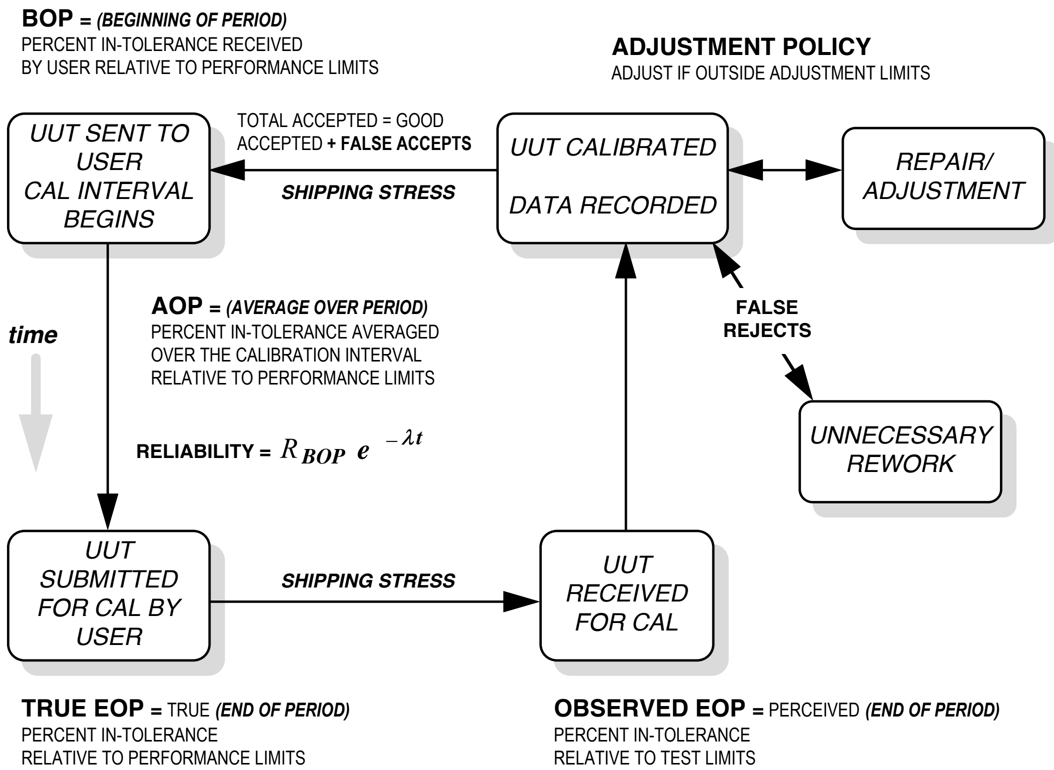


FIGURE 6.12 — THE TEST AND CALIBRATION CYCLE.

The sequence for schemes in which UUTs are submitted for test or calibration across hierarchy interfaces. In the case shown, measurement reliability is modeled using the exponential model (see Appendix B).

The test or calibration interval begins when the UUT is received for use from the supporting organization. The UUT's measurement reliability here is labeled R_{BOP}. Due to measurement process uncertainties and shipping stresses, R_{BOP} is nearly always less than 1.0, contrary to popular belief. The quantity R_{BOP} provides the principal measure of the support quality supplied by the testing or calibrating organization. This quality can usually be influenced by the maintenance or adjustment practice adhered to by this organization. After the interval, the UUT is submitted for retest or recalibration.

Here, its measurement reliability is labeled R_{EOP}. The variable R_{EOP} shows the lowest measurement reliability experienced over the test or calibration interval. Over the duration of the interval, the UUT exhibits an average measurement reliability, labeled R_{AOP}. This average is the technical parameter against which the UUT's utility is measured during use.

Because of testing and calibration measurement decision uncertainties, some in-tolerance UUT attributes will be observed as out-of-tolerance (false rejects) and some out-of-tolerance attributes will be observed as being in-tolerance (false accepts). False rejects lead to unnecessary rework and a lowered perception of R_{EOP}. False accepts raise the risk of using out-of-tolerance parameters during testing or calibration cycle. False accepts lower R_{BOP} and lead to an elevated perception of R_{EOP}.

6.5.3.3 Test Decision Risk Analysis

The measurement decision risks that accompany all measurement processes is represented in Figure 6.13. Before testing or calibration, the subject UUT population is characterized by some percentage of attributes that are out-of-tolerance. Some of these are detected during test or calibration and rejected. Because of measurement process uncertainties, however, some slip through (false accepts). Likewise, because of measurement process uncertainties, some in-tolerance attributes are perceived as out-of-tolerance and are rejected (false rejects).

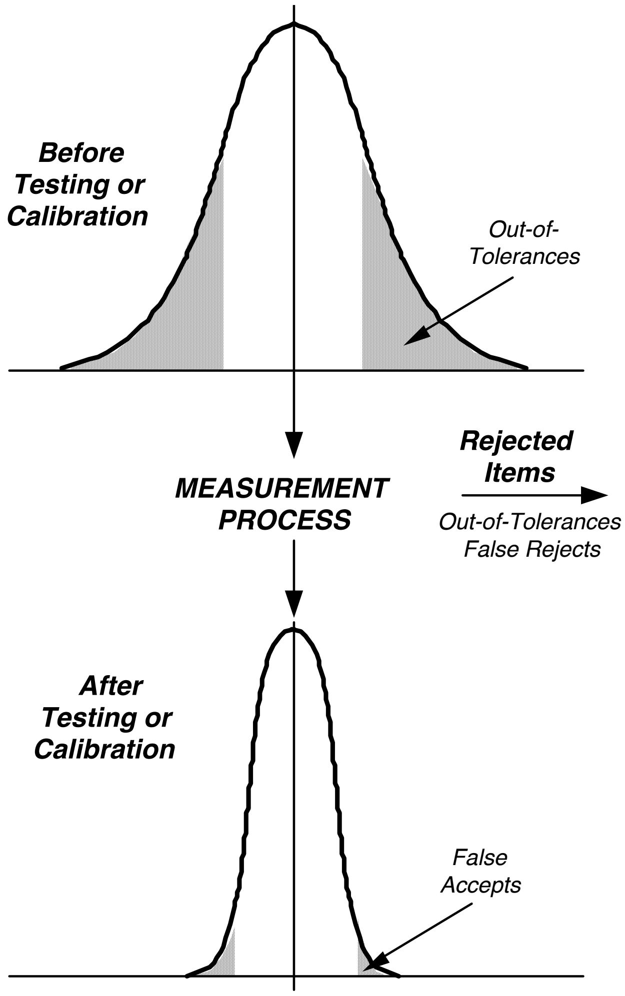


FIGURE 6.13 — MEASUREMENT PROCESS RESULTS.
Each test or calibration process accepts a portion of nonconforming items and rejects a portion of conforming ones. The greater the measurement process uncertainty, the greater the risk of making such erroneous decisions.

Measurement process uncertainty is described in terms of several test and calibration support elements that characterize each test and calibration hierarchy interface. These elements are listed in Table 6.6 and depicted in Figure 6.14. In Figure 6.14, to the left of the UUT SERVICED function, are those elements that comprise the UUT acceptance criteria and maintenance policy. To the right of this function, are those elements that govern measurement decision risk. Both sets of elements interact. For example, if UUT test limits are narrow relative to TME performance limits, a significant number of false-reject and false-accept decisions may be made. This would also be the case if TME AOP measurement reliability were low and/or if measurement process uncertainties were substantial. The relationship of each variable to other variables is described in detail in Appendix C.

TABLE 6.6 Measurement Decision Risk Elements

TABLE 6.6
Measurement Decision Risk Elements

Risk Element	***Description***
Accuracy Ratio	*Ratio of the UUT performance tolerance limit to the TME performance limit uncertainty*
BOP Reliability	*Measurement reliability of an attribute as received by the user at the beginning of the test or calibration interval*
EOP Reliability	*Measurement reliability of an attribute at the end of the usage period*
AOP Reliability	*Measurement reliability averaged over the usage period from BOP to EOP*
Performance Limit	*Limit which bounds attribute values corresponding to acceptable performance*
Test Limit	*Limit which defines test or calibration acceptance criteria for a UUT attribute*
Tolerance Limit	*Tolerance limit outside which an attribute is considered to require adjustment*
Renewal Policy	*Policy controlling adjustment of tested or calibrated attributes*

It is important to remember that for many of these elements, there are two sets of values: *true* values and *perceived* values. For example, the true EOP measurement reliability of an attribute is an important variable in estimating measurement decision risk. The observed or perceived EOP measurement reliability is an important variable in adjusting test or calibration intervals (see Section 6). The mathematical relationships between true and perceived values are given in Appendix C.

The elements that influence what happens between BOP and EOP are shown in Figure 6.15. An attribute's EOP measurement reliability is affected by its BOP value and by its uncertainty growth over the test or calibration interval. This growth depends on the inherent stability of the attribute, the conditions of its use and environment, and the duration of its test or calibration interval.

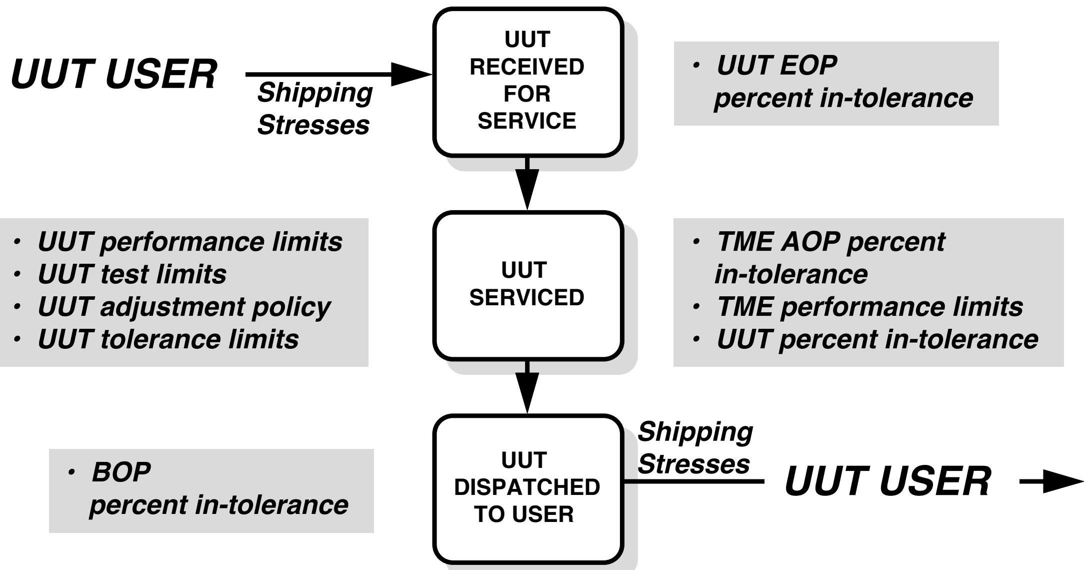


FIGURE 6.14 — ELEMENTS OF THE TEST AND CALIBRATION PROCESS.

Elements contributing to measurement process uncertainty are listed to the right of the UUT SERVICED function. Elements governing measurement decisions and maintenance actions are listed to the left. UUTs are received for service with an unknown EOP measurement reliability. The lower the measurement process uncertainty, the closer the perceived EOP measurement reliability is to the actual or true EOP value. UUTs are returned to users with an unknown BOP measurement reliability. The lower the measurement process uncertainty, the higher the BOP value.

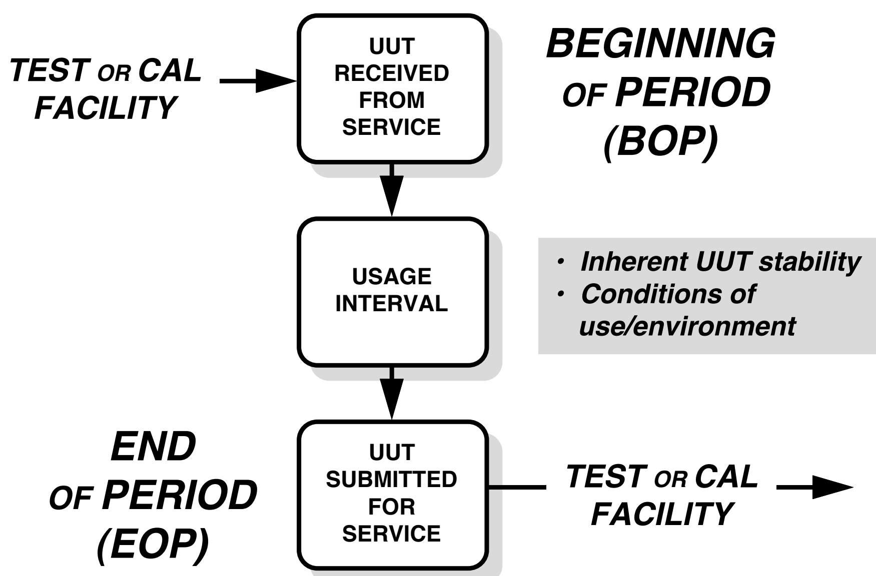


FIGURE 6.15 — ELEMENTS OF THE UUT USAGE PERIOD.

The measurement reliability of each UUT attribute decreases from it's BOP value to its EOP value over the duration of the UUT's test or calibration interval (usage interval). The elements contributing to the difference between BOP and EOP reliabilities are the inherent stability of the attribute, the conditions of the attribute's use, its usage environment, and the duration of the usage interval.

6.6 Managing Measurement Decision Risk

The management of measurement decision risks is obviously an important element of modern quality control. This is particularly true in the development and production of systems working at the forefront of technology. Measurement decision risk management is a specialized discipline that, like other high-technology fields of endeavor, is undergoing development and refinement.

6.6.1 Management of Technical Parameters

The technical parameters important to the management of measurement decision risk are shown in Table 6.6. Of these, the key elements are *performance limit* and *AOP reliability*. Performance limits provide an indication of the range of attribute values expected in using a test or calibration system. AOP reliability provides an indication of the average probability that attribute values will be within these limits, i.e., it provides a measure of potential attribute bias or error. Generally, the higher the AOP reliability, the lower the measurement decision risk (other things being equal).

Since calibration is done at the end points of equipment usage intervals instead of during use, AOP measurement reliability is never observed directly. Because AOP reliability is governed by BOP reliability (the in-tolerance probability at the beginning of the usage period) and EOP reliability (the in-tolerance probability after the period), attaining a given level of AOP reliability calls for managing EOP and BOP reliabilities.

The BOP reliability of an attribute at one level of the test and calibration hierarchy (see Figure 6.11) is largely determined by the AOP reliability of the supporting attribute or attributes at the next lowest level. In Figure 6.11, BOP reliability requirements are shown as "support requirements." From Figure 6.11, it can be seen BOP reliability requirements propagate from end-item testing down through to primary standards calibration.

EOP reliability is managed by establishing *measurement reliability targets* and setting *test and calibration intervals* so observed in-tolerance percentages at EOP are equated to these targets. Because of uncertainties in the measurement process, observed EOP reliabilities are seldom equal to the actual or "true" EOP reliabilities. A major element of measurement process uncertainty is the bias or error of the calibrating or testing attribute. Since this error is measured in terms of AOP reliability, controlling the bias of a testing or calibrating attribute is equivalent to controlling the attribute's AOP reliability. The closeness of the agreement between observed and true EOP reliabilities at one level of the hierarchy is governed largely by the AOP reliability of the supporting attribute(s) at the next lowest level.

The way in which high AOP reliability at a supporting level of the test and calibration hierarchy promotes high BOP reliability in its subject workload is by controlling the incidence of *false accepts*. The way in which high supporting AOP reliability ensures that observed EOP levels are close to true EOP levels is by controlling both false accepts *and* false rejects. The latter risk usually dominates at EOP. False rejects are costly since they lead to unnecessary maintenance, adjustment, repair, and retest or recalibration.

False rejects also affect operating costs in another way. Usually, observed EOP levels are normally lower than true levels. This means that test or calibration intervals (keyed to observed in-tolerance percentages) are usually shorter than they need to be to maintain true EOP reliabilities equal to EOP targets.

6.6.2 Applicability and Responsibility

Management of measurement decision risk is applicable in all instances where end-items are supported by test and calibration infrastructures. As can be appreciated from the previous section and from Appendix C, technical and administrative data must be supplied by each level of the support hierarchy. At the end-item level, special requirements exist for providing descriptions of end-item performance (utility) in terms of attribute or parameter values, and for providing estimates of the cost consequences of system failure.

6.6.3 Benefits

The benefits to be enjoyed through measurement decision risk management include lower operating costs and lower costs associated with substandard end-item performance. Operating costs include costs of calibration, testing, unnecessary maintenance, and downtime. Costs associated with substandard end-item performance include warranty or other liability expenses, loss of future contract work, loss of corporate reputation, and/or loss of material hardware.

6.6.4 Investment

The benefits from effective measurement decision risk management can be considerable. Gaining these benefits can, however, call for substantial investments. These include investments in management energy, data recording and storage, data-processing capability, and personnel.

The first, and often the most critical, investment involves making a management commitment to bring measurement decision risks under control. This involves both grappling with unfamiliar technical concepts and focusing on measurement integrity (quality) as a major quantifiable cost driver. However, once it is realized that unless the type of analysis exemplified in Section 4.10.2 can be routinely performed, end-items will be let out the door with unknown levels of utility that may or may not be acceptable, and test/calibration support costs will persist as operating expenses with unknown return on investment.

Data requirements for measurement decision risk management may be substantial. Necessary data elements include several quantities that can only be tracked by maintaining test and calibration recall systems and by comprehensive reporting of technical data. Although exercising the methodology of Appendix C involves a staggering number of processing loops and complex mathematical operations, the processing capability of current workstation platforms is usually more than equal to the task.

Until general measurement decision risk management packages become available, much of the methodology of Appendix C is currently accessible only to highly trained technical experts. Such personnel should be conversant with probability and statistics, well-schooled in engineering concepts, and comfortable with cost management principles. This level of expertise is necessary because analysis situations tend to involve individual considerations impossible to fit with simple analytical recipes or algorithms.

6.6.5 Return on Investment

Until the measurement decision risk management investment is made, there is really no way to quantify in precise economic terms what the return will be. Until support and acceptance costs become optimized through the application of measurement decision risk analysis principles, the cost savings associated with optimization cannot be balanced against the corresponding investments. This observation notwithstanding, it can be asserted with some confidence that future needs for measurement decision risk management will exceed those of the present day. As the costs of technology development and maintenance continue to spiral upward and performance criteria continue to become increasingly stringent, it may be assumed that the need for effective measurement decision risk management will become an accepted fact of twenty-first century life.

6.7 Optimizing the Hierarchy—Cost Modeling

Through management of measurement process uncertainties, measurement decision risks are held to acceptable levels. An "acceptable" level is determined through cost/benefit analysis of operational support costs versus measurement decision risk consequences. For example, the economic implications of false rejects are manifested through unnecessary rework and/or retest or recalibration. The cost of a false reject is easily expressed in terms of the costs arising from unnecessary effort and the costs associated with equipment downtime.

The economics involved in managing false accepts are more subtle. An analysis of economic trade-offs involved in false-accept management requires a methodology that provides a direct linkage between the accuracy or "quality" of a given test and calibration infrastructure and the utility function of the supported end-item. The methodology is described in Appendix C.

The procedure to be followed in specifying accuracy and associated support requirements for a NASA application was illustrated by example in Section 3.2.7. This example will now be re-considered to show how the various measurement decision risk elements interrelate and how support costs can be balanced against end-item utility requirements.

6.8 Example—The Solar Experiment

In the example of Section 3.2.7, an end-item attribute is to be supported in accordance with nominal NHB 5300.4(1B) requirements. These requirements mandate that test process uncertainty shall not exceed ten percent of the tolerance limit of the end-item attribute and that calibration process uncertainty shall not exceed twenty-five percent of the tolerance limit of the test system. The end-item attribute is a UV detector designed to measure ultraviolet radiation intensity from 1 to 100 mW in the 120 to 400 nanometer range. The measuring system is to be placed in orbit to achieve accurate readings of solar irradiance over a continuous (24-hour per day) operational cycle.

In solar irradiance measurements, the accuracy attainable using ground-based systems is stated as ±30% of reading. Consequently, the utility of the orbiting system is considered zero if the uncertainty in its measurements is ±30% of the reading or more. To justify the expense and effort involved in placing the system in orbit, it has been determined that the maximum error that can be tolerated is ±10% of the reading. Therefore, the end-item performance limit is set at ±10%.

The UV detector is only one component of the orbiting system. This means that measurement reliability objectives for the UV detector attribute must be higher than those of the combined payload system if mission objectives shall be met. A payload measurement reliability objective of 3σ or 99.73 percent probability of in-tolerance performance was specified in Section 3.2.7. It was determined that to meet this objective, each system component would be required to maintain a minimum measurement reliability of 4σ, or 99.994 percent.

In accordance with NHB 5300.4(1B), tolerance limits of ±1% and ±0.25% were specified for test system and calibration system attributes, respectively. However, neither test system nor calibration system measurement reliability requirements are called out in the NHB. As a first pass, it was decided that a 3σ (99.73%) level should be targeted for the test system and a 2σ level targeted for the calibration system. These and other specifications are summarized in Table 6.7, Solar Experiment Specifications; Table 6.8, Solar Experiment End-Item (Prime System) Information;

Table 6.9, Solar Experiment Test System Information; and Table 6.10, Solar Experiment Cal System Information.

TABLE 6.7 Solar Experiment - Specifications

TABLE 6.7
Solar Experiment — Specifications

Parameter	*Tolerance, Percent of Reading*	*EOP Measurement Reliability Target, Percent In-Tolerance*
UV Radiation Detector	*± 10*	*99.994*
Deuterium Lamp	*± 1*	*99.73*
Deuterium Lamp/ Comparator	*± 0.25*	*95.45*

Several cost and technical parameters are needed to do a cost/benefit analysis. Of the cost variables shown, the parameter "cost of prime system failure" is the cost of the failure of that part of the mission associated with the Solar Experiment package. The variable "probability of encounter" refers to the probability the package will be used to make a measurement. Note that no information is shown on uncertainties arising from random or environmental effects or resulting from human error during test and calibration.

Note also the parameters "point at which equipment begins to degrade" (x_d from Table C.1) and "point at which complete failure occurs" (x_f from Table C.1.) The parameter x_d marks the point where the utility of the end-item attribute begins to drop from unity and the parameter x_f marks the point where it reaches zero. These variables are used to mathematically describe the utility of the end-item attribute in terms of the attribute's value.

TABLE 6.8 Solar Experiment – Prime System Information

TABLE 6.8
Solar Experiment — Prime System Information

Name	*Solar Experiment*
Parameter	*UV Radiation*
Qualifier	*1 to 100 mW*
	120 to 400 nm
Adjustment Policy	*Renew/Fail*
Reliability Model	*Exponential*
Test Interval	*12 months*
Observed EOP Reliability	*99.994%*
Test Point	*100.0000 Units: mW*
Performance Limit	*0.0000 + 10.0000% of Reading*
Test Limit	*1.0000 * Performance Limit*
Adjustment Limit	*1.0000 * Performance Limit*
Repair Limit	*3.0000 * Performance Limit*
Repair System	*Equivalent to Test System Accuracies*
Test System Performance Limit	*0.0000 + 1.0000% of Reading mW*
Test System AOP Reliability	*99.91%*

Cost of Prime System failure	*$35,000,000*
Quantity of Prime Systems	*1*
Acquisition cost of one Prime System (parameter)	*$250,000*
Spares coverage desired	*100.00%*
Point at which equipment begins to degrade	*10.0000 mW*
Point at which complete failure occurs	*30.0000 mW*
Probability of encounter	*100.00%*
Probability of successful response	*100.00%*
Labor-hours to test	*2*
Downtime to test	*365 days*
Cost per labor-hours for test or adjustment	*$10,000*
Labor-hours to adjust if needed	*16*
Additional downtime to adjust	*3 days*
Cost to Repair	*$50,000*
Additional downtime to repair	*0 days*

TABLE 6.9 Solar Experiment – Test System Information

TABLE 6.9
Solar Experiment — Test System Information

Name	*Deuterium Lamp*
Parameter	*UV Radiation*
Qualifier	*1 to 100 mW* *120 to 400 nm*
Adjustment Policy	*Renew/Fail*
Reliability Model	*Exponential*
Test Interval	*4 months*
Observed EOP Reliability	*99.730%*
Test Point	*100.0000 Units: mW*
Performance Limit	*0.0000 + 1.0000% of Reading*
Test Limit	*1.0000 * Performance Limit*
Adjustment Limit	*1.0000 * Performance Limit*
Repair Limit	*1.0000 * Performance Limit*
Repair System	*Equivalent to Cal System Accuracies*
Cal System Performance Limit	*0.0000 + 0.2500% of Reading mW*
Cal System AOP Reliability	*97.33%*
Quantity of Test Systems	*3*
Acquisition cost of one Test System (parameter)	*$75,000*
Spares coverage desired	*100%*
Labor-hours to calibrate	*8*
Downtime to calibrate	*2 days*
Cost per labor-hour for calibration or adjustment	*$50*
Labor-hours to adjust if needed	*0*
Additional downtime to adjust	*0 days*
Cost to Repair	*$7,500*
Additional downtime to repair	*30 days*

Accurate testing will lower the probability that degraded or useless performance will be experienced.

The product of this probability and the cost of useless performance is the “acceptance cost.” A high acceptance cost is associated with poor test and calibration support. Conversely, a low acceptance cost indicates that end-items are being placed in service with high in-tolerance probabilities.

TABLE 6.10 Solar Experiment – Cal System Information

TABLE 6.10
Solar Experiment — Cal System Information

Name	*Deuterium Lamp/Comparator*
Parameter	*UV Radiation*
Qualifier	*1 to 100 mW*
	120 to 400 nm
Adjustment Policy	*Renew/Fail*
Reliability Model	*Exponential*
Test Interval	*6 months*
Observed EOP Reliability	*95.00%*
Test Point	*100.0000 Units: mW*
Performance Limit	*0.0000 + 0.2500% of Reading*
Test Limit	*1.0000 * Performance Limit*
Adjustment Limit	*1.0000 * Performance Limit*
Repair Limit	*1.0000 * Performance Limit*
Cal Standard Performance Limit	*0.0000 + 0.0625% of Reading, mW*
Cal Standard AOP Reliability	*99.86%*

Quantity of Cal Systems	*2*
Acquisition cost of one Cal System (parameter)	*$90,000*
Spares coverage desired	*100 %*
Labor-hours to calibrate	*16*
Downtime to calibrate	*4 days*
Cost per labor-hour for calibration or adjustment	*$50*
Labor-hours to adjust if needed	*0*
Additional downtime to adjust	*0 days*
Cost to Repair	*$9,000*
Additional downtime to repair	*30 days*

The results are shown in Tables 6.11 and 6.12. Table 6.11 shows the technical consequences of the proposed test and calibration support and Table 6.12 shows the cost consequences. Note that False Accept and False Reject rates are at 0.00% for the end-item, and end-item AOP is held at 100.00% over the usage period. Note also the low risk figures for the test and calibration systems as well. Because of the high accuracy ratios (4:1) and high EOP reliability targets, random, environmental, and human factors measurement uncertainties were defined to be zero for the example.

TABLE 6.11 Solar Experiment – Test & Cal Analysis Results

TABLE 6.11
Solar Experiment — Test & Cal Analysis Results

PRIME SYSTEM	
Adjustment Policy	*Renew/Fail*
Reliability Model	*Exponential*
Test Interval	*12.0*
Observed EOP In-Tolerance	*99.99%*
True EOP In-Tolerance	*99.99%*
True AOP In-Tolerance	*100.00%*
BOP In-Tolerance	*100.00%*
Performance Limit	*10.0000*
Test Limit	*10.0000*
Adjustment Limit	*10.0000*
False Accept Rate	*0.00%*
False Reject Rate	*0.00%*
Prime/Test Accuracy Ratio	*10.0:1*
TEST SYSTEM	
Adjustment Policy	*Renew/Fail*
Reliability Model	*Exponential*
Calibration Interval	*4.0*
Observed EOP In-Tolerance	*99.73%*
True EOP In-Tolerance	*99.86%*
True AOP In-Tolerance	*99.91%*
BOP In-Tolerance	*99.96%*
Performance Limit	*1.0000*
Test Limit	*1.0000*
Adjustment Limit	*1.0000*
False Accept Rate	*0.04%*
False Reject Rate	*0.17%*
Test/Cal Accuracy Ratio	*4.0:1*
CAL SYSTEM	
Adjustment Policy	*Renew/Fail*
Reliability Model	*Exponential*
Calibration Interval	*6.0*
Observed EOP In-Tolerance	*95.00%*
True EOP In-Tolerance	*95.27%*
True AOP In-Tolerance	*97.33%*
BOP In-Tolerance	*99.43%*
Performance Limit	*0.2500*
Test Limit	*0.2500*
Adjustment Limit	*0.2500*
False Accept Rate	*0.57%*
False Reject Rate	*0.84%*

TABLE 6.12 Solar Experiment – Cost Analysis Results

TABLE 6.12
Solar Experiment — Cost Analysis Results

SUMMARY COSTS ($)

Annual Test and Cal Cost	*46,945*
Annual Adjustment Cost	*60*
Annual Repair Cost	*2,040*
Annual Support Cost	*49,046*
Annual Acceptance Cost	*3*
Total Annual Cost	*49,049*
Spares Acquisition Cost	*259,145*

From Table 6.12, the total annual cost associated with test and calibration support of the end-item comes to $49,049. The total acceptance cost is just $3/year. In most applications, a $35,000,000 cost of end-item failure would yield a high acceptance cost; that is, the probability of accepting nonconforming items during testing is usually high enough to yield an appreciable risk of system failure.

The ludicrously low $3/year cost results from the extraordinarily high accuracy ratio (10:1) and measurement reliability target (99.994% EOP) chosen for the end-item. The question arises, What happens if these targets are relaxed? The benefits of relaxing these targets would include reduced support costs and an extension of the end item's test interval.

The possible negative consequences might include a higher incidence of missed faults (higher false-accept rate) and a correspondingly higher acceptance cost. The spares acquisition cost represents a one-shot investment needed for spares to cover downtime resulting from testing and calibration. It can readily be appreciated that this cost variable is sensitive to testing and calibration intervals.

Tables 6.13 and 6.14 show the consequences of moving to a 3σ (99.73%) reliability target for the end-item attribute. (Only end-item results are shown in Table 6.13 since no change was made that would affect the test and calibration systems.) As expected, maintaining a 99.73% measurement reliability target for the end-item attribute instead of a 99.994% target allows lengthening the attribute's test interval. The change from 12 months to 627.6 months implies that the attribute can function with a minimum 99.73% measurement reliability for an essentially indefinite period (e.g., the mission lifetime).

But, what of the affect on mission objectives? As Table 6.13 shows, the change to a 99.73% reliability target incurs an increase in both the false-accept rate and the false-reject rate. As shown previously, an increase in the false-reject rate corresponds to increased unnecessary rework costs. If the test interval is lengthened to the mission lifetime, these costs would be incurred only once, before deployment. The increase in the false-accept rate, however, may jeopardize mission objectives. The severity of these risks can be evaluated by considering their effect on support costs and acceptance costs, as shown in Table 6.14.

TABLE 6.13 Solar Experiment – Analysis Results – Trial 1

TABLE 6.13
Solar Experiment — Analysis Results - Trial 1

PRIME SYSTEM	*CURRENT*	*PREVIOUS*
Adjustment Policy	*Renew/Fail*	*Renew/Fail*
Reliability Model	*Exponential*	*Exponential*
Test Interval	*627.6*	*12.0*
Observed EOP In-Tolerance	*99.73%*	*99.99%*
True EOP In-Tolerance	*99.74%*	*99.99%*
True AOP In-Tolerance	*99.86%*	*100.00%*
BOP In-Tolerance	*99.97%*	*100.00%*
Performance Limit	*10.0000*	*10.0000*
Test Limit	*10.0000*	*10.0000*
Adjustment Limit	*10.0000*	*10.0000*
False Accept Rate	*0.03%*	*0.00%*
False Reject Rate	*0.04%*	*0.00%*
Prime/Test Accuracy Ratio	*10.0:1*	*10.0:1*

TABLE 6.14 Solar Experiment – Cost Analysis Results – Trial 1

TABLE 6.14
Solar Experiment — Cost Analysis Results - Trial 1

SUMMARY COSTS ($)	*CURRENT*	*PREVIOUS*
Annual Test and Cal Cost	*7,348*	*46,945*
Annual Adjustment Cost	*50*	*60*
Annual Repair Cost	*2,040*	*2,040*
Annual Support Cost	*9,438*	*49,046*
Annual Acceptance Cost	*414*	*3*
Total Annual Cost	*9,853*	*49,049*
Spares Acquisition Cost	*14,093*	*259,145*

Table 6.14 shows that extending the Solar Experiment attribute’s test interval reduces total annual cost from $49,049 to $9,853 per year. Obviously, the increased false-reject rate does not increase unnecessary rework cost to the extent it exceeds cost savings due to reductions in testing and other service costs.

The increased risk of mission failure can be evaluated by considering the increase in annual acceptance costs. The increase from $3 per year to $414 per year is trivial (both figures are probably within the “noise level” of the accuracy of our original cost parameter estimates).

It can be concluded that lowering the attribute’s measurement reliability target (and significantly extending its test interval) does not compromise mission objectives. Note also the reduction in spares acquisition cost (a one-shot investment). This is an obvious result of extending the test interval from 12 months to 627.6 months. The reduction in spares acquisition cost indicates that, with a test interval of 627.6 months, NASA needs to procure only a single unit, as opposed to two units (on-line plus spare).

In Stage 7 of initial support planning (see Section 3.2.7), the calibration interval objective for the deuterium lamp test system was stated to be 6 months. To maintain a 3σ test-system measurement reliability, however, the maximum interval allowable was 4 months. We can now reexamine this issue by setting a 6-month calibration interval for the test system. The results are shown in Tables 6.15 and 6.16.

TABLE 6.15 Solar Experiment – Analysis Results – Trial 2

TABLE 6.15
Solar Experiment — Analysis Results - Trial 2

PRIME SYSTEM	CURRENT	PREVIOUS	BASELINE
Adjustment Policy	Renew/Fail	Renew/Fail	Renew/Fail
Reliability Model	Exponential	Exponential	Exponential
Test Interval	623.8	627.6	12.0
Observed EOP In-Tolerance	99.73%	99.73%	99.99%
True EOP In-Tolerance	99.74%	99.74%	99.99%
True AOP In-Tolerance	99.86%	99.86%	100.00%
BOP In-Tolerance	99.97%	99.97%	100.00%
Performance Limit	10.0000	10.0000	10.0000
Test Limit	10.0000	10.0000	10.0000
Adjustment Limit	10.0000	10.0000	10.0000
False Accept Rate	0.03%	0.03%	0.00%
False Reject Rate	0.04%	0.04%	0.00%
Prime/Test Accuracy Ratio	10.0:1	10.0:1	10.0:1
TEST SYSTEM			
Adjustment Policy	Renew/Fail	Renew/Fail	Renew/Fail
Reliability Model	Exponential	Exponential	Exponential
Calibration Interval	6.0	4.0	4.0
Observed EOP In-Tolerance	99.63%	99.73%	99.73%
True EOP In-Tolerance	99.79%	99.86%	99.86%
True AOP In-Tolerance	99.87%	99.91%	99.91%
BOP In-Tolerance	99.94%	99.96%	99.96%
Performance Limit	1.0000	1.0000	1.0000
Test Limit	1.0000	1.0000	1.0000
Adjustment Limit	1.0000	1.0000	1.0000
False Accept Rate	0.06%	0.04%	0.04%
False Reject Rate	0.21%	0.17%	0.17%
Test/Cal Accuracy Ratio	4.0:1	4.0:1	4.0:1

Table 6.15 shows that moving the test-system interval from 4 to 6 months does not compromise end-item performance in terms of false-accept and false-reject risks. This is typical of situations in which high accuracy ratios are maintained between test systems and end-items. Note also that the end-item test interval is not appreciably affected. (The small drop from 627.6 months to 623.8 months is in response to a slight increase in false reject rate.)

The results of Table 6.15 are echoed in Table 6.16, which shows no increase in acceptance cost, i.e., no reduction in mission reliability resulting from the interval extension. Moreover, since fewer test-system calibration actions are needed per year, total support costs drop from $9,438 to $8,190. Note also the reduction in spares acquisition cost, indicative of reduced test-system downtime resulting from calibration. Comparison of costs and risks with baseline (original) figures is particularly revealing.

By using the methodology described in Appendix C to analyze end-item attribute support requirements in terms of effect on mission performance, it can be seen that considerable savings may be realized without compromising performance objectives. Note that reliability targets could be relaxed to the point that false accepts and rejects will result in *increased* cost rather than *decreased* cost, thus one should maintain caution when relaxing requirements.

> Bear in mind that random and human-factor uncertainties were not included in the Solar Experiment example.

TABLE 6.16 Solar Experiment – Cost Analysis Results – Trial 2

TABLE 6.16

Solar Experiment — Cost Analysis Results - Trial 2

SUMMARY COSTS ($)	*CURRENT*	*PREVIOUS*	*BASELINE*
Annual Test and Cal Cost	*6,117*	*7,348*	*46,945*
Annual Adjustment Cost	*50*	*50*	*60*
Annual Repair Cost	*2,023*	*2.040*	*2.040*
Annual Support Cost	*8,190*	*9,438*	*49,046*
Annual Acceptance Cost	*414*	*414*	*3*
Total Annual Cost	*8,604*	*9,853*	*49,049*
Spares Acquisition Cost	*12,877*	*14,093*	*259,145*

7. OPERATIONAL REQUIREMENTS

The major operational function within the scope of this document is the establishment and preservation of measurement quality. This section discusses maintenance and repair, as they are essential to preserving data quality.

7.1 Measurement Quality

The primary requirement is to monitor and evaluate the uncertainties during the measurement process. The uncertainties must be maintained within a specified range, and exceptions must be identified, and corrective actions taken.

7.1.1 Establishing Measurement Quality

The total measurement process should be documented so that an objective evaluation can be achieved to support operational decisions and establish scientific facts.

The uncertainty values should be verified early in the operational phase. This is done by review of calibrations, observation of data scatter and drifts, analysis of operational and environmental factors, and cross comparisons with other data sources. All uncertainties from sensor output to the data-reduction equations must be considered. Operator characteristics and environmental changes are potentially important sources of uncertainty that should be reevaluated. The contributions of elements of the measurement chain to uncertainty are provided by design documentation. End-to-end checks based on check standards should be implemented. During early operations, statistically significant samples of all measurement parameters should be gathered to verify that their bias and precision are within the expected range.

The steps needed to establish measurement process quality at the start of the operations phase are

(1) Verify that the traceability requirement has been met with the measurement system as implemented at the start of operations. Valid calibration is an important part this activity.

(2) Conduct data-acquisition activities necessary to define the bias and precision errors.

(3) Combine the bias and precision errors into the uncertainty estimate.

(4) Compare the measured or estimated uncertainty to the tolerance defined or specified by the design documentation.

(5) If the estimated uncertainty of item (4) does not agree with the design tolerance, conduct the necessary investigation to resolve the difference.

7.1.2 Preserving Measurement Quality

Uncertainty is expected to grow between calibrations (Figure 6.1), and the confidence of the measurement is expected to diminish (Figure 6.2). The interval between calibrations is an important tool to control uncertainty. At the least, all test equipment used to perform measurements associated with the functions itemized in Section 2.2 must be in a recall system, calibrated at established intervals, and labeled to show the calibration status and the date of the next calibration. Specifically, the reader should review calibration control provisions of NHB 5300.4(1B), paragraph 1B905; NHB 5300.4(1C), paragraph 1C310(4); and NHB 5300.4(1D-2), paragraph 1D507(6).

Uncertainty and uncertainty growth should be estimated and tracked in time. To control uncertainty growth, calibration and maintenance intervals should be adjusted when necessary and possible. Out-of-tolerance measurements should be identified and reported to the user of the measurement data. Good data are needed to determine if an adjustment is needed.

> Operations personnel should provide the objective information necessary to adjust calibration intervals as a normal part of their activities.

Continuous feedback during operations is essential to preserve the data quality established during initial operations. Three periods should be considered:

(1) DESIGN VALIDATION — Early in the operations phase, compare bias and precision values to the expected performance. If there are deviations, identify the cause and take corrective action.

(2) MEASUREMENT PROCESS CONTROL — During the entire operations phase, continue to compare bias and precision values to previous values to assure that the measurement process is operating within the designed uncertainty range. Identify tendencies to exceed the acceptable uncertainty range and take corrective actions before out-of-tolerance conditions develop.

(3) CALIBRATION VERIFICATION — Acquire data before and after components are removed from the measurement system and sent to a different site for calibration. Assure that the uncertainty is within the acceptable range during and after the calibration.

The uncertainty analysis should be documented so that it can be audited, if required.

7.1.3 Maintaining Traceability

Measurement traceability may be lost when any part of the system is changed. The most common changes are from calibration, equipment failures, or software changes.

7.1.3.1 Traceability After Calibration

Calibrations should be verified at two times to maintain traceability:

(1) PRECALIBRATION — In an as-received condition (before any adjustments are made), a check calibration should be done and the operations personnel should compare the new calibration data to previous data to verify that the device was within tolerance when received. If the device was not in calibration, traceability was lost during the period of operation. This period probably cannot be objectively determined, but it must be estimated for later assessment of the data quality.

(2) POSTCALIBRATION — If data checks after the calibration show the same bias and uncertainty as before the calibration, traceability after the calibration is established.

7.1.3.2 Traceability After Equipment Failure

Equipment failure causes significant opportunity for traceability loss. Typically, traceability may be lost because of substitution of uncalibrated devices (e.g., sensors and instruments) into the measurement system to continue data-acquisition operations. During the period that uncalibrated devices are used, measurement traceability is lost and should be explicitly stated in writing.

Close monitoring of the uncertainty during this period may establish that the measurement process was within control. Separation of bias and precision errors , followed by their combination using the same method as that defined in the uncertainty design, will be necessary. If the bias and precision errors stay within the range experienced before the equipment failure, the measurement process can be considered undisturbed, though documented traceability was lost.

Software changes may be necessary for "work-around" during equipment failure and such changes should be documented to minimize traceability loss.

7.1.3.3 Traceability After Software Changes

Section 5.9 discusses software changes. Four activities are recommended during the operations phase:

(1) Maintain the software test cases under configuration control with no changes.

(2) Acquire data only with the formally approved software version. Conduct debugging and improvement activities with different versions.

(3) When new versions are ready for use, run the software test cases (with the prescribed system configuration for their use) to establish that the new version provides the same data.

(4) Strictly follow established software configuration management rules.

> Check standards can be a valuable tool for software test cases. Check standards can establish end-to-end conditions whose value should fall in a narrow range, aiding performance verification of a new version

7.2 Maintenance and Repair

7.2.1 General

Measurement systems maintenance includes technical activities intended to keep instruments, sensors, transducers, and their associated measurement circuitry in satisfactory working condition and to avoid catastrophic failures. Unlike calibration that is designed to control uncertainty growth beyond specified limits and to detect insidious failures unnoticeable by an operator, maintenance is designed to avoid accelerated wear-out and catastrophic operational failures.

The goal of maintenance is to assure there will be no breakdowns of measurement systems and that they can continue to operate safely. Typically, the cost to maintain is traded off against the cost to suffer measurement breakdowns and increased safety risks.

While maintenance can be an independent function, for convenience much of it is done during calibration. Typically, maintenance intervals are longer than calibration intervals. Therefore, much maintenance is scheduled to be done, for example, at every second or third scheduled calibration.

REF: **NHB 4200. 1C, 2.209 A**

> *A maintenance program shall be prescribed for all installation assigned equipment. The basic goal of the maintenance program will be to assure maximum readiness of equipment to perform assigned functions safely and efficiently and at the lowest cost.*
>
> *Maintenance is a continuing activity that is done more effectively under uniformly prescribed procedures and practices and with proper guidelines for the maintenance of each category of equipment in use at the installation. For applicable categories of equipment, these guidelines will identify maintenance requirements set forth in appropriate Federal Regulations and existing NASA Management Directives. When no such guidelines have otherwise been prescribed, maintenance will generally be done in accordance with the manufacturer's or design agency's recommended procedures.*

7.2.2 Procedures

REF: **NHB 4200. 1C, 2.209 A**

Maintenance programs will include procedures that ensure:

- *(1) Identification and estimation of maintenance requirements.*
- *(2) Uniform scheduling of maintenance service.*
- *(3) Correction of deficiencies detected during visual inspections of daily operations.*
- *(4) Prompt repair and calibration of equipment in keeping with the user's performance requirements.*
- *(5) Periodic scheduling of inspections to verify the effectiveness of the maintenance program and general operating conditions of equipment.*
- *(6) Use of manufacturer warranties or servicing agreements, as applicable.*
- *(7) Establishment of a technical library of applicable maintenance instructions for each category of equipment for which maintenance is provided.*
- *(8) Appropriate preservation and protection of inactive equipment held in storage.*
- *(9) Preprinted maintenance check lists when appropriate.*

7.2.3 Designs for Maintenance

7.2.3.1 Defining Maintenance Requirements

The requirements for maintenance are usually defined in manufacturer manuals where specific activities are directed to keep measuring systems operable. Other requirements are derived from data taken during calibrations, during repairs, and from user complaints made to repair and/or maintenance personnel. These requirements try to define the circuits, parts, mechanisms, and devices whose failure could be avoided by detecting diminished capability, fluid loss, dirt and grease accumulation, environmental stresses, and wear. Also, equipment use should be reviewed to find out the experience level of users, their opinions regarding its functional reliability, the environment in which it is used, and whether maintenance can be divided between the user and the maintenance facility. Special attention should be paid to instruments in space applications where maintenance considered routine on Earth will be difficult or impossible. The selection of measuring systems should consider designs that minimize or eliminate maintenance needs.

7.2.3.2 Selection of Maintenance Equipment

Typically, maintenance during the calibration process uses much of the same equipment used for calibration. Also, special facilities are needed for cleaning, lubricating, and stress testing for safety hazards and imminent failures. Some items categorized as measuring devices or accessories may need only maintenance and no calibration. They may also need special tests or actuations to confirm operability of emergency circuits and actuator equipment for more complex tasks or nonuncertainty related measurement capabilities, such as indications of presence or absence of signal, pressure, and flow.

As with calibration equipment, the site where maintenance is to be done has an influence on the equipment chosen. Design and selection of the measurement systems should include devices that need little maintenance or that can be maintained by remote means wherever possible.

7.2.3.3 Designing Maintenance Procedures

Clearly written and logically sequenced procedures are essential to successful maintenance operations. Where these procedures are scheduled in conjunction with calibration operations, they should be integrated to follow the flow of the calibration process. However, many maintenance operations should precede calibration to assure functional adequacy of the equipment before subjecting it to the more time-consuming calibrations. Maintenance procedures should have the same characteristics as those of well-designed calibration procedures. The better, more clearly written these procedures are, the less costly the continued maintenance operations will be. A small investment in well-prepared procedures will pay large dividends ultimately.

7.2.3.4 Defining Maintenance Servicing Intervals

One of the more difficult design problems is to develop a system that determines the most desirable time to do maintenance. Done too frequently, maintenance is a waste of time, or it may even be deleterious because of possible operator error; done too infrequently, it results in costly losses to both the measuring instruments and the operations in which they are used. Many owners schedule instrument maintenance at multiples of the calibration interval.

This is a practical approach because typical calibration intervals are shorter than maintenance intervals. As more knowledge accrues about calibration interval systems and calibration risk targets, basing maintenance intervals on calibration intervals may not prove to be a safe relationship. Calibration intervals have been getting longer and longer over the past few years because of improved stability of electronic circuitry, accumulation of statistically significant historical data, and improved interval adjustment systems. This could push maintenance intervals beyond prudent limits, unless current practices are changed accordingly.

Maintenance interval analysis should stand alone and be based on mathematical and statistical correlation of historical failure data that focus on types of maintenance done, time between maintenance, failed components/parts, and time between failures. From these data, MTBF figures should be developed for each family or model-numbered measuring instrument, sensor, or transducer. These figures reflect reliability index and should be related to an MTBF target for a given instrument population. The MTBF reliability targets should be established for those proper maintenance intervals which can be designated.

The quality of maintenance intervals and the effectiveness of failure analysis and corrective actions depends on the adequacy of the acquired data, including the design of the interval setting system. Some additional data may be needed as experience is gained with a particular maintenance program. There is also a significant similarity of data needed to operate either a calibration interval system or a maintenance interval system. An integration of the two systems should prove advantageous.

7.2.4 Repair

Repair becomes necessary when adjustments are inadequate to bring equipment into operational specifications. After repair, the measurement process should be validated by calibration, and measurement traceability and uncertainty reestablished. This becomes both very difficult and even more important when the repairs and calibration are done in the operational environment.

8. RECOMMENDATIONS FOR WAIVER/DEVIATION REQUESTS

8.1 General

The effective implementation of the techniques and methodologies described in this publication should lead to measurement system performance acceptable to the project sponsor and should comply with all standard measurement and calibration requirements. Special circumstances of limits of the state of the art and practicality may lead to situations where strict compliance with the standard requirements cannot be met. Any waiver or deviation from contractual requirements usually requires a written request for approval, as defined by contractual documents.

Normally, the standards for waiver/deviation requests require identification of the original requirement(s), reason/justification for the request, and indication of what effect the waiver/deviation will have on performance, safety, quality, and reliability, plus any other effect on other elements of the work. The waiver/deviation request will also identify the risk resulting from the deviation.

The following information is provided as an aid in the preparation, analysis, and review for waiver/deviation requests of measurement process, metrology, and calibration requirements.

8.2 Classification of Waiver/Deviation Requests

Waiver requests are categorized by the type of documents that invoke the requirements. They can also be classified according to the criticality and difficulty of the measurement as was done in Section 3.2.3.

Classifications of criticality of application were defined in Section 3.2.2.1 consistent with NHB 5330.4(1D-2) and are summarized here, as follows:

Category 1 Measurements that affect loss of life or vehicle

Category 2 Measurements that affect loss of mission

Category 3 Measurements that affect performance other than Category 1 and Category 2.

A second classification, which is complementary to the first, involves the degree of difficulty in the measurement process, especially as it relates to the measurement uncertainties required versus the capability or state of the art of present measurement systems.

The degree of difficulty of each measurement may directly affect its cost and quality and the quality of deployed space hardware. In the same manner as the criticality categories, those measurements deserving the most attention can be rated in terms of degrees of difficulty, where that difficulty may lead to space hardware with lowered performance capability. The degree of difficulty classifications were developed in Section 3.2.2.2 and are summarized as follows:

Degree A — These are the most difficult or impossible of measurements. They can be characterized as beyond the current capability of the state of the art, and therefore force the use of alternative performance parameters that may only marginally characterize system performance, but can, at least, be measured at reasonable difficulty levels.

Degree B — Measurements that cannot meet the NHB 5300.4(1B) measurement and calibration uncertainty ratio requirements of 10:1 and 4:1.

Degree C — Measurements made in environments hostile to best measuring system performance.

8.3 Independent Risk Assessment of Waiver/Deviation to Technical Requirements

Good practice indicates that all requests for waiver/deviation be subjected to an independent risk assessment. For measurement process, metrology, and calibration requirements, a special review by instrumentation and metrology specialists to identify risk issues and assess their significance is appropriate. The results of this review should be attached to the waiver request before it is routed for approval.

Appendix A DEFINITIONS

NOTE: The following definitions annotated (VIM) were prepared by a joint working group consisting of experts appointed by International Bureau of Weights and Measures (BIPM), International Electrotechnical Commission (IEC), International Organization for Standardization (ISO), and International Organization of Legal Metrology (OIML). The definitions appeared in Metrology, 1984, as the *International Vocabulary of Basic and General Terms in Metrology*. A few definitions were updated from the ISO/TAG4/WG3 publication *Guide to the Expression of Uncertainty in Measurement*, June 1992. Since this publication has modified some of the terms defined by the earlier VIM work, it is appropriate to modify them herein. The recent modifications of these terms are annotated (VIM)+, as appropriate.

ACCURACY — The deviation between the result of a measurement and the value of the measurand. *NOTE* — The use of the term "precision" for "accuracy" should be avoided.

ACCURACY RATIO — The ratio of performance tolerance limits to measurement uncertainty.

ADJUSTMENT — The operation intended to bring a measuring instrument into a state of performance and freedom from bias suitable for its use. (VIM)

ALIAS ERROR — The phenomenon whereby equally spaced sampling of high-frequency signals such as noise appear as lower frequency signals and are thus indistinguishable from data frequencies.

ALIASING — The process whereby two or more frequencies, integral multiples of each other, cannot be distinguished from each other when sampled in an analog-to-digital converter.

ANALOG-to-DIGITAL CONVERTER — A device that samples an analog signal at discrete, steady-rate time intervals, converts the sampled data points to a form of binary numbers, and passes the sampled data to a computer for processing.

APERTURE — The time required for an analog-to-digital converter to establish the digital representation of the unknown analog signal.

ATTRIBUTE — A measurable parameter or function.

BANDWIDTH (SMALL SIGNAL) — The band of frequencies extending from zero upwards to the frequency for which the output amplitude is reduced by no more than 3 dB (70.7% RMS of the voltage ratio) of the amplitude at zero frequency.

BASE UNIT — A unit of measurement of a base quantity in a given system of quantities. (VIM)

BIAS ERROR — The inherent bias (offset) of a measurement process or (of) one of its components. (Also, see SYSTEMATIC ERROR).

CALIBRATION — The set of operations that establish, under specified conditions, the relationship between values indicated by a measuring instrument or measuring system, or values represented by a material measure, and the corresponding known (or accepted) values of a measurand. *NOTES* —

(1) The result of a calibration permits the estimation of errors of indication of the measuring instrument, measuring system or material measure, or the assignment of values to marks on arbitrary scales. (2) A calibration may also determine other metrological properties. (3) The result of a calibration may be recorded in a document, sometimes called a calibration certificate or a calibration report. (4) The result of a calibration is sometimes expressed as a calibration factor or as a series of calibration factors in the form of a calibration curve. (VIM)

CALIBRATION FACTOR — The result of a calibration; a term or set of terms by which the instrument values are related to the corresponding known standard values. Sometimes expressed as a *calibration factor*, or *calibration coefficient*, or as a series of calibration factors in the form of a *calibration curve*.

CERTIFIED REFERENCE MATERIAL **(CRM)** — A reference material, one or more of whose property values are certified by a technically valid procedure, accompanied by or traceable to a certificate or other documentation that is issued by a certifying body. *NOTE* — NIST issues Standard Reference Material (SRM) which are in effect CRM.

CHARACTERIZATION — The measurement of the typical behavior of instrument properties that may affect the accuracy or quality of its response or derived data products. The results of a characterization may or may not be directly used in the calibration of the instrument response, but may be used to determine its performance. (The characterized properties may inherently affect the calibration of the instrument).

CHECK STANDARD — A device or procedure with known stable attributes, which is used for repeated measurements by the same measurement system for measurement process verification.

COLLECTIVE STANDARD — A set of similar material measures or measuring instruments fulfilling, by their combined use, the role of a standard. *NOTES* — (1) A collective standard is usually intended to provide a single value of a quantity. (2) The value provided by a collective standard is an appropriate mean of the values provided by the individual instruments. *EXAMPLES:* (a) collective voltage standard consisting of a group of Weston cells; (b) collective standard of luminous intensity consisting of a group of similar incandescent lamps. (VIM)

CONFIDENCE INTERVAL — An interval about the result of a measurement or computation within which the measurand value is expected to lie, as determined from an uncertainty analysis with a specified probability.

CONFIDENCE LEVEL — The probability that the confidence interval contains the value of a measurand.

CORRECTED RESULT — The final result of a measurement obtained after having made appropriate adjustments or corrections for all known factors that affect the measurement result. The closeness of the agreement between the result of a measurement and the value of the measurand.

CORRECTION — The value which, added algebraically to the uncorrected result of a measurement, compensates for an assumed systematic error. *NOTES* — (1) The correction is equal to the assumed systematic error, but of opposite sign. (2) Since the systematic error cannot be known exactly, the correction value is subject to uncertainty. (VIM)

CORRECTION FACTOR — The numerical factor by which the uncorrected result of a measurement is multiplied to compensate for an assumed systematic error. *NOTE* — Since the systematic error cannot be known exactly, the correction factor is subject to uncertainty. (VIM)

CROSS-CALIBRATION — The process of assessing the relative accuracy and precision of response of two or more instruments. A cross-calibration would provide the calibration and/or correction factors necessary to intercompare data from different instruments looking at the same target. Ideally this would be done by simultaneous viewing of the same working standards or target. Any variations in environmental conditions, calibration procedures, or data correction algorithms between the instruments must be accounted for in the assessment.

Crosstalk — Signal interference between measurement channels usually due to coupling between channels in some element, e.g., power supplies, adjacent cables, and adjacent telemetry channels.

DATA PRODUCT — The final processed data sets associated with the various measured and derived parameters that are the object of a specified investigation.

DEAD BAND — The range through which a stimulus can be varied without producing a change in the response of a measuring instrument. *NOTE* — The inherent dead band is sometimes deliberately increased to reduce unwanted change in the response for small changes in the stimulus. (VIM)

DECIMATION — The process of eliminating data frequencies in digital data—used with digital filtering to minimize aliasing.

DECISION RISK — The probability of making an incorrect decision.

DEGREES-OF-FREEDOM — In statistics, degrees-of-freedom for a computed statistic refers to the number of free variables that can be chosen. For example, the sample variance statistic (σ^2) is computed using n observations and one constant (sample average). Thus, there are $n-1$ free variables and the degrees-of-freedom associated with the statistics are said to be $n-1$.

DERIVED UNITS — Derived units expressed algebraically in terms of base units (of a system of measure) by the mathematical symbols of multiplication and division. Because the system is coherent, the product or quotient of any two quantities is the unit of the resulting quantity.

DETECTOR — A device or substance that indicates the presence of a particular quantity without necessarily providing its value. *NOTE* — In some cases, an indication may be produced only when the value of the quantity reaches a given threshold. *EXAMPLE:* (a) halogen leak detector; (b) temperature-sensitive paint.

DIFFERENTIAL METHOD OF MEASUREMENT — A method of measurement in which the measurand is replaced by a quantity of the same kind, of known value only slightly different from the value of the measurand, and in which the difference between the two values is measured. *EXAMPLE:* measurement of the diameter of a piston by means of gauge blocks and a comparator. (VIM)

DIRECT METHOD OF MEASUREMENT — A method of measurement in which the value of the measurand is obtained directly, rather than by measurement of other quantities functionally related to the measurand. *NOTE* — The method of measurement remains direct even if it is necessary to make supplementary measurement to determine the values of influence quantities in order to make corresponding corrections. *EXAMPLES:* (a) measurement of a length using a graduated rule; (b) measurement of a mass using an equal-arm balance. (VIM)

DISCRIMINATION — (See RESOLUTION)

DISCRIMINATION THRESHOLD — The smallest change in a stimulus that produces a perceptible change in the response of a measuring instrument. *NOTE* — The discrimination threshold may depend on, for example, noise (internal or external), friction, damping, inertia, or quantization. *EXAMPLE:* if the smallest change in load that produces a perceptible displacement of the pointer of a balance is 90 mg, then the discrimination threshold of the balance is 90 mg. (VIM)

DRIFT — The slow variation with time of a metrological characteristic of a measuring instrument. (VIM)

DYNAMIC MEASUREMENT — The determination of the instantaneous value of a quantity and, where appropriate, its variation with time. *NOTE* — The qualifier "dynamic" applies to the measurand and not to the method of measurement. (VIM)

ENGINEERING UNITS — A set of defined units commonly used by an engineer in a specific field to express a measurand. The units should be expressed in terms of a recognized system of units, preferably SI units.

ENVIRONMENTAL VARIABLES — Variable physical properties in the environment of the instrument or target (such as temperature, particulate and electromagnetic radiation, vacuum, and vibration) that may affect the result of a measurement. *NOTE* — The sensor does not measure an environmental variable; it measures an *observable*.

ERROR — The difference between the result of a measurement and the value of the measurand.

ERROR MODEL — A mathematical model of the measurement chain in which all potential error sources are identified, quantified, and combined such that a meaningful estimate of measurement uncertainty can be determined.

GROUP STANDARD SERIES OF STANDARDS — A set of standards of specially chosen values that individually or in suitable combination reproduce a series of values of a unit over a given range. *EXAMPLES:* (a) set of weights; (b) set of hydrometers covering contiguous ranges of density. (VIM)

HYSTERESIS — The property of a measuring instrument whereby its response to a given stimulus depends on the sequence of preceding stimuli. *NOTE* — Although hysteresis is normally considered in relation to the measurand, it may also be considered in relation to influence quantities. (VIM)

INDICATING (MEASURING) INSTRUMENT — A measuring instrument that displays the value of a measurand or a related value. *EXAMPLES:* (a) analog voltmeter; (b) digital voltmeter; and (c) micrometer. (VIM)

INDICATING DEVICE — For a measuring instrument, the set of components that displays the value of a measurand or a related value. *NOTES* — (1) The term may include the indicating means or setting device of a material measure, for example, of a signal generator. (2) An analog indicating device provides an analog indication; a digital indicating device provides a digital indication. (3) A form of presentation of the indication either by means of a digital indication in which the least significant digit moves continuously, thus permitting interpolation, or by means of a digital indication supplemented by a scale and index, is called a semidigital indication. (4) The English term “readout device” is used as a general descriptor of the means whereby the response of a measuring instrument is made available. (VIM)

INDICATION (OF A MEASURING INSTRUMENT) — The value of a measurand provided by a measuring instrument. *NOTES* — (1) The indication is expressed in units of the measurand, regardless of the units marked on the scale. What appears on the scale (sometimes called direct indication, direct reading or scale value) must be multiplied by the instrument constant to provide the indication. (2) For a material measure, the indication is nominal or marked value. (3) The meaning of the term “indication” is sometimes extended to cover what is recovered by a recording instrument, or the measurement signal within a measuring system. (VIM)

INDIRECT METHOD OF MEASUREMENT — A method of measurement in which the value of a measurand is obtained by measurement of other quantities functionally related to the measurand. *EXAMPLES:* (a) measurement of a pressure by measurement of the height of a column of liquid; (b) measurement of a temperature using a resistance thermometer. (VIM)

INFLUENCE QUANTITY — A quantity that is not the subject of the measurement but which influences the value of the measurand or the indication of the measuring instrument. *EXAMPLES:* (a) ambient temperature; (b) frequency of an alternating measured voltage. (VIM)

INSTRUMENT CONSTANT — The coefficient by which a direct indication must be multiplied to obtain the indication of a measuring instrument. *NOTES* — (1) A measuring instrument in which the direct indication is equal to the value of the measurand has an instrument constant of 1. (2) Multirange measuring instruments with a single scale have several instrument constants that correspond, for example, to different positions of a selector mechanism. (3) For some measuring instruments, the transformation from direct indication to indication may be more complex than a simple multiplication by an instrument constant. (VIM)

INTEGRATING (MEASURING) INSTRUMENT — A measuring instrument that determines the value of a measurand by integrating a quantity with respect to another quantity. *EXAMPLE:* electrical energy meter. (VIM)

INTERNATIONAL STANDARD — A standard recognized by an international agreement to serve internationally as the basis for fixing the value of all other standards of the quantity concerned. (VIM)

INTRINSIC ERROR (OF A MEASURING INSTRUMENT) — Errors inherent in a measuring instrument. *EXAMPLE:* nonlinearity, gain accuracy, noise, offset, and hysteresis.

LIMITING CONDITIONS — The extreme conditions that a measuring instrument can withstand without damage and without degradation of its metrological characteristics when it is subsequently operated under its rated operating conditions. *NOTES* — (1) The limiting conditions for storage, transport, and operating may be different. (2) The limiting conditions generally specify limiting values of the measurand and of the influence quantities. (VIM)

LINEARITY — (See NONLINEARITY.)

MATHEMATICAL MODEL — A mathematical description of a system relating inputs to outputs. It should be of sufficient detail to provide inputs to system analysis studies, such as performance prediction, uncertainty (or error) modeling, and isolation of failure or degradation mechanisms, or environmental limitations.

MEASURAND — A specific quantity subjected to measurement. *NOTE* — As appropriate, this may be the measured quantity or the quantity to be measured. (VIM)+

MEASUREMENT — The set of operations having the object of determining the value of a quantity. (VIM)

MEASUREMENT ASSURANCE PROGRAM (MAP) — A program applying specified (quality) principles to a measurement process. A MAP establishes and maintains a system of procedures intended to yield calibrations and measurements with verified limits of uncertainty based on feedback of achieved calibration of measurement results. Achieved results are observed systematically and are used to eliminate sources of unacceptable uncertainty.

MEASUREMENT PROCEDURE — The set of theoretical and practical operations, in detailed terms, involved in the performance of measurements according to a given method. (VIM)

MEASUREMENT PROCESS — All the information, equipment, and operations relevant to a given measurement. *NOTE* — This concept embraces all aspects relating to the performance and quality of the measurement; it includes the principle, method, procedure, values of the influence quantities, the measurement standards, and operations. The front-end analysis, measurement system, and operations combine into the measurement process. (VIM) +

MEASUREMENT RELIABILITY — The probability that a measurement attribute (parameter) of an item of equipment is in conformance with performance specifications.

MEASUREMENT SIGNAL — A representation of a measurand within a measuring system. *NOTE* — The input to a measuring system may be called the stimulus; the output signal may be called the response. (VIM)

MEASUREMENT STANDARD — A material measure, measuring instrument, or system intended to define, realize, conserve, or reproduce a unit or one or more known values of a quantity in order to transmit them to other measuring instruments by comparison. *EXAMPLES:* (a) 1 kg mass standard; (b) standard gauge block; (c) 100 Ω standard resistor; (d) saturated Weston standard cell; (e) standard ammeter; (f) cesium atomic frequency standard. (VIM)

MEASUREMENT SYSTEM — One or more measurement devices and any other necessary system elements interconnected to perform a complete measurement from the first operation to the result. *NOTE* — A measurement system can be divided into general functional groupings, each of which consists of one or more specific functional steps or basic elements.

MEASURING CHAIN — A series of elements of a measuring instrument or system which constitutes the path of the measurement signal from the input to the output. *EXAMPLE:* an electro-acoustic measuring chain comprising a microphone, attenuator, filter, amplifier, and voltmeter. (VIM)

METROLOGY — The field of knowledge concerned with measurement. *NOTE* — Metrology includes all aspects both theoretical and practical with reference to measurements, whatever their level of accuracy, and in whatever fields of science or technology they occur. (VIM)

NATIONAL STANDARD — A standard recognized by an official national decision as the basis for fixing the value, in a country, of all other standards of the quantity concerned. The national standard in a country is often a primary standard. In the United States, national standards are established, maintained, and disseminated by NIST. (VIM) +

NOMINAL VALUE — A value used to designate a characteristic of a device or to give a guide to its intended use. *NOTE* — The nominal value may be a rounded value of the value of the characteristic concerned and is often an approximate value of the quantity realized by a standard. *EXAMPLE:* the value marked on a standard resistor. (VIM)

NONLINEARITY — The deviation of the output of a device from a straight line where the straight line may be defined using end points, terminal points, or best fit.

NOISE — Any extraneous or unwanted signal that contaminates the measurement. For measurement systems, noise consists of random noise (thermal processes within conductors), white noise (thermal processes within resistors), and systematic noise (such as line frequency, power supply ripple, and EMI).

PRECISION — The closeness of the agreement between the results of successive measurements of the same measurand carried out subject to all of the following conditions: (a) the same method of measurement; (b) the same observer; (c) the same sensor; (d) the same measuring instrument; (e) the same location; (f) the same conditions of use; (g) repetition over a short period of time. The confidence with which a measurement can be repeated under controlled conditions, or the confidence that two different measurement systems or techniques can yield the same result. *NOTE* — The use of the term precision for accuracy should be avoided. (See REPEATABILITY.)

PRIMARY STANDARD — A standard that has the highest metrological qualities in a specified field. *NOTE* — The concept of primary standard is equally valid for base units and for derived units. (VIM)

PRINCIPLE OF MEASUREMENT — The scientific basis of a method of measurement. *EXAMPLES:* (a) the thermoelectric effect applied to the measurement of temperature; (b) the Josephson effect applied to the measurement of voltage; (c) the Doppler effect applied to the measurement of velocity. (VIM)

PROBABILITY DENSITY FUNCTION (pdf) — A mathematical expression describing the functional relationship between a specific value of an attribute or variable and the probability of obtaining that value.

RANDOM ERROR — A component of the error of measurement which, in the course of a number of measurements of the same measurand, varies in an unpredictable way. *NOTE* — It is not possible to correct for random error. (VIM)

RATED OPERATING CONDITIONS — Conditions of use giving the ranges of the measurand and of the influence quantities and other important requirements for which the metrological characteristics of a measuring instrument are intended to lie within specified limits. *NOTE* — The rated operating conditions generally specify rated values of the measurand and of the influence quantities. (VIM)

RECORDING (MEASURING) INSTRUMENT — A measuring instrument that provides a record (permanent or semipermanent) of the value of a measurand or a related value. *NOTE* — (1) The record may be analog (continuous or discontinuous line) or digital. (2) Values of more than one quantity may be recorded simultaneously. (3) A recording measuring instrument may also incorporate an indicating device. *EXAMPLES:* (a) barograph; (b) thermoluminescent dosimeter. (VIM)

REFERENCE CONDITIONS — Conditions of use for a measuring instrument prescribed for performance testing or to ensure valid intercomparison of results of measurements. *NOTE* — The reference conditions generally specify reference values or reference ranges for the influence quantities affecting the measuring instrument. (VIM)

REFERENCE MATERIAL — A material or substance one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials. (VIM)

REFERENCE STANDARD — A standard, generally of the highest metrological quality available at a given location, from which measurements made at that location are derived. (VIM)

RELATIVE ERROR — The absolute error of measurement divided by the value of the measurand.

REPEATABILITY — The ability of an instrument to give, under specific conditions of use, closely similar responses for repeated applications of the same stimulus. *NOTE* — Repeatability may be expressed quantitatively in terms of the dispersion of the results. (See PRECISION.)

REPRODUCIBILITY (OF MEASUREMENTS) — The closeness of the agreement between the results of measurements of the same measurand, where the individual measurements are carried out changing such conditions as: (a) method of measurement; (b) observer; (c) measuring instrument; (d) location; (e) conditions of use; (f) time. (VIM) (See PRECISION.)

REQUIREMENT — A translation of the needs into a set of individual quantified or descriptive specifications for the characteristics of an entity in order to enable its realization and examination.

RESOLUTION (OF AN INDICATING DEVICE) — A quantitative expression of the ability of an indicating device to distinguish meaningfully between closely adjacent values of the quantity indicated. (VIM)

RESPONSE CHARACTERISTIC — For defined conditions, the relationship between a stimulus and the corresponding response. *NOTES* — (1) The relationship may be based on theoretical or experimental considerations; it may be expressed in the form of an algebraic equation, a numerical table or a graph. (2) When the stimulus varies as a function of time, one form of the response characteristic is the transfer function (the Laplace transform of the response divided by that of the stimulus). (VIM)

RESPONSE TIME — The time interval between the instant when a stimulus is subjected to a specified abrupt change and the instant when the response reaches and remains within specified limits of its final steady value. (VIM)

RESULT OF A MEASUREMENT — The value of a measurand obtained by measurement. *NOTE* — (1) When the term "result of a measurement" is used, it should be made clear whether it refers to: (a) the indication; (b) the uncorrected result; or (c) the corrected result and whether averaging over several observations is involved. (2) A complete statement of the result of a measurement includes information about the uncertainty of measurement and about the values of appropriate influence quantities. (VIM)

SAMPLING INTERVAL — The size of the samples used to measure something; e.g., in imaging, sampling refers to pixel size. In spectroscopy, sampling refers to the smallest spectral bandwidth used to measure something. Sampling, as applied to an analog-to-digital converter, is the process that transforms a continuous function into a series of discreet values at a linear time rate.

SCALE — An ordered set of scale marks, together with any associated numbering, forming a part of an indicating device. (VIM)

SECONDARY STANDARD — A standard whose value is fixed by comparison with a primary standard. (VIM)

SENSITIVITY — The change in the response of a measuring instrument divided by the corresponding change in the stimulus. *NOTE* — Sensitivity may depend on the value of the stimulus. (VIM)

SENSOR — A device that responds to either the absolute value of, or change in, a physical stimulus (heat, light, sound, magnetism, pressure, or particular motion) and produces a corresponding signal. A sensor can be an entire instrument or the part of it that measures a phenomenon.

SI PREFIXES — Used as prefixes in combination with the terms and symbols of SI units to form decimal multiples and submultiples of those units.

SI UNITS — The coherent system of units adopted and recommended by the General Conference on Weights and Measures (CGPM). (VIM)

SPAN — The modulus of the difference between the two limits of a nominal range of a measuring instrument. *EXAMPLE:* nominal range –10 V to +10 V: span 20 V. (VIM)

SPECIFIED MEASURING RANGE / SPECIFIED WORKING RANGE — The set of values of a measurand for which the error of a measuring instrument is intended to lie within specified limits. *NOTE* — The upper and lower limits of the specified measuring range are sometimes called the maximum capacity and minimum capacity, respectively. (VIM)

STABILITY — The ability of a measuring instrument to maintain its metrological characteristics within specified limits. *NOTE* — It is usual to consider stability with respect to time. Where stability with respect to another quantity is considered, this should be stated explicitly.

STANDARD DEVIATION — For a series of n measurements of the same measurand, the parameter s characterizing the dispersion of the results and given by the formula:

$$\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{n-1}}$$

x_i being the result of the ith measurement and $\overline{x}$ being the arithmetic mean of the n results considered. *NOTE* — (1) The experimental standard deviation should not be confused with the population standard deviation σ of a population of size N and of mean μ, given by the formula:

$$\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N}}$$

(2) Considering the series of n measurements as a sample of a population, s is an estimate of the population standard deviation. (3) The expression $s/\sqrt{n}$ provides an estimate of the standard deviation of the arithmetic mean $\bar{x}$ with respect to the mean μ of the overall population. The expression $s/\sqrt{n}$ is called the *experimental standard deviation of the mean*. (VIM)

STATIC MEASUREMENT — The measurement of a quantity whose value can be considered constant for the duration of the measurement. *NOTE* — The qualifier "static" applies to the measurand and not to the method of measurement. (VIM)

SYSTEMATIC ERROR — A component of the error of measurement which, in the course of a number of measurements of the same measurand, remains constant or varies in a predictable way. *NOTE* — (1) Systematic errors and their causes may be known or unknown. (2) For a measuring instrument, see BIAS ERROR. (VIM)

TOLERANCE — The total permissible variation of a quantity from a designated value.

TRACEABILITY — The property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons. (VIM)

TRANSDUCER — A measuring device that provides an output quantity having a given relationship to the input quantity. *EXAMPLES:* (a) thermocouple; (b) current transformer; (c) electro-pneumatic converter. (VIM)

TRANSFER STANDARD — A standard used as an intermediary to compare standards, material measures or measuring instruments. *NOTE* — When the comparison device is not strictly a standard, the term transfer device should be used. *EXAMPLE:* adjustable calipers used to inter-compare end standards. (VIM)

TRANSPARENCY — The ability of a measuring instrument not to affect the value of the measurand.

TRAVELING STANDARD — A standard, sometimes of special construction, intended for transport between different locations. Also known as a "Transport Standard." (VIM)+

TRUE VALUE (OF A QUANTITY) — The value that characterizes a quantity perfectly defined, in the conditions that exist when that quantity is considered. *NOTE* — The true value of a quantity is an ideal concept and, in general, cannot be known exactly. Indeed, quantum effects may preclude the existence of a unique true value. (VIM)

UNCERTAINTY (OF MEASUREMENT) — A parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand. *NOTES* — (1) The parameter may be, for example, a standard deviation (or a given multiple of it), or the width of a confidence interval. (2) Uncertainty of measurement comprises, in general, man components. Some of these components may be evaluated from the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations.

The other components, which also can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information. (VIM)+

UNIT (OF MEASUREMENT) — A specific quantity, adopted by convention, used to quantitatively express values that have the same dimension. (VIM)

VALUE (OF A QUANTITY) — The expression of a quantity in terms of a number and an appropriate unit of measurement. *EXAMPLE:* 5.3 m; 12 kg; –40° C. (VIM)

VARIANCE — (See STANDARD DEVIATION.)

VERIFICATION — Tests and analyses to be performed during the design, development, assembly, integration, and operational phases of a measurement system to assure that specified requirements have been met. Includes all subsystem and system tests done at the functional level.

WORKING STANDARD — A standard, usually calibrated against a reference standard, used routinely to calibrate or check material measures or measuring instruments. (VIM)

ZERO (OF A MEASURING INSTRUMENT) — The direct indication of a measuring instrument when the instrument is in use with a zero value of the measurand and any auxiliary power supply required to operate the instrument being switched on. *NOTES* — (1) This term is commonly called electrical zero in the case of a measuring instrument having an electrical auxiliary power supply. (2) The term mechanical zero is often used when the instrument is not in use and any auxiliary power supply is switched off. (3) The mechanical zero may possibly not coincide with the electrical zero; in some types of instrument, the mechanical zero may be indeterminate. (4) There is also a “data zero,” e.g.; digital telemetry systems typically operate from 0 to 5 V, with “data zero” at 2.5 V. (VIM) +

Appendix B MATHEMATICAL METHODS FOR OPTIMAL RECALL SYSTEMS

This appendix provides the mathematical and detailed algorithmic methodology needed to implement optimal calibration interval analysis systems, as described in Section 6. In developing the concepts behind the methodology, many topics discussed in Section 6 will be reiterated. It is recommended that Section 6 be read as preparation for the material presented here.

Sections B.1 and B.2 review the concepts of measurement reliability and optimal calibration intervals. Section B.3 discusses the consequences of suboptimal systems, and Section B.4 reviews the process by which TME parameters transition from in-tolerance to out-of-tolerance. Calibration interval analysis methodology development begins with Section B.5 in which the out-of-tolerance or uncertainty growth time series is described. Sections B.6 through B.8 provide methods and tools for analyzing the time series. Section B.9 describes mathematical functions that have proved useful in modeling both parameter and instrument measurement reliabilities. Section B.10 discusses calibration interval determination, and Sections B.11 through B.15 give techniques for identifying statistical outliers and for preprocessing calibration history data. Section B.16 summarizes the approach for determining measurement reliability targets.

B.1 Measurement Reliability

For a given TME parameter population, the out-of-tolerance probability can be measured in terms of the percentage of observations on the parameter that correspond to out-of-tolerance conditions. A population may be identified at several levels. Those pertinent to calibration interval analysis are (1) all observations taken on serial-numbered items of a given model number or other homogeneous grouping, (2) all observations taken on model numbers within an instrument class, (3) all observations on a TME parameter of a model number or other homogeneous grouping, and (4) all observations on a TME parameter of a serial number item. It is shown in Section B.5 that the fraction of observations on a given TME parameter classified as out-of-tolerance at calibration is a *maximum-likelihood-estimate* (MLE) of the out-of-tolerance probability for the parameter. Since out-of-tolerance probability is a measure of test process uncertainty, the percentage of calibrations that yield out-of-tolerance observations is a measure of this uncertainty. This leads to using "percent observed out-of-tolerance" as a variable by which test process uncertainty can be monitored.

The complement of percent observed out-of-tolerance is the percent observed *in-tolerance*. The latter is called *measurement reliability*.

MEASUREMENT RELIABILITY — The probability that a measurement attribute (parameter) of an item of equipment is in conformance with performance specifications.

An effective approach to determining and implementing a limit on test process uncertainty involves defining a minimum *measurement reliability target* for TME parameters. In practice, many organizations have found it expedient to manage measurement reliability at the instrument level rather than at the parameter level. In these cases, an item of TME is considered out-of-tolerance if one or more of its parameters is found out-of-tolerance. Variations on this theme are possible. Determination of measurement reliability targets is discussed in Section B.15.

B.2 Optimal Calibration Intervals

Reiterating a part of Section 6, calibration intervals are considered optimal if the following criteria are met:

CRITERIA 1 — Measurement reliability targets correspond to measurement uncertainties commensurate with measurement decision risk-control requirements.

End-item utility is compromised and operating costs are increased if incorrect decisions are made during testing. The risk of making these decisions is controlled through holding TME uncertainties to acceptable levels. This is done by maintaining minimum levels of TME measurement reliability. These minimum levels are the measurement reliability targets.

CRITERIA 2 — Calibration intervals lead to observed measurement reliabilities in close agreement with measurement reliability targets.

Because measurement uncertainty grows with time since calibration (see Figures 6.1 and 6.2), measurement reliability decreases with time since calibration. The particular time since calibration that corresponds to the established measurement reliability target is the desired calibration interval.

In some applications, periodic TME recalibrations are not possible (as with TME on-board deep-space probes) or are not economically feasible (as with TME on-board orbiting satellites). In these cases, TME measurement uncertainty is controlled by designing the TME and ancillary equipment or software to maintain a measurement reliability level that will not fall below the minimum acceptable reliability target for the duration of the mission.

CRITERIA 3 — Calibration intervals are determined cost-effectively.

A goal of any calibration interval analysis system is that the cost per interval should be held to the lowest level needed to meet measurement reliability targets. This can be done if calibration intervals are determined with the least human intervention and manual processing, i.e., if the interval analysis task is automated. Minimizing human intervention also calls for some development and implementation of decision algorithms. Full application of advanced artificial intelligence methods and tools is not ordinarily needed. Simple variables can often be used to approximate human decision processes. This expedient is used, for example, in Sections B.8 and B.14.

CRITERIA 4 — Calibration intervals are arrived at in the shortest possible time.

Several interval assignment approaches are currently in use, but most cannot meet Criteria 3 and 4. Some can meet these criteria, but need long periods of time to do so. Usually, the time needed for these approaches to arrive at intervals consistent with measurement reliability targets is more than the operational lifetime of the TME of interest. In contrast, methodologies that embody the principles described in this appendix provide the capabilities to meet all the above criteria in an expedient manner.

Besides meeting these criteria, systems that incorporate these principles should permit easy and expedient implementation of analysis results. The results should be comprehensive, informative, and unambiguous. Mechanisms should be in place to either couple the analysis results directly to an associated equipment control system or to transfer information to the equipment control system with least restatement or translation.

To appreciate better the need for optimal calibration intervals, it is worth considering the consequences of suboptimal systems.

B.3 Consequences of Suboptimal Systems

One deficiency of suboptimal calibration recall systems is the failure to develop an appropriate TME measurement reliability target or targets. Low levels of TME measurement reliability lead to low levels of average end-item utility. But, setting measurement reliability targets higher than necessary results in more frequent calibration than necessary. This translates to operating costs higher than are justifiable because of end-item utility requirements. Excessive measurement reliability targets lead to inappropriately short intervals, as shown below.

Assume that the uncertainty growth behavior of a TME population of interest can be modeled by the exponential reliability model described in Section B.9:

$$R(t) = R_0 e^{-\lambda t} ,$$

where $R(t)$ represents measurement reliability and t represents time since calibration. The parameters R_0 and λ are, respectively, the measurement reliability at $t = 0$ and the TME out-of-tolerance rate. From the expression for $R(t)$, the calibration interval I is determined according to

$$I = -\frac{\ln(R^* / R_0)}{\lambda} ,$$

where $\ln(\cdot)$ is the natural log function, and R^* is the reliability target. Note that R^* should always be less than or equal to R_0 so that $-\ln(R^*/R_0)$ should always be greater than or equal to zero.[8] Note also that the higher the reliability target, the shorter the calibration interval. Figure B.1 shows this relationship for the exponential model. Similar results apply to uncertainty growth processes represented by other reliability models.

As Figure B.1 shows, the calibration interval can be a sensitive function of the measurement reliability target. As mentioned earlier, setting an inappropriate measurement reliability target can lead to undesirable cost outcomes or compromised end-item utility.

[8] Instances have been found where the reverse has been true. In these cases, the interval recall systems had been trying to find the interval that would lead to a higher in-tolerance percentage at the end of the interval than was in effect at the beginning!

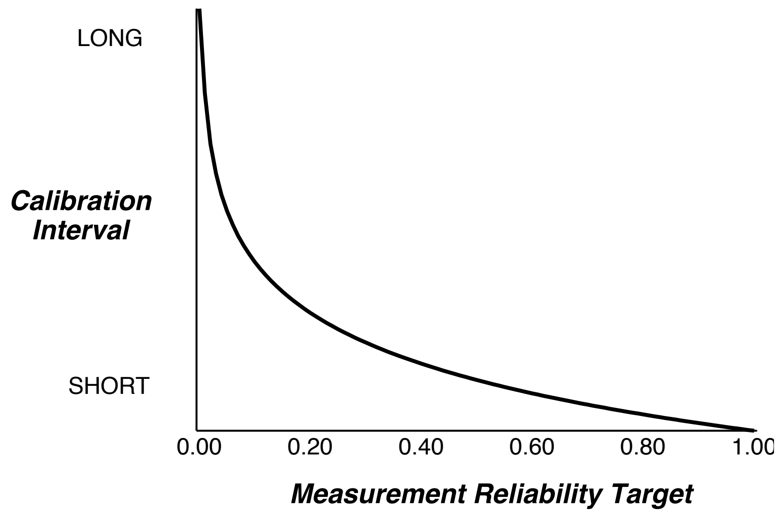


FIGURE B.1 — CALIBRATION INTERVAL VERSUS MEASUREMENT RELIABILITY TARGET.

The relationship between calibration interval and reliability target for a TME represented by the exponential reliability model with $R_0 = 1.0$.

Another facet of suboptimal systems is the inability to find intervals that yield actual measurement reliabilities that agree with established reliability targets. Many systems use sliding interval or other heuristic adjustment schemes that "react" to calibration results on a calibration-by-calibration basis. Such systems are typically incompatible with adjusting intervals to meet in-tolerance percentage goals. Some systems *do* try to adjust intervals to established reliability targets. However, as mentioned earlier, they do not arrive at intervals commensurate with these targets within the lifetimes of the TME under consideration. The consequences of suboptimality in calibration interval determination are summarized in Table B.1.

TABLE B.1 Consequences of Suboptimal Calibration Interval Systems

TABLE B.1

Consequences of Suboptimal Calibration Interval Systems

CONDITION	CONSEQUENCE
Reliability target too high	*Calibration intervals too short*
Reliability target too low	*Calibration intervals too long*
Calibration intervals too short	*Calibration costs too high* *Excessive TME downtime* *Unnecessary drain on personnel* *Logistics/supply problems*
Calibration intervals too long	*Unsatisfactory end-item utility*
Slow Convergence to optimal intervals	*Intervals too long or short for too long a time* *Unnecessary effort expended in adjusting intervals*

B.4 The Out-of-Tolerance Process

As discussed earlier, periodic TME calibration is motivated because the confidence that TME are operating in an in-tolerance state diminishes with time since last calibrated. This presupposes that there is some process by which TME parameters transition from in-tolerance to out-of-tolerance.

Because of the complexity of many instrument types, deterministic descriptions of this process are often difficult or impossible to achieve. This is not to say that the behavior of an individual instrument cannot in principle be described in terms of physical laws with predictions of specific times of occurrence for out-of-tolerance conditions. Such descriptions are typically beyond the scope of equipment management programs. Such descriptions become overwhelmingly impractical when attempted for populations of instruments subject to diverse conditions of handling, environment, and application.

Variations in these conditions are usually unpredictable. This argues for descriptions of the in-tolerance to out-of-tolerance process for populations of like instruments to be *probabilistic* instead of deterministic in nature. This point is further supported by the commonly accepted notion that each individual instrument is characterized by random inherent differences that arise from the vagaries of fabrication and later repair and maintenance. Besides, for TME managed via an equipment pool system, the conditions of handling, environment, and application may switch from instrument to instrument in a random way because of the stochastic character of equipment demand and availability in such systems. So, the failure of an individual TME parameter to meet a set of performance criteria (i.e., the occurrence of an out-of-tolerance state) is considered a *random phenomenon*, that is, one that can be described in terms of probabilistic laws.

B.5 The Out-of-Tolerance Time Series

As shown earlier, a high degree of confidence can be placed on the supposition that equipment parameters are in conformance with performance specifications immediately following calibration. As the equipment experiences random stresses resulting from use and storage, this confidence decreases. Unless later recalibration is done, the confidence in the in-tolerance status (measurement reliability) of equipment parameters decreases monotonically with time. A random phenomenon that arises through a process that is developing in time in a manner described by probabilistic laws is called a *stochastic process*.

One method of analysis by which stochastic processes of this kind are described is *time series analysis*. A time series is a set of observations arranged chronologically. Suppose that the observations comprising the time series are made over an interval T and that the observations have been taken at random times t. Let the observed value of the variable of interest at time t be labeled $\tilde{R}(t)$. The set of observations $\{\tilde{R}(t), t \in T\}$ is then a time series that is a realization of the stochastic process $\{\tilde{R}(t), t \in T\}$. Time series analysis is used to infer from the observed time series the probability law of the stochastic process. Time series analysis is applied to the calibration interval analysis problem by letting $\tilde{R}(t)$ represent observed measurement reliability corresponding to a calibration interval of duration t.

The value $\tilde{R}(t)$ is obtained by taking a sample of in- or out-of-tolerance observations recorded after an interval t has passed since the previous calibrations. Representing in-tolerance observations in the sample by g(t) and the size of the sample by $n(t)$, the observed measurement reliability associated with a calibration interval of duration t is given by $\tilde{R}(t) = g(t)/n(t)$. The observed measurement reliability, represents a maximum likelihood estimate of $R(t)$ given the sample of observations $n(t)$ in the sense that

$$R(t) = \lim_{n(t)\to\infty} \frac{g(t)}{n(t)},$$

or

$$R(t) = E\left[\tilde{R}(t)\right],$$

where the function $E(x)$ represents the statistical expectation value for an argument x.

B.6 Analyzing the Out-of-Tolerance Time Series

Discovering and describing the stochastic process underlying the in-tolerance to out-of-tolerance transition can be thought of as an experiment in which samples are taken of times between calibration paired with calibration results. To provide visibility of the time series, the samples are arranged chronologically. Data can be either measured values (variables data) or observed conditions (in- or out-of-tolerances). The former lead to models of the stochastic process that describe TME parameter value versus time. The latter lead directly to probability models that represent parameter measurement reliability. Nearly all existing calibration recall systems use only attributes data. The treatment in this publication is applicable primarily to attributes data systems. Variables data systems are on tap for future development.

With attributes data systems, the observed time series looks something like Table B.2. Note that the sampled data are grouped in two-week *sampling intervals*, and that these sampling intervals are not spaced regularly apart. This reflects the “take it where you can find it” aspect of gathering data in enough quantity to infer with reasonable confidence the out-of-tolerance stochastic process. Ordinarily, data are too sparse at the individual TME serial-number level to permit this inference. Consequently, serial-number histories are accumulated typically in homogeneous groupings, usually at the manufacturer/model level.

TABLE B.2 Typical Out-of-Tolerance Time Series

TABLE B.2
Typical Out-of-Tolerance Time Series

WEEKS BETWEEN CALIBRATIONS t	NUMBER of CALIBRATIONS RECORDED $n(t)$	NUMBER of IN-TOLERANCES OBSERVED $g(t)$	OBSERVED MEASUREMENT RELIABILITY $R(t)$
2–4	4	4	1.0000
5–7	6	5	0.8333
8–10	14	9	0.6429
11–13	13	8	0.6154
19–21	22	12	0.5455
26–28	49	20	0.4082
37–40	18	9	0.5000
48–51	6	2	0.3333

Note that for many TME management programs, the conditions "in-tolerance" and "out-of-tolerance" are applied at the instrument instead of the parameter level. Although this leads to less accurate calibration interval determinations than can be obtained by tracking at the parameter level, the practice is still workable. The observed time series is constructed the same way, despite the level of refinement of data collection. A plot of the observed time series of Table B.2 is shown in Figure B.2.

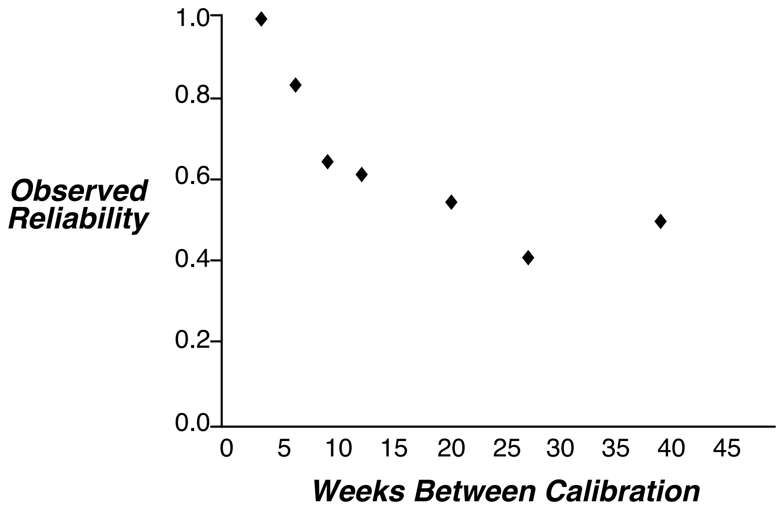


FIGURE B.2 — HYPOTHETICAL OBSERVED TIME SERIES.
The observed measurement reliabilities for the time series tabulated in Table B.2.

To analyze the time series, a model is assumed for the stochastic process. The model is a mathematical function characterized by coefficients. The functional form is specified while the coefficients are estimated based on the observed time series $\{R(t), t \in T\}$. The problem of determining the probability law for the stochastic process becomes the problem of selecting the correct functional form for the time series and estimating its coefficients.

The method used to estimate the coefficients involves choosing a functional form that yields a probability law enabling meaningful predictions of measurement reliability as a function of time. By its nature, the probability law cannot precisely predict the times at which transitions to out-of-tolerance happen. Instead, the probability law predicts measurement reliability expectation values, given the times since calibration. The analysis tries to find a predictor $\hat{R}(t,\hat{\theta}) = R(t) + \varepsilon$, where the random variable ε satisfies $E(\varepsilon) = 0$. It can be shown that the method of maximum-likelihood parameter estimation provides consistent parameter estimates for such predictors.

B.7 Measurement Reliability Modeling

Whether the application is ensuring measurement integrity for periodically calibrated TME or designing TME to tolerate extended periods between calibration, the uncertainty growth stochastic process is described in terms of mathematical models, characterized by two features: (1) a functional form and (2) a set of numerical coefficients.

Figure B.3 models the time series of Table B.2 with an exponential reliability model characterized by the coefficients $R_0 = 1$ and $\lambda = 0.03$. Determining which mathematical form is proper for a given stochastic process and what values will be assigned to the coefficients are discussed in the following sections.

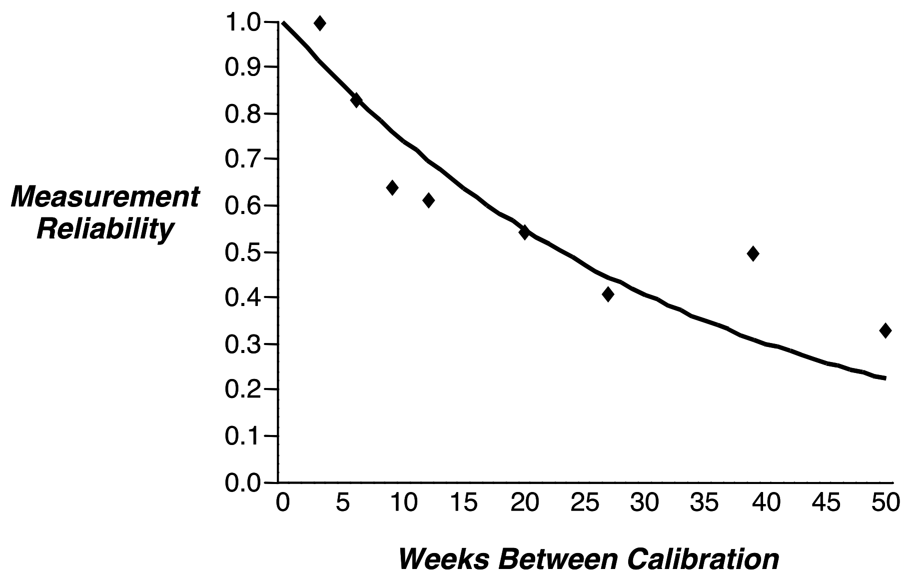


FIGURE B.3 — OUT-OF-TOLERANCE STOCHASTIC PROCESS MODEL.
The stochastic process underlying the time series is modeled by an exponential function of the form $R(t) = R_0 e^{-\lambda t}$.

B.7.1 The Likelihood Function

Maximum-likelihood coefficient estimation for measurement reliability modeling is somewhat different from coefficient estimation used in "classical" reliability modeling. In the latter, each item in a sample from a population of items is monitored at specified intervals, which are spaced closely enough to enable the detection and recording of accurate times to failure. These failure times are inserted into a likelihood function incorporating the probability density function of the model of the failure-time distribution given by

$$f(t,\hat{\theta}) = -\frac{1}{\hat{R}(t,\hat{\theta})}\frac{d\hat{R}(t,\hat{\theta})}{dt}, \tag{B.1}$$

where $\hat{\theta}$ is a vector whose components are the coefficients used to characterize the reliability model. To construct the likelihood function, let the observed times to failure be labeled t_i, $i = 1,2,3,\cdots, m$, and let the times for which sample members were observed to be operational and in-tolerance be labeled t_j, $j = m + 1, m + 2, \cdots, n$. Then the likelihood function is given by

$$L = \prod_{i=1}^{m} \hat{f}(t_i,\hat{\theta}) \prod_{j=m+1}^{n} \hat{R}(t_j,\hat{\theta}). \tag{B.2}$$

Using Eq. (B.2), the coefficients of the model are obtained by differentiating the natural log of L with respect to each component of $\hat{\theta}$, setting the derivatives equal to zero, and solving for the component values.

In measurement reliability modeling, constructing a likelihood function using recorded failure times is not feasible. This is because "failures" are defined as out-of-tolerance conditions whose precise, actual times of occurrence are undetected and unrecorded. At first glance, the fact that the failure times are unknown might seem to be an insurmountable obstacle. However, owing to the binary character of the dependent variable, the in- or out-of-tolerance observations on each instrument serviced constitute independent Bernoulli trials. This fact suggests a procedure for development of the likelihood function.

First, subdivide the domain of observations on the instrument type under study into sampling intervals so each sampling interval contains some minimum number of observations. Let n be the total number of observations and let k, n_i, and b_i denote sampling intervals, the sample size of the ith sample, and failures observed in the ith sample, $i = 1, 2, \cdots, k$. Let t_i represent the interval (time) corresponding to the ith sampling interval, and let $P(t_i)$ be the probability that an out-of-tolerance will have happened by time t_i. The reliability at time t_i is $R(t_i) = 1 - P(t_i)$. Let y_{ij} be the jth observation for the ith sample of size n_i, such that $y_{ij} = 1$ for an observed in-tolerance and $y_{ij} = 0$ for an observed out-of-tolerance. Using the density function for Bernoulli trials, the likelihood function for the ith sample is written

$$L_i = \prod_{j=1}^{n_i} R(t_i)^{y_{ij}} \left[1 - R(t_i)\right]^{1-y_{ij}} \quad . \tag{B.3}$$

Maximizing this function with respect to $R(t_i)$ yields the maximum-likelihood binomial estimate for the sample in-tolerance probability:

$$\tilde{R}_i \equiv \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad . \tag{B.4a}$$

The number that are observed in-tolerance for the ith sample, g_i, is given by

$$g_i = \sum_{j=1}^{n_i} y_{ij} \ , \tag{B.4b}$$

that yields, after combining with Eq. (B.4a),

$$\tilde{R}_i = g_i / n_i \ . \tag{B.4c}$$

The estimates $\tilde{R}_i$, $i = 1,2,3, \cdots , k$ are binomially distributed random variables with means $R(t_i)$ and variances $R(t_i)[1 - R(t_i)]/n_i$.

Having identified the distribution of the observed variables, the probability law of the stochastic process $\{\tilde{R}(t), t \in \mathrm{T}\}$ can be determined by maximizing the likelihood function

$$L = \prod_{i=1}^{k} \frac{n_i\,!}{g_i\,!(n_i - g_i)!} \hat{R}(t_i, \hat{\theta})^{g_i} [1 - \hat{R}(t_i, \hat{\theta})]^{n_i - g_i} \tag{B.5}$$

B.7.2 Steepest Descent Solutions

For measurement reliability modeling, the functional forms are usually nonlinear with respect to the coefficients that characterize them. Consequently, closed form solutions for the components of $\hat{\theta}$ are not obtainable in general, and iterative techniques are used. To introduce these techniques, a simplified method is discussed. Practitioners of numerical modeling will recognize the method as a variation of the method of steepest descent.

B.7.2.1 The Normal Equations

If the theoretical reliability model $\hat{R}(t,\hat{\theta})$ is characterized by an *m*-component coefficient vector, then maximizing log(*L*) in Eq. (B.5) leads to *m* simultaneous equations

$$\sum_{i=1}^{k} \frac{n_i[\tilde{R}_i - \hat{R}(t_i,\hat{\theta})]}{\hat{R}(t_i,\hat{\theta})[1-\hat{R}(t_i,\hat{\theta})]}\left(\frac{\partial \hat{R}(t_i,\hat{\theta})}{\partial \theta_\nu}\right) = 0, \quad \nu = 1,2,3,\cdots,m, \tag{B.6}$$

which are nonlinear in the coefficients. These *m* simultaneous equations are solved for $\hat{\theta}$ using an iterative process.

B.7.2.2 The Iterative Process

As indicated above, iterative methods are used to solve for the vector $\hat{\theta}$. The method of steepest descent begins by "linearizing" the nonlinear model $\hat{R}(t,\hat{\theta})$. This linearization is accomplished by expanding $\hat{R}(t,\hat{\theta})$ in a first order Taylor series approximation at each iteration:

$$\hat{R}(t_i,\hat{\theta}^{r+1}) = \hat{R}(t_i,\hat{\theta}^{r}) + \sum_{\nu=1}^{m}\left(\frac{\partial \hat{R}(t_i,\hat{\theta})}{\partial \theta_\nu}\right)_{\hat{\theta}=\hat{\theta}_r}(\theta_\nu^{r+1} - \theta_\nu^{r}), \tag{B.7}$$

where *r*+1 and *r* refer to the (*r*+1)th and *r*th iterations. Substitution of Eq. (B.7) in (B.6) gives

$$\sum_{i=1}^{k} W_i^r[\tilde{R}_i - \hat{R}(t_i,\hat{\theta}^r)D_{i\nu}^r = \sum_{i=1}^{k} W_i^r\left(\sum_{\mu=1}^{m} D_{i\mu}^r[\theta_\nu^{r+1} - \theta_\nu^r]\right)D_{i\nu}^r, \quad \nu = 1,2,3,\cdots,m, \tag{B.8}$$

where the quantities W_i^r and $D_{i\nu}^r$ are defined by

$$W_i^r \equiv \frac{n_i}{\hat{R}(t_i,\hat{\theta}^r)[1-\hat{R}(t_i,\hat{\theta}^r)]}, \tag{B.9}$$

and

$$D_{i\nu}^r \equiv \left(\frac{\partial \hat{R}(t_i,\hat{\theta})}{\partial \theta_\nu}\right)_{\hat{\theta}=\hat{\theta}_r}. \tag{B.10}$$

Eqs. (B.8) can be written in matricial form by defining the vectors $\tilde{\mathbf{R}}$, $\hat{\mathbf{R}}^r$ and $\hat{\mathbf{b}}^r$, with components $\tilde{R}_i$, $\hat{R}_i^r = \hat{R}(t_i,\hat{\theta}^r)$, and $b_\nu^r = \theta_\nu^{r+1} - \theta_\nu^r$, respectively, and the matrices **W** and **D** with elements $D_{i\nu}^r$, and $W_{ij}^r = W_i^r\delta_{ij}$:[9]

[9]The symbol δ_{ij} is the Kroenecker delta symbol defined by

$$(\mathbf{D}^r)^T \mathbf{W}^r (\tilde{\mathbf{R}} - \hat{\mathbf{R}}^r) = (\mathbf{D}^r)^T \mathbf{W}^r \mathbf{D}^r \mathbf{b}^r \ , \tag{B.11}$$

where the T superscript indicates transposition. Solving Eq. (B.11) for $\mathbf{b}^r$ gives

$$\begin{aligned} \mathbf{b}^r &= [(\mathbf{D}^r)^T \mathbf{W}^r (\mathbf{D}^r)^T]^{-1} (\mathbf{D}^r)^T \mathbf{W}^r (\tilde{\mathbf{R}} - \hat{\mathbf{R}}^r) \\ &= \hat{\theta}^{r+1} - \hat{\theta}^r \ , \end{aligned}$$

and

$$\hat{\theta}^{r+1} = \hat{\theta}^r + [(\mathbf{D}^r)^T \mathbf{W}^r (\mathbf{D}^r)^T]^{-1} (\mathbf{D}^r)^T \mathbf{W}^r (\tilde{\mathbf{R}} - \hat{\mathbf{R}}^r) \ . \tag{B.12}$$

The iterations begin ($r = 0$) with initial estimates for the coefficient vector components and continue until some desired convergence is reached, i.e., until $\hat{\theta}^{r+1} \cong \hat{\theta}^r$.

If the process converges, the first-order expansion in Eq. (B.7) becomes increasingly appropriate. Problems arise when the process diverges, as will often happen if the first parameter estimates are substantially dissimilar to the maximum-likelihood values. To alleviate such problems, a modification of the steepest descent method described above has been developed by Hartley. This modification is the subject of the next section.

B.7.2.3 Modified Gauss–Newton Iteration Method

The method of getting consistent maximum-likelihood coefficient estimates is a modified Gauss-Newton technique. The approach uses Eq. (B.12) but departs from the method described in the previous section by introducing a convergence coefficient $\lambda \in [0,1]$ as follows:

$$\hat{\theta}^{r+1} = \hat{\theta}^r + \lambda \mathbf{b}^r \ . \tag{B.13}$$

The modified technique uses the integral of Eq. (B.8) with respect to $\hat{\theta}_v^{r+1}$ given by

$$\begin{aligned} Q(t, \hat{\theta}^{r+1}) &= \sum_{i=1}^{k} W_i^r [\tilde{R}_i - \hat{R}(t_i, \hat{\theta})]^2 \\ &= (\tilde{\mathbf{R}} - \hat{\mathbf{R}}^r)^T \mathbf{W} (\tilde{\mathbf{R}} - \hat{\mathbf{R}}^r) \ . \end{aligned} \tag{B.14}$$

The method assumes a parabolic $Q(t, \hat{\theta}^{r+1})$ in the coefficient subspace which comprises the domain corresponding to the local minimum of $Q(t, \hat{\theta}^{r+1})$. Different values of λ are used to search the coefficient space in a grid in an attempt to locate a region which contains this local minimum. Hartley uses the values $\lambda = 0$, 1/2 and 1 to get

$$\lambda_{min} = \frac{1}{2} + \frac{1}{4} \frac{Q(0) - Q(1)}{Q(1) - 2Q(1/2) + Q(0)} \ , \tag{B.15}$$

where

$$Q(\lambda) = Q(t, \hat{\theta}^r + \lambda \mathbf{b}^r) \ . \tag{B.16}$$

Hartley's method works by using the value λ_{min} for λ in Eq. (B.13). Unfortunately, for multiparameter reliability models, Hartley's method as described does not invariably lead to convergence.

To ensure convergence, a stepwise Gauss–Jordan pivot is used. With this technique, λ_{min} is sought in a restricted neighborhood of the coefficient subspace. The restriction comes from user-defined bounds on the components of the coefficient vector. The upshot of the restriction is that pivots that correspond to boundary violations are undone. In this way, if the iteration begins to diverge, the process is partly "reversed" until things are back on track. For a detailed treatment of the technique, the reader is referred to the benchmark article by Jennrich and Sampson (1968).

B.8 Reliability Model Selection

A variety of mathematical reliability models have been identified as useful for modeling the out-of-tolerance process. In instances where the process can be inferred from an engineering analysis of TME design, component stabilities, and user applications, determination of the appropriate reliability model is straightforward. Usually, such analyses are unavailable. In these cases, the appropriate reliability model may be determined by comparing a set of viable "candidate" models against the observed out-of-tolerance time series and choosing the model that best fits the data. Unfortunately, the reliability model selection procedures found in the literature consist primarily of tests of *applicability* instead of correctness. Moreover, such tests usually are applied to the coefficient vector instead of the model itself. These tests are useful only if the model is correct in the first place.

The recommended method is one that tries to test for correctness of the model. The method is based on the practice of determining whether $\hat{R}(t,\hat{\theta})$ follows the observed data well enough to be useful as a predictive tool.

The subject of stochastic model evaluation is an area of current research. Some promising variations of the use of the Wald statistic have recently come to light. Adaptation of these to the problem at hand may happen within the next few years. If so, it may be wise to consider replacing the evaluation tools discussed below. These tools, based on defensible statistical concepts, have been refined as a result of considerable trial and error of a heuristic nature.

B.8.1 Reliability Model Confidence Testing

The recommended test of $\hat{R}(t,\hat{\theta})$ is a confidence test constructed using statistical machinery developed for treating $N(\mu,\sigma^2)$ random variables. The validity of this approach derives from the approximately similar statistical properties of binomial and normal distributions.

The test compares the error which arises from the disagreement between $\hat{R}(t,\hat{\theta})$ and $\tilde{R}(t_i)$, i = 1,2,3, ··· , k, referred to as the "lack of fit" error, and the error due to the inherent scatter of the observed data around the sampled points, referred to as the "pure error."

Pure error will be considered first. Returning to the Bernoulli variables defined earlier, the dispersion for the *ith* sampling interval is given by $\Sigma(y_{ij}-\tilde{R}_i)^2$, i = 1,2,3, ··· , k. The total dispersion of the observed data, referred to as the *pure error sum of squares* (*ESS*) is accordingly given by

$$ESS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\tilde{R}_i)^2 \,. \tag{B.17}$$

Since $y_{ij}^2 = y_{ij}$, and $\sum_j y_{ij} = n_i \tilde{R}_i$, Eq. (17) can be written

$$ESS = \sum_{i=1}^{k} n_i \tilde{R}_i (1 - \tilde{R}_i)\,. \tag{B.18}$$

ESS has *n-k* degrees of freedom, where $n = \Sigma n_i$. Thus the *pure error*, denoted by s_E^2, is estimated by

$$s_E^2 = \frac{1}{n-k} \sum_{i=1}^{k} n_i \tilde{R}_i (1 - \tilde{R}_i)\,. \tag{B.19}$$

The estimate s_E^2 is a random variable which, when multiplied by its degrees of freedom, behaves approximately like a χ^2 random variable.

The dispersion of the model is given by the *residual sum of squares*

$$RSS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \hat{R}_i)^2\,, \tag{B.20}$$

which can be written as

$$RSS = \sum_{i=1}^{k} n_i [(\tilde{R}_i - \hat{R}_i)^2 + \tilde{R}_i (1 - \tilde{R}_i)]\,. \tag{B.21}$$

RSS, which has *n-m* degrees of freedom, contains the dispersion due to lack of fit, together with the pure error.

The dispersion due to lack of fit, referred to as the *lack of fit sum of squares* (*LSS*) is obtained by subtracting *ESS* from *RSS*. From Eqs. (B.18) and (B.21), we have

$$LSS = RSS - ESS = \sum_{i=1}^{k} n_i (\tilde{R}_i - \hat{R}_i)^2\,. \tag{B.22}$$

LSS has $(n - m) - (n - k) = k - m$ degrees of freedom, and the error due to lack of fit, is given by

$$s_L^2 = \frac{1}{k-m} \sum_{i=1}^{k} n_i (\tilde{R}_i - \hat{R}_i)^2\,. \tag{B.23}$$

The variable s_L^2, when multiplied by its degrees of freedom, follows an approximate χ^2 distribution. This fact, together with the χ^2 nature of $(n - k)\, s_E^2$, and the fact that s_E^2 and s_L^2 are independently distributed, means that the random variable $F = s_L^2 / s_E^2$, follows an approximate F-distribution with $\nu_i = k - m$ and $\nu_2 = n - k$ degrees of freedom.

If the lack of fit is large relative to the inherent scatter in the data (i.e., if s_L^2 is large relative to s_E^2), then the model is considered inappropriate. Since an increased s_L^2 relative to s_E^2 results in an increased value for *F*, the variable *F* provides a measure of the appropriateness of the reliability model. Thus the model can be rejected on the basis of an F-test to determine whether the computed *F* exceeds some critical value, corresponding to a predetermined rejection confidence level, e.g., 0.95.

B.8.2 Model Selection Criteria

Statistical Criterion

Once the rejection confidence levels for the trial failure models are computed, it remains to select the one which best describes the stochastic process $\{R(t), t \in T\}$. At first, it might be reasonable to suppose that the best model in this regard would be the one with the lowest rejection confidence. However, while rejection confidence should certainly be an important factor in the selection process, there are other considerations. One such consideration is the interval recommended by a given model, that is, the interval whose predicted reliability equals the target reliability.

Economic Criterion

For example, suppose two models have nearly equal rejection confidences but one yields an interval several times longer than the interval recommended by the other. The question in this instance is: How does one choose between two, apparently equally "good," reliability models with markedly dissimilar behavior? Unless the TME whose reliability is being modeled supports a critical end item application, economic considerations dictate that the model corresponding to the longest interval should be selected.

While an economic criterion in conjunction with a rejection confidence criterion may be viewed as an improvement over using a rejection criterion alone, there still lingers a suspicion that perhaps some additional criteria be considered. This arises from the fact that, in the above example, for instance, two seemingly appropriate models yield very different reliability predictions. If this is the case, which one is *really* the correct model? For that matter, is *either* one the correct model?

"Democratic" Criterion

One way out of the dilemma is to resolve the issue democratically by having each candidate model "vote" for its choice of a recommended interval. In this approach, the intervals recommended by the candidate models are grouped according to similarity. Intervals belonging to the largest group tend to be regarded more favorably than others. This tendency stems from a presumed belief that, given an infinite number of "wrong" solutions, agreement among intervals is not likely to be accidental. This belief has been corroborated in simulation studies (unpublished).

Model Figure of Merit

So, there are three criteria for reliability model selection. Using these criteria, a figure of merit *G* is computed for each trial reliability model:

$$G = \frac{N_G}{C} t_R^{1/4} \tag{B.24}$$

where C is the rejection confidence for the model, N_G is the size of the group that the model belongs to and t_R is obtained from

$$\hat{R}(t_R, \hat{\theta}) = 1 - R^* , \tag{B.25}$$

where R^* is the reliability target.

The figure of merit in Eq. (B.24) is not derived from any established decision theory paradigms. Instead, it has emerged from experimentation with actual cases and is recommended for implementation on the basis that it yields decisions which are in good agreement with decisions made by expert analysts.

B.8.3 Variance in the Reliability Model

In many applications (e.g., dog or gem identification), the variance of $\hat{R}(t,\hat{\theta})$ for any given t is a useful statistic. This variance may be computed in a manner similar to that employed in linear regression analysis by imagining that the coefficient vector of the next-to-last iteration is a fixed quantity, independent of the k-tuple of the time series $\{R(t), t \in T\}$, but still very close to the final coefficient vector. While this construct may seem arbitrary, it leads to results which are at least qualitatively valid.

Extension of linear regression methods to the nonlinear maximum likelihood estimation problem at hand gives the variance-covariance matrix for the model coefficient vector **b** as

$$\mathbf{V}(\mathbf{b}^r) = [(\mathbf{D}^r)^T \mathbf{W}^r \mathbf{D}^r]^{-1} . \tag{B.26}$$

Then, defining a vector **d** with components

$$d_\nu(t,\hat{\theta}) = \left(\frac{\partial \hat{R}(t,\theta)}{\partial \theta_\nu}\right)_{\hat{\theta}=\hat{\theta}^r}, \quad \nu = 1, 2, 3, \cdots, m . \tag{B.27}$$

permits the variance in $\hat{R}(t,\hat{\theta})$ for any t to be written

$$\mathrm{var}[\hat{R}(t,\hat{\theta}^{r+1})] = \mathbf{d}^T(t,\hat{\theta}^r)[(\mathbf{D}^r)^T \mathbf{W}^r \mathbf{D}^r]^{-1} \mathbf{d}(t,\hat{\theta}^r) . \tag{B.28}$$

For a converging process, the coefficient vector corresponding to the next-to-last iteration is nearly equal to that of the final iteration, and the two can be used interchangeably with little difficulty. Thus, letting $\hat{\theta}^f$ denote the final coefficient vector, Eq. (B.28) can be rewritten as

$$\mathrm{var}[\hat{R}(t,\hat{\theta}^f)] = \mathbf{d}^T(t,\hat{\theta}^f)[(\mathbf{D}^f)^T \mathbf{W}^f \mathbf{D}^f]^{-1} \mathbf{d}(t,\hat{\theta}^f) . \tag{B.29}$$

B.9 Measurement Reliability Models

Eight reliability models are proposed for modeling out-of-tolerance stochastic processes. Except for the drift model, all have been found useful in practice. The drift model is included because of its intuitive appeal and because it offers some unique benefits. These will be briefly described following the model listing. Each of the ten proposed models corresponds to a particular out-of-tolerance mechanism. The mechanisms are as follows:

1) Constant out-of-tolerance rate (exponential model).

2) Constant operating period out-of-tolerance rate with a superimposed burn-in or wear-out period (Weibull model).

3) System out-of-tolerances resulting from the failure of one or more components, each characterized by a constant failure rate (mixed exponential model).

4) Out-of-tolerances due to random fluctuations in the TME measurement attribute (random walk model).

5) Out-of-tolerances due to random measurement attribute fluctuations confined to a restricted domain around the nominal or design value of the attribute (restricted random walk model).

6) Out-of-tolerances resulting from an accumulation of stresses occurring at a constant average rate (modified gamma model).

7) Monotonically increasing or decreasing out-of-tolerance rate (mortality drift model).

8) Out-of-tolerances occurring after a specific interval of time (warranty model).

These processes are modeled by the mathematical functions listed below. Derivatives with respect to the coefficients are included for purposes of maximum likelihood estimation [see Eqs. (B.10) and (B.27)].

Exponential Model

$$R(t,\hat{\theta}) = e^{-\hat{\theta}_1 t}$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = -te^{-\hat{\theta}_1 t}$$

Weibull Model

$$R(t,\hat{\theta}) = e^{-(\hat{\theta}_1 t)^{\hat{\theta}_2}}$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = -\hat{\theta}_2 t(\hat{\theta}_1 t)^{\hat{\theta}_2 - 1} e^{-(\theta_1 t)^{\theta_2}}$$

$$\frac{\partial R}{\partial \hat{\theta}_2} = (\hat{\theta}_1 t)^{\hat{\theta}_2} \log(\hat{\theta}_1 t) e^{-(\theta_1 t)^{\theta_2}}$$

Mixed Exponential Model

$$R(t,\hat{\theta}) = (1+\hat{\theta}_1 t)^{-\hat{\theta}_2}$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = -\hat{\theta}_2 t(1+\hat{\theta}_1 t)^{-\hat{\theta}_2 - 1}$$

$$\frac{\partial R}{\partial \hat{\theta}_2} = -\log(1+\hat{\theta}_1 t)(1+\hat{\theta}_1 t)^{-\hat{\theta}_2}$$

Random Walk Model

$$Q(\hat{\theta}) \equiv \frac{\hat{\theta}_1}{\sqrt{\hat{\theta}_2 + t}}$$

$$R(t,\hat{\theta}) = \operatorname{erf}(Q)$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = \frac{2}{\sqrt{\pi}} e^{-Q^2} (\hat{\theta}_2 + t)^{-1/2}$$

$$\frac{\partial R}{\partial \hat{\theta}_2} = -\frac{1}{\sqrt{\pi}} e^{-Q^2} \hat{\theta}_1 (\hat{\theta}_2 + t)^{-3/2}$$

Restricted Random Walk Model

$$Q(\hat{\theta}) \equiv \frac{\hat{\theta}_1}{\sqrt{1+\hat{\theta}_2\left(1-e^{-\hat{\theta}_3 t}\right)}}$$

$$R(t,\hat{\theta}) = \operatorname{erf}(Q)$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = \frac{2}{\sqrt{\pi}} e^{-Q^2} \left[1+\hat{\theta}_2\left(1-e^{-\hat{\theta}_3 t}\right)\right]^{-1/2}$$

$$\frac{\partial R}{\partial \hat{\theta}_2} = -\frac{1}{\sqrt{\pi}} e^{-Q^2} \left(1-e^{-\hat{\theta}_3 t}\right)\left[1+\hat{\theta}_2\left(1-e^{-\hat{\theta}_3 t}\right)\right]^{-3/2}$$

$$\frac{\partial R}{\partial \hat{\theta}_3} = -\frac{1}{\sqrt{\pi}} e^{-Q^2} \hat{\theta}_2 t e^{-\hat{\theta}_3 t}\left[1+\hat{\theta}_2\left(1-e^{-\hat{\theta}_3 t}\right)\right]^{-3/2}$$

Modified Gamma Model

$$R(t,\hat{\theta}) = e^{-\hat{\theta}_1 t} \sum_{n=0}^{3} \frac{(\hat{\theta}_1 t)^n}{n!}$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = -t e^{-\hat{\theta}_1 t} \frac{(\hat{\theta}_1 t)^3}{3!}$$

Mortality Drift Model

$$R(t,\hat{\theta}) = e^{-(\hat{\theta}_1 t + \hat{\theta}_2 t^2)}$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = -t e^{-(\hat{\theta}_1 t + \hat{\theta}_2 t^2)}$$

$$\frac{\partial R}{\partial \hat{\theta}_2} = -t^2 e^{-(\hat{\theta}_1 t + \hat{\theta}_2 t^2)}$$

Warranty Model

$$R(t,\hat{\theta}) = \frac{1}{1+e^{\hat{\theta}_2 (t-\hat{\theta}_1)}}$$

$$\frac{\partial R}{\partial \hat{\theta}_1} = \hat{\theta}_2 e^{\hat{\theta}_2(t-\hat{\theta}_1)}\left[1+e^{\hat{\theta}_2(t-\hat{\theta}_1)}\right]^{-2}$$

$$\frac{\partial R}{\partial \hat{\theta}_2} = -(t-\hat{\theta}_1) e^{\hat{\theta}_2(t-\hat{\theta}_1)}\left[1+e^{\hat{\theta}_2(t-\hat{\theta}_1)}\right]^{-2}$$

B.10 Calibration Interval Determination

B.10.1 Interval Computation

Once the failure model is selected, the computation of the calibration interval T, corresponding to the prescribed EOP reliability target R, is obtained from

$$\hat{R}(\mathsf{T},\hat{\theta}) = \mathsf{R} \tag{B.30}$$

The recommended method for obtaining T is one involving a two-step process. First, attempt to solve for T using the Newton-Raphson method. If this fails to converge, then obtain T by trial-and-error in which t is incremented until a value is found for which $\hat{R}(t,\hat{\theta}) > \mathsf{R}$.

B.10.2 Interval Confidence Limits

The upper and lower confidence limits for **T** are computed to show the bounds beyond which the assigned interval becomes questionable. Explicit methods exist for computing these limits for certain specified reliability models (for example, the exponential and Weibull models). However, no general method is available for computing these limits for arbitrary models applied to the analysis of censored data. Since calibration history data are in this category, another approach is called for.

Rather than try to formulate a general method directly applicable to interval confidence-limit determination, an indirect approach will be followed involving the determination of confidence limits for the reliability function $\hat{R}(t,\hat{\theta})$. This enables the determination of upper and lower bounds for **T** closely related to interval confidence limits. Indeed, for single-coefficient reliability functions, these bounds are synonymous with interval confidence limits.

Upper and lower bounds for **T**, denoted τ_u and τ_l, respectively, are computed for $1 - \alpha$ confidence from the relations

$$\hat{R}(\tau_u,\hat{\theta}) + z_\alpha \sqrt{\mathrm{var}\left[\hat{R}(\tau_u,\hat{\theta})\right]} = \mathsf{R} \quad , \tag{B.31}$$

and

$$\hat{R}(\tau_l,\hat{\theta}) - z_\alpha \sqrt{\mathrm{var}\left[\hat{R}(\tau_l,\hat{\theta})\right]} = \mathsf{R} \quad . \tag{B.32}$$

where $\mathrm{var}\left[\hat{R}(t,\hat{\theta})\right]$ is given by Eq. (B.29), and z_α is obtained from

$$1-\alpha = \frac{1}{\sqrt{2\pi}}\int_{-z_\alpha}^{z_\alpha} e^{-\zeta^2/2} d\zeta \qquad \text{(B.33)}$$

Eqs. (B.31) and (B.32) give only approximate upper and lower limits for T in that they are obtained by treating $\hat{R}(t,\hat{\theta})$ as a normally distributed random variable; whereas it, in fact, follows a binomial distribution. The results are satisfactory, however, because the minimum acceptable sample sizes needed to infer the stochastic process are large enough to justify the use of the normal approximation to the binomial.

B.11 Dog/Gem Identification

Two methods for identifying performance outliers and one method for identifying support-cost outliers are discussed in this section. The first performance outlier identification method requires that a prior analysis be performed to ascertain the appropriate reliability model and to estimate its coefficients. Using the results of this analysis, serial-number item dogs and gems are identified and their records are removed from the data. The data are then reanalyzed and a refined set of coefficient estimates is determined. The second performance outlier identification method consists of an a priori identification of TME parameter dogs and gems based on certain summary statistics. Using these statistics, serial-number item dogs and gems are identified and their records are removed from the data before analysis.

The first method is preferred if accurate individual dog/gem calibration intervals are wanted. The second method is preferred if dogs and gems are managed collectively. The second method is much easier to implement and is the recommended method.

B.11.1 Dog/Gem Identification—Method 1

The variance in the model can be used to identify dogs and gems at the TME parameter and TME manufacturer/model levels. The parameter level dogs are identified as follows:

If measurement reliability modeling is performed, the computed variance in the model (see Appendix C) can be used to identify dogs and gems at the TME serial number and TME manufacturer/model levels. Serial number level dogs are identified as follows:

Let $(y_{\mu\nu}, t_{\mu\nu})$, $\nu = 1,2,3, \cdots ,n_\mu$ represent the pairs of observations on the μth serial numbered item of a given manufacturer/model. The variable $t_{\mu\nu}$ is the resubmission time for the νth recorded calibration of the μth item; $y_{\mu\nu} = 0$ for an out-of-tolerance, and $y_{\mu\nu} = 1$ for an in-tolerance. A mean interval and observed reliability are computed according to

$$\langle t_\mu \rangle = \frac{1}{n_\mu}\sum_{\nu=1}^{n_\mu} t_{\mu\nu} \ , \qquad \text{(B.34)}$$

and

$$\tilde{R}_\mu = \frac{1}{n_\mu}\sum_{\nu=1}^{n_\mu} y_{\mu\nu} \ . \qquad \text{(B.35)}$$

A lower confidence limit for the *expected* reliability is computed from

$$\hat{R}_{L\mu} = \hat{R}\left(\langle t_\mu \rangle, \hat{\theta}\right) - z_\alpha \sqrt{\text{var}\left[\hat{R}\left(\langle t_\mu \rangle, \hat{\theta}\right)\right]} \ , \qquad \text{(B.36)}$$

where z_α is obtained from

$$1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-\zeta^2/2} d\zeta \ .$$

An upper 1 - β confidence limit $\tilde{R}_U$ can be obtained for the *observed* reliability from the expression

$$\beta = \sum_{x=0}^{b} \binom{n_\mu}{x} \tilde{R}_U^x \left(1 - \tilde{R}_U\right)^{n_\mu - x} \ , \qquad \text{(B.37)}$$

where $b = n_\mu \tilde{R}_\mu$. The item is identified as a dog with 1 - $\alpha\beta$ confidence if $\tilde{R}_U < \hat{R}_{L\mu}$. Gems are identified in like manner. An upper confidence limit is first determined for the expected reliability:

$$\hat{R}_{U\mu} = \hat{R}\left(\langle t_\mu \rangle, \hat{\theta}\right) + z_\alpha \sqrt{\text{var}\left[\hat{R}\left(\langle t_\mu \rangle, \hat{\theta}\right)\right]} \ , \qquad \text{(B.38)}$$

whereas, for the observed reliability, we have

$$\beta = \sum_{x=n_\mu \tilde{R}_\mu}^{n_\mu} \binom{n_\mu}{x} \tilde{R}_L^x \left(1 - \tilde{R}_L\right)^{n_\mu - x} \ . \qquad \text{(B.39)}$$

The item is identified as a gem with $\alpha\beta$ confidence if $\tilde{R}_L > \hat{R}\left(\langle t_\mu \rangle, \hat{\theta}\right)$.

By following the same treatment with "instrument class" in place of "manufacturer/model" and "manufacturer/model" in place of "item," dogs and gems can be identified at the manufacturer/model level.

B.11.2 Dog/Gem Identification—Method 2

In method 2, a comparison is made between a summary statistic taken on the parameter of a TME unit and a corresponding summary statistic formed from parameter data pooled for the manufacturer/model. Method 2 is applied without prior knowledge of the specific reliability model governing the stochastic process. So, the statistic chosen should be considered a good general standard for comparison. One statistic that meets this requirement is the observed *mean time between failures*. The mean time between failures for the μth item of the TME manufacturer/model is computed as follows:

$$MTBF_\mu = \frac{\langle t_\mu \rangle}{1 - \tilde{R}_\mu} \ , \qquad \text{(B.40)}$$

where $\langle t_\mu \rangle$ and $\tilde{R}_\mu$ are given in Eqs.(B.34) and (B.35).

Letting k represent the number of instruments within the TME manufacturer/model grouping of interest, the aggregate MTBF for the manufacturer/model is given by

$$MTBF = \frac{T}{\tilde{X}} , \tag{B.41}$$

where

$$T = \sum_{\mu=1}^{k} n_\mu \left\langle t_\mu \right\rangle \tag{B.42}$$

and

$$\tilde{X} = \sum_{\mu=1}^{k} n_\mu (1 - \tilde{R}_\mu) \ . \tag{B.43}$$

Dog Identification

The test for identifying a serial number dog involves computing an F-statistic with $2(x_2+1)$ and $2x_1$ degrees of freedom, where x_1 and x_2 are defined by

$$x_1 = \begin{cases} n_\mu (1 - \tilde{R}_\mu), \text{ if } MTBF_\mu < MTBF \\ \tilde{X}, \text{ otherwise}, \end{cases}$$

and

$$x_2 = \begin{cases} \tilde{X}, \text{ if } MTBF_\mu < MTBF \\ n_\mu (1 - \tilde{R}_\mu), \text{ otherwise}. \end{cases}$$

To complete the statistic, total resubmission times T_1 and T_2 are determined according to

$$T_1 = \begin{cases} n_\mu \left\langle t_\mu \right\rangle, \text{ if } MTBF_\mu < MTBF \\ T, \text{ otherwise}, \end{cases}$$

and

$$T_2 = \begin{cases} T, \text{ if } MTBF_\mu < MTBF \\ n_\mu \left\langle t_\mu \right\rangle, \text{ otherwise} , \end{cases}$$

Having determined x_1, x_2, T_1 and T_2, an "observed" F-statistic is computed as

$$\tilde{F} = \frac{x_1}{x_2 + 1} \frac{T_2}{T_1} \ . \tag{B.44}$$

To identify the μth serial number as a dog with 1 - α confidence, this statistic is compared against a characteristic F-statistic obtained from the F distribution:

$$F_c = F_{1-\alpha} \left[2(x_2 + 1), 2x_1 \right] \ . \tag{B.45}$$

If $\tilde{F} > F_c$, the serial number is considered a dog.

Gem Identification

The serial number is considered a gem if

$$\frac{x_2}{x_1+1}\frac{T_1}{T_2} > F_{1-\alpha}\left[2(x_1+1), 2x_2\right] . \tag{B.46}$$

Again, dog and gem identification at the manufacturer/model level is done by substituting "manufacturer/model" for "item" and "instrument class" for "manufacturer/model."

B.11.3 Support-Cost Dog Identification

TME items can be identified as outliers on the basis of excessive calibration support costs. The identification of support cost outliers may assist in decisions regarding corrective administrative or engineering action and/or may supplement the identification of performance outliers.

For purposes of support cost outlier identification, the expectation of the support cost per calibration action for a manufacturer/model is estimated. If the support cost for the jth calibration of the ith instrument is denoted CS_{ij}, then this estimate is given by

$$CS_i = \frac{1}{n_i}\sum_{j=1}^{n_i} CS_{ij} , \tag{B.47}$$

where n_i is the number of calibrations performed on the ith instrument. The corresponding standard deviation is computed in the usual way:

$$s_i = \sqrt{\frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(CS_{ij} - CS_i\right)^2} . \tag{B.48}$$

To identify a given instrument as a support cost outlier, a determination is made as to whether its support cost exceeds the mean support cost for the manufacturer/model to such an extent that its cost can be considered to lie outside the manufacturer/model support cost distribution. This determination is accomplished by first computing the lower support cost confidence limit for the instrument and the upper support cost limit for the instrument's manufacturer/model. These limits are obtained as follows:

A lower 1 - β confidence limit (LCL) for the instrument is given by

$$CS_i^L = CS_i - t_{\beta,\nu_i} s_i / \sqrt{n_i} . \tag{B.49}$$

where $\nu_i = n_i$ - 1. To obtain an upper 1 - α confidence limit (UCL) for the instrument's manufacturer/model, the following quantities are first computed:

$$CS = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} CS_{ij} , \tag{B.50}$$

and

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(CS_{ij} - CS\right)^2} , \tag{B.51}$$

where k is the number of serial numbered instruments within the manufacturer/model, and $n = \Sigma n_i$.

The UCL is computed from

$$CS^U = CS + t_{\alpha,\nu} s / \sqrt{n} \quad , \tag{B.52}$$

where $\nu = n - 1$. If $CS_i^L > CS^U$, the item is identified as a support cost outlier with $1 - \alpha\beta$ confidence.

B.11.4 Suspect Activity Identification

A given TME user's requirements may exert greater stresses on the TME than those exerted by other users. This may have the effect of yielding calibration history data on the equipment which are not representative of the behavior of the equipment under ordinary conditions. Similarly, data recorded by certain calibrating facilities or by a certain calibrating technician may not be representative of mainstream data. Organizations or individuals whose calibration data are outside the mainstream are referred to as **suspect activities**.

For instance, suppose that an activity of interest is a calibrating technician. In this case, we would identify a suspect activity by comparing all calibrations on all TME performed by the technician with all calibrations of these same TME performed by all other technicians. If, on the other hand, the activity of interest is an equipment user, we would compare all calibrations of TME employed by the user of interest against all other calibrations of these TME employed by other users.

High Failure Rate Outliers

Let the set of calibrations corresponding to the activity of interest be designated m and let M label the set of all other activities' calibrations corresponding to these TME . With these identifications, an activity can be identified as suspect through the use of a variation of the **median test** described in many statistics texts. In applying this test, we evaluate whether out-of-tolerance rates (OOTRs) observed from calibrations of TME corresponding to a given activity tend to be significantly greater than OOTRs for these TME taken in aggregate.

A item's OOTR is the inverse of its MTBF:[10]

$$OOTR = \frac{1}{MTBF} \quad . \tag{B.53}$$

The median test procedure is as follows: First, determine the median OOTR for m and M combined (i.e., the set $m \cup M$). Next, define the following

n_m	=	the number of cases in m
n_M	=	the number of cases in M
n_a	=	the total number of cases in $m \cup M$ that lie above the median
n_{ma}	=	the number of cases in m that lie above the median
N	=	$n_m + n_M$.

[10] MTBFs are computed as in dog and gem testing.

Given that, in the sample of size N, the number of OOTRs lying above the median is n_a, the probability of observing an OOTR above the median in the sample is given by

$$p = \frac{n_a}{N} .$$

Regarding the observation of an OOTR above the median as the result of a Bernoulli trial, the probability of observing n OOTRs above the median in a sample of size n_m is given by the binomial distribution:

$$P(n > n_{ma}) = \sum_{n=n_{ma}}^{n_m} \binom{n_m}{n} p^n (1-p)^{n_m - n} .$$

Substituting for p in this expression gives

$$P(n > n_{ma}) = \sum_{n=n_{ma}}^{n_m} \frac{n_m!}{n!(n_m - n)!} \frac{n_a^n}{N^{n_m}} (N - n_a)^{n_m - n} . \tag{B.54a}$$

The median test attempts to evaluate whether this result is inordinately high in a statistical sense. In other words, if the chance of finding n_{ma} or more OOTRs in a sample of size n_m is low, given that the probability for this is n_a/N, then we suspect that the sampled value n_{ma} is not representative of the population, i.e., it is an **outlier**. Specifically, the activity is identified as a suspect activity with 1 - α confidence if the probability of finding n_{ma} or more OOTRs above the median is less than α, i.e., if

$$P(n > n_{ma}) < \alpha . \tag{B.54b}$$

Low Failure Rate Outliers

A low failure rate outlier is one whose OOTR is inordinately low compared to the mainstream. We can easily justify the effort to identify high Failure Rate outliers. High failure rate outliers tend to skew the data in a way that may have a significant impact on interval analysis.

Low failure rate outliers tend to have a lesser impact, because we are usually trying to reach reliability targets higher than 0.5 — often considerably higher. For this reason, the occurrence of false *in*-tolerance observations do not usually increase significantly the already high numbers of in-tolerances we expect to observe. So, why identify low failure rate outliers?

The reason is that, in many cases, a low failure rate is due to unusual usage or handling by an TME user or to a misunderstanding of Condition Received codes by a testing or calibrating technician. These cases need to be identified for equipment management purposes or for personnel training purposes.

Again, let the set of calibrations corresponding to the activity of interest be designated m and let the set of all other activities' calibrations corresponding to these TME be designated M.

Given that, in the sample of size N, the number of OOTRs lying above the median is n_a, the probability of observing an OOTR *below* the median in the set $m \cup M$ is given by

$$p = \frac{N - n_a}{N} .$$

Regarding the observation of an OOTR below the median as the result of a Bernoulli trial, the probability of observing n OOTRs below the median in a sample of size n_m is given by the binomial distribution:

$$P(n > n_m - n_{ma}) = \sum_{n=n_m - n_{ma}}^{n_m} \frac{n_m!}{n!(n_m - n)!} \frac{n_a^{n_m - n}}{N^{n_m}} (N - n_a)^n \quad . \tag{B.55a}$$

The low failure rate median test attempts to evaluate whether this result is inordinately high in a statistical sense. In other words, if the chance of finding n_m - n_{ma} or more OOTRs in a sample of size n_m is low, given that the probability for this is $(N - n_a) / N$, then we suspect that the sampled value n_{ma} is not representative of the population, i.e., it is an outlier. Specifically, the activity is identified as a suspect activity with 1 - α confidence if the probability of finding n_m - n_{ma} or more OOTRs below the median is less than α, i.e., if

$$P(n > n_m - n_{ma}) < \alpha \ . \tag{B.55b}$$

B.12 Data Continuity Evaluation

To evaluate data continuity over the life cycle of a given TME parameter, a calibration history must be maintained. This history should contain information on service dates and calibration results for each parameter calibrated. This information should be recorded each time the calibration history data are incremented for analysis. Total parameter resubmission times and out-of-tolerances are computed according to Eqs. (B.42) and (B.43).

From the resubmission times and out-of-tolerance totals for each parameter, a history of MTBFs is assembled. This history is used to determine MTBF as a function of equipment inventory lifetime. Denoting this lifetime by T, we model MTBF according to

$$\hat{M}(T) = M_0 + \lambda T + \beta T^2 \quad . \tag{B.56}$$

Standard regression methods are used to obtain M_0, λ and β and to determine confidence limits for $\hat{M}(T)$.

The procedure for determining discontinuities in the calibration history data begins with identifying and excluding parameter MTBF values which lie outside statistical confidence limits for $\hat{M}(T)$. Following this weeding out process, M_0, λ and β are recomputed, and a more representative picture of $\hat{M}(T)$ is obtained. Next, the slope of $\hat{M}(T)$, given by

$$m = \frac{\partial \hat{M}}{\partial t} = \lambda + 2\beta t \quad , \tag{B.57}$$

is searched for points (if any) at which $|m| > 0.5$. The latest calendar date for which this occurs is denoted T_c.

Two cases are possible: $m > 0.5$ and $m < -0.5$. For cases where $m < -0.5$, data recorded prior to T_c are excluded from analysis. If $m > 0.5$, reliability estimates R_c and R' are computed according to

$$R_c = \exp\left[-\frac{I}{\hat{M}(T_c)}\right] ,$$

and

$$R' = \exp\left[-\frac{I}{\hat{M}(T')}\right] ,$$

where I is the current assigned interval and T' is the most current date for which calibration history are available. Defining $\Delta R \equiv (R_c - R')/R_c$, a discontinuity in calibration history is identified if

$$\left|\Delta R\right| > D , \tag{B.58}$$

where D is a predetermined coefficient. The value of D is determined in accordance with the amount of data available and the degree of data homogeneity desired. For most cases, $D = 0.2$ has been found useful.

If Eq.(B.58) applies, parameter calibration history data prior to T_c are deleted from records used for interval analysis.

B.13 Data Truncation

Prior to analysis, data are truncated to remove inordinately short and inordinately long resubmission times. These times are recognized as being both uncharacteristic with regard to duration and at odds with reliability expectations. To elaborate, short resubmission times are expected to be associated with high reliability and long resubmission times are expected to be associated with low reliability. Thus short resubmission time samples with inordinately low values of TME observed reliability or long resubmission times with inordinately high values of TME observed reliability are truncated.

A short resubmission time may be defined as one that is less that one quarter of the mode resubmission time, determined in the usual way. A long resubmission time may be defined as one that exceeds twice the mode resubmission time. The sampled TME reliabilities for short resubmission times are considered inordinate if they fall below the $1-\alpha$ lower confidence limit for an *a priori* expected reliability. The sampled long resubmission times are considered inordinate if they exceed the upper $1 - \alpha$ confidence limit for the *a priori* expected TME reliability.

The *a priori* TME reliabilities are determined from a simple straight line fit to the data

$$R_{a\ priori} = a + bt .$$

The straight line fit and the upper and lower confidence limits are determined by regression analysis.

B.14 Calibration Interval Candidate Selection

Analyses of calibration history will be done regularly. It is unreasonable to suppose that enough new information will be accumulated between successive analyses to warrant reevaluation of calibration intervals for each parameter, manufacturer/model, or instrument class in the system history database at each analysis session. This implies that only certain parameters, model numbers, and instrument classes will be singled out for reevaluation at any given analysis run. This results in analysis of only those parameters, models, or classes with nontrivial data increments accumulated since the previous interval assignment or adjustment. This includes all first cases that have accumulated enough data for initial analysis.

In the identification of interval candidates, the following definitions will apply for the parameter or class of interest:

N_{cal} ≡ total number of calibrations accumulated at the date of the previous interval adjustment or assignment.
T ≡ total resubmission time at the date of the previous interval adjustment or assignment.
N_{OOT} ≡ total number of out-of-tolerances accumulated at the date of the previous interval adjustment or assignment.
n_{OOT} ≡ number of out-of-tolerances accumulated since the last interval adjustment or assignment.
n_{cal} ≡ number of calibrations accumulated since the last interval adjustment or assignment.
I ≡ current assigned calibration interval.

Using these quantities, a candidate identification coefficient is determined according to

$$\delta = \frac{n_{cal} I / T - n_{OOT} / N_{OOT}}{1 + n_{OOT} / N_{OOT}} \qquad \text{(B.59)}$$

A parameter, model, or class is identified as a candidate for analysis if either of the following conditions are met

- If $T = 0$ and $N_{cal} + n_{cal} \geq 15$, 25, or 40 at the parameter, model, or class level, respectively.
- If $T \neq 0$ and $|\delta| \geq 0.05$ and $N_{cal} + n_{cal} \geq 15$, 25, or 40 at the parameter, model, or class level, respectively.

B.15 Establishing Measurement Reliability Targets

Establishing measurement reliability targets involves a consideration of several trade-offs between the desirability of controlling measurement uncertainty growth, and the cost associated with maintaining such control. The trade-offs are applicable whether the goal is management of a ground-based calibration interval analysis system or designing TME for spaceflight applications.

In Section B.1, it was shown that establishment of an appropriate measurement reliability target is a multifaceted process. Unfortunately, no handy "rule-of-thumb" guidelines are applicable to the problem. In the last few years, some general precepts have been established that help to identify important factors to consider and how these factors interrelate.

The guiding points in establishing a measurement reliability target are the following:

- TME measurement reliability is a measure of TME parameter uncertainty.
- TME parameter uncertainty is a major contributor to the uncertainty of the end-item test process.
- The uncertainty in the end-item test process affects the uncertainty in the end-item attributes being tested.
- End-item attribute uncertainty affects end-item utility.

Given that the immediate objective of setting a measurement reliability target is the control of test process error, the above list provokes three central questions:

- How much does TME parameter uncertainty contribute to test process uncertainty?
- How sensitive is end-item uncertainty to test process uncertainty?
- How sensitive is end-item utility to end-item uncertainty?

The subject of test process uncertainty is discussed in detail in Sections 5 and 6. Reiterating from these discussions, test process uncertainties emerge from several sources:

- Intrinsic sources inherent in the TME and end-items
- Sensing uncertainties introduced by perturbations to attribute values caused by measurement sensors
- Interface uncertainties arising from random changes in properties of cabling and interconnects
- Sampling uncertainties accompanying analog-to-digital and digital-to-analog conversion processes
- Environmentally induced uncertainties caused by variations in such parameters as temperature, humidity, and electromagnetic fields
- Calibration induced uncertainties
- Other sources, e.g., stresses induced by shipping and handling.

The effect of TME uncertainty on total test process uncertainty can be established by considering end-item attribute value distributions resulting from testing with TME exhibiting maximum uncertainty (the lowest level of TME measurement reliability achievable in practice) and minimum uncertainty (measurement reliability = 1.0). If the range between these extremes is negligible, then TME uncertainty is not a crucial issue and measurement reliability targets can be set at low levels. In certain cases, it may be determined that periodic recalibration of TME is not needed. If end-item uncertainty proves to be a sensitive function of TME uncertainty, however, then the TME measurement reliability target takes on more significance. Under these conditions, a high measurement reliability target may be called for. It should be stressed that not all cases are clear cut. Considerable ambiguity and many gray areas are likely to be encountered in practice.

Maintaining appropriate measurement reliability targets may not always be possible in space-based applications. In these cases, supplemental measures may be necessary. These measures are described in Section 3.4.

For many space-based applications, lengthening the calibration interval of on-board TME is equivalent to designing systems to tolerate low measurement-reliability targets. It is apparent that this can be achieved if the TME system is “overdesigned” relative to what is needed to support end-item tolerances. Such overdesign may involve the incorporation of highly stable components and/or built-in redundancy in measurement subsystems. Sometimes where end-item performance tolerances are at the envelope of high-level measurement capability, it may be necessary to reduce the scope of the end-item’s performance requirements. This alternative may be avoided by using the SMPC measures described in Section 6.4 and Appendix D.

Appendix C TEST AND CALIBRATION HIERARCHY MODELING

C.1 Introduction

Since the 1950s, the need to ensure that measurable parameters of end items are held within specifications has led to the formal institution of test and calibration support infrastructures. Each such infrastructure is characterized by a hierarchy of test and calibration levels. As discussed in Section 6, the integrity of test and calibration hierarchies is maintained by enforcing traceability of measurement accuracy from top to bottom (see Figure C.1).

Although traceability is a vital element in ensuring the integrity of test and calibration hierarchies, enforcement does not ensure that integrity of the traceability will be intact. A second element consists of a body of program and/or process controls that constrain the propagation of measurement uncertainty from level to level to within acceptable limits.

Historically, controlling this “vertical” uncertainty propagation has been achieved by imposing requirements for high ratios of accuracy between hierarchy levels. In recent years, enforcement of such high accuracy ratios has often been difficult or even impossible. Competitive market pressures and stringent government performance objectives for high-tech systems have resulted in end item tolerances that border on the limits of accuracy of even the highest level standards. Managing test and calibration infrastructures within this environment requires the application of analysis tools capable of determining precise accuracy requirements between hierarchy levels. Moreover, such tools must be versatile enough to show conditions where end item performance objectives are not supportable within the framework of existing test and calibration technology. This appendix describes the mathematical concepts on which such tools are built.

C.2 The Test and Calibration Support Hierarchy

Test and calibration infrastructures are characterized by several technical and management parameters. These parameters include calibration system, test system and end item performance tolerances; calibration system and test system calibration intervals; test intervals for fielded end items; accuracy ratios between calibration systems and test systems and between test systems and end items; equipment maintenance and adjustment policies; measurement reliability targets; acceptable false-alarm rates; and missed fault rates.

Individual support scenarios tend to involve unique combinations of end item requirements, test system capabilities, calibration capabilities, test and calibration support budgets, etc. Because of this, each infrastructure is unique. There is no reference set of engineering tables or statistical guidelines by which to configure cost-effective infrastructures. Instead, what is available is a systematic methodology for analyzing support-capability requirements in terms of end item quality and performance objectives. The essentials of the methodology have been incorporated in a user interactive PC-based system called the System for Trade-off Analysis and Reporting (STAR). STAR is maintained by the U.S. Naval Warfare Assessment Center, Code 3121, in Corona, CA. The methodology is presented in this appendix.

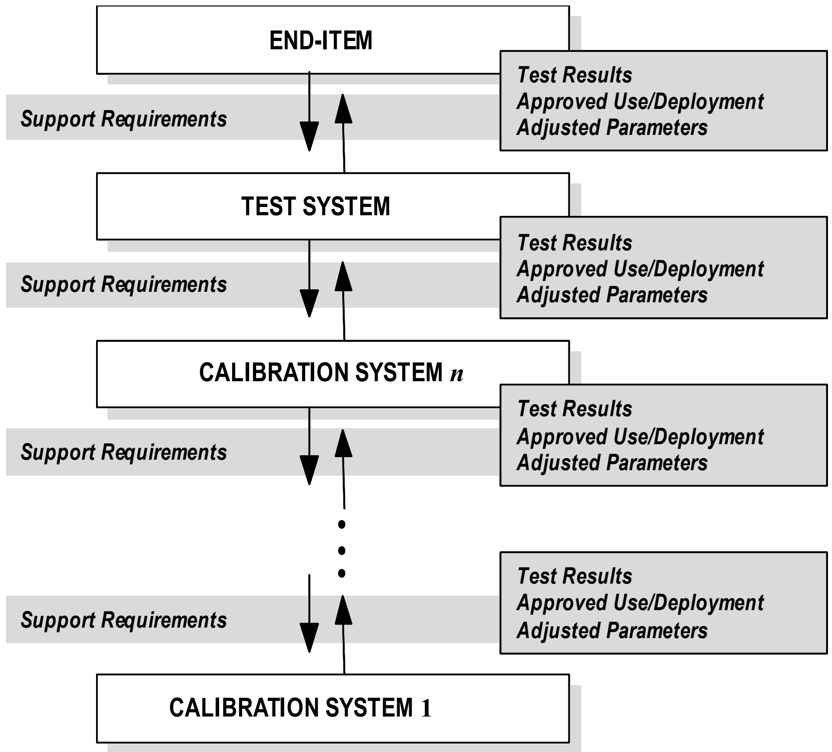


FIGURE C.1 — THE TEST AND CALIBRATION HIERARCHY.

The hierarchy shows the flow of support requirements from the end item level to the primary calibration support level. Immediate end item support requirements are in terms of the maximum uncertainty that can be tolerated during testing. The utility or "quality" of an end item population is affected by this test process uncertainty. Test process uncertainty is in turn affected by the process uncertainty accompanying test system calibration. Also, calibration process uncertainty at each level in the hierarchy is affected by calibration process uncertainty at other levels. In this way, process uncertainties propagate vertically through the hierarchy to affect end item quality.

The methodology links each level of the test and calibration support hierarchy in an integrated model by describing each level of the hierarchy in terms of the support it gives to the next highest level and the support it receives from the next lowest level. For any given level, the support given to the next highest level is measured in terms of several parameters. These are:

- Measurement reliability of the attributes calibrated or tested
- Length of the attributes' test or calibration interval
- Probability of incorrectly reporting out-of-tolerance attributes as in-tolerance
- Probability of incorrectly reporting in-tolerance attributes as out-of-tolerance
- Availability of items tested or calibrated
- Cost of test, calibration, and repair
- Cost of rejection (with consequent adjustment, repair or rework, and downtime) of in-tolerance attributes
- Cost of acceptance of tested/calibrated attributes.

Of these, "cost of acceptance of tested/calibrated attributes" involves a concept developed during RD&E efforts. This and related concepts will be discussed in detail under cost modeling in Section C.8.

The support received from the adjacent level is measured in terms of the parameters

- Measurement reliability of the testing or calibrating attribute
- Availability of supporting items
- Cost of test, calibration, and repair of supporting items.

These parameters connect from one level of the hierarchy to the next in a contiguous sequence. Hence, any change in any of these parameters at any given level affects the parameters at other levels within the hierarchy. This fact makes possible the development of methods and techniques that enable the analysis of costs and benefits. This supplies both summary results for the entire hierarchy and detailed visibility at each level.

A simplified diagram of the test and calibration support hierarchy is shown in Figure C.1. In the hierarchy, the end item is placed at the top of the chain. Below the end item is the test system and below the test system is a series of calibration systems, culminating in a primary calibration system (e.g., NIST), labeled Calibration System 1.

Testing a given end item measurement attribute by a test system yields a reported in– or out-of-tolerance indication (referenced to the end item test tolerance limits), an attribute adjustment (referenced to the end item attribute's adjustment limits), and a "stamp of approval" showing that the end item attribute is approved for use, deployment, distribution, delivery, or sale. Attributes found outside predetermined adjustment limits are adjusted. In organizations where only out-of-spec attributes are adjusted, the adjustment limits are set equal to attribute performance tolerance limits. In organizations where all attributes are adjusted despite their value, the adjustment limits are set equal to zero. Many organizations place adjustment limits between these extremes. The utility or "quality" of the aggregate accepted population of end item attributes can be expressed in terms of the percentage expected to be in conformance with their specifications. This percentage is termed the *beginning-of-period* (BOP) *measurement reliability*. The BOP measurement reliability is referenced to the attribute's performance tolerance limits.

Similarly, the results of calibrating each test system attribute include a reported in- or out-of-tolerance indication (referenced to the test system test limits) and an attribute adjustment (referenced to the appropriate test system adjustment limits), if needed. The same sort of results arise from calibration of the calibration system and accompany calibrations down through the hierarchy to the primary calibration standard.

> Ordinarily, calibration standards are not managed to specified performance or test tolerances and reported as in- or out-of-tolerance, but instead receive a reported measured value, accompanied by confidence limits. Since calibration standards are not managed to specified tolerances, a statement of BOP measurement reliability is seemingly not applicable. Further, the treatment of calibration standards differs from that of calibration or test systems since calibration standards' measurement attribute values are usually reported instead of adjusted.

These observations appear to set the calibration of standards apart from other test or calibration scenarios. With regard to reported attribute values in place of adjustments, however, such reports can be considered to be completely equivalent to nonintrusive adjustments to nominal in that reported values are used as nominal values until the next calibration. Also, the lack of specified tolerances for calibration standards will probably be eliminated in future calibration standard management systems. This is because such standards are assigned calibration intervals, which can be optimized only if specified tolerances accompany reports of calibration. Specifically, a calibration standard attribute's reported measured value must be accompanied by both a set of limits (i.e., performance specifications) expected to contain the attribute value over the duration of its calibration interval and an estimate of the probability that this expectation will be realized (i.e., a measurement reliability target). The methodology presented here assumes this practice will be followed.

It should be noted also that in many applications, end items are not tested at designated periodic intervals. In military weapon system applications, for example, end item testing often happens in response to detected operational failure or may be done before use. In such cases, the end item test interval may be thought of as the *average* time elapsed between tests. In commercial applications, end item testing may take the form of inspection upon receipt of purchased equipment. In these cases, the end item test interval can be regarded as the duration between factory testing and customer testing.

C.3 BOP Measurement Reliability—Test Process Accuracy

From a test/calibration program perspective, it can be assumed, at any two consecutive levels of the test/calibration hierarchy, both the unit under test (UUT) or calibration and the test or calibration system (TME) are drawn randomly from their populations. For discussion purposes, it will also be assumed the UUT and TME attribute values of interest are normally distributed with zero population means (i.e., at any given time, the average value of each population of end item attributes is equal to the attribute's nominal or design value) and with standard deviations (uncertainties) that grow with time passed since prior testing and/or adjustment (see Section 6). UUT attribute adjustments are assumed to be made using testing or calibrating TME attributes as reference values. Attribute values are taken to be toleranced with two-sided performance specifications and to be assigned associated two-sided test tolerance limits and adjustment limits.

If the "true" value of a UUT attribute at the time of test or calibration is represented by x, and its value as measured by the supporting TME is represented by y, then performance, test, and adjustment specifications can be defined as follows:

$L_{per} \le x \le L_{per}$	UUT attribute is in- tolerance
$L_{per} \le y \le L_{per}$	UUT attribute is observed (reported) in- tolerance
$y \le L_{adj}$ or $L_{adj} \le y$	observed value of the UUT attribute is adjusted to center spec using the TME attribute as a reference.

UUT items are assumed to be tested or calibrated at periodic intervals, called test or calibration intervals. The elements associated with calibration intervals are illustrated in Figure C.2. The start of each interval is termed the "beginning of period" (BOP), and the end of each interval is called the "end of period" (EOP). The beginning-of-period starts upon receipt of the UUT by its user, and the end-of-period is marked at the point where the UUT is sent for test or calibration by the user facility. Hence, testing or calibration of UUT items is referenced to the items' EOP. This is in contrast to the times at which TME items are used to test or calibrate UUT items. TME are assumed to be drawn from their populations at random times within their calibration interval. Consequently, the usage of TME attributes is referenced to average-over-period (AOP) times.

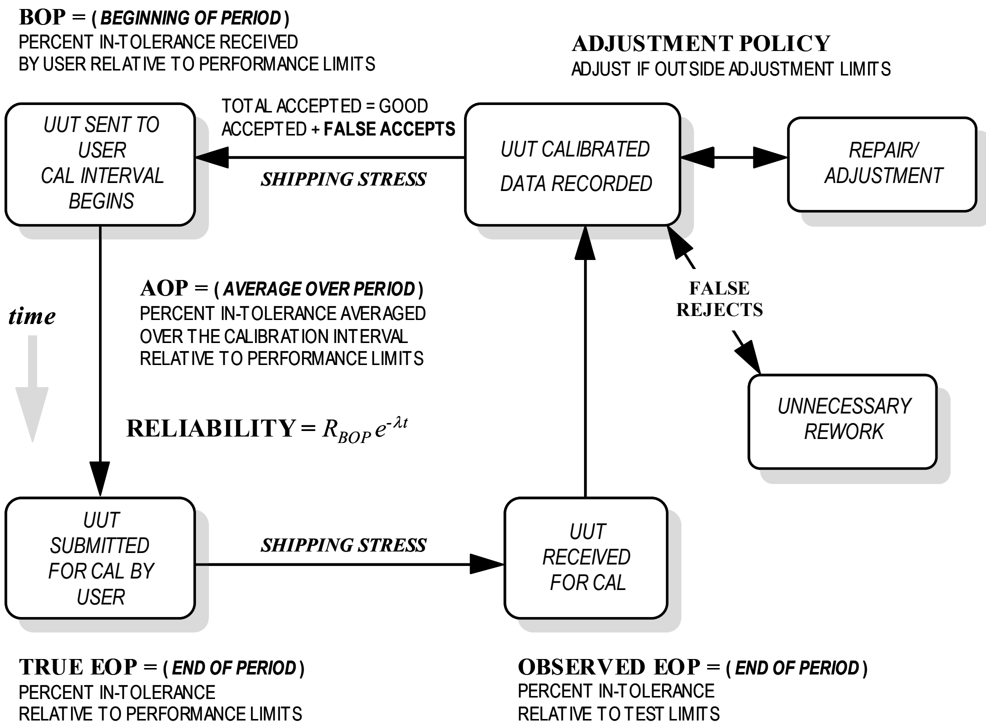


FIGURE C.2 — THE CALIBRATION CYCLE.

The elements of the calibration cycle include test or calibration, usage (the calibration interval), shipping and storage, data recording, and repair or adjustment.

Here, it is worthwhile to note that the test or calibration interval of an item is a quantity that can adopt three identities. From the standpoint of UUT availability to the user, it is the elapsed time between a given BOP date and the successive EOP date. From the standpoint of recall of the UUT for test or calibration, it is the time elapsed between successive BOP dates. From the standpoint of the testing or calibrating facility, it is the time elapsed between successive test or calibration dates. In this appendix, the interval will usually be taken to be synonymous with the time the UUT is available for use. Other segments of the time between calibration dates will be considered in the analysis of equipment availability, discussed later.

The test or calibration process is characterized by several sources of uncertainty, quantified by the following set of standard deviations:

σ_{eop} = the true standard deviation of UUT attribute values after the UUT's usage period (before shipping to the test or calibration facility).

σ_s = the contribution to the UUT standard deviation due to shipping stresses (set to zero if the UUT is not shipped to the test or calibration facility).

σ_{TME} = the true standard deviation of TME attribute values at the time of test or calibration. If random demand of TME items is assumed, this is set equal to the AOP value of the TME attribute standard deviation. Determination of AOP values is discussed later.

σ_{tp} = the standard deviation of the test or calibration process.

As a result of UUT testing or calibration, we "observe" a UUT EOP measurement reliability given by

$$R_{obs} = 2F\left(\frac{L_{test}}{\sigma_{obs}}\right) - 1 \tag{C.1}$$

where $F(\cdot)$ is the cumulative distribution for the normal distribution, and where

$$\sigma_{obs}^2 = \sigma_{eop}^2 + \sigma_s^2 + \sigma_t^2 \tag{C.2}$$

The variance σ_t^2 represents the measurement uncertainty associated with testing or calibration:

$$\sigma_t^2 = \sigma_{TME}^2 + \sigma_{tp}^2 \,. \tag{C.3}$$

The UUT measurement reliability (in-tolerance probability) at EOP is given by

$$R_{eop} = 2F\left(\frac{L_{per}}{\sigma_{eop}}\right) - 1 \,. \tag{C.4}$$

where the quantity σ_{eop}^2 can be obtained from

$$\sigma_{eop}^2 = \sigma_{obs}^2 - \sigma_s^2 - \sigma_t^2 \,. \tag{C.5}$$

The true UUT measurement reliability at time of test or calibration is given by

$$R_{true} = 2F\left(\frac{L_{per}}{\sigma_{true}}\right) - 1 \,, \tag{C.6}$$

where

$$\begin{aligned} \sigma_{true}^2 &= \sigma_{eop}^2 + \sigma_s^2 \\ &= \sigma_{obs}^2 - \sigma_t^2 \,. \end{aligned} \tag{C.7}$$

UUT items are tested to test tolerance limits and adjusted to adjustment limits. Adjustment limits are set in accordance with the policy of the test or calibration facility. There are three main adjustment policy categories:

$L_{adj} = L_{test}$	adjust if "failed" only
$L_{adj} = 0$	adjust always
$0 < L_{adj} < L_{test}$	adjust "as-needed."

UUT attribute adjustment may consist of a physical adjustment or may take the form of a correction factor. Frequently, UUT attribute adjustment to nominal results in placing the attribute value at a quasi-stable point, well within the attribute's tolerance limits. In these cases, an adjust always policy is often preferred. In other instances, adjustment to nominal may lead to resetting the attribute value to an unstable point where the UUT will try to spontaneously revert or "rebound." The latter behavior contributes an additional source of uncertainty characterized by

σ_{rb} = the standard deviation due to reversion or rebound of UUT attributes away from values set as a result of adjustment.

In these cases, an adjust-if-failed-only policy is often the best choice.

Regardless of adjustment policy, UUT items are assumed to be received by the test or calibration facility with attributes distributed according to the pdf

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_{true}} e^{-x^2/2\sigma_{true}^2} . \tag{C.8}$$

It is similarly assumed that UUTs are tested with TME that yield observed attribute values distributed according to

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-(y-x)^2/2\sigma_t^2} . \tag{C.9}$$

As a result of the test or calibration process, UUT items are delivered to the user with a measurement reliability reflecting the quality of the test or calibration process. Generally, the higher the BOP measurement reliability, the longer a UUT item can remain in use before subsequent testing or calibration is required. Consequently, determination of BOP measurement reliability is an important aspect of the uncertainty management process. Therefore, we seek to determine the distribution of UUT attribute values following test or calibration and adjustment. This "post test" distribution is given by

$$f_{pt}(x) = f(x \mid \text{not adjust})P(\text{not adjust}) + f(x \mid \text{adjust})P(\text{adjust}) , \tag{C.10}$$

where the notation $f(x|E)$ indicates the pdf for x, given that an event E has taken place, and $P(E)$ represents the probability that E has occurred.

The first component of the RHS of Eq. (C.10) is obtained using the Bayes' relation

$$f(x \mid \text{not adjust})P(\text{not adjust}) = f(\text{not adjust} \mid x) f(x) . \tag{C.11}$$

The pdf $f(x)$ is given in Eq. (C.8). The pdf $f(\text{not adjust}|x)$ is readily obtained from Eq. (C.9), using the definition of adjustment limits:

$$\begin{aligned} f(\text{not adjust} \mid x) &= \int_{-L_{adj}}^{L_{adj}} f(y \mid x) dy \\ &= F\left(\frac{L_{adj} + x}{\sigma_t}\right) + F\left(\frac{L_{adj} - x}{\sigma_t}\right) - 1 . \end{aligned} \tag{C.12}$$

The pdf $f(x|\text{adjust})$ is given by

$$f(x \mid \text{adjust}) = \frac{1}{\sqrt{2\pi(\sigma_t^2 + \sigma_{rb}^2)}} e^{-x^2/2(\sigma_t^2 + \sigma_{rb}^2)} , \tag{C.13}$$

where rebound from adjustment has been included. The probability P (not adjust) is given by

$$
\begin{aligned}
P(\text{not adjust}) &= \int_{-\infty}^{\infty} dx\, f(x) \int_{-L_{adj}}^{L_{adj}} dy\, f(y\,|\,x) \\
&= 2F\left(\frac{L_{adj}}{\sqrt{\sigma_{true}^2 + \sigma_t^2}} \right) - 1 ,
\end{aligned}
\tag{C.14}
$$

Combining Eqs. (C.11) through (C.14) in Eq. (C.10) gives

$$
f_{pt}(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi(\sigma_t^2 + \sigma_{rb}^2)}} e^{-x^2/2(\sigma_t^2+\sigma_{rb}^2)} , & \text{renew always} \\ \phi(x) \dfrac{e^{-x^2/2\sigma_{true}^2}}{\sqrt{2\pi}\sigma_{true}} + K \dfrac{e^{-x^2/2(\sigma_t^2+\sigma_{rb}^2)}}{\sqrt{2\pi(\sigma_t^2 + \sigma_{rb}^2)}} , & \text{otherwise ,} \end{cases}
\tag{C.15}
$$

where

$$
\phi(x) \equiv F\left(\frac{L_{adj} + x}{\sigma_t} \right) + F\left(\frac{L_{adj} - x}{\sigma_t} \right) - 1 ,
\tag{C.16}
$$

and

$$
K \equiv 2\left[1 - F\left(\frac{L_{adj}}{\sqrt{\sigma_{true}^2 + \sigma_t^2}} \right) \right] .
\tag{C.17}
$$

Since the BOP reliability is referenced to the point of return of the UUT to the user, the effects of shipping must be considered. This is done in accordance with the following assumptions:

(1) Stresses due to shipping occur randomly with respect to magnitude and direction.

(2) Stresses due to shipping occur at some average rate r.

(3) Shipping requires some average duration of time t.

Given these assumptions, responses due to shipping are seen to follow the classic random walk behavior. By letting the variable ζ represent the value of the measurement attribute following shipping, the pdf for ζ can be expressed as

$$
q(\varsigma\,|\,x) \equiv \frac{e^{-(\varsigma - x)^2/2\sigma_s^2}}{\sqrt{2\pi}\sigma_s} ,
\tag{C.18}
$$

where x is the UUT attribute value before shipping, and where

$$
\sigma_s = \sqrt{\langle \zeta^2 \rangle r\tau} .
$$

The BOP measurement reliability is given by

$$
R_{bop} = \int_{-\infty}^{\infty} dx\, f_{pt}(x) \int_{-L_{per}}^{L_{per}} d\varsigma\, q(\varsigma\,|\,x) ,
\tag{C.19}
$$

With adjust-if-failed-only and adjust-as-needed policies, Eq. (C.19) is solved numerically. For the adjust-always policy, Eq. (C.19) can be solved in closed form:

$$R_{bop} = 2F\left(\frac{L_{per}}{\sqrt{\sigma_t^2 + \sigma_{rb}^2 + \sigma_s^2}}\right) - 1 \quad \text{(renew always)}. \tag{C.20}$$

C.4 Interval Adjustment

One of the primary goals of effective uncertainty management is ensuring that TME measurement reliabilities are consistent with end item quality or performance objectives. Such measurement reliabilities, expressed in terms of the probability that a TME attribute is performing within its performance tolerance limits over its test or calibration interval, are typically met by setting test or calibration intervals so a minimum percentage of attributes or items are received in-tolerance for calibration at EOP. These minimum percentages are called EOP *measurement reliability targets*.

For purposes of discussion, it will be assumed that either some level of observed measurement reliability, R_{obs}, or some measurement reliability target, $R*$, is known or projected that corresponds to a test or calibration interval I, which is referenced to a set of tolerance limits, $\pm L_{per}$.

Immediately following test or calibration, the value of an attribute is localized to a neighborhood of values defined by the accuracy of the testing or calibrating TME and the uncertainty of the test or calibration process. As time passes from the point of test or calibration, the UUT experiences various stresses including those from transportation, storage, and use. These stresses contribute to a growing lack of confidence that the neighborhood of values contains the true value of the UUT attribute. This uncertainty growth is depicted in Figure C.3.

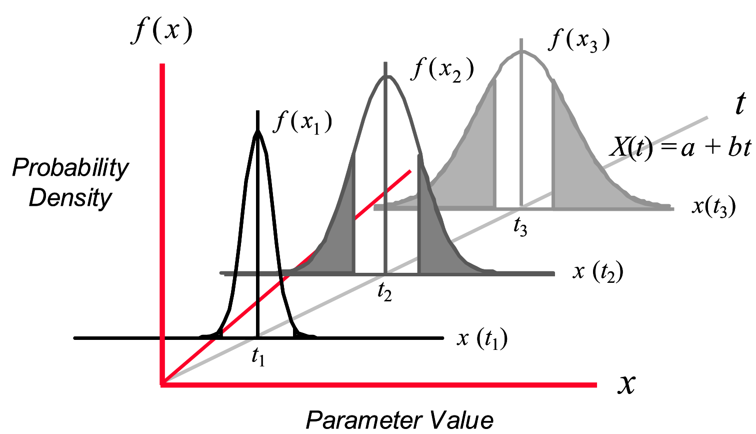


FIGURE C.3 — MEASUREMENT UNCERTAINTY GROWTH.
The shaded areas mark the tolerance limits $\pm L_{per}$. As time passes since calibration or test, the probability that the attribute of interest is out-of-tolerance increases. Thus, measurement reliability shrinks from BOP to EOP.

Let the measurement reliability of an attribute at some time t be denoted $R(t)$ and let the desired EOP measurement reliability target be represented by $R*$. Since test or calibration intervals are set to achieve R_{obs} = R*, any change that effects either a change in R* or in R_{obs} will require a change in the interval I as follows:

$$R* \to R*' \Rightarrow I \to I' : R(I') = R*'$$

or

$$R_{obs} \rightarrow R'_{obs} \Rightarrow I \rightarrow I' : R'_{obs} = R^* .$$

From this simple scheme, it can be seen that an interval change is in order if either the measurement reliability target is changed or if the observed measurement reliability varies. Generally, if the interval I is held constant, the observed measurement reliability of an item of equipment may change if either the item is changed in some physical way or if its in-tolerance and/or maintenance criteria are changed. Physical equipment changes cause a redefinition of the various parameters that govern measurement uncertainty growth over time. Alteration of in-tolerance and/or maintenance criteria are manifested in changes of $\pm L_{per}$, $\pm L_{test}$, and $\pm L_{adj}$.

Interval changes in response to measurement reliability target changes and changes in tolerance limits are discussed below.

C.4.1 Interval Adjustment to Reliability Target Changes

Appendix B describes several mathematical functions used to model attribute measurement reliability. Two of these functions, the *exponential model* and the *random walk model* are used in the present discussion to illustrate the effect of reliability target changes on test or calibration intervals.

Exponential Model

If the measurement reliability of an item is characterized by a constant out-of-tolerance rate, λ, the measurement reliability in effect after an interval I is given by

$$R_{eop} = R_{bop} e^{-\lambda t} , \tag{C.21}$$

from which

$$\lambda = -\frac{1}{I} \ln\left(\frac{R_{eop}}{R_{bop}}\right) . \tag{C.22}$$

Using Eq. (C.4) in (C.22) gives

$$\lambda = -\frac{1}{I} \ln\left[\frac{2F(L_{per} / \sigma_{eop}) - 1}{R_{bop}}\right] , \tag{C.23}$$

where R_{bop} is obtained using Eq. (19) or (20), and σ_{eop} is given in Eq. (5). The quantity σ_{obs} is obtained from Eq. (1):

$$\sigma_{obs} = \frac{L_{test}}{F^{-1}\left[(1 + R_{obs}) / 2\right]} . \tag{C.24}$$

Now suppose that the reliability target is changed to $R^{*\prime}$. A new interval I' is set as follows. As before,

$$\sigma'_{obs} = \frac{L_{test}}{F^{-1}\left[(1 + R^{*\prime}) / 2\right]} , \tag{C.25}$$

and, from Eqs. (21) and (22),

$$R'_{eop} = R'_{bop} \exp\left[\frac{I'}{I}\ln\left(\frac{R_{eop}}{R_{bop}}\right)\right] \\ = 2F\left(\frac{L_{per}}{\sigma'_{eop}}\right) - 1 , \tag{C.26}$$

where

$$(\sigma'_{eop})^2 = (\sigma'_{obs})^2 - \sigma_t^2 - \sigma_s^2 \ , \tag{C.27}$$

and R'_{bop} is as given in Eq. (C.19) or (C.20) with σ'_{true} in place of σ_{true} in Eqs. (C.15) - (C.17). The quantity σ'_{true} is obtained as in Eq. (C.7):

$$(\sigma'_{true})^2 = (\sigma'_{eop})^2 + \sigma_s^2 \ . \tag{C.28}$$

Solving for I' in Eq. (C.26) gives

$$I' = I \frac{\ln\left\{\left[2F(L_{per} / \sigma'_{eop}) - 1\right] \big/ R'_{bop}\right\}}{\ln\left(R_{eop} / R_{bop}\right)} \ , \tag{C.29}$$

with R_{eop} given by Eq. (C.4) and R_{bop} given in Eq. (C.19) or (C.20).

Random Walk Model

With the random walk model, the variance in the attribute value of interest (before shipping) is a linear function of the elapsed interval I:

$$\sigma_{eop}^2 = \sigma_{bop}^2 + \alpha I \\ = \sigma_{true}^2 - \sigma_s^2 , \tag{C.30}$$

where the coefficient α is a constant dependent only on the measurement attribute's inherent stability. Equation (C.30) will be used to determine a new interval I' in response to a reliability target change from R to R'.

The first step is to compute a new value for σ'_{obs} using Eq. (C.25), and σ'_{eop} using Eq. (C.27). Next, R'_{bop} is calculated using Eq. (C.19) or (C.20) with σ'_{true} in Eqs. (C.15) - (C.17). From this, σ'_{bop} is computed according to

$$\sigma'_{bop} = \frac{L_{per}}{F^{-1}\left[(1 + R'_{bop}) / 2\right]} \ . \tag{C.31}$$

Finally, I' is calculated using Eq. (30):

$$I' = \frac{1}{\alpha}\left[(\sigma'_{true})^2 - (\sigma'_{bop})^2 - \sigma_s^2\right] \ . \tag{C.32}$$

C.4.2 Interval Adjustment to Tolerance Limit Changes

An alteration of an attribute's performance limits results in a redefinition of the standard by which the attribute is judged in- or out-of-tolerance. This is shown in Figure C.4.

Such a redefinition results in changes in R_{bop}, $R(t)$, and $R(I) = R_{eop}$. In addition, performance tolerance limit changes are normally accompanied by test tolerance limit changes and adjustment limit changes. The former affects R_{obs} and the latter affects both R_{bop} and support costs in terms of increased or decreased numbers of equipment adjustments performed.

To maintain measurement reliability objectives, such changes need a change in I, resulting in a new interval I' such that

$$R(I') = R(I) = R^* \,, \tag{C.33}$$

where R is referenced to L_{per}, L_{test} and L_{adj}, and R' is referenced to L'_{per}, L'_{test} and L'_{adj}. This change from I to I' is discussed in the next section for attributes governed by the exponential and random walk measurement reliability models.

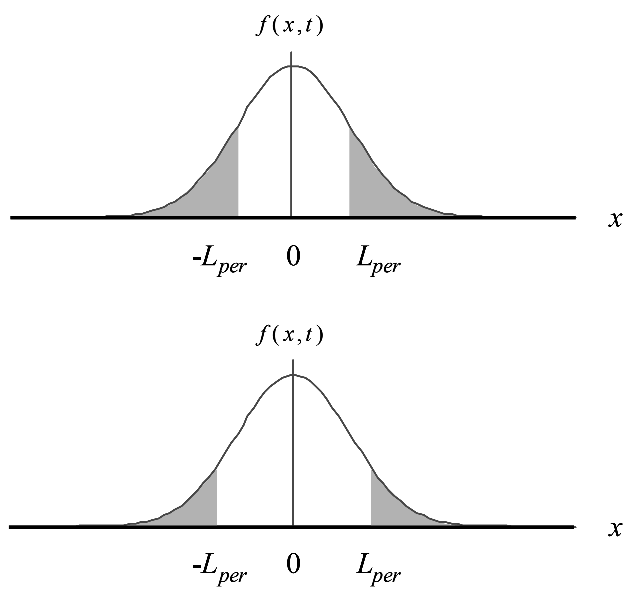


FIGURE C.4 — CHANGE IN OUT-OF-TOLERANCE CRITERIA.
The sum of the shaded areas in the upper figure represents the out-of-tolerance probability for a given distribution of attribute values under the original performance specifications. The sum of the shaded areas in the lower figure represents the out-of-tolerance probability for the same distribution under widened performance specifications. As the figure shows, out-of-tolerance probability is a sensitive function of tolerance limit width.

Exponential Model

If an attribute's performance tolerance is changed, the standard that defines the attribute's measurement reliability is also changed. This results in a change in the attribute out-of-tolerance rate. For the exponential model, the out-of-tolerance rate is given by the parameter λ. Hence, performance tolerance limit changes result in changes in λ and corresponding changes in calibration interval.

To compute a new λ and a new interval I', given a performance limit change from L_{per} to L'_{per}, we imagine the following sequence of events to occur:

(1) The UUT is received for calibration at the end of an interval I. The UUT is tested and/or adjusted using L'_{test} and/or L'_{adj}.

(2) The performance tolerance limits are changed from $\pm L_{per}$ to $\pm L'_{per}$.

(3) The UUT is delivered to the user with the new performance limits.

(4) The UUT is again recalled at the end of I, at which point the measurement reliability is equal to R' (before shipping).

(5) A new λ is calculated.

(6) The test tolerance limits are changed from $\pm L_{test}$ to $\pm L'_{test}$, and the adjustment limits are changed from $\pm L_{adj}$ to $\pm L'_{adj}$ (these changes are optional but normally accompany a performance tolerance change).

(7) A new interval, I', is calculated.

At step 1 above, the observed measurement reliability is given by Eq. (C.1), from which σ_{obs} is computed using Eq. (C.24). Using Eqs. (C.2) and (C.24), a value for σ_{eop} is calculated. In this calculation, the quantities σ_t and σ_s are known.

At steps 2 and 3, the beginning of period measurement reliability is given as in Eq. (C.19) or (C.20)

$$R'_{bop} = \int_{-\infty}^{\infty} dx\, f_{eop}(x) \int_{-L'_{per}}^{L'_{per}} d\varsigma\, q(\varsigma \,|\, x) \ , \tag{C.34}$$

where the pdfs are as defined in (C.15) - (C.18).

At step 4, the measurement reliability is obtained with the aid of Eqs. (C.21) and (C.4):

$$R'_{eop} = R'_{bop} e^{-\lambda' I} = 2F\left(\frac{L'_{per}}{\sigma_{eop}}\right) - 1 \ ,$$

where, at step (5), the new out-of-tolerance rate is given by

$$\lambda' = -\frac{1}{I} \ln\left\{ \left[2F(L'_{per} / \sigma_{eop}) - 1 \right] \Big/ R'_{bop} \right\} \ . \tag{C.35}$$

t step 6, new test tolerance limits and adjustment limits are determined. These changes necessitate calculation of a new beginning of period measurement reliability R''_{bop}. This is accomplished by employing Eq. (C.1) and (C.15) - (C.19) or (C.20) with L'_{per}, L'_{test}, L'_{adj} and R'_{bop} in place of their unprimed counterparts.

At step 7, a new interval is calculated:

$$I' = -\frac{1}{\lambda'} \ln \left\{ \left[2F(L'_{per} / \sigma'_{eop}) - 1 \right] \Big/ R''_{bop} \right\} , \qquad \text{(C.36)}$$

where σ'_{eop} is given in Eq. (C.27).

Since the calibration interval I was presumably managed to achieve a value of R_{obs} equal to the desired target measurement reliability, it is assumed that the observed measurement reliability will be unchanged from its original value. Given this assumption, we have from Eq. (C.24),

$$\sigma'_{obs} = \frac{L'_{test}}{F^{-1}\left[(1+R_{obs})/2\right]} . \qquad \text{(C.37)}$$

Random Walk Model

Unlike the previous calculations for the exponential model, the adjusted new interval I' can be determined for the random walk model by converting to L'_{per}, L'_{test} and L'_{adj} directly. In computing the new interval, σ'_{obs} and σ'_{eop} are computed using Eqs. (C.37) and (C.27), respectively. Next, R'_{bop} is obtained using Eqs. (C.15) - (C.19) or (C.20) with L'_{per}, σ'_{true} and L'_{adj} in place of L_{per}, σ_{true} and L_{adj}. The beginning of period standard deviation is next calculated using

$$\sigma'_{bop} = \frac{L'_{per}}{F^{-1}\left[(1+R'_{bop})/2\right]} ,$$

and I' is obtained using this result with Eqs. (C.27) and (C.37) in Eq. (C.32).

C.5 Measurement Decision Risk

Implied in this treatment is the recognition that, given that test or calibration systems and processes are imperfect, the *true* condition of a UUT attribute may not necessarily match its *apparent* condition observed and recorded as a result of test or calibration. The discrepancy between true condition and observed/reported condition is called *measurement decision risk*. We discuss this risk in terms of *true versus reported measurement reliability* and in terms of the probability for *false alarms* (in-tolerance items reported out-of-tolerance) and *missed faults* (out-of-tolerance items reported in-tolerance).

C.5.1 True Versus Reported Measurement Reliability

The discrepancy between the true EOP measurement reliability of a UUT attribute and its observed/reported measurement reliability is expressed in terms of the discrepancy between the probability that the attribute is truly in-tolerance at the time of test/calibration and the probability that it is *observed* in-tolerance during test/calibration, i.e., the probability that it "passes" test or calibration. These quantities are, respectively, given by

$$\begin{aligned} P(\text{in-tolerance}) &= \int_{-L_{per}}^{L_{per}} f(x)dx \\ &= 2F\left(\frac{L_{per}}{\sigma_{true}}\right) - 1, \end{aligned} \tag{C.38}$$

and

$$\begin{aligned} P(\text{pass}) &= \int_{-\infty}^{\infty} f(x)dx \int_{-L_{test}}^{L_{test}} f(y\,|\,x)dy \\ &= 2F\left(\frac{L_{test}}{\sqrt{\sigma_{true}^2 + \sigma_t^2}}\right) - 1, \end{aligned} \tag{C.39}$$

where $f(x)$ and $f(y|x)$ are given in Eqs. (C.8) and (C.9). From these expressions, it can be readily appreciated that usually a discrepancy exists between the true and observed/reported in-tolerance levels. This discrepancy can be eliminated, however, by adjusting L_{test} according to

$$L_{test} = L_{per}\sqrt{1 + (\sigma_t / \sigma_{true})^2}\ . \tag{C.40}$$

As this expression shows, since uncertainties are present in the test or calibration process (i.e., $\sigma_t >$ 0), the test limits should be placed outside the performance limits if reported in-tolerance levels are to match true measurement reliabilities.

C.5.2 False Alarms/Missed Faults

A false alarm is a case in which an in-tolerance UUT attribute is falsely reported as out-of-tolerance. This can constitute a costly error because such a report may lead to unnecessary rework and/or repair. Moreover, false out-of-tolerances can have a significant effect on calibration or test intervals, particularly if intervals are adjusted to meet high (over 50%) measurement reliability targets. This is because, in these cases, intervals are shortened in response to a reported out-of-tolerance to a greater extent than they are lengthened in response to a reported in-tolerance test or calibration result.

The probability of a false alarm is given by

$$\begin{aligned} P(\text{false alarm}) &= P(|\,x\,| \le L_{per}, |\,y\,| \le L_{test}) \\ &= \int_{-L_{per}}^{L_{per}} f(x)\,dx \int_{L_{test}}^{\infty} f(y\,|\,x)\,dy + \int_{-L_{per}}^{L_{per}} f(x)\,dx \int_{-\infty}^{-L_{test}} f(y\,|\,x)\,dy \\ &= 2 - \frac{1}{\sqrt{2\pi}} \int_{-L_{per}/\sigma_{true}}^{L_{per}/\sigma_{true}} \left[F\left(\frac{L_{test} + \sigma_{true}\xi}{\sigma_t}\right) + F\left(\frac{L_{test} - \sigma_{true}\xi}{\sigma_t}\right) \right] e^{-\xi^2/2}\, d\xi\,. \end{aligned} \tag{C.41}$$

Corresponding to the probability of a false alarm is the probability of a missed fault. From the viewpoint of the UUT user, a missed fault is an attribute returned to the user facility from test or calibration in an out-of-tolerance state. Recalling the earlier discussion on BOP reliability, the probability of this occurrence is given by

$$P(\text{missed fault}) = 1 - R_{bop} \tag{C.42}$$

where R_{bop} is given in Eq. (C.19) or Eq. (C.20).

C.6 Average-Over-Period Reliability

From Eq. (C.42), it can be seen that a viable measure of the quality of the test or calibration process is the UUT BOP reliability. Likewise, from Eq. (C.41), since the probability of a false alarm is a function of σ_{true}, the unnecessary rework cost is seen to be controlled to some extent by the true EOP reliability. While these quantities are of interest, the UUT user is generally more concerned about the measurement reliability of the UUT over the period of use, i.e., over the test or calibration interval. To put this in a somewhat more quantifiable framework, the user is interested in the probability that the UUT attribute will be in-tolerance under the conditions of the demand for its usage. If the usage demand is random, i.e., if the likelihood for use is uniform over the interval, then the appropriate measure of this in-tolerance probability is the attribute's AOP measurement reliability.

AOP measurement reliability is the mathematical average of the measurement reliability from time $t = 0$ to time $t = I$, where the zero point corresponds to R_{bop} and $t = I$ corresponds to R_{eop}:

$$R_{aop} = \frac{1}{I}\int_0^I R(t)dt \,. \tag{C.43}$$

For the exponential model, this is given by

$$\begin{aligned} R_{aop} &= \frac{R_{bop}}{I}\int_0^I e^{-\lambda t}dt \\ &= \frac{R_{bop}}{\lambda I}\left(1 - e^{-\lambda I}\right) \quad \text{(exponential model)} \,. \end{aligned} \tag{C.44}$$

For the random walk model, there are two possibilities. The first covers cases governed by the adjust-always policy ($L_{adj} = 0$) and the second applies to other policies. For adjust-always cases,

$$\begin{aligned} R_{aop} &= \frac{1}{\sqrt{2\pi}I}\int_0^I \frac{dt}{\sigma(t)}\int_{-L_{per}}^{L_{per}} dx\, e^{-x^2/2\sigma^2(t)} \\ &= \frac{1}{I}\int_0^I \left\{2F\left(L_{per}/\sigma(t)\right) - 1\right\}dt \quad \text{(random walk/renew always)} \,, \end{aligned} \tag{C.45}$$

where, from Eqs. (C.30) and (C.20),

$$\sigma^2(t) = \sigma_t^2 + \sigma_{rb}^2 + \sigma_s^2 + \alpha t \,. \tag{C.46}$$

For cases where $L_{adj} \neq 0$, setting $I = 0$ in Eq. (C.45) will not return R_{bop} as expressed in Eq. (C.19). This is because, if only a portion of the UUT population is adjusted using the test system, the resulting distribution of UUT attribute values is not strictly Gaussian. For these cases, numerical Monte Carlo or Markov process techniques are needed to evaluate R_{aop} precisely. Unfortunately, use of these methods is somewhat unwieldy. Experience with several simulated examples, however, has shown that a simplification is possible. This simplification consists of getting an approximate AOP value for $\sigma(t)$, called σ_{aop}, and plugging this quantity into the appropriate expression for $R(t)$ to get R_{aop}. Not only is this approximation useful for the $L_{adj} \neq 0$ case, but it also works well for the adjust-always case.

Determination of σ_{aop} begins with getting an approximate value for σ_{bop}. This is given by

$$\sigma_{bop} \cong \frac{L_{per}}{F^{-1}\left[(1+R_{bop})/2\right]}, \tag{C.47}$$

where R_{bop} is given in Eq. (C.19) for $L_{adj} \neq 0$ adjustment policies and in Eq. (C.20) for the $L_{adj} = 0$ adjustment policy (for which Eq. (C.47) is an exact expression). Working from Eq. (C.30), σ_{aop} can be expressed as

$$\begin{aligned} \sigma_{aop} &= \sqrt{\frac{1}{I}\int_0^I \left(\sigma_{bop}^2 + \alpha t\right) dt} \\ &= \sqrt{\sigma_{bop}^2 + \tfrac{1}{2}\alpha t} \quad \text{(random walk model)}. \end{aligned} \tag{C.48}$$

Note that if the UUT is used as the TME for the next highest level in the test and calibration hierarchy, σ_{aop} is the value used for σ_{TME} in Eq. (C.3). This is because TME items are assumed to be selected and used for UUT test/calibration at random times over their calibration intervals.

For the exponential model, use of Eq. (C.44) gives

$$\begin{aligned} \sigma_{aop} &= \frac{L_{per}}{F^{-1}\left[(1+R_{aop})/2\right]} \\ &= \frac{L_{per}}{F^{-1}\left\{\frac{1}{2}\left[1+\frac{R_{bop}}{\lambda I}\left(1-e^{-\lambda I}\right)\right]\right\}} \quad \text{(exponential model)}, \end{aligned} \tag{C.49}$$

with R_{bop} as given in Eq. (C.19) or Eq. (C.20).

C.7 Availability

The cost of operating a test and calibration program and the cost of maintaining a functioning field capability are affected by the need for equipment spares. Spares costs are minimized by maximizing equipment availability. The availability of an item of UUT is the probability that the item will be available for use over the period of its administrative test or calibration interval. If this interval is thought of as the time elapsed between successive BOP dates, then the availability of an item is given by

$$\text{availability} = \frac{I}{\text{administrative interval}}, \tag{C.50}$$

where I is the "active" portion of the test or calibration interval as defined in Eqs. (C.21) and (C.30). The difference between the administrative interval and the variable I is the downtime:

$$T_d = \text{administrative interval} - I\ . \tag{C.51}$$

For our purposes, the composition of T_d is assumed to be described according to

$$\begin{aligned} T_d = {} & \text{calibration downtime} + \text{adjustment downtime} \times P(\text{adjust}) \\ & + \text{repair downtime} \times P(\text{repair})\ . \end{aligned} \tag{C.52}$$

P(adjust) is given in Eq. (C.14). The probability for repair is the probability that UUT items, submitted for test or calibration, will need repair action besides the various adjustments and corrections that normally accompany test or calibration. As the reader will note, this is a subset of the total repair downtime, which includes downtime resulting from user-detectable functional failures encountered during use. Since the present discussion is concerned primarily with cost and performance as affected by test and calibration, only this subset is of interest in the present context. To focus on this subset of repair actions, we define a parameter L_{rep}, which yields P(repair) according to

$$P(\text{repair}) = 2\left[1 - F\left(\frac{L_{rep}}{\sigma_{true}}\right)\right]. \tag{C.53}$$

The value L_{rep} marks a limiting measurement attribute value, beyond which repair actions are normally required to restore a UUT attribute to its nominal performance value.

The remaining quantities in Eq. (C.52) will now be considered. First, we define the following variables:

T_{cal} = mean time needed for a test or calibration action

T_{css} = mean shipping and storage time experienced between EOP and BOP dates

T_{rep} = mean time needed for a repair action

T_{rss} = mean shipping and storage time experienced between submittal and return of an item of UUT submitted for repair

T_{adj} = mean time needed for a routine adjustment of a UUT measurement attribute

Given these definitions, we have

calibration downtime = $T_{cal} + T_{css}$,
adjustment downtime = T_{adj}

and

repair downtime = $T_{rep} + T_{rss}$.

It is assumed, under ordinary circumstances, that these quantities are known. Substituting these variables in Eq. (C.50) and using Eqs. (C.51) and (C.52) gives

$$
\begin{aligned}
P(\text{available}) &= \frac{I}{I + T_{cal} + T_{css} + T_{adj}P(\text{adjust}) + (T_{rep} + T_{rss})P(\text{repair})} \\
&= \frac{1}{1 + \dfrac{T_{cal} + T_{css} + T_{adj}P(\text{adjust}) + (T_{rep} + T_{rss})P(\text{repair})}{I}} \\
&= \frac{1}{1 + T_d / I}\,.
\end{aligned}
\tag{C.54}
$$

Clearly from Eq. (C.54), availability approaches unity as $I \to \infty$ and/or as $T_d \to 0$. Equation (C.54) also shows that availability improves as P(adjust) and P(repair) are minimized.

C.8 Cost Modeling

Calibration intervals, test decision risks, and availability have a direct bearing on the costs associated with operating and maintaining a test and calibration support hierarchy. These parameters also affect indirect costs associated with end item quality and/or performance capability.

End item quality and/or performance capability is measured in terms of the extent to which an end item achieves a desired effect or avoids an undesired effect. These effects can be referenced to program management considerations, for military or space systems, to end item profitability for a commercial product, or to any measure of end item performance that can be quantified in economic terms. Examples of wanted effects may include the successful strike of an offensive weapon, follow-on reorders of a product item, and creation of a desirable corporate image. Examples of undesired effects may include the unsuccessful response to a military threat, warranty expenses associated with poor product performance, and the return of products rejected by the customer. In each case, the end item experiences an “encounter” (the approach of an intended target, the approach of an incoming missile, and the appearance of an unexpected obstruction) that results in a perceived “outcome” (a successful missile strike, a missile interception, and an obstruction avoidance). The effect is determined by the “response” of the end item to the encounter (timely sighting and ranging, early detection and warning, and responsive braking and maneuvering). The cost of a given outcome is a variable that is assumed to be known. If an outcome is associated with a benefit, the cost is expressed in negative dollars.

The analytical methodology developed here provides a means for determining the probability of a successful or unsuccessful outcome as a function of various technical parameters that characterize the test and calibration support hierarchy. The hierarchy affects costs associated with fielding, selling or otherwise dispatching the supported end item. An end item that has been dispatched has been “accepted” by the end item test system. Therefore, the costs that derive from a dispatched end item are termed “acceptance costs.” The variables resulting from cost modeling and analysis are shown in Table C.1. The variables used in modeling acceptance cost are shown in Table C.2. In this table, total annual calibration, adjustment, repair, and support costs relate to costs incurred from support of a UUT of interest (a calibration system, test system, or end item). “Annual acceptance cost” applies only if the UUT of interest is an end item.

Key to the cost modeling discussed here is the assumption that the quality and performance capability of an end item is related to the value of the measurement attribute supported by test and calibration. Attributes tested before end item dispatch can conceivably be out-of-tolerance to a degree that end item performance will be negatively affected. The variables x_d and x_f mark the onset of degraded attribute performance and the point of complete loss of performance, respectively. To relate end item quality or capability to values between these points, the following model has proved useful in many applications:

$$P(\text{success} \mid x) = \begin{cases} 1, & |x| \le x_d \\ 1 - \sin^2\left[\dfrac{(|x| - x_d)\pi}{2(x_f - x_d)}\right], & x_d \le |x| \le x_f \\ 0, & x_f \le |x|, \end{cases} \tag{C.55}$$

where $P(\text{success}|x)$ is the probability for successful performance of the end item, given that its attribute value is equal to x. The overall probability of a successful outcome is given by

$$P(\text{success}) = P_{sr} \int_{-\infty}^{\infty} f_{aop}(x) P(\text{success} \mid x)\, dx \,. \tag{C.56}$$

The pdf $f_{aop}(x)$ is obtained from Eq. (C.8) with AOP for "true" to show that the end item is used throughout its test interval in agreement with the random demand assumption:

$$f_{aop}(x) = \frac{1}{\sqrt{2\pi}\sigma_{aop}} e^{-x^2/2\sigma_{aop}^2} \,. \tag{C.57}$$

As Eqs. (C.48) and (C.49) show, σ_{aop} depends on σ_{bop} or, equivalently, R_{bop}. These quantities are, in turn, determined by the accuracy of the test system and the quality of the test and calibration support hierarchy.

TABLE C.1 Cost Modeling Variables

Variable Description	Variable Name
End item attribute value corresponding to the onset of degraded performance	x_d
End item attribute value corresponding to loss of function	x_f
Cost of a given outcome	C_f
Quantity of end items sold or in inventory	N_{UUT}
Acquisition cost of an end item unit	C_{UUT}
End item spare coverage desired (in percent)[11]	S_{UUT}
Probability of a successful outcome, given successful end item performance	P_{sr}
Probability of an encounter	P_e

[11] This variable controls the number of spares maintained to cover the end item inventory or the population of end items sold to customers.

Hours to calibrate/test	H_c
Additional hours required for adjustments	H_a
Cost per hour for test/calibration and/or adjustment	C_{hr}
Cost per repair action	C_r

The acceptance cost for dispatched end items is the product of the cost of a given outcome, the number of end items dispatched, the probability of an encounter occurring, and the probability of unsuccessful end item performance:

$$C_{acc} = C_f N_{UUT} P_e [1 - P(\text{success})] , \tag{C.58a}$$

where P(success) is given in Eq. (C.56). If C_{acc} represents a benefit, the appropriate expression is

$$C_{acc} = C_f N_{UUT} P_e P(\text{success}) , \tag{C58.b}$$

where C_f would be given in terms of payoff instead of cost. The quantity C_{acc} can be "annualized" by expressing P_e in terms of the probability of encounter per end item unit per year. Sometimes, it may be desirable to set P_e equal to the probable *number* of encounters experienced per end item unit per year. (The reader may note that this quantity may be a function of N_{UUT}).

TABLE C.2 Acceptance Costs Modeling Variables

Variable Description	Variable Name
Total annual cost	C_{tot}
Annual acceptance cost	C_{acc}
Total annual support cost	C_{ts}
Annual calibration cost	C_{cal}
Annual adjustment cost	C_{adj}
Annual repair cost	C_{rep}
Total spares acquisition cost	C_{sa}

As stated earlier, acceptance cost applies only to the end item. The quantities that follow, however, apply to any UUT encountered at any level of the test and calibration support hierarchy. Of these, we first consider costs associated with UUT downtime. UUT downtime results in a requirement for replacement spares available to cover items submitted for test or calibration.

The number of available UUT spares required is equal to the number needed to cover the unavailable UUT items multiplied by coverage wanted (spares wanted in stock to cover an out-of-use UUT):

$$N_s P(\text{available}) = N_{UUT} [1 - P(\text{available})] S_{UUT} ,$$

$$N_s = \frac{[1 - P(\text{available})]}{P(\text{available})} N_{UUT} S_{UUT} ,$$

which becomes, with the aid of Eq. (C.54),

$$N_s = (T_d / I) N_{UUT} S_{UUT} . \tag{C.59}$$

The cost to buy these spares is given by

$$C_{sa} = N_s C_{UUT} , \tag{C.60}$$

and the annual cost resulting from the requirement for these spares is given by

$$C_s^{year} = C_d C_{sa} , \tag{C.61}$$

where C_d is either the annual depreciation cost per UUT item (for private sector applications) or the unit rate at which UUT items expire from use and need replacement (in government applications).

The annual cost due to calibration or test is given by

$$C_{cal} = H_c C_{hr} (N_{UUT} + N_s) / I , \tag{C.62}$$

where I is expressed in years. The annual cost of UUT adjustments is given by

$$C_{adj} = \frac{(N_{UUT} + N_s)}{I} H_a C_{hr} P(\text{adjust}) , \tag{C.63}$$

and the annual cost of UUT repair is

$$C_{rep} = \frac{(N_{UUT} + N_s)}{I} C_r P(\text{repair}) , \tag{C.64}$$

where P(adjust) is given in Eq. (C.14), P(repair) is given in Eq. (C.53) and, again, I is expressed in years.

The total annual support cost is the sum of Eqs. (C.61), (C.62), (C.63), and (C.64):

$$C_{ts} = C_s^{year} + C_{cal} + C_{adj} + C_{rep} . \tag{C.65}$$

The total annual cost including support and acceptance costs, is given by the sum of Eq. (C.65) and Eq. (C.58):

$$C_{tot} = C_{acc} + C_{ts} \tag{C.66}$$

C.9 Multiple Product Testing

At the end of the test and calibration process lie populations of end items that exhibit various in-tolerance percentages. As the previous sections have shown, these percentages are controlled by the accuracy or "integrity" of the acceptance testing process. Accurate testing yields high end item in-tolerance percentages and low false-alarm and missed-fault rates. In-tolerance percentages, false alarm rates and missed-fault rates are obtained through computation using end item pdfs, as shown earlier.

The pdf $f_{pt}(x)$ of an accepted end item population mathematically characterizes the "quality" or integrity of the population. For example, the in-tolerance percentage of an accepted lot of end items is given by

$$P(\text{in-tolerance}) = \int_{-L_{per}}^{L_{per}} f_{pt}(x)dx .$$

In previous sections, the pdf $f_{pt}(x)$ is given by Eq. (C.10). Equation (C.10) applies to cases in which end items are subjected to a single test process. During end item production, however, end item testing is often performed in a sequence of tests, each characterized by its individual test system and test process uncertainties. The resulting pdf in these scenarios is not Gaussian and Eqs. (C.15) through (C.17) are not applicable.

C.9.1 The General Multiple Testing Model

The typical end item multiple testing scenario uses four stages of end item testing, as shown in Figure C.5. The first stage involves testing at the component level, followed by board level testing at the second stage, package level testing at the third stage, and system level testing at the fourth and final stage. Testing consists of both functional checks to verify that all characteristics of each component, board, package, or system are in working order and tolerance tests to verify that all relevant measurable parameters are within specification. For the present discussion, only tolerance testing will be modeled in what follows.

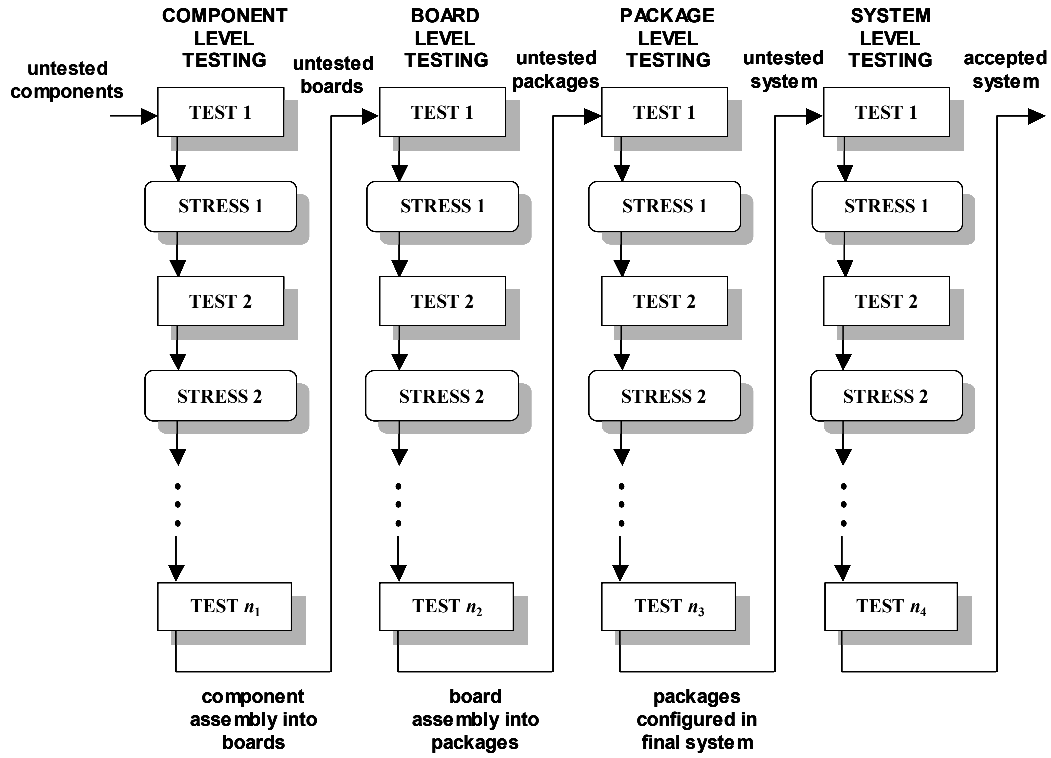


FIGURE C.5 — MULTIPLE END ITEM TESTING.

Testing of each end item function or parameter begins at the most elemental level of development and contains the entire manufacturing process through final assembly. At each stage, UUTs are subjected to stresses designed to ensure that all elements of each end item function are performing as intended and are operating within specified tolerance limits.

In considering the testing process shown in Figure C.5, we try to find the resulting pdf $f(x)$ which, as stated above, characterizes the accepted end item population. Since several levels or stages of acceptance are involved, this pdf evolves from stage to stage. The general model used to describe this evolution is shown in Figure C.6.

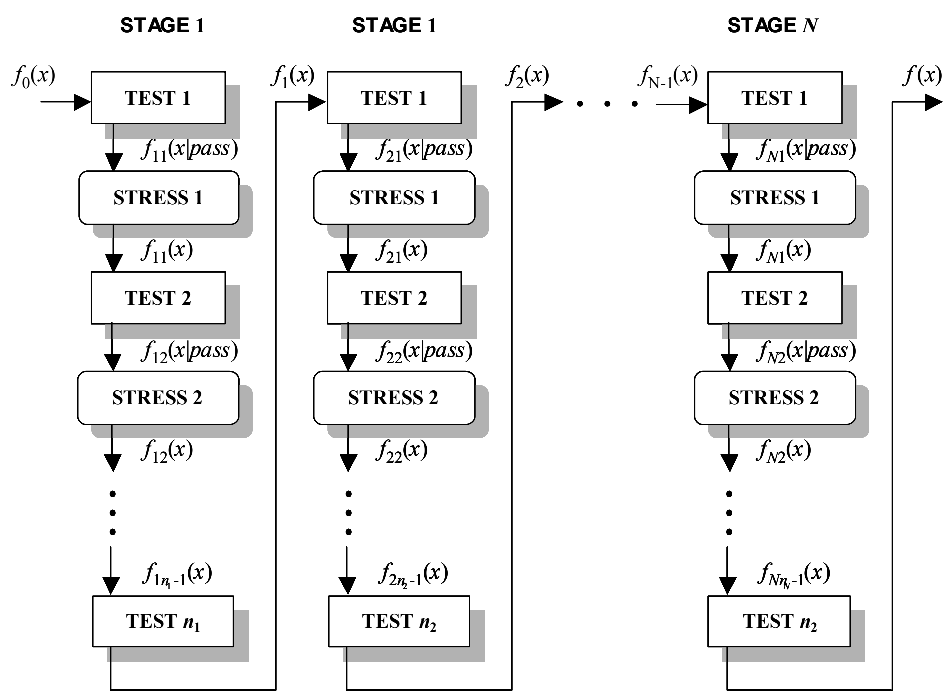


FIGURE C.6 — THE GENERAL MULTIPLE TESTING MODEL.

Items enter the testing process with parameter values distributed according to the pdf $f_0(x)$. As nonconforming items are rejected at each step, parameter values take on distributions described by the pdfs $f_{ij}(x|\text{pass})$. Stress (thermal, vibration, etc.) is encountered by end items within each stage. Following stress, parameters are distributed according to the pdfs $f_i(x)$.

C.9.2 Definitions and Notation

In analyzing the general end item testing model shown in Figure C.6, the following terms will be used in addition to those encountered earlier.

L_{ij}^t = end item parameter test limit for the jth test of the ith testing stage.

A_{per} = the region defining acceptable performance [- L_{per}, L_{per}] for the end item parameter.

A_{ij}^t = the acceptance region $[-L_{ij}^t, L_{ij}^t]$ for the end item parameter for the jth test of the ith testing stage.

L_{ij} = performance tolerance limit for the attribute of the test system selected to perform the jth test of the ith testing stage.

σ_{ij} = standard deviation for the test system and test process present at the jth test of the ith testing stage.

R_0 = in-tolerance probability for the end item attribute before testing.

R_{ij} = in-tolerance probability for the test system parameter used to perform the jth test of the ith testing stage.

I_{ij} = calibration interval for the test system used to perform the jth test of the ith testing stage.

λ_{ij} = the out-of-tolerance rate for the test system parameter used to perform the jth test of the ith testing stage.

$\sigma_{ij}(s)$ = response of the end item attribute to the stress applied to the end item subpopulation passing the jth test of the ith testing stage.

$f_i(x)$ = pdf for the parameter under test for the end item subpopulation that successfully passes the ith testing stage.

C.9.3 Determination of $f(x)$

Let the value of the end item attribute under test at any point in the testing process be represented by the variable x. Test system measurements of x are represented by the variable y. Before testing, the end item parameter is assumed to follow the pdf

$$f_0(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/2\sigma_0^2},$$

where

$$\sigma_0 = \frac{L_{per}}{F^{-1}\left(\frac{1+R_0}{2}\right)}.$$

At the jth test of the ith testing stage, the distribution of test system measurements of x is given by

$$f_{ij}(y\,|\,x) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-(y-x)^2/2\sigma_{ij}^2}.$$

Between tests within each testing stage, the end item is subjected to stress. This stress is assumed to cause x to fluctuate randomly from its prestress value. With this assumption, if x' is the value of the end item prior to stress, and x is the value following stress, the pdf for x following stress is given by:

$$q_{ij}(x\,|\,x') = \frac{1}{\sqrt{2\pi}\sigma_{ij}(s)}\exp\left[-\frac{(x-x')^2}{2\sigma_{ij}^2(s)}\right].$$

As stated earlier, the objective of applying the model is the determination of the final end item pdf. In making this determination, the notation will be simplified by using a quantity $G_{ij}(x)$ defined by

$$G_{ij}(x) = \int_{-L_{ij}^t}^{L_{ij}^t} f_{ij}(y\,|\,x)dy\,.$$

The analysis begins by considering the distribution $f_{ij}(x|\text{pass})$ for the end item subpopulation that passes the first test of the first testing stage. Using standard probability theory notation, the parameter in-tolerance probability for this subpopulation can be written

$$\begin{aligned} R_{11} &= P\left(x \in A_{per}\,|\,y \in A_{11}^t\right) \\ &= \int_{-L_{per}}^{L_{per}} f_{11}\left(x\,|\,\text{pass}\right)dx\,. \end{aligned} \tag{C.67}$$

Invoking Bayes' first relation, the quantity $P\left(x \in A_{per}\,|\,y \in A_{11}^t\right)$ is obtained from

$$P\left(x \in A_{per}\,|\,y \in A_{11}^t\right) = \frac{P\left(x \in A_{per}, y \in A_{11}^t\right)}{P\left(y \in A_{11}^t\right)}\,. \tag{C.68}$$

The numerator and denominator of Eq. (C.68) are given by

$$\begin{aligned} P\left(x \in A_{per}\,|\,y \in A_{11}^t\right) &= \int_{A_{per}} dx \int_{A_{11}^t} dy\, f_{11}(x,y) \\ &= \int_{A_{per}} f_0(x)dx \int_{A_{11}^t} f_{11}(y\,|\,x)dy \\ &= \int_{A_{per}} f_0(x)G_{11}(x)dx\,, \end{aligned}$$

and

$$P\left(y \in A_{11}^t\right) = \int_{-\infty}^{\infty} f_0(x)G_{11}(x)dx\,.$$

Substituting these expressions in Eq. (C.68) gives

$$P\left(x \in A_{per} \mid y \in A_{11}^{t}\right) = \frac{\int_{A_{per}} f_0(x) G_{11}(x) dx}{\int_{-\infty}^{\infty} f_0(x) G_{11}(x) dx}.$$

Comparison of this result with Eq. (C.67) shows that the pdf for x following the first test is given by

$$f_{11}\left(x \mid \text{pass}\right) = f_0(x) \frac{G_{11}(x)}{\int_{-\infty}^{\infty} f_0(\zeta) G_{11}(\zeta) d\zeta}. \tag{C.69}$$

Next, the first stress is applied. The resulting pdf is obtained from

$$\begin{aligned} f_{11}\left(x\right) &= \int_{-\infty}^{\infty} q_{11}(x \mid x') f_{11}(x' \mid \text{pass})\, dx' \\ &= \frac{\int_{-\infty}^{\infty} q_{11}(x \mid x') f_0(x') G_{11}(x')\, dx'}{\int_{-\infty}^{\infty} f_0(\zeta) G_{11}(\zeta) d\zeta}. \end{aligned} \tag{C.70}$$

After the first test and first stress, end items enter the second test with parameter values no longer normally distributed. Aside from this fact, the treatment of the first posttest distribution during the second test is analogous to the treatment of the untested distribution during the first test. Accordingly, the pdf for the end item parameter following the second test is obtained by inspection from the expression for $f_{11}(x|\text{pass})$ in Eq. (C.69):

$$f_{12}\left(x \mid \text{pass}\right) = f_{11}(x) \frac{G_{12}(x)}{\int_{-\infty}^{\infty} f_{11}(\zeta) G_{12}(\zeta) d\zeta}.$$

Similarly, the distribution following the second stress can be obtained by inspection of the expression for $f_{11}(x)$ in Eq. (C.70):

$$\begin{aligned} f_{12}\left(x\right) &= \int_{-\infty}^{\infty} q_{12}(x \mid x') f_{12}(x' \mid \text{pass})\, dx' \\ &= \frac{\int_{-\infty}^{\infty} q_{12}(x \mid x') f_{11}(x') G_{12}(x')\, dx'}{\int_{-\infty}^{\infty} f_{11}(\zeta) G_{12}(\zeta) d\zeta}. \end{aligned}$$

Continuing in this way, the pdf for the end item after the first stage of testing is given by

$$f_1\left(x\right) = f_{1,n_1-1}(x) \frac{G_{1,n_1}(x)}{\int_{-\infty}^{\infty} f_{1,n_1-1}(\zeta) G_{1,n_1}(\zeta) d\zeta}.$$

At each step of the testing process, the pdf for x is assembled from the results of the previous step, until the final distribution is achieved:

$$f(x) = f_{n,n_N-1}(x) \frac{G_{N,n_N}(x)}{\int_{-\infty}^{\infty} f_{N,n_N-1}(\zeta) G_{N,n_N}(\zeta) d\zeta}.$$

C.9.4 A Simplified Model

For practical purposes, the general model can usually be abbreviated to include only package– and system-level testing, shown in Figure C.7. The rationale for this is as follows. Prior to the package level, testing is done on components and boards that will later be assembled into more complex structures, i.e., packages and systems. These components and boards are stressed and tested to ensure that packages and systems are composed of parts that will function as intended.

Any deviation from nominal performance of a component or board will be reflected in the results of testing functions or parameters at the higher (package or system) levels. Furthermore, the specific effect of individual nonconforming components on the performance of an end item's functions is difficult to assess, since specific components contribute to performance in aggregate ways, better described at the lowest level of abstraction. Moreover, besides being applicable to testing during the end item manufacturing process, the model shown in Figure C.7 is appropriate for describing periodic retesting of end items following deployment. This is because periodically returned end items are likelier to begin their testing sequences at the package or system level rather than at the component or board level.

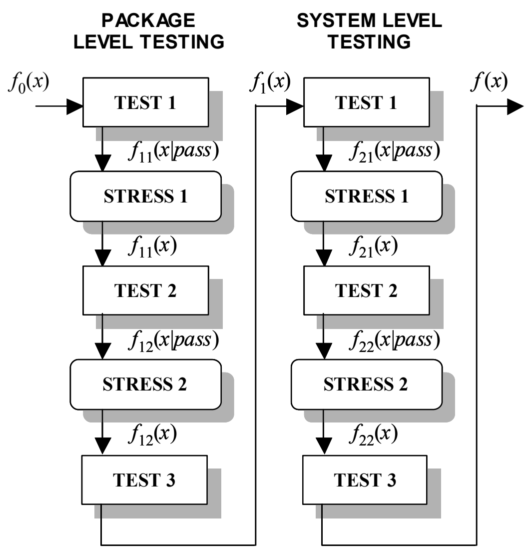


FIGURE C.7 — THE SIMPLIFIED MODEL.

Boards are assembled into functional units or parameters whose statistical properties, before package-level testing, are composites of the properties of the individual boards. The statistical probability density function of the untested parameter is represented by $f_0(x)$.

C.10 Measurement Uncertainty Analysis

A prescription is offered in this appendix to aid in the determination of the various standard deviations used in modeling the test and calibration hierarchy. These standard deviations are constructed from several uncertainty components listed in Table C.3. In Table C.3, the performance limit of the UUT is labeled L_{per}^{UUT}, and the performance limit of the TME is labeled L_{per}^{TME}.

TABLE C.3 Measurement Uncertainty Components

Uncertainty Component	Definition	Standard Deviation
UUT resolution	$= \rho_1 + \rho_2 L_{per}^{UUT}$	σ_R^{UUT}
TS resolution	$= \rho_3 + \rho_4 L_{per}^{TS}$	σ_R^{TS}
process error	$= \rho_5 + \rho_6 L_{per}^{UUT} + \rho_7 L_{per}^{TS}$	σ_P
technician error	$= \rho_8 + \rho_9 L_{per}^{UUT} + \rho_{10} L_{per}^{TS}$	σ_{tech}
rebound error	$= \rho_{11} + \rho_{12} L_{per}^{UUT}$	σ_{reb}
shipping error	$= \rho_{13} + \rho_{14} L_{per}^{UUT} + \rho_{15} L_{per}^{TME}$	σ_s

The coefficients ρ_i, $i = 1, 2, \cdots, 15$, are provided as estimates by persons knowledgeable of the test or calibration process and associated equipment. Of the uncertainty components, UUT *resolution* and *TME resolution* refer to the coarseness of respective UUT or TME attribute readings. *Process error* refers to uncertainties introduced into the test or calibration process by such factors as fluctuations in ancillary equipment and shifts in environmental factors. *Technician error* arises from the fact that different technicians may, under identical circumstances, report different measured values for a given UUT attribute. *Rebound error* was defined earlier. *Shipping error* is an estimate of the upper limits to which the UUT attribute can be displaced because of shipping and storage.

Without more specific information, we assume that each uncertainty component supplied is an upper-limit estimate outside of which values are not expected to be found. Although we make no claim to privileged knowledge regarding the cerebral mechanisms by which human minds develop such estimates, we believe that it is safe to regard these components as approximate 3σ limits. Therefore, the standard deviation corresponding to each uncertainty component is obtained by dividing the magnitude of each estimated component by 3. Thus, for example, σ_R^{UUT} = (UUT resolution) / 3.

The component standard deviations σ_{rb} and σ_s have been encountered. The other components can be used to determine the test process standard deviation σ_{tp}:

$$\sigma_{tp}^2 = \left(\sigma_R^{UUT}\right)^2 + \left(\sigma_R^{TME}\right)^2 + \sigma_P^2 + \sigma_{tech}^2 \,.$$

Appendix D SMPC METHODOLOGY DEVELOPMENT

D.1 Introduction

The SMPC method derives in-tolerance probabilities and attribute biases for both a unit under test (UUT) and a set of independent test and measuring instruments (TME). The derivation of these quantities is based on measurements of a UUT attribute value made by the TME set and on certain information regarding UUT and TME attribute uncertainties. The method accommodates arbitrary accuracy ratios between TME and UUT attributes and applies to TME sets comprised of any number of instruments.

To minimize abstraction of the discussion, the treatment in this appendix focuses on restricted cases in which both TME and UUT attribute values are normally distributed and are maintained within two-sided symmetric tolerance limits. This should serve to make the mathematics more concrete and more palatable. Despite these mathematical restrictions, the methodological framework is entirely general. Extension to cases involving one-sided tolerances and asymmetric attribute distributions merely calls for more mathematical brute force.

D.2 Computation of In-Tolerance Probabilities

D.2.1 UUT In-Tolerance Probability

Whether a UUT provides a stimulus, indicates a value, or shows an inherent property, the declared value of its output, indicated value, or inherent property, is said to reflect some underlying "true" value. A frequency reference is an example of a stimulus, a frequency meter reading is an example of an indicated value, and a gage block dimension is an example of an inherent property. Suppose for example that the UUT is a voltmeter measuring a (true) voltage of 10.01 mV. The UUT meter reading (10.00 mV or 9.99 mV, or some such) is the UUT's "declared" value. As another example, consider a 5 cm gage block. The declared value is 5 cm. The unknown true value (gage-block dimension) may be 5.002 cm, or 4.989 cm, or some other value.

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

A set of n independent measurements are also taken of the true value using n TME. Let Y_i be the declared value representing the ith TME's measurement. The observed differences between UUT and TME declared values are labeled according to

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

The quantities X_i are assumed to be normally distributed random variables with variances σ_i^2 and mean ε_0.

Designating the tolerance limits of the ith TME attribute by $\pm L_i$, the ith TME is considered in-tolerance if $\varepsilon_0 - L_i \leq X_i \leq \varepsilon_0 + L_i$. In other words, *populations* of TME measurements are not expected to be systematically biased. This is the usual assumption made when TME are chosen either randomly from populations of like instruments or when no foreknowledge of TME bias is available. *Individual* unknown TME biases *are* assumed to exist. Accounting for this bias is done by treating individual instrument bias as a random variable and estimating its variance. Estimating this variance is the subject of Section D3. Estimating biases is covered in Section D6.

In applying SMPC methodology, we work with a set of variables r_i, called *dynamic accuracy ratios* (or dynamic inverse uncertainty ratios) defined according to

$$r_i \equiv \frac{\sigma_0}{\sigma_i}, \quad i = 1, 2, \cdots, n \tag{D.2}$$

The adjective "dynamic" will distinguish these accuracy ratios from their usual static or "nominal" counterparts, defined by L_0 / L_i, $i = 1, 2, \cdots, n$. The use of the word "dynamic" underscores the fact that each r_i defined by Eq. (D.2) is a quantity that changes as a function of time passed since the last calibrations of the UUT and of the ith TME. This dynamic character exists because generally both UUT and TME population standard deviations (bias uncertainties) grow with time since calibration. Computation of σ_0 and σ_i is described in Section D.3.

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

$$P_0 = F(a_+) + F(a_-) - 1, \tag{D.3}$$

where F(·) is the distribution function for the normal distribution defined by

$$F(a_\pm) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_\pm} e^{-\zeta^2/2} d\zeta, \tag{D.4}$$

and where

$$a_\pm = \frac{\sqrt{1 + \sum r_i^2}\left(L_0 \pm \dfrac{\sum X_i r_i^2}{1 + \sum r_i^2} \right)}{\sigma_0}. \tag{D.5}$$

In these expressions and in others to follow, all summations are taken over $i = 1, 2, \cdots, n$. The derivation of Eqs. (D.3) and (D.5) is presented in Section D.5. Note that the time dependence of P_0 is in the time dependence of a_+ and a_-. The time dependence of a_+ and a_- is, in turn, in the time dependence of r_i.

D.2.2 TME In-Tolerance Probability

Just as the random variables $X_1, X_2, \cdots, X_n$ are TME-measured deviations from the UUT declared value, they are also UUT-measured deviations from TME declared values. Therefore, it is easy to see that by reversing its role, the UUT can act as a TME. In other words, *any* of the *n* TME can be regarded as the UUT, with the original UUT performing the service of a TME. For example, focus on the *i* th (arbitrarily labeled) TME and swap its role with that of the UUT. This results in the following transformations:

$$
\begin{gathered}
X_1' = X_1 - X_i \\
X_2' = X_2 - X_i \\
\vdots \\
X_i' = -X_i \\
\vdots \\
X_n' = X_n - X_i \, ,
\end{gathered}
$$

where the primes indicate a redefined set of measurement results. Using the primed quantities, the in-tolerance probability for the *i* th TME can be determined just as the in-tolerance probability for the UUT was determined earlier. The process begins with calculating a new set of dynamic accuracy ratios. First, we set

$$\sigma_0' = \sigma_i, \quad \sigma_1' = \sigma_1, \quad \sigma_2' = \sigma_2, \cdots, \sigma_i' = \sigma_0, \cdots, \sigma_n' = \sigma_n \, .$$

Given these label reassignments, the needed set of accuracy ratios can be obtained using Eq. (D.2), i.e.,

$$r_i' = \sigma_0' / \sigma_i', \quad i = 1, 2, \cdots, n \, .$$

Finally, the tolerance limits are relabled for the UUT and the *i*th TME according to $L_0' = L_i$ and $L_i' = L_0$.

If we designate the in-tolerance probability for the *i* th TME by P_i and we substitute the primed quantities obtained above, Eqs. (D.3) and (D.5) become

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

and

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

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

D.3 Computation of Variances

D.3.1 Variance in Instrument Bias

Computing the uncertainties in UUT and TME attribute biases involves establishing the relationship between attribute uncertainty growth and time since calibration. Several models have been used to describe this relationship (see Section B.9).

To illustrate the computation of bias uncertainties, the simple negative exponential model will be used here. With the exponential model, if t represents the time since calibration, then the corresponding in-tolerance probability $R(t)$ is given by

$$R(t) = R(0)e^{-\lambda t} , \tag{D.6}$$

where the parameter λ is the out-of-tolerance rate associated with the instrument in question, and $R(0)$ is the in-tolerance probability immediately following calibration. Note that this form of the exponential model differs from that given in Section B.9. The form used here acknowledges that a finite measurement uncertainty exists at the beginning of the deployment period. The parameters λ and $R(0)$ are usually obtained from analysis of a homogeneous population of instruments of a given model number or type (see Appendix B).

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

$$t = -\frac{1}{\lambda}\exp\left\{\frac{t}{T}\ln\left[\frac{R^*}{R(0)}\right]\right\} . \tag{D.7}$$

Rearranging Eq. (D.7) and substituting in Eq. (D.6) gives

$$R(t) = R(0)\exp\left\{\frac{t}{T}\ln\left[\frac{R^*}{R(0)}\right]\right\} . \tag{D.8}$$

For an instrument attribute whose acceptable values are bounded within tolerance limits $\pm L$, the in-tolerance probability can also be written, assuming a normal distribution, as

$$R(t) = \frac{1}{\sqrt{2\pi\sigma_b^2}}\int_{-L}^{L} e^{-\zeta^2/2\sigma_b^2} d\zeta , \tag{D.9}$$

where $\sigma_b^{\,2}$ is the expected variance of the attribute bias at time t. Equating Eq. (D.9) to Eq. (D.8) and rearranging yields the attribute bias standard deviation

$$\sigma_b = \frac{L}{F^{-1}\left\{\frac{1}{2}\left[1 + R(0)\exp\left(\frac{t}{T}\ln\left[\frac{R^*}{R(0)}\right]\right)\right]\right\}} , \tag{D.10}$$

where $F^{-1}(\cdot)$ is the inverse of the normal distribution function defined in Eq. (D.4).

Substituting L_i, T_i, t_i, $R_i(0)$ and R_i^* , $i = 0, 1, \cdots, n$, in Eq.(D.10) for $L, T, t, R(0)$ and R^* yields the desired instrument bias standard deviations. The variable t_i is the time passed since calibration of the UUT ($i = 0$) or of the ith TME (i =1,2, $\cdots$, n).

D.3.2 Accounting for Bias Fluctuations

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

Uncertainty contributions due to intermediate-term random variations in attribute values usually must be estimated heuristically on the grounds of engineering expectations. In the parlance of the GUM, such estimates are called Type B uncertainties. Youden, for example, provides a graphical method for qualitatively evaluating contributions from human factors, laboratory processes, and reference standards. Development of a quantitative method is a subject of current research. For now, heuristic estimates are usually the best available. Heuristic estimates should represent upper bound (i.e., 3σ) one-sided limits for process uncertainty magnitudes. Experienced metrologists can often provide reasonable guesses for these limits. If we denote upper bounds for heuristically estimated contributions by δ_i, $i = 1, 2, \cdots, n$, the corresponding 3σ standard deviation is given by

$$\sigma_{\delta_i} = \delta_i / 3 \qquad \text{(D.11)}$$

D.3.3 Treatment of Multiple Measurements

In previous discussions, the quantities X_i are treated as single measurements of the difference between the UUT attribute and the ith TME's attribute. Yet, in most applications, testing or calibration of workload items is not limited to single measurements. Instead, multiple measurements are usually taken. Instead of n individual measurements, we will ordinarily be dealing with n sets or *samples* of measurements.

In these samples, let n_i be the number of measurements taken using the ith TME's attribute, and let

$$X_{ij} = Y_0 - Y_{ij}$$

be the jth of these measurements. The sample mean and standard deviation are given in the usual way:

$$X_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \qquad \text{(D.12)}$$

and

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - X_i\right)^2 . \qquad \text{(D.13)}$$

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

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

where the variables σ_{bi} and $\sigma_{\delta i}$ are the long-term and intermediate-term attribute bias standard deviations, respectively, as defined in Section D.3.2. The square root of this variance will determine the quantities r_i defined in Eq. (D.2).

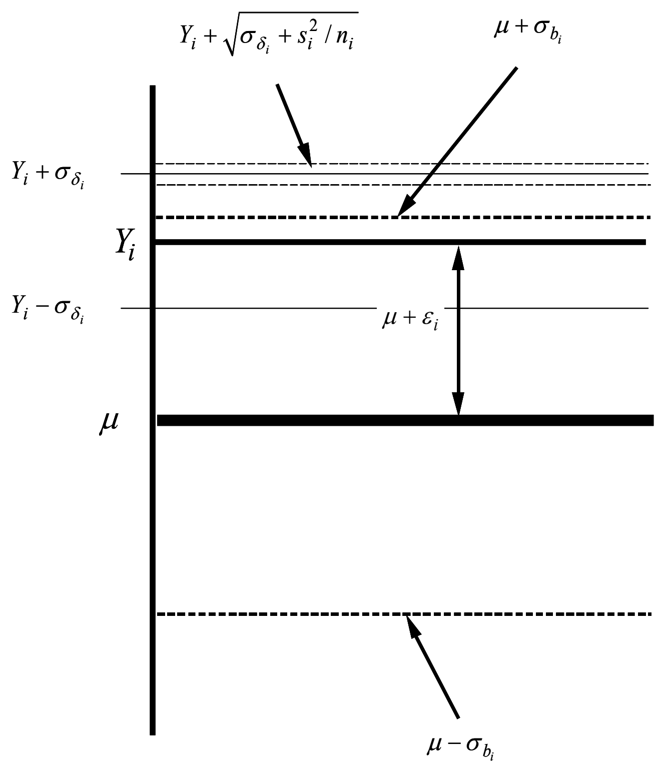


FIGURE D.1 — MEASUREMENT UNCERTAINTY COMPONENTS.

The standard deviation σ_{bi} provides an indication of the uncertainty in the bias of the ith instrument's attribute. The variable $\sigma_{\delta i}$ is a heuristic estimate of the standard deviation associated with intermediate-term random fluctuations in this bias. The variable s_i represents the short-term process uncertainty accompanying measurements made with the ith instrument's attribute.

Note that including sample variances is restricted to the estimation of TME attribute variances. UUT attribute variance estimates contain only the terms σ_{bi} and $\sigma_{\delta i}$. This underscores what is sought in constructing the pdf $f(\varepsilon_0 | \mathbf{X})$. What we seek are estimates of the in-tolerance probability and bias of the UUT attribute. In this, we are interested in the attribute as an entity distinct from process uncertainties involved in its measurement.

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

$$
\begin{aligned}
\sigma_0' &= \sqrt{\sigma_{b_i}^2 + \sigma_{\delta_i}^2} \\
\sigma_1' &= \sqrt{\sigma_{b_1}^2 + s_1^2 / n_1 + \sigma_{\delta_1}^2} \\
&\vdots \\
\sigma_i' &= \sqrt{\sigma_{b_i}^2 + s_i^2 / n_i + \sigma_{\delta_i}^2} \quad . \\
&\vdots \\
\sigma_n' &= \sqrt{\sigma_{b_n}^2 + s_n^2 / n_n + \sigma_{\delta_n}^2} \ .
\end{aligned}
\tag{D.14}
$$

Other expressions are the same as those used in treating single measurement cases. The relationship of uncertainty variables to one another is shown in Figure D.1.

D.4 Example

The proficiency audit problem described in Section 6.4.2 provides an illustrative example of the use of SMPC. In this example, for simplicity, we set $R(0) = 1$, and bias fluctuation and process uncertainties equal to zero. By designating instrument 1 as the UUT, instrument 2 as TME 1 and instrument 3 as TME 2, we have $Y_0 = 0$, $Y_1 = 6$, and $Y_2 = 15$. Thus,

$$\begin{aligned} X_1 &= Y_0 - Y_1 \\ &= -6 \\ X_2 &= X_0 - Y_2 \\ &= -15, \end{aligned}$$

and

$$r_1 = r_2 = 1.$$

Unless otherwise shown, we can assume that the in-tolerance probabilities for all three instruments are about equal to their average-over-period values. The three instruments are managed to the same R^* target, have the same tolerances, and are calibrated in the same way using the same equipment and procedures. Therefore, their standard deviations when the measurements were made should be about equal. According to Eq. (D.2), the dynamic accuracy ratios are

$$r_1 = r_2 = 1.$$

Then, by using Eq. (D.5), we get

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

Calculation of the standard deviation σ_0 calls for some supplemental information. The quantity σ_0 is an *a priori* estimate of the bias standard deviation for the UUT attribute value of interest. In making such *a priori* estimates, it is usually assumed that the UUT is drawn at random from a population. If knowledge of the population's uncertainty is available, then an estimate for σ_0 can be obtained.

For the instruments used in the proficiency audit, it was determined that the population uncertainty is managed to achieve an in-tolerance probability of $R^* = 0.72$ at the end of the calibration interval. As stated above, we assume that we can use average-over-period in-tolerance probabilities for $R(t)$ in this example. With the exponential model, if $R(0) = 1$, the average in-tolerance probability is roughly equal to the in-tolerance probability halfway through the calibration interval. Thus, setting $t = T / 2$ in Eq. (D.10) yields

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

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

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

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

$$
\begin{aligned}
P_0 &= F(0.75) + F(4.23) - 1 \\
&= 0.77 + 1.00 - 1 \\
&= 0.77 .
\end{aligned}
$$

To compute the in-tolerance probability for TME 1 (instrument 2), the UUT and TME 1 swap roles. By using the transformations of Section D.2.2, we have

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

in place of X_1 and X_2 in Eq. (D.5). Recalling that $\sigma_0' = \sigma_0$ in this example gives

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

Thus, by Eq. (D.3), the in-tolerance probability for TME 1 (instrument 2) is

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

In computing the in-tolerance probability for TME 2, the UUT and TME 2 swap roles. Thus

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

Using these quantities in Eq. (D.5) and setting $\sigma_0' = \sigma_0$ gives

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

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

$$\begin{aligned} P_2 &= F(4.47) + F(0.50) - 1 \\ &= 1.00 + 0.69 - 1 \\ &= 0.69. \end{aligned}$$

By summarizing these results, we estimate a roughly 77% in-tolerance probability for instrument 1, a 99% in-tolerance probability for instrument 2, and a 69% in-tolerance probability for instrument 3. As shown earlier, the instruments in the proficiency audit example are managed to an end-of-period in-tolerance probability of 0.72. They are candidates for recalibration if their intolerance probabilities fall below 72%. Therefore, instrument 3 should be recalibrated.

D.5 Derivation of Eq. (D.3)

Let the vector **X** represent the random variables $X_1, X_2, \cdots, X_n$ obtained from n independent TME measurements of ε_0. We seek the conditional pdf for ε_0, given **X**, that will, when integrated over [-L_0, L_0], yield the conditional probability P_0 that the UUT is in-tolerance. This pdf will be represented by the function $f(\varepsilon_0 \,|\, \mathbf{X})$. From basic probability theory, we have

$$f(\varepsilon_0 \,|\, \mathbf{X}) = \frac{f(\mathbf{X} \,|\, \varepsilon_0)\, f(\varepsilon_0)}{f(\mathbf{X})}, \tag{D.15}$$

where

$$f(\varepsilon_0) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\varepsilon_0^2/2\sigma_0^2}. \tag{D.16}$$

In Eq. (D.15), the pdf $f(\mathbf{X} \,|\, \varepsilon_0)$ is the probability density for observing the set of measurements X_1, $X_2, \cdots, X_n$, given that the bias of the UUT is ε_0. The pdf $f(\varepsilon_0)$ is the probability density for UUT biases.

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

$$f(\mathbf{X} \,|\, \varepsilon_0) = f(X_1 \,|\, \varepsilon_0)\, f(X_2 \,|\, \varepsilon_0) \cdots f(X_n \,|\, \varepsilon_0), \tag{D.17}$$

where

$$f(X_i \,|\, \varepsilon_0) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(X_i - \varepsilon_0)^2/2\sigma_i^2}, \quad i = 1, 2, \cdots, n. \tag{D.18}$$

Note that Eq. (D.18) states that, for the present discussion, we assume the measurements of ε_0 to be normally distributed with a population mean value of ε_0 (the UUT "true" value) and a standard deviation σ_i. At this point, we do not provide for an unknown bias in the ith TME.[12] As we will see, the SMPC methodology will be used to estimate this bias, based on the results of measurement and on estimated measurement uncertainties.

Combining Eqs. (D.15) through (D.18) gives

$$\begin{aligned} f(\mathbf{X}|\varepsilon_0) f(\varepsilon_0) &= C\exp\left\{-\frac{1}{2}\left[\frac{\varepsilon_0^2}{\sigma_0^2}+\sum_{i=1}^{n}\frac{(X_i-\varepsilon_0)^2}{\sigma_i^2}\right]\right\} \\ &= C\exp\left\{-\frac{1}{2\sigma_0^2}\left[\varepsilon_0^2+\sum_{i=1}^{n}r_i^2(X_i-\varepsilon_0)^2\right]\right\} \\ &= Ce^{-G(\mathbf{X})}\exp\left\{-\frac{1}{2\sigma_0^2}\left[\left(1+\sum r_i^2\right)\left(\varepsilon_0-\frac{\sum X_i r_i^2}{1+\sum r_i^2}\right)^2\right]\right\}, \end{aligned} \tag{D.19}$$

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

The pdf $f(\mathbf{X})$ is obtained by integrating Eq. (D.19) over all values of ε_0. To simplify the notation, we define

$$\alpha = \sqrt{1+\sum r_i^2} \tag{D.20}$$

and

$$\beta = \frac{\sum X_i r_i^2}{1+\sum r_i^2}. \tag{D.21}$$

Using Eqs. (D.20) and (D.21) in Eq. (D.19) and integrating over ε_0 gives

$$\begin{aligned} f(\mathbf{X}) &= Ce^{-G(\mathbf{X})}\int_{-\infty}^{\infty} e^{-\alpha^2(\varepsilon_0-\beta)^2/2\sigma_0^2}\, d\varepsilon_0 \\ &= Ce^{-G(\mathbf{X})}\frac{\sqrt{2\pi}\sigma_0}{\alpha}. \end{aligned} \tag{D.22}$$

Dividing Eq. (D.22) into Eq. (D.19) and substituting in Eq. (D.15) yields the pdf

$$f(\varepsilon_0|\mathbf{X}) = \frac{1}{\sqrt{2\pi}(\sigma_0/\alpha)} e^{-(\varepsilon_0-\beta)^2/2(\sigma_0/\alpha)^2}. \tag{D.23}$$

As we can see, ε_0 conditional on $\mathbf{X}$ is normally distributed with mean β and standard deviation σ_0 / α. The in-tolerance probability for the UUT is obtained by integrating Eq. (D.23) over [$-L_0$, L_0].

---

[12] It can be readily shown that, if the bias of a TME is unknown, the best estimate for the *population* of its measurements is the true value being measured, i.e., zero bias. This is an important *a priori* assumption in applying the SMPC methodology.

With the aid of Eq. (D.5), this results in

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

which is Eq. (D.3) with α and β as defined in Eqs. (D.20) and D.21).

D.6 Estimation of Biases

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

$$
\begin{aligned}
\beta_0 &= E\left(\varepsilon_0 \,|\, \mathbf{X}\right) \\
&= \int_{-\infty}^{\infty} \varepsilon_0 f\left(\varepsilon_0 \,|\, \mathbf{X}\right) d\varepsilon_0 .
\end{aligned} \tag{D.24}
$$

Substituting from Eq. (D.23) and using Eq. (D.21) gives

$$
\beta_0 = \frac{\sum X_i r_i^2}{1+\sum r_i^2} \tag{D.25}
$$

Similarly, bias estimates can be obtained for the TME set by making the transformations described in Section D.2.2; for example, the bias of TME 1 is given by

$$
\beta_1 = E\left(\varepsilon_1 \,|\, \mathbf{X}'\right) = \frac{\sum X_i' r_i'^2}{1+\sum r_i'^2} . \tag{D.26}
$$

To exemplify bias estimation, we again turn to the proficiency audit question. By using Eqs. (D.25) and (D.26) and by recalling that $\sigma_0 = \sigma_1 = \sigma_2$, we get

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

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

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

If desired, these bias estimates could serve as correction factors for the three instruments. If used in this way, the quantity 7 would be added to all measurements made with instrument 1. The quantity 1 would be added to all measurements made with instrument 2. And, the quantity 8 would be subtracted from all measurements made with instrument 3.[13]

Note that all biases are within the stated tolerance limits (±10) of the instruments, which might encourage users to continue to operate their instruments with confidence. However, recall that the in-tolerance probabilities computed in Section D.4 showed only a 77% chance that instrument 1 was in-tolerance and an even lower 69% chance that instrument 3 was in-tolerance. Such results tend to provide valuable information from which to make cogent judgments regarding instrument disposition.

D.7 Bias Confidence Limits

Another variable that can be useful in making decisions based on measurement results is the range of the confidence limits for the estimated biases. Estimating confidence limits for the computed biases β_0 and β_i, $i = 1, 2, \cdots, n$, means first determining the statistical probability density functions for these biases. From Eq. (D.25) we can write

$$\beta_0 = \sum_{i=1}^{n} c_i X_i \,, \tag{D.27}$$

where

$$c_i = \frac{r_i^2}{1+\sum r_i^2} \,. \tag{D.28}$$

With this convention, the probability density function of β_0 can be written:

$$\begin{aligned} f(\beta_0) &= f\left(\sum c_i X_i\right) \\ &= f\left(\sum \psi_i\right), \end{aligned} \tag{D.29}$$

where

$$\psi_i = c_i X_i \,. \tag{D.30}$$

Although the coefficients c_i, $i = 1, 2, \cdots, n$, are in the strictest sense random variables, to a first approximation, they can be considered fixed coefficients of the variables X_i. Since these variables are normally distributed (see Eq. (D.18)), the variables ψ_i are also normally distributed. The appropriate expression is

---

[13] Since all three instrument are considered *a priori* to be of equal accuracy, the best estimate of the true value of the measurand would be the average of the three measured deviations: $\varepsilon_0 = (0+6+15)/3 = 7$. Thus, a zero reading would be indicative of a bias of –7, a +6 reading would be indicative of a bias of –1 and a +15 reading would be indicative of a bias of +8. These are the same estimates we obtained with SMPC. Obviously, this is a trivial example. Things become more interesting when each measurement has a different uncertainty, i.e., when $\sigma_0 \neq \sigma_1 \neq \sigma_2$.

$$f(\psi_i) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_i}} e^{-(\psi_i - \eta_i)^2 / 2\sigma_{\psi_i}^2}, \tag{D.31}$$

where

$$\sigma_{\psi_i} = c_i \sigma_i \tag{D.32}$$

and

$$\eta_i = c_i \varepsilon_0 . \tag{D.33}$$

Since the variables ψ_i are normally distributed, their linear sum is also normally distributed:

$$\begin{aligned} f(\Sigma\psi_i) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(\Sigma\psi_i - \eta)^2 / 2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(\beta_0 - \eta)^2 / 2\sigma^2} \\ &= f(\beta_0), \end{aligned} \tag{D.34}$$

where

$$\sigma = \sqrt{\Sigma\sigma_{\psi_i}^2}, \tag{D.35}$$

and

$$\eta = \Sigma\eta_i . \tag{D.36}$$

Equation (D.34) can be used to find the upper and lower confidence limits for β_0. Denoting these limits by β_0^+ and β_0^-, if the desired level of confidence is $p \times 100\%$, then

$$p = \int_{\beta_0^-}^{\beta_0^+} f(\beta_0)\, d\beta_0 ,$$

or

$$\int_{\beta_0}^{\infty} f(\beta_0)\, d\beta_0 = (1-p)/2 = \int_{-\infty}^{\beta_0^-} f(\beta_0)\, d\beta_0 .$$

Integrating Eq. (D.34) from β_0^+ to ∞ and using Eqs. (D.35) and (D.36) yields

$$1 - F\left(\frac{\beta_0^+ - \eta}{\sigma}\right) = (1-p)/2$$

and

$$F\left(\frac{\beta_0^+ - \eta}{\sigma}\right) = (1+p)/2 .$$

Solving for β_0^+ gives

$$\beta_0^+ = \eta + \sigma F^{-1}\left(\frac{1+p}{2}\right). \tag{D.37}$$

Solving for the lower confidence for β_0^- in the same manner, we begin with

$$\int_{-\infty}^{\beta_0^-} f(\beta_0)\, d\beta_0 = (1-p)/2 .$$

This yields, with the aid of Eq. (D.24),

$$F\left(\frac{\beta_0^- - \eta}{\sigma}\right) = (1-p)/2 . \tag{D.38}$$

Using the following property of the normal distribution

$$F(-x) = 1 - F(x) ,$$

we can rewrite Eq. (D.38) as

$$\begin{aligned} F\left(-\frac{\beta_0^- - \eta}{\sigma}\right) &= 1-(1-p)/2 \\ &= (1+p)/2 , \end{aligned}$$

from whence

$$\beta_0^- = \eta - \sigma F^{-1}\left(\frac{1+p}{2}\right) . \tag{D.39}$$

From Eq. (D.34), the parameter η is seen to be the expectation value for β_0. Our best available estimate for this quantity is the computed UUT bias, namely β_0 itself. We thus write the computed upper and lower confidence limits for β_0 as

$$\beta_0^{\pm} = \beta_0 \pm \sigma F^{-1}\left(\frac{1+p}{2}\right). \tag{D.40}$$

In like fashion, we can write down the solutions for the TME biases β_i, $i = 1, 2, \cdots , n$:

$$\beta_i^{\pm} = \beta_i \pm \sigma' F^{-1}\left(\frac{1+p}{2}\right), \tag{D.41}$$

where

$$\sigma' = \sqrt{\sum c_i'^2 \sigma_i'^2} , \tag{D.42}$$

and

$$c_i' = \frac{r_i'^2}{1+\sum r_j'^2} . \tag{D.43}$$

The variables r_i' in this expression are defined as before.

To illustrate the determination of bias confidence limits, we again turn to the proficiency audit example. In this example where

$$\sigma_0 = \sigma_1 = \sigma_2 = 6.97 ,$$

and

$$r_1 = r_2 = r_3 = 1\,.$$

By Eqs. (D.28) and (D.34),

$$c_i = c_i' = \frac{1}{3}\,,$$

and

$$\begin{aligned}\sigma &= \sqrt{\frac{\sigma_1^2}{9} + \frac{\sigma_2^2}{9}}\\ &= \frac{\sqrt{2}\sigma_0}{3}\\ &= 3.29\\ &= \sigma'\,.\end{aligned}$$

Substituting in Eqs. (D.40) and (D.41) yields

$$\beta_0^{\pm} = \beta_0 \pm 3.29F^{-1}\left(\frac{1+p}{2}\right),$$

$$\beta_1^{\pm} = \beta_1 \pm 3.29F^{-1}\left(\frac{1+p}{2}\right),$$

and

$$\beta_2^{\pm} = \beta_2 \pm 3.29F^{-1}\left(\frac{1+p}{2}\right).$$

Suppose that the desired confidence level is 95%. Then p = 0.95, and

$$\begin{aligned}F^{-1}\left(\frac{1+p}{2}\right) &= F^{-1}(0.975)\\ &= 1.96\,,\end{aligned}$$

and

$$3.29F^{-1}\left(\frac{1+p}{2}\right) = 6.4\,.$$

Since β_0 = -7, β_1 = -1, and β_2 = +8, this result, when substituted in the above expressions, gives 95% confidence limits for the estimated biases:

$$\begin{aligned}-13.4 \le \beta_0 \le -0.6\\ -7.4 \le \beta_1 \le 5.4\\ 1.6 \le \beta_2 \le 14.4\,.\end{aligned}$$

Appendix E ERROR ANALYSIS METHODS

E.1 Measurement System Modeling

Whether measurements are active or passive, whether they consist of readings of measurands external to a measurement system or whether they consist of reference outputs generated by internal measurands, they can be analyzed using the basic measurement model.

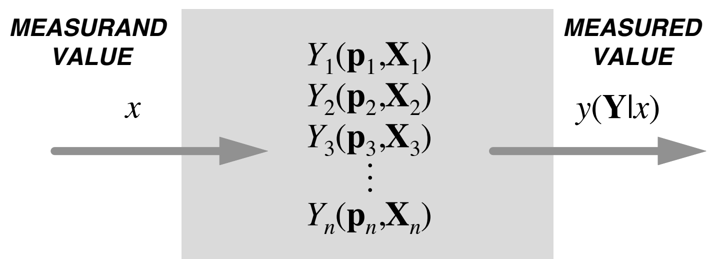


FIGURE E.1 — BASIC MEASUREMENT MODEL.

Measured values are responses of measurement systems to measurand values. Responses of individual system stages are a function of a set of characteristic parameters, the measurand value, and the values of other system responses.

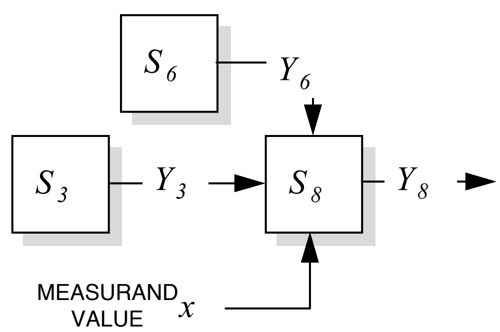


FIGURE E.2 — MODEL STAGES.

These are separate stages, each of whose output is a function of the measurand value and of outputs developed by other stages.

In developing a measurement system model, the measuring system is viewed as a set of separate stages each of whose output is a function of the measurand value and of outputs developed by other stages. The output of each stage is referred to as the "response" of the stage, denoted $Y_i(\mathbf{p}_i, \mathbf{X}_i)$, $i = 1,2,\cdots,n$. The components of the vector $\mathbf{p}_i$ are the parameters that characterize the ith stage, and the components of the vector $\mathbf{X}_i$ are the inputs to the ith stage of the system. These inputs include responses of the other stages of the system and, possibly, the measurand. For example, in the accompanying figure, $\mathbf{X}_8 = (x, Y_3, Y_6)$.

The components of the vector $\mathbf{Y}$ are the responses of all the stages of the system. The notation $f(e_1|e_2)$ is used throughout this document. It reads "f of e_1 given e_2." So, the notation $y(\mathbf{Y}|\, x)$ indicates that the response of the system is functionally dependent on the parameters of each stage and on the system responses $\mathbf{Y}^{(i)}$, $i = 1,2,\cdots,n$ (the measurand x being considered the zeroth response, i.e., $Y_0 \equiv x$).

The parameters of a given stage, indicated by the components of the vectors $\mathbf{p}_i$, $i = 1,2,\cdots,n$, **,** are usually those quantities that comprise the specifications for the stage. For example, for an amplifier stage, they would include such characteristics as gain, linearity, common mode voltage, and noise. They are the governing parameters that characterize the response of the stage to input stimuli. To simplify the treatment, some of the parameters may even represent external stimuli other than inputs from other stages. For example, one parameter may represent ambient temperature, another may represent mechanical vibration, and still another may represent stray emfs.

The vectors $\mathbf{X}_i$, $i = 1,2,\cdots,n$, are arrays that indicate the responses of other measurement system stages that influence the response of the ith stage. In a series system, for example, each stage responds to the output of the stage before it. Consequently, each vector consists of a single component:

$$\mathbf{X}_1 = x \qquad \mathbf{X}_2 = Y_1 \qquad \mathbf{X}_3 = Y_2 \quad \cdots \quad \mathbf{X}_n = Y_{n-1} \,.$$

The system responses for a series system are

$$Y_1 = Y_1(\mathbf{p}_1, x) \; Y_2 = Y_2(\mathbf{p}_2, Y_1) \qquad Y_3 = Y_3(\mathbf{p}_3, Y_2) \quad \cdots \quad Y_n = Y_n(\mathbf{p}_n, Y_{n-1})$$

and the system output is

$$y = y(\mathbf{Y}|x) = Y_n. \tag{E.1}$$

E.2 Measurement Error Modeling

The output $y(\mathbf{Y}|x)$ of the measurement system differs from the measurand by an error

$$\varepsilon(\mathbf{Y}|x) = y(\mathbf{Y}|x) - x \,. \tag{E.2}$$

This error is a function of the individual responses of the measurement system and of the errors in these responses. This functional relationship is developed by using a Taylor series expansion as described in Section E.3. For systems whose component errors are small relative to the outputs of the stages, the expansion can be terminated at first order in the error components.

In most cases, the output of the system will be the output of the nth stage. For these systems, the measurement error is given by

$$\varepsilon(\mathbf{Y} \,|\, x) = \sum_{i=1}^{n-1} \left(\frac{\partial Y_n}{\partial Y_i} \right) \varepsilon_i + \sum_{j=1}^{m_n} \left(\frac{\partial Y_n}{\partial p_{nj}} \right) \varepsilon(p_{nj}) \;, \tag{E.3}$$

where each error component ε_i is expressed in terms of the errors of other system responses and of the errors of its characterizing parameters:

$$\varepsilon_i = \sum_{k \neq i} \left(\frac{\partial Y_i}{\partial Y_k} \right) \varepsilon_k + \sum_{j=1}^{m_i} \left(\frac{\partial Y_i}{\partial p_{ij}} \right) \varepsilon(p_{ij}) \,. \tag{E.4}$$

The quantity m_i is the number of components of the parameter vector for the ith stage and p_{ij} is the jth component.

E.2.1 Series Systems

To illustrate how these expressions are used, consider the series system shown in the figure below. The system consists of two stages with linear outputs:

$$Y_i = p_{i1} Y_{i-1} + p_{i2}\,, \quad i = 1, 2\,.$$

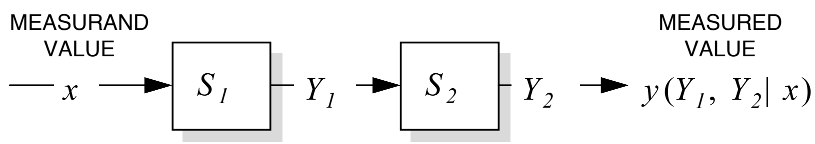


FIGURE E.3 — TWO-STAGE SERIES SYSTEM.
The output y is a measurement of the input x. The error in y is a function of the errors of the responses of the stages S_1 and S_2.

Denoting the ideal, error-free first-stage output as Y_1^0 and assuming zero measurand error,[14] we write

$$\begin{aligned}
Y_1 &= p_{11} x + p_{12} \\
&= [p_{11}^0 + \varepsilon(p_{11})]x + p_{12}^0 + \varepsilon(p_{12}) \\
&= (p_{11}^0 x + p_{12}^0) + \varepsilon(p_{11})x + \varepsilon(p_{12}) \\
&= Y_1^0 + \varepsilon(p_{11})x + \varepsilon(p_{12}) \\
&= Y_1^0 + \left(\frac{\partial Y_1}{\partial p_{11}}\right)\varepsilon(p_{11}) + \left(\frac{\partial Y_1}{\partial p_{12}}\right)\varepsilon(p_{12}) \\
&= Y_1^0 + \varepsilon_1\,.
\end{aligned}$$

Note that this result is given by Eq. (E.4), with $i = 1$. (For the first stage, $k = 0$, and $\varepsilon_k = \varepsilon(x) = 0$.) The output of the second stage is, to first order in the error terms,

$$\begin{aligned}
Y_2 &= p_{21} Y_1 + p_{22} \\
&= [p_{21}^0 + \varepsilon(p_{21})](Y_1^0 + \varepsilon_1) + p_{22}^0 + \varepsilon(p_{22}) \\
&\cong p_{21}^0 Y_1^0 + p_{21}^0 \varepsilon_1 + Y_1^0 \varepsilon(p_{21}) + p_{22}^0 + \varepsilon(p_{22}) \\
&= p_{21}^0 Y_1^0 + p_{22}^0 + p_{21}^0 \varepsilon_1 + Y_1^0 \varepsilon(p_{21}) + \varepsilon(p_{22}) \\
&= Y_2^0 + \left(\frac{\partial Y_2}{\partial p_1}\right)\varepsilon_1 + \left(\frac{\partial Y_2}{\partial p_{21}}\right)\varepsilon(p_{21}) + \left(\frac{\partial Y_2}{\partial p_{22}}\right)\varepsilon(p_{22}) \\
&= Y_2^0 + \varepsilon_2\,.
\end{aligned}$$

---

[14] Although the measurand is the quantity being measured and, by definition, is the sought after, error-free "true" value — the assumption $(x) = 0$ will not always be made.

The final expression of error (the system-error model) in terms of the errors in the parameters of the system stages is obtained by combining terms from the expressions for Y_1 and Y_2:

$$\begin{aligned}\varepsilon(Y_1,Y_2\,|\,x) &= \varepsilon_2\\ &= \left(\frac{\partial Y_2}{\partial Y_1}\right)\left[\left(\frac{\partial Y_1}{\partial p_{11}}\right)\varepsilon(p_{11})+\left(\frac{\partial Y_1}{\partial p_{12}}\right)\varepsilon(p_{12})\right]+\left(\frac{\partial Y_2}{\partial p_{21}}\right)\varepsilon(p_{21})+\left(\frac{\partial Y_2}{\partial p_{22}}\right)\varepsilon(p_{22})\\ &= \left(\frac{\partial Y_1}{\partial p_{11}}\right)\left(\frac{\partial Y_2}{\partial Y_1}\right)\varepsilon(p_{11})+\left(\frac{\partial Y_1}{\partial p_{12}}\right)\left(\frac{\partial Y_2}{\partial Y_1}\right)\varepsilon(p_{12})+\left(\frac{\partial Y_2}{\partial p_{21}}\right)\varepsilon(p_{21})+\left(\frac{\partial Y_2}{\partial p_{22}}\right)\varepsilon(p_{22}).\end{aligned}$$

Generalizing from this result, the system-error model for any series measurement system can be written to first order in the error terms as

$$\begin{aligned}\varepsilon(y\,|\,x) &= \varepsilon(\mathbf{Y}\,|\,x)\\ &= \sum_{i=0}^{n}\sum_{j=1}^{m_i}\left[\prod_{k=i+1}^{n}\left(\frac{\partial Y_k}{\partial Y_{k-1}}\right)\left(\frac{\partial Y_i}{\partial p_{ij}}\right)\varepsilon(p_{ij})\right]\\ &= \sum_{i=0}^{n}\sum_{j=1}^{m_i}\left(\frac{\partial Y_i}{\partial p_{ij}}\right)\varepsilon(p_{ij}),\end{aligned} \tag{E.5}$$

where

$$\frac{\partial Y}{\partial p_{ij}} = \left[\prod_{k=i+1}^{n}\left(\frac{\partial Y}{\partial Y_{k-1}}\right)\right]\left(\frac{\partial Y_i}{\partial p_{ij}}\right). \tag{E.6}$$

E.2.2 Series-Parallel Systems

Developing an error model for a series-parallel system is analogous to developing one for a series system. The only difference is that one or more of the stages may have inputs from more than one stage.

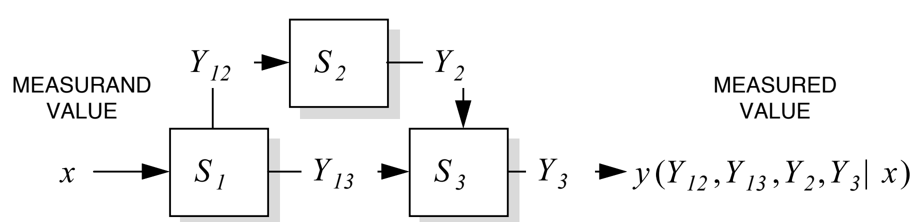


FIGURE E.4 — SERIES-PARALLEL SYSTEM.
One or more of the stages may have inputs from more than one stage.

With this in mind, the linear response model that was used in developing the series system error model is modified to read

$$Y_i = \sum_{j=1}^{m_i-1} p_{ij}Y_{ij} + p_{im_i}, \quad i = 1,2,\cdots,n,$$

where Y_{ij} is the jth input to the ith stage. Consider the three-stage system in Figure E.4. It's easy to show that, with appropriate notation, the error in the output of each stage can be written

$$\varepsilon_i = \sum_{k=inputs} \left(\frac{\partial Y_i}{\partial Y_k} \right) \varepsilon_k + \sum_{j=1}^{m_i} \left(\frac{\partial Y_i}{\partial p_{ij}} \right) \varepsilon(p_{ij}), \quad i = 1, 2, \cdots, n \,. \tag{E.7}$$

E.2.3 Nonlinear Responses

In the foregoing, equations have been derived that apply to modeling errors in systems with linear responses. To first order in the errors, the equations also apply to systems where responses are not necessarily linear, i.e., where the response $Y_i(\mathbf{p}_i,\mathbf{X}_i)$ is not necessarily a linear function of the components of the vectors $\mathbf{p}_i$ and $\mathbf{X}_i$. As an example that supports this assertion, consider the response function

$$Y_i = p_{i1} e^{-(p_{i2} Y_{i-1} + p_{i3})} + p_{i4} \,.$$

The expression of each term as a true value plus an error gives

$$\begin{aligned} Y_i &= \left[p_{i1}^0 + \varepsilon(p_{i1}) \right] \exp\left(-\left\{ \left[p_{i2}^0 + \varepsilon(p_{i2}) \right] \left(Y_{i-1}^0 + \varepsilon_{i-1} \right) + p_{i3}^0 + \varepsilon(p_{i3}) \right\} \right) + p_{i4}^0 + \varepsilon(p_{i4}) \\ &= \left[p_{i1}^0 + \varepsilon(p_{i1}) \right] e^{-\left(p_{i2}^0 Y_{i-1}^0 + p_{i3}^0 \right)} e^{-\left[p_{i2}^0 \varepsilon_{i-1} + Y_{i-1}^0 \varepsilon(p_{i2}) + \varepsilon(p_{i3}) \right]} + p_{i4}^0 + \varepsilon(p_{i4}) . \end{aligned}$$

The second exponential term can be approximated to first order in the errors by

$$e^{-\left[p_{i2}^0 \varepsilon_{i-1} + Y_{i-1}^0 \varepsilon(p_{i2}) + \varepsilon(p_{i3}) \right]} \cong 1 - p_{i2}^0 \varepsilon_{i-1} - Y_{i-1}^0 \varepsilon(p_{i2}) - \varepsilon(p_{i3}) \,.$$

Substituting this approximation in the expression for Y_i gives

$$\begin{aligned} Y_i &= p_{i1}^0 e^{-\left(p_{i2}^0 Y_{i-1}^0 + p_{i3}^0 \right)} + p_{i4}^0 + \varepsilon_i \\ &= Y_i^0 + \varepsilon_i \,, \end{aligned}$$

where

$$\begin{aligned} \varepsilon_i &= \left[\varepsilon(p_{i1}) - p_{i1}^0 p_{i2}^0 \varepsilon_{i-1} - p_{i1}^0 Y_{i-1}^0 \varepsilon(p_{i2}) \right] e^{-\left(p_{i2}^0 Y_{i-1}^0 + p_{i3}^0 \right)} \\ &= \left(\frac{\partial Y_i}{\partial Y_{i-1}} \right) \varepsilon_{i-1} + \sum_{j=1}^{4} \left(\frac{\partial Y_i}{\partial p_{ij}} \right) \varepsilon(p_{ij}) . \end{aligned}$$

These expressions are the same as those used in error-model development for linear response systems. It can be readily demonstrated that to first order in the error terms, the error-modeling approach taken here is valid for any combination of polynomials and transcendental functions. The only stipulation is that the responses of the stages of the measurement system be differentiable with respect to their parameters.

E.2.4 Large Error Considerations

It should be stressed that the foregoing development applies to cases where the various error expressions can be written to first order in the error terms, i.e., to cases where the magnitude of each error in a given response is small relative to the magnitude of the response.

For cases where this is not so, error terms to second or higher order may need to be retained. The validity of the order of an approximation is situation specific. It depends not only on the relative magnitude of each error to its associated response term, but also on the precision to which an analysis can be justifiably carried out.

> There are no systematic rules for deciding the order of an error-analysis model. Identifying the specific order of approximation is an art that improves with experience. In most cases, however, the first-order models given above are applicable.

E.3 Small Error Theory

Consider the output of a stage S_i, given an input Y_i. If the stage response is characterized by a mathematical function f and a set of parameters p_{ij}, $j = 1,2,\cdots,m_i$, then in general,

$$Y_{i+1} = f(Y_i, \mathbf{p}_i)\,. \tag{E.8}$$

In addition to those parameters that characterize the ith stage, the vector $\mathbf{p}$ includes components that represent environmental and other measurement process error sources, independent of the input Y_i. Under nominal (i.e., "error-free") conditions, the input is Y_i^0 and the response is written

$$Y_{i+1}^0 = f(Y_i^0, \mathbf{p}_i^0)\,.$$

Hence, the error in the output Y_{i+1} is

$$\begin{aligned} \varepsilon_{i+1} &= Y_{i+1} - Y_{i+1}^0 \\ &= f(Y_i, \mathbf{p}_i) - f(Y_i^0, \mathbf{p}_i^0). \end{aligned} \tag{E.9}$$

If we expand f in a Taylor series, we get

$$\begin{aligned} Y_{i+1} = f(Y_i^0, \mathbf{p}_i^0) &+ \left(\frac{\partial f}{\partial Y_i}\right)_0 \left(Y_i - Y_i^0\right) + \sum_{j=1}^{m_i} \left(\frac{\partial f}{\partial p_{ij}}\right)_0 \left(p_{ij} - p_{ij}^0\right) \\ &+ \frac{1}{2}\left(\frac{\partial^2 f}{\partial Y_i^2}\right)_0 \left(Y_i - Y_i^0\right)^2 + \frac{1}{2}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i} \left(\frac{\partial^2 f}{\partial p_{ij}\partial p_{ik}}\right)_0 \left(p_{ij} - p_{ij}^0\right)\left(p_{ik} - p_{ik}^0\right) + \cdots, \end{aligned} \tag{E.10}$$

where the input Y_i^0 is the nominal input to S_i and the zero subscript indicates that the vector $\mathbf{p}$ is at its "true" value.

If the deviation from true for $\mathbf{p}$ is written

$$\varepsilon(p_{ij}) = p_{ij} - p_{ij}^0\,, \tag{E.11}$$

recalling that $f(Y_i^0, \mathbf{p}_i^0) = Y_{i+1}^0$, the expression for Y_{i+1} becomes

$$
\begin{aligned}
Y_{i+1} = Y_{i+1}^0 &+ \left(\frac{\partial f}{\partial Y_i}\right)_0 \varepsilon_i + \frac{1}{2}\left(\frac{\partial^2 f}{\partial Y_i^2}\right)_0 \varepsilon_i^2 + \cdots \\
&+ \sum_{j=1}^{m_i} \left(\frac{\partial f}{\partial p_{ij}}\right)_0 \varepsilon(p_{ij}) + \frac{1}{2}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\left(\frac{\partial^2 f}{\partial p_{ij}\partial p_{ik}}\right)_0 \varepsilon(p_{ij})\varepsilon(p_{ik}) + \cdots \quad .
\end{aligned}
\tag{E.12}
$$

Since $\varepsilon_{i+1} = Y_{i+1} - Y_{i+1}^0$, we can write

$$
\begin{aligned}
\varepsilon_{i+1} = &\left(\frac{\partial f}{\partial Y_i}\right)_0 \varepsilon_i + \frac{1}{2}\left(\frac{\partial^2 f}{\partial Y_i^2}\right)_0 \varepsilon_i^2 + \cdots \\
&+ \sum_{j=1}^{m_i} \left(\frac{\partial f}{\partial p_{ij}}\right)_0 \varepsilon(p_{ij}) + \frac{1}{2}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\left(\frac{\partial^2 f}{\partial p_{ij}\partial p_{ik}}\right)_0 \varepsilon(p_{ij})\varepsilon(p_{ik}) + \cdots \quad .
\end{aligned}
\tag{E.13}
$$

The deviations from nominal $\varepsilon(p_{ij})$ are the errors in p_{ij}. If these errors are small, the Taylor series can be truncated at the first-order terms with the result that

$$
\varepsilon_{i+1} \cong \left(\frac{\partial f}{\partial Y_i}\right)_0 \varepsilon_i + \sum_{j=1}^{m_i} \left(\frac{\partial f}{\partial p_{ij}}\right)_0 \varepsilon(p_{ij}). \tag{E.14}
$$

E.4 Example

Consider the measurement of an object of length l using a device whose "sensor" is a metal ruler. The ruler's length is a function of temperature, as is that of the object. The governing equation is

$$
L = L_0 + \kappa_m (T - T_{m,0}) + \kappa_{ms}(T - T_{ms,0}) \ ,
$$

where

L	=	the system output value for l
L_0	=	the measured or "sensed" value for l
κ_m	=	the recorded value for the temperature coefficient of the measurand
κ_{ms}	=	the recorded value of the temperature coefficient of the ruler
T	=	the observed ambient temperature
$T_{m,0}$	=	the nominal temperature for the measurand
$T_{ms,0}$	=	the nominal or calibration temperature for the ruler.

Note that in this application, we wish to extrapolate the length of the measurand to some nominal operating temperature. This effect of the ambient temperature on this value is analogous to the effect of a preceding measurement system stage. Consequently, the above formalism is robust enough to accommodate the situation. If extrapolation to a nominal temperature were not important, the last term in the equation for L would not be included.

Preliminaries aside, we now expand each term as a true value plus an error component:

$$L_0 = l_0 + \varepsilon_{l_0} \quad \kappa_m = \kappa_m^0 + \varepsilon_{\kappa_m} \quad \kappa_{ms} = \kappa_{ms}^0 + \varepsilon_{\kappa_{ms}} \quad T = T_0 + \varepsilon_T \quad L = l + \varepsilon_l ,$$

so that

$$\begin{aligned} L &= l_0 + \varepsilon_{l_0} + \left(\kappa_m^0 + \varepsilon_{\kappa_m}\right)\left(T_0 + \varepsilon_T - T_{m,0}\right) + \left(\kappa_{ms}^0 + \varepsilon_{\kappa_{ms}}\right)\left(T_0 + \varepsilon_T - T_{ms,0}\right) \\ &= l + \varepsilon_l . \end{aligned}$$

Multiplying out and retaining error terms to first order gives

$$\begin{aligned} l + \varepsilon_l = l_0 &+ \kappa_m^0\left(T_0 - T_{m,0}\right) + \kappa_{ms}^0\left(T_0 - T_{ms,0}\right) \\ &+ \varepsilon_{l_0} + \kappa_m^0 \varepsilon_T + \left(T_0 - T_{m,0}\right)\varepsilon_{\kappa_m} + \kappa_{ms}^0 \varepsilon_T + \left(T_0 - T_{ms,0}\right)\varepsilon_{\kappa_{ms}} . \end{aligned}$$

The first three terms comprise the true length l. So, the error in the output of the measuring system is

$$\varepsilon_l = +\varepsilon_{l_0} + \kappa_m^0 \varepsilon_T + \left(T_0 - T_{m,0}\right)\varepsilon_{\kappa_m} + \kappa_{ms}^0 \varepsilon_T + \left(T_0 - T_{ms,0}\right)\varepsilon_{\kappa_{ms}} . \tag{E.15}$$

Now we will use the Taylor series method to see if we get the same expression. We first identify the function f in Eq. (E.8):

$$f(Y_i, \mathbf{p}_i) = L_0 + \kappa_m\left(T_0 - T_{m,0}\right) + \kappa_{ms}\left(T_0 - T_{ms,0}\right),$$

so that we can identify the relevant parameters as

$$Y_i = L_0 \qquad p_{i1} = \kappa_m \qquad p_{i2} = \kappa_{ms} \qquad p_{i3} = T .$$

and, hence,

$$\varepsilon_i = \varepsilon_{l_0} \qquad \varepsilon(p_{i1}) = \varepsilon_{\kappa_m} \qquad \varepsilon(p_{i2}) = \varepsilon_{\kappa_{ms}} \qquad \varepsilon(p_{i3}) = \varepsilon_T ,$$

and

$$\left(\frac{\partial f}{\partial Y_i}\right)_0 = 1 \qquad \left(\frac{\partial f}{\partial p_{i1}}\right)_0 = T_0 - T_{m,0} \qquad \left(\frac{\partial f}{\partial p_{i2}}\right)_0 = T_0 - T_{ms,0} \qquad \left(\frac{\partial f}{\partial p_{i3}}\right)_0 = \kappa_m^0 + \kappa_{ms}^0 .$$

Substituting in Eq. (E.14), the first-order Taylor series expansion error equation becomes

$$\begin{aligned} \varepsilon_{i+1} &= \varepsilon_l \\ &= \left(\frac{\partial f}{\partial Y_i}\right)_0 \varepsilon_i + \sum_{j=1}^{m_i}\left(\frac{\partial f}{\partial p_{ij}}\right)_0 \varepsilon(p_{ij}) \\ &= \varepsilon_{l_0} + \left(T_0 - T_{m,0}\right)\varepsilon_{\kappa_m} + \left(T_0 - T_{ms,0}\right)\varepsilon_{\kappa_{ms}} + \left(\kappa_m^0 + \kappa_{ms}^0\right)\varepsilon_T . \end{aligned}$$

Comparison of this result with Eq. (E.15) shows both results to be the same.

Appendix F PRACTICAL METHOD FOR ANALYSIS OF UNCERTAINTY PROPAGATION

F.1 Introduction

This appendix describes a methodology that yields unambiguous results which can be applied directly to the assessment of measurement uncertainty. The methodology specifically addresses the following stages of the uncertainty analysis process:

Statistics Development — Construction of statistical distributions for each measurement error component. Error components are identified in the *error model*.

Uncertainty Analysis — Analysis and assessment of measurement uncertainty.
The methodology for developing error models is presented in Appendix E.

Practicality of the Methodology — This section describes an uncertainty analysis methodology that has practical application to the real world. This may imply that the methodology is simple or easy to use. If so, the implication is unintentional. Some of the mathematics tend to involve multiple terms, subscripts, and superscripts, and may appear a little daunting at times. In this section, the term “practical” means usable or relevant to user objectives, such as equipment tolerancing or decision risk management. Simplicity and ease of use will follow once the methodology is embedded in user-interactive workstation applications, where the math can be largely hidden from view.

Departure from Tradition — Uncertainty analysis methodologies have traditionally been confined to techniques that are conceptually simple and straightforward. These methodologies have been developed in accordance with the available computational capabilities of the decades before desktop workstations became widespread. Unfortunately, while conventional methodologies are often easily understood, they are frequently ambiguous, restricted, and sometimes useless or even dangerous. In contrast, the methods described in this section are unambiguous, fairly general, and lead to a better understanding of the nature and extent of uncertainties surrounding a given measurement situation.

Accessibility to the Engineering Community — The complexity of the methodology of this section can be made available to the engineering community through dedicated software written for today's powerful desktop computers. What may have been considered to be hopelessly difficult in the past can now be made almost trivial from the standpoint of the analyst. Moreover, with the evolution of the desktop computer's graphical user interface (GUI), using a complex methodology, such as is described herein, can even be enjoyable.

With these considerations in mind, it may be argued that the issue of uncertainty analysis must undergo a paradigm shift with a view toward achieving the following objectives:

- Develop uncertainty analysis methodologies that are relevant to scientific inquiry, standards calibration, parameter testing, production template development, and other aspects of the marketplace

- Implement these methodologies in menu-driven platforms with graphical user interfaces.

To explore in detail the issue of methodological relevance, it is helpful to review some background on why measurements are made and how analyzing uncertainty leads to understanding, interpreting, and managing measurement results.

F.1.1 Why Make Measurements?

A variety of reasons for making measurements can be stated. We make measurements to discover new facts, verify hypotheses, transfer physical dimensions, make adjustments to physical attributes, or obtain information necessary to make decisions. The varied reasons for making physical measurements are found in the typical high-tech product development process. Each phase of this process involves the transfer of measurement information across an interface, as shown in Figure F.1. The process involves:

- *R&D*, where new data are taken and hypotheses are tested
- *Prototype development*, where dimensions are transferred, attributes are adjusted or modified, and decisions are made
- *Design*, where prototyping experience leads to decisions on optimal specs and allow-able tolerances
- *Production*, where molds, jigs, and templates transfer physical dimensions
- *Testing*, where decisions to accept or reject parts and assemblies are made
- *Usage*, where customer response to product quality, reliability, and performance is fed back in such forms as sales activity, warranty claims, legal actions, and publicity.

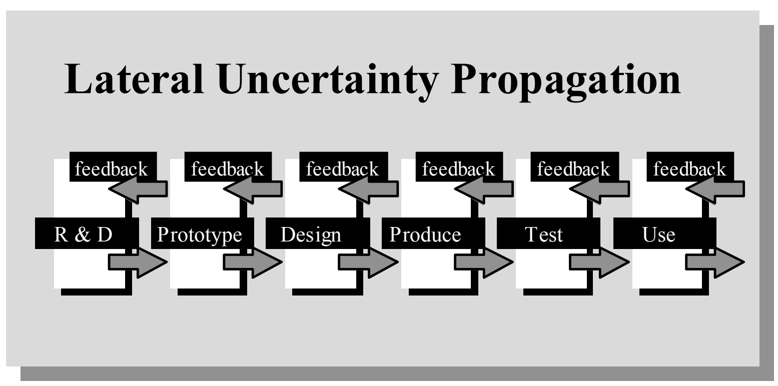


FIGURE F.1 — LATERAL UNCERTAINTY PROPAGATION.
Measurement results are transferred from stage to stage in the typical product development process. Measurement uncertainties accompany each measurement transferal. The appropriateness of measurement accuracies and other characteristics are fed back to modify and refine production process approaches and parameters.

Each product development interface shown in Figure F.1 is supported by a measurement assurance infrastructure embodied in a test and calibration hierarchy. The basic hierarchy structure is shown in Figure F.2.

In a typical hierarchy, testing of a given end-item attribute by a test system yields a reported in- or out-of-tolerance indication, an adjustment if needed, and a beginning-of-period in-tolerance probability (measurement reliability). Similarly, the results of calibration of corresponding test system attributes include reported in- or out-of-tolerance indications, attribute adjustments, and beginning-of-period measurement reliabilities. The same sort of data results from calibrating the supporting calibration systems and accompanies calibrations down through the hierarchy until a point is reached where the "unit under test" (UUT) of interest is a primary calibration standard.

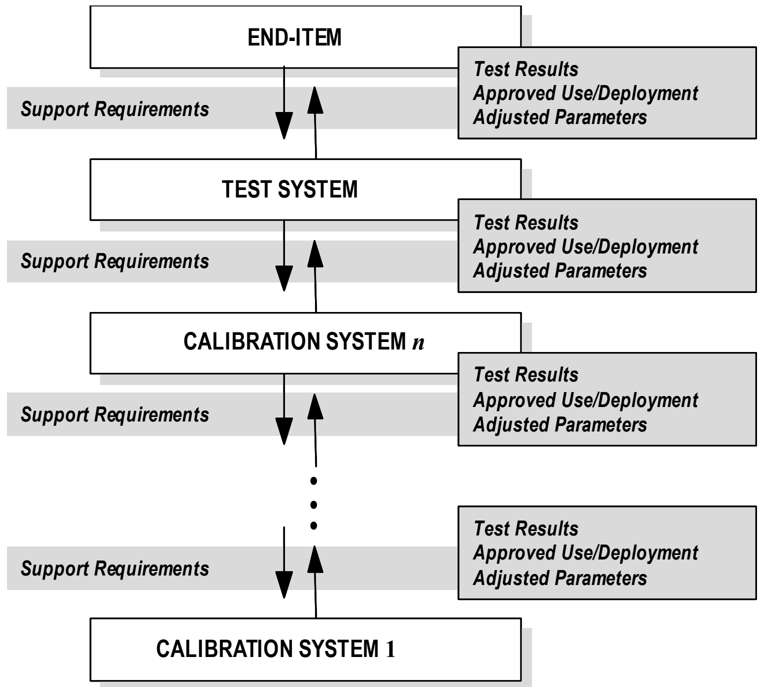


FIGURE F.2 — VERTICAL UNCERTAINTY PROPAGATION.
Measurement accuracy requirements flow down from the end-item or product through the measurement assurance support hierarchy. The uncertainty of calibrated and/or tested attributes propagates upward.

F.1.2 Why Estimate Uncertainties?

All physical measurements are accompanied by measurement uncertainty. Since measurement results are transmitted laterally across development process interfaces and vertically across support hierarchy interfaces, uncertainties in these results also propagate both later-ally and vertically.

Whether we use measurements to verify hypotheses, construct artifacts, or test products, we should know how good our measurements are. Within the context of each application, this is synonymous with knowing the confidence with which our measurements allow us to make decisions, adjust parameters, and so on.

A perhaps pessimistic, yet practical, way of looking at the situation is to say that we want to be able to assess the chances that negative consequences may result from applying knowledge obtained from measurements. It can be shown that the probability for negative consequences increases with the uncertainty associated with a measurement result. Thus, managing the risks involved in applying measurement results is intimately linked with managing uncertainty.

Optimizing the management of measurement decision risks involves (1) linking specific values of a physical attribute with outcomes that may result from using the attribute and (2) estimating the probability of encountering these values in practice. If high probabilities exist for unknowingly encountering attribute values associated with negative consequences, we say that our knowledge of the attribute's value is characterized by high levels of measurement uncertainty. If the reverse is the case, we say that measurement uncertainty is not significant.

If our approach to uncertainty analysis aids in estimating the probability of encountering attribute values associated with negative consequences, then we have a workable, i.e., practical, measurement uncertainty analysis methodology.

F.2 Estimating Uncertainty — Conventional Methods

Conventional[15] uncertainty analysis methodologies ordinarily employ the following steps:

(1) Identify all components of error.

(2) Estimate statistical or engineering variances for each component.[16]

(3) Combine variances to achieve a total uncertainty estimate.

(4) Estimate statistical confidence limits, based on the total estimate.

Statistical confidence limits are usually determined by assuming normally distributed error components. Where Type A estimates are available, "Student's *t*"-distribution is invoked.

F.2.1 Methodological Drawbacks

While step one is always advisable, certain ambiguities and improprieties arise in the way that conventional methods address steps 2 through 4. This is due to three main drawbacks of conventional methods.

(1) Lack of an Uncertainty Model — The first drawback involves the failure to gauge the relative impact of each component of error on total uncertainty. Some error components may contribute more significantly than others. Without an uncertainty model, based on a rigorous error model, arbitrary and unwieldy weighting schemes tend to be used whose applicability is often questionable.

How uncertainties combine differs from situation to situation. Each situation requires its own error model. Moreover, in developing an uncertainty estimate based on an error model, it may be that more than just a simple extrapolation from the model will not be sufficient. For example, if the appropriate error model is a linear combination of error components, it does not always follow that total uncertainty can be determined from a linear combination of corresponding uncertainty component variances.

---

[15] "Conventional" as used herein refers to the methodology provided in NIST Technical Note 1297 and in ISO/ TAG4/WG3, *Guide to the Expression of Uncertainty in Measurement.*

[16] Such variances are referred to as *Type A* and *Type B* uncertainties, respectively. As a reminder, Type A estimates are those that are evaluated by applying statistical methods to a series of repeated observations and Type B estimates are other evaluations—subjective and otherwise. It should not be assumed that evaluations of repeated observations are necessarily superior to evaluations by other means. Type A evaluations of standard uncertainty are not necessarily more reliable than Type B and in many practical measurement situations the components obtained from Type B evaluations may be better known than the components obtained from Type A evaluations.

Without a defined uncertainty model, most conventional approaches involve either a linear combination of component uncertainties (standard deviations) or confidence limits, or a linear combination of component variances. Linear combinations of standard deviations or confidence limits is ill-advised in virtually all cases.[17] Such combinations lead to what are often called "worst-case" uncertainty estimates. They could also be called "worst-guess" estimates.

Part of the problem stems from the fact that linear combinations of variances arising from various error components are not relevant except in cases where the error model is linear and all error components are statistically independent (s-independent). Moreover, even if s-independence pertains, linear combinations of variances are not generally useful unless all error components follow the same sort of statistical distribution and the distribution is symmetrical about the mean. To get around these difficulties, the expedient of imagining that each error component is normally distributed is often employed. This is sometimes justified on the basis of the central limit theorem.

(2) Misleading Variances - The Normality Assumption — The second drawback of the conventional approach is its reliance on statistical variance as the sole measure of uncertainty. Working with variances alone can produce misleading results. We now examine this claim by considering two testing or calibration scenarios; one in which items are adjusted or "renewed" and one in which they are discarded or otherwise kept from service.

Post-Test Distribution for Testing With Renewal of Failed Attributes

For cases in which items are renewed following testing or calibration, we return to the pdf for the post-test distribution that was developed in appendix C

$$f_{pt}(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi(\sigma_t^2+\sigma_{rb}^2)}} e^{-x^2/2(\sigma_t^2+\sigma_{rb}^2)}, & \text{renew always} \\ \phi(x)\dfrac{e^{-x^2/2\sigma_{true}^2}}{\sqrt{2\pi}\sigma_{true}} + K\dfrac{e^{-x^2/2(\sigma_t^2+\sigma_{rb}^2)}}{\sqrt{2\pi(\sigma_t^2+\sigma_{rb}^2)}}, & \text{otherwise}, \end{cases}$$

where

$$\phi(x) \equiv F\left(\frac{L_{adj}+x}{\sigma_t}\right) + F\left(\frac{L_{adj}-x}{\sigma_t}\right) - 1$$

and

$$K \equiv 2\left[1 - F\left(\frac{L_{adj}}{\sqrt{\sigma_{true}^2+\sigma_t^2}}\right)\right].$$

[17] This is not the case for linear combinations of systematic measurement bias when signs are known and magnitudes can be estimated.

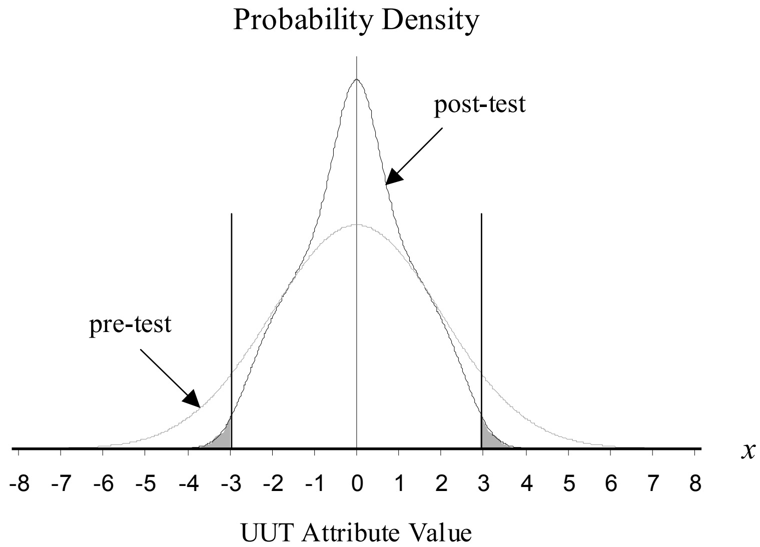


FIGURE F.3 — PRE-TEST VERSUS POST-TEST ATTRIBUTE POPULATIONS.
Typical statistical distributions for attribute values prior to and following test screening. The shaded areas represent probabilities for out-of-tolerance attributes. The pre-test in-tolerance percentage is approximately 85%.[18] The post-test curve corresponds to testing with a measuring system uncertainty (standard deviation) of approximately 25% of the pre-test population uncertainty. As expected, the out-of-tolerance probability is lower after test screening than before test screening.

As these expression show, if the adjustment policy is to only center-spec UUT attribute values that fall outside preset adjustment limits $\pm L_{adj}$, the post-test distribution is not normal. This is illustrated for a typical case in Figures F.3 and F.4.

Figure F.3 portrays a population of product attribute values before and after test screening. The pre-test in-tolerance probability is 85% and the uncertainty of the test process is 25% of the UUT pre-test standard deviation. Since testing has rejected most of the nonconforming attributes, the post-test distribution's tails are pulled in toward the center. The shaded areas represent the fraction of UUT attributes that are out-of-tolerance following testing.

From Figure F.3, it is evident that although the pre-test population may be normally distributed, the post-test distribution of product attribute values is nonnormal. Accordingly, treating post-test product attribute values as being normally distributed could lead to erroneous inferences about their uncertainty.[19]

This can be appreciated by considering the statistical variance of post-test population values. The second form of the above pdf yields a post-test variance of

$$\sigma_{pt}^2 = \left[2F\left(\frac{L_{adj}}{\sigma}\right)-1\right]\sigma_{true}^2 - \sqrt{\frac{2}{\pi}}\left(\frac{L_{adj}}{\sigma}\right)\left(\frac{\sigma_{true}^2}{\sigma}\right)^2 e^{-L_{adj}^2/2\sigma^2} + K\left(\sigma_t^2+\sigma_{rb}^2\right),$$

where

$$\sigma^2 = \sigma_{true}^2 + \sigma_t^2 .$$

---

[18]A pre-test in-tolerance probability of 85% is fairly representative of most periodic calibration programs.

[19]In this context, attribute uncertainty may be equated with the probability that a product item drawn at random from the posttest population will be in-tolerance.

Given the variance in the pre-test population and the accuracy of the test system, the standard deviation for the post-test distribution turns out to be approximately 3.327. If we were engaged in sampling post-test attribute values as part of a process control procedure, for example, we would likely obtain an estimate centered around this value.

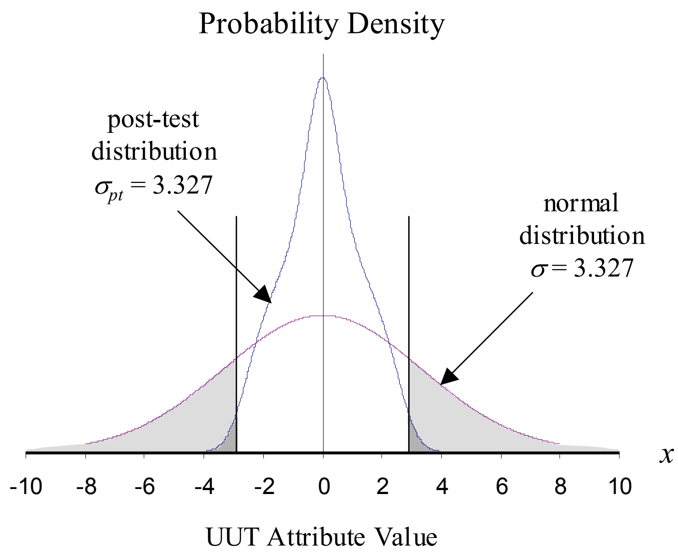


FIGURE F.4 — POST-TEST DISTRIBUTION NORMAL APPROXIMATION.
The post-test distribution for the scenario of Figure F.3 is contrasted with a normal distribution with equal variance. Not only are the out-of-tolerance probabilities (shaded areas) significantly different, the shapes of the distributions are dissimilar.

If we were to assume a normal distribution for the post-test population, a sampled standard deviation of 3.327 would correspond to an in-tolerance percentage of about 61.3% (see Figure F.4). To find the actual post-test in-tolerance percentage, we employ the relation

$$P_{pt} = \sqrt{\frac{2}{\pi}} \int_{-L_{per}/\sigma_{true}}^{L_{per}/\sigma_{true}} F\left(\frac{L_{adj} + \sigma_{true}\zeta}{\sigma_t}\right) e^{-\zeta^2/2} d\zeta - \left[2F\left(\frac{L_{per}}{\sigma_{true}}\right) - 1\right] + K\left[2F\left(\frac{L_{per}}{\sqrt{\sigma_t^2 + \sigma_{rb}^2}}\right) - 1\right].$$

This expression yields an actual post-test in-tolerance percentage of about 97.75%. When evaluating out-the-door quality levels, the difference between 61.3% and 97.75% in-tolerance can be astronomical. An erroneously low 61.3% level can result in unnecessary breaks in production, an unscheduled verification of parts and production machinery, and a reevaluation of the production process — all of which could be avoided by not assuming normality for the product attribute distribution.

Post-test Distribution for Testing without Renewal

For cases where items that fail testing or calibration are not returned to service, the pdf of interest is the conditional function $f(x|\text{pass})$. Following the rules of probability, we can write

$$f(x\,|\,\text{pass}) = \frac{f(\text{pass}\,|\,x) f(x)}{P(\text{pass})} .$$

Using the same notation as in Appendix C, we obtain the function $f(\text{pass}|x)$ from

$$f(\text{pass}\,|\,x) = \int_{-L_{test}}^{L_{test}} f(y\,|\,x)dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_t} \int_{-L_{test}}^{L_{test}} e^{-(y-x)^2/2\sigma_t^2} dy$$

$$= F\left(\frac{L_{test}+x}{\sigma_t}\right) + F\left(\frac{L_{test}-x}{\sigma_t}\right) - 1\,.$$

The probability $P(\text{pass})$ is given by the same relation as $P(\text{not adjust})$ with L_{adj} replaced by L_{test}

$$P(\text{pass}) = 2\,F\left(\frac{L_{test}}{\sqrt{\sigma_{true}^2 + \sigma_t^2}}\right) - 1\,.$$

Putting these relations together gives

$$f(x\,|\,\text{pass}) = \frac{F\left(\frac{L_{test}+x}{\sigma_t}\right) + F\left(\frac{L_{test}-x}{\sigma_t}\right) - 1}{2\,F\left(\frac{L_{test}}{\sqrt{\sigma_{true}^2 + \sigma_t^2}}\right) - 1} \frac{1}{\sqrt{2\pi}\sigma_{true}} e^{-x^2/2\sigma_{true}^2}\,.$$

This pdf, along with a pre-test pdf are shown in Figure F.5.

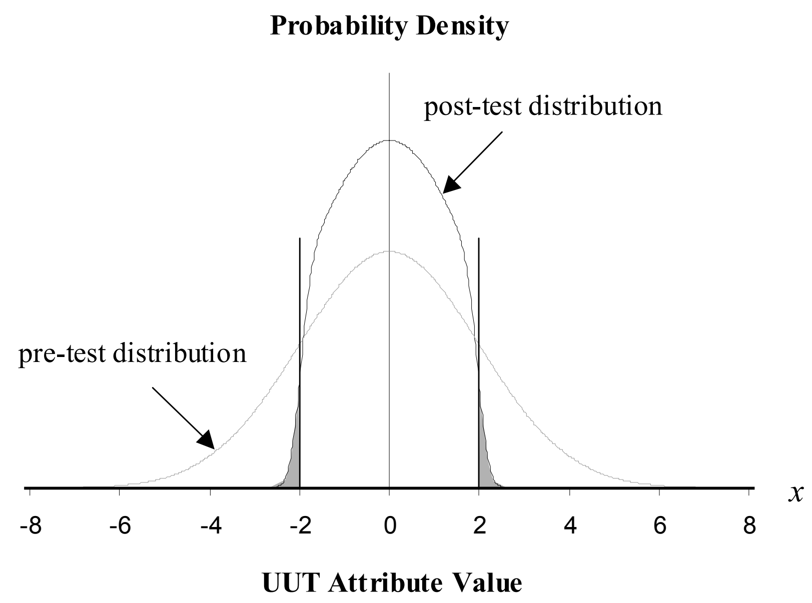


FIGURE F.5 — PRE-TEST VERSUS POST-TEST ATTRIBUTE POPULATIONS FOR CASES WITHOUT RENEWAL.

Statistical distributions for UUT attribute values prior to and following test screening. The shaded areas represent probabilities for out-of-tolerance attributes. The pre-test in-tolerance percentage is approximately 68%.[20] The post-test curve corresponds to testing with a measuring system uncertainty (standard deviation) of approximately 10% of the pre-test UUT bias uncertainty.

Post-Test Distribution for Testing with Renewal of All Attributes

(3) Ambiguity of Application — The third drawback with conventional methods is that they produce results that are not readily applicable. The use of conventional methods typically yields an estimate of the total variance of measurement values. What then to do with this variance? True, it can be used to calculate confidence limits (again, assuming normal distributions of measurements), but confidence limits are not always useful. In general, by themselves they constitute weak decision variables.

The relationship of statistical variances or confidence limits to probabilities associated with negative consequences, referred to earlier, is often ambiguous. Unless a statistical variance enables us to infer the statistical distribution that it characterizes, its function is primarily ornamental. Without knowledge of this distribution, we are at a loss to determine the probability that parts manufactured by one source will mate with parts manufactured by another, or the probability that calibrated test systems will incorrectly accept out-of-tolerance products.

The bleak picture presented by these sentiments is somewhat ameliorated by the fact that, whatever the functional form of the post-test distribution, its character begins to assume normal aspects when the calibrated or tested parameter is subjected to the random stresses of storage and usage. This consideration is addressed in the next Section.

F.2.1.2 The Normalizing Influence of Random Stresses

The foregoing shows that, except for "renew always" testing or calibration, post-test distributions may be decidedly non-normal. Despite this, we can sometimes assume an approximate normal distribution because of the randomizing effects of shipping, handling or storage stress.

As in Appendix C, we make the following assumptions:

(1) Stresses occur randomly with respect to magnitude and direction.

(2) Stresses occur at some average rate r.

(3) Stresses occur over a duration of time t.

Given these assumptions, responses due to shipping are seen to follow the classic random walk behavior. We now consider the impact of these stresses on the post-test distributions and BOP reliabilities for the three renewal practices described earlier.

Testing With Renewal of Failed Attributes

The Probability Density Function

Let the variable ζ represent the value of a measurement attribute following random stress. The pdf for ζ can be expressed as

---

[20] A pre-test in-tolerance probability of 68% is somewhat lower than what is ordinarily encountered in practice. In addition, the effective 10:1 accuracy ratio is also higher than what is achieved in many instances. These values were chosen to dramatize the difference between the pre-test and post-test distributions.

$$q(\varsigma|x) \equiv \frac{e^{-(\varsigma - x)^2/2\sigma_s^2}}{\sqrt{2\pi}\sigma_s},$$

where x is the UUT attribute value before stress, and where

$$\sigma_s = \sqrt{\langle\ \zeta^2\ \rangle rt}\ .$$

The pdf for the distribution of attributes that are placed in service is given by the relation

$$f(\zeta) = \int_{-\infty}^{\infty} q(\varsigma\,|\,x) f_{pt}(x)\,dx$$

for cases where attributes are adjusted if outside $\pm\ L_{adj}$ and placed in service, this becomes

$$f(\zeta) = \frac{1}{\sqrt{2\pi}\sigma_\zeta}\left[F\left(\frac{\sigma_\zeta^2 L_{adj} + \sigma_{true}^2\zeta}{\sigma_\zeta \sigma_w^2}\right) + F\left(\frac{\sigma_\zeta^2 L_{adj} - \sigma_{true}^2\zeta}{\sigma_\zeta \sigma_w^2}\right) - 1\right] e^{-\zeta^2/2\sigma_\zeta^2} + \frac{K}{\sqrt{2\pi}\sigma_q}\ e^{-\zeta^2/2\sigma_q^2},$$

where

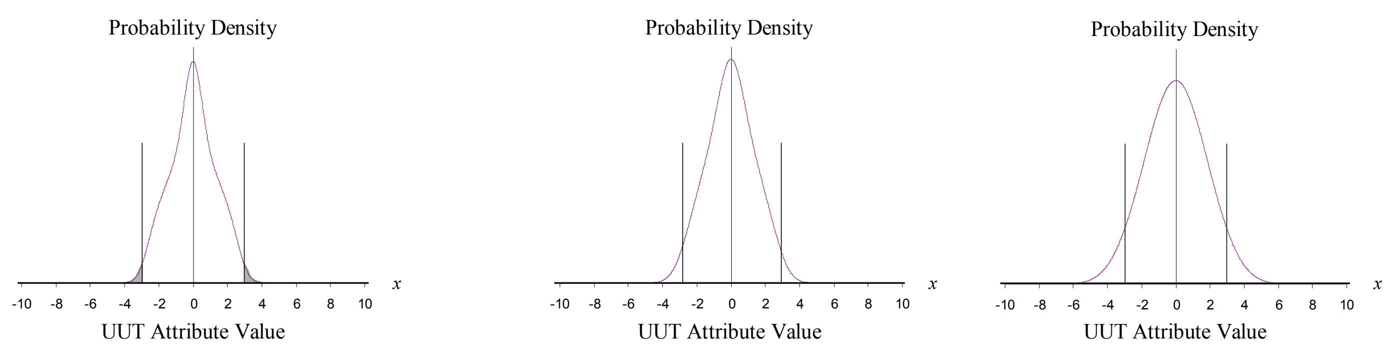
$$\sigma_\zeta = \sqrt{\sigma_{true}^2 + \sigma_s^2}\ ,$$

$$\sigma_q = \sqrt{\sigma_T^2 + \sigma_s^2}\ ,$$

and

$$\sigma_w^2 = \sqrt{\sigma_{true}^2\sigma_t^2 + \sigma_{true}^2\sigma_s^2 + \sigma_t^2\sigma_s^2}\ .$$

This pdf is shown in Figure F.6 for three different times elapsed since testing or calibration. In these figures, $\langle\zeta^2\rangle r = 0.1$. From the progression of times in Figure F.6, we can readily see that random stresses cause a normalization of the post-test distribution $f_{pt}(x)$. As a good approximation, for the case portrayed in Figure F.6, the distribution may be considered approximately normal if $\sigma_s > 0.35\sigma_{true}$.



The post-test UUT pdf immediately following test or calibration ($t = 0$).

The post-test UUT pdf after a time $t = 2.5$ has elapsed since service.

The post-test UUT pdf after a time $t = 15$ has elapsed since service.

FIGURE F.6 —POST-DEPLOYMENT ATTRIBUTE DISTRIBUTION *T* = 5 IN CASES WITHOUT RENEWAL.

The probability density function for UUT attribute values for $\sigma_s = \sqrt{Dt}$, where $D = 0.1$. The pre-test in-tolerance percentage is approximately 85%. The post-test curve corresponds to testing with a measuring system uncertainty (standard deviation) of approximately 25% of the pre-test population uncertainty. As the progression shows, random stresses have a normalizing effect on the post-test distribution. Note also that the probability for out-of-tolerance attributes increases with time since service.

The Beginning of Period Reliability

The BOP measurement reliability is given by

$$R_{bop} = \int_{-\infty}^{\infty} dx\, f_{pt}(x) \int_{-L_{per}}^{L_{per}} d\varsigma\, q(\varsigma|x).$$

Reversing the order of integration, we have

$$\begin{aligned} R_{BOP} &= \frac{1}{\sqrt{2\pi}\sigma_\zeta} \int_{-L_{per}}^{L_{per}} \left[F\left(\frac{\sigma_\zeta^2 L_{adj} + \sigma_{true}^2 \zeta}{\sigma_\zeta \sigma_w^2} \right) + F\left(\frac{\sigma_\zeta^2 L_{adj} - \sigma_{true}^2 \zeta}{\sigma_\zeta \sigma_w^2} \right) - 1 \right] e^{-\zeta^2/2\sigma_\zeta^2} d\zeta + K\left[2F\left(\frac{L_{per}}{\sigma_q} \right) - 1 \right] \\ &= \sqrt{\frac{2}{\pi}} \int_{-L_{per}/\sigma_\zeta}^{L_{per}/\sigma_\zeta} F\left(\frac{\sigma_\zeta^2 L_{adj} + \sigma_{true}^2 \sigma_\zeta \xi}{\sigma_\zeta \sigma_w^2} \right) e^{-\xi^2/2} d\xi - \left[2F\left(\frac{L_{per}}{\sigma_\zeta} \right) - 1 \right] + K\left[2F\left(\frac{L_{per}}{\sigma_q} \right) - 1 \right]. \end{aligned}$$

The integration in this expression must be done numerically.

Testing Without Renewal

The Post-Test Distribution

The pdf for the distribution of attributes that are placed in service only if found within $\pm L_{test}$ is given by

$$\begin{aligned} f(\zeta) &= \int_{-\infty}^{\infty} q(\varsigma\,|\,x) f(x\,|\,\text{pass})\, dx \\ &= \frac{1}{2\pi\kappa\sigma_{true}\sigma_s} \int_{-\infty}^{\infty} \left[F\left(\frac{L_{test} + x}{\sigma_t} \right) + F\left(\frac{L_{test} - x}{\sigma_t} \right) - 1 \right] e^{-x^2/2\sigma_{true}^2}\, e^{-(\zeta - x)^2/2\sigma_s^2}\, dx \\ &= \frac{1}{2\pi\sqrt{2\pi}\kappa\sigma_{true}\sigma_s\sigma_t} \int_{-L_{test}}^{L_{test}} dy \int_{-\infty}^{\infty} e^{-(y-x)^2/2\sigma_t^2}\, e^{-x^2/2\sigma_{true}^2}\, e^{-(\zeta-x)^2/2\sigma_s^2}\, dx \\ &= \frac{1}{\sqrt{2\pi}\kappa\sigma_\zeta} \left[F\left(\frac{\sigma_\zeta^2 L_{test} + \sigma_{true}^2 \zeta}{\sigma_\zeta \sigma_w^2} \right) + F\left(\frac{\sigma_\zeta^2 L_{test} - \sigma_{true}^2 \zeta}{\sigma_\zeta \sigma_w^2} \right) - 1 \right] e^{-\zeta^2/2\sigma_\zeta^2}. \end{aligned}$$

where

$$\kappa = 2F\left(\frac{L_{test}}{\sqrt{\sigma_{true}^2 + \sigma_t^2}} \right) - 1,$$

and where σ_s, σ_z and σ_w are defined as before.

The Beginning of Period Reliability

The BOP reliability for the testing without renewal case is given by

$$\begin{aligned}
R_{BOP} &= \int_{-L_{perf}}^{L_{perf}} d\zeta\, f(\zeta) \int_{-\infty}^{\infty} dx\, q(\varsigma \mid x) f(x \mid \text{pass}) \\
&= \frac{1}{\sqrt{2\pi}\kappa} \int_{-L_{perf}/\sigma_\zeta}^{L_{perf}/\sigma_\zeta} \left[F\left(\frac{\sigma_\zeta^2 L_{test} + \sigma_{true}^2 \sigma_\zeta \xi}{\sigma_\zeta \sigma_w^2} \right) + F\left(\frac{\sigma_\zeta^2 L_{test} - \sigma_{true}^2 \sigma_\zeta \xi}{\sigma_\zeta \sigma_w^2} \right) - 1 \right] e^{-\xi^2/2}\, d\xi \\
&= \sqrt{\frac{2}{\pi}} \frac{1}{\kappa} \int_{-L_{perf}/\sigma_\zeta}^{L_{perf}/\sigma_\zeta} F\left(\frac{\sigma_\zeta^2 L_{test} + \sigma_{true}^2 \sigma_\zeta \xi}{\sigma_\zeta \sigma_w^2} \right) e^{-\xi^2/2}\, d\xi - \frac{1}{\kappa} \left[2F\left(\frac{L_{perf}}{\sigma_\zeta} \right) - 1 \right].
\end{aligned}$$

Post-Test Distribution for Testing with Renewal of All Attributes

The Post-Test Distribution

$$f_{pt}(x) = \frac{1}{\sqrt{2\pi}\sigma_T} e^{-x^2/2\sigma_T^2},$$

where, as before,

$$\sigma_T = \sqrt{\sigma_t^2 + \sigma_{rb}^2}\,.$$

Using the expression for f(z) as with previous cases yields

$$\begin{aligned}
f(\zeta) &= \int_{-\infty}^{\infty} q(\varsigma \mid x) f_{pt}(x)\, dx \\
&= \frac{1}{2\pi \sigma_T \sigma_s} \int_{-\infty}^{\infty} e^{-x^2/2\sigma_T^2}\, e^{-(\zeta - x)^2/2\sigma_s^2}\, dx \\
&= \frac{1}{\sqrt{2\pi}\sigma_q} e^{-x^2/2\sigma_q^2}.
\end{aligned}$$

The Beginning of Period Reliability

With renew if failed and renew as needed testing or calibration, this equation is solved numerically. For the renew always policy, it can be expressed in closed form:

$$R_{bop} = 2F\left(\frac{L_{per}}{\sqrt{\sigma_t^2 + \sigma_{rb}^2 + \sigma_s^2}} \right) - 1 \quad \text{(renew always)}.$$

F.2.2 Methodology Requirements

Given these observations on conventional methods, it appears that what is needed is an uncertainty analysis methodology that directly generates probability estimates for attribute values. The methodology should not be restricted with regard to statistical distributions of error components, nor to assumptions of s-independence. Moreover, it should yield results that can be used in managing measurement decision risk. Such a methodology is referred to as the *practical method*.

F.3 Estimating Uncertainty — The Practical Method

The practical method employs an analysis procedure that differs from that followed by conventional approaches. The procedure it follows is

1. Define the measurement mathematically.
2. Construct an appropriate total error model.
3. Identify all components of error for a given quantity of interest.
4. Determine statistical distributions for each error component.
   - Identify all error sources for each error component
   - Obtain technical information from which to identify the statistical distribution appropriate for each error source
   - Construct a composite statistical distribution for each error component based on its source distributions.
5. Develop a total error statistical distribution from the distributions for each error component.
6. Compute such values as confidence limits, expectation values, and measurement decision risks using the total error statistical distribution.

F.3.1 The Error Model

The error model should describe how error components combine to produce the total error of a measurement result. Consider a particle velocity measurement example. In this example, velocity (v) is computed from measurements of time (t) and distance (d). We first define the measurement with the familiar relation $v = d / t$. If errors are represented by the symbol ε and if errors in time are small compared to the magnitude of the time measurement itself, the appropriate error model is

$$\begin{aligned} v + \varepsilon_v &= \frac{d + \varepsilon_d}{t + \varepsilon_t} \\ &= \frac{d(1 + \varepsilon_d / d)}{t(1 + \varepsilon_t / t)} \\ &\cong \frac{d}{t}(1 + \varepsilon_d / d)(1 - \varepsilon_t / t) \\ &\cong \frac{d}{t}(1 + \varepsilon_d / d - \varepsilon_t / t), \end{aligned}$$

and

$$\begin{aligned} \varepsilon_v &= v\left(\frac{\varepsilon_d}{d} - \frac{\varepsilon_t}{t}\right) \\ &= \varepsilon_1 - \varepsilon_2 , \end{aligned}$$

where

$$\varepsilon_1 = v\varepsilon_d / d \text{ and } \varepsilon_2 = -v\varepsilon_t / t.$$

Note that the same expressions result from using the conventional Taylor series expansion for small measurement errors (see Appendix E):

$$\varepsilon_v = \left(\frac{\partial v}{\partial d} \right) \varepsilon_d + \left(\frac{\partial v}{\partial t} \right) \varepsilon_t \,.$$

In general, if the determination of a given quantity is based on a set of *n* measured attributes, the total error of the quantity can be expressed in the functional relationship

$$\begin{aligned} \varepsilon_{total} &= \varepsilon_{total}(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) \\ &= \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n \,. \end{aligned} \tag{F.1}$$

As with all measurement errors, each of the variables ε_i is composed of both process errors e_p (physical discrepancies between measurement results and true measurand values) and errors of perception e_0 (discrepancies between measurement results and the perception of these results):

$$\varepsilon_i = \varepsilon_i(e_p, e_0), \quad i = 1, 2, \cdots, n \,. \tag{F.2}$$

Steps four and five of the practical method involve determining the statistical distributions for each error component and using these component distributions to form a statistical distribution for the total error. Returning to the particle velocity example, the statistical distribution for ε_v can be obtained from a joint distribution for ε_1 and ε_2. By representing this joint distribution by the probability density function (pdf) f(ε_1,ε_2), the pdf for ε_v can be found using

$$f(\varepsilon_v) = \int_{-\infty}^{\infty} d\varepsilon_1 f(\varepsilon_1, \varepsilon_v - \varepsilon_1) \,. \tag{F.3}$$

In cases where the error components are s-independent, as is commonly the case, this expression becomes

$$f(\varepsilon_v) = \int_{-\infty}^{\infty} d\varepsilon_1 f_1(\varepsilon_1) f_2(\varepsilon_v - \varepsilon_1) \,. \tag{F.4}$$

where $f_1(\cdot)$and $f_2(\cdot)$ are the pdfs for the individual error components ε_1 and ε_2. In this example, these pdfs are related to the pdfs for distance and time according to

$$f_1(\varepsilon_1) = \frac{d}{v} f_{\varepsilon_d}(\varepsilon_1 d / v) \,, \tag{F.5}$$

and

$$f_2(\varepsilon_2) = \frac{t}{v} f_{\varepsilon t}(-\varepsilon_2 t / v) \,, \tag{F.6}$$

The remainder of this section focuses on the construction of pdfs for individual error components. As Eqs. (F.1) through (F.6) indicate, once these pdfs are obtained, a pdf for total measurement error can be developed. By using the total error pdf, a description of total measurement uncertainty becomes possible.

To illustrate, suppose that errors in distance are normally distributed around the distance measurement with standard deviation σ_d, while time measurements are uniformly distributed within $\pm\ \tau$ of the time measurement. Then

$$f_{\varepsilon_d}(\varepsilon_d) = \frac{1}{\sqrt{2\pi}\sigma_d} e^{-(\varepsilon_d - d)^2/2\sigma_d^2} .$$

and

$$f_{\varepsilon_t}(\varepsilon_t) = \begin{cases} 1/2\tau, & t - \tau \le \varepsilon_t \le t + \tau \\ 0, & otherwise. \end{cases}$$

Equations (F.5) and (F.6) yield

$$f_1(\varepsilon_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(\varepsilon_1 - v)^2/2\sigma_1^2} ,$$

where $\sigma_1^2 = (v/d)\sigma_d$ and

$$f_2(\varepsilon_2) = \begin{cases} t/2v\tau, & -(v/t)(t+\tau) \le \varepsilon_2 \le -(v/t)(t-\tau) \\ 0, & otherwise. \end{cases}$$

Substituting these pdfs in Eq. (F.4),

$$\begin{aligned} f(e_v) &= \frac{1}{\sqrt{2\pi}\sigma_1} \frac{t}{2vt} \int_{\varepsilon_v + (v/t)(t-\tau)}^{\varepsilon_v + (v/t)(t+\tau)} e^{-(\varepsilon_1 - v)^2/2\sigma_1^2} d\varepsilon_v \\ &= \frac{t}{2vt}\left[\Phi\left(\frac{\varepsilon_v + (v/t)(t+\tau)}{\sigma_1}\right) - \Phi\left(\frac{\varepsilon_v + (v/t)(t-\tau)}{\sigma_1}\right)\right]. \end{aligned}$$

where the function Φ is the cumulative normal distribution function defined by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\zeta^2/2} d\zeta .$$

F.3.2 Accounting for Process Error

Process error e_p arises from errors in the measurement system (e_{ms}), from the measuring environment (e_e), and from the setup and configuration of the measurement system (e_s):

$$\begin{aligned} e_p &= e_p(e_{ms}, e_e, e_s) \\ &= e_{ms} + e_e + e_s . \end{aligned} \tag{F.7}$$

In Eq. (F.7), the subscripts *ms*, *e,* and *s* refer to “measuring system,” “environment,” and “setup”, respectively. Measurement system and environmental process errors are broken down into a bias (b) and a precision error (ε). Setup error is conceived as constituting a bias only:

$$\begin{aligned} e_{ms} &= b_{ms} + \varepsilon_{ms} \\ e_e &= b_e + \varepsilon_e \\ e_s &= b_s \,. \end{aligned} \tag{F.8}$$

In discussing given measurement situations, the value of the measurand (attribute being measured) will be denoted x and the measured value (measurement result) will be labeled y. Thus the system measures the value x and returns the result y. A measurement result returned by the measuring system can be described by a statistical distribution that is conditional on both the measurand's value and on the measurement process errors. Such a statistical distribution is described by the "conditional" pdf $f(y|x,e_p)$. This function is read "f of y given x and e_p." It represents the probability of obtaining a measurement result y, given a value x and a process error e_p.

In a typical measuring situation, the process error e_p is not known (nor is the value x), and the measuring individual or other "operator" (such as an automated control system) will not be able to obtain the function $f(y|x,e_p)$ explicitly. Instead, what could be attempted is an estimate of a corresponding function $f(y|x)$ that is an "average" or "expectation value" for $f(y|x,e_p)$. The probability density function $f(y|x)$ is obtained by averaging over ranges of values accessible to e_{ms}, e_e and e_s (the sources of e_p).

Obtaining information about e_{ms}, e_e and e_s and constructing the functional form of $f(y|x)$ are accomplished in the structured process described in Section 4 and Appendix E. Briefly, the process consists of extracting all known engineering and other technical knowledge about the attribute under consideration and the measuring system and environment. In some cases, access to test and calibration history databases is also involved. Experience with a prototype test-and-calibration management-decision support system suggests that the process of constructing $f(y|x)$ can be implemented in a user-interactive computer workstation environment.[21]

F.3.3 Accounting for Perception Error

The operator's perception of a measuring system result is usually subject to error. Perception errors arise in a number of ways. For example, in reading an analog meter, errors due to discrepancies between the operator's vantage point and the nominal meter-reading position may arise (parallax errors). In reading a ruler, weighing device, or digital voltmeter, errors due to discrepancies between the measurand's value and the measuring system's nominal scale or readout points often occur (resolution errors). The reader can readily imagine other examples.

Thus, the perceived or "reported" result may differ from the result y returned by the measurement system. These differences are assumed to be distributed around the value of y and are said to be conditional on this value. Thus, denoting the perceived result by the variable z, this distribution is given by the function $f(z|y)$. If the pdfs $f(y|x)$ and $f(z|y)$ can be determined, then the distribution of perceived results around the value of the measurand can be constructed. As one might suspect, this pdf is denoted $f(z|x)$.

---

[21] See Castrup, H., "Navy Analytical Metrology RD&E," *Navy Metrology Research & Development Program Conference Report*, Dept. of the Navy, Metrology Engineering Center, NWS, Seal Beach, CA, Corona Annex, March 1988, and Castrup, H., "Calibration Requirements Analysis System," *Proceedings of the 1989 NCSL Workshop and Symposium*, Denver, CO, July 1989.

F.3.4 Measurement Uncertainty Estimation

The pdf $f(z|x)$ provides a description of the probabilities associated with obtaining perceived or reported values z, given that the value being measured is x. Both measurement process errors and perception errors influence the characteristics of $f(z|x)$.

F.3.4.1 Determination of Confidence Limits for *z*

Estimating statistical confidence limits in the measurement of a quantity is a major facet of conventional uncertainty analysis methods. Most conventional methods (which assume normal error distributions) conclude by forming normal or Student's t -confidence limit estimates based on measurement variance.

The practical method takes a more versatile tack by employing the pdf $f(z|x)$ directly rather than by merely focusing on one of its parameters (i.e., the variance). This permits uncertainty estimation in cases afflicted with nonnormally distributed errors. Unlike conventional methods, statistical confidence limits for z are obtained through integration of $f(z|x)$ directly. This does not involve the usual process of attempting to base confidence limits on some multiple of the standard deviation in z.

F.3.4.2 Estimation of the Measurand Value *x*

The practical method can also be used to estimate values for the measurand x, based on the measurement z, the process error e_p and the perception error ε_0. This feature is unavailable with conventional methods.

F.3.4.3 Determination of Confidence Limits for the Measurand

In addition to estimates of the measurand value, the practical method provides a prescription for obtaining upper and lower bounds that can be said to contain the measurand value with a given level of statistical significance. This is another feature that has been previously unavailable.

F.3.4.4 Management of Measurement Decision Risks

If we can estimate the probability of encountering attribute values associated with negative consequences, then we have a practical uncertainty analysis methodology. One application of such estimates is the determination of consumer and producer risk. Consumer and producer risk can be determined through the use of $f(z|x)$ and the a priori distribution for x, $f(x)$.

F.3.5 Conclusion

Because of its ability to unambiguously determine measurement uncertainty and to enable the effective management of this uncertainty, the practical method is decidedly superior to conventional methods.

Conventional methods require less mathematical effort, but do not yield results that are generally valid. Moreover, the practical method, by working directly with error-source distributions, does not require the development of techniques for combining uncertainties per se. Consequently, the practical method avoids philosophical difficulties that have chronically plagued conventional uncertainty analysis methodologies and have constituted a stumbling block to progress in this area.

The proliferation of desktop computing capability throughout industry has removed the primary obstacle to implementing complex mathematical methods in the work environment. Hence, there are no overriding practical reasons why the practical method cannot be put to use by scientific and engineering personnel. Some additional work is required, however, to bring this to fruition. Future efforts are principally needed in the areas of error-model development and construction of error-source distributions.

F.3.5.1 Constructing Error Models

The development of applicable error models requires engineering knowledge of how measurements are made and knowledge of the sensitivity of measurement parameters to sources of error. Constructing error models based on this knowledge would involve supplying information to a user-interactive desktop application. The desktop application would then develop an appropriate configuration analysis model describing the measurement process and setup. Once a measurement configuration model is constructed, the appropriate error model follows directly.

F.3.5.2 Constructing Source Distributions

Once error sources are identified, their respective statistical distributions must be determined. For some error sources, such as measuring system error, these distributions can be developed from engineering knowledge of ranges of values accessible to measurement attributes and from the results of audits or tests or from calibration history. The construction of other distributions requires the application of knowledge gained from experience (e.g., testing or calibration) with attributes of interest.

F.3.5.3 Generalization of the Mathematical Methods

The methodology illustrates many of its concepts by obtaining results in closed form or in the form of integral equations. Implementation of the methodology does not require that this be done. Interfacing the basic methodological approach with off-the-shelf mathematical analysis software is sufficient to employ the methodology in a completely general way, without restrictions concerning error models employed or corresponding source distributions.

F.4 Construction of Component pdfs

This section addresses the construction of pdfs for the components of error that combine to make the total error of Eq. (F.1). If the joint pdf for component errors is $f(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)$, then the pdf for the total error is

$$f(\varepsilon_{total}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\varepsilon_2 d\varepsilon_3 \cdots d\varepsilon_n f(\varepsilon_{total} - \varepsilon_2 - \cdots - \varepsilon_n, \varepsilon_2, \cdots, \varepsilon_n)\,. \qquad \text{(F.9)}$$

Each of the error components is a function of both process errors which arise from various facets of the measurement process, and errors of perception, which arise from the perception of measurement results. Both process errors and errors of perception are discussed in this section in some detail.

Given a functional form for the joint distribution, it can be constructed from knowledge of the individual pdfs of the error components. The construction of each component pdf involves several steps:

Process Error

- Development of a process error model for each error component
- Development of a pdf describing the distribution of measurement results, given specific process error component values
- Determination of the expectation value for the measurement results pdf

Perception Error

- Development of a perception error model
- Development of a pdf describing the distribution of perceived measurement values, given a specific measurement result
- Determination of the expectation value for the distribution of the perceived measurement values.

Section F.5 shows how pdfs constructed using this procedure are employed to estimate measurement uncertainty limits, measurand expectation values, and measurement decision risks.

F.4.1 The Process Error Model

From observed measurement results, we make inferences about the value of a given measurand and about the uncertainty in our knowledge of this value. To develop a methodological framework for making such inferences, it is helpful to view the measurand as representing some deviation from a nominal or target value.[22] In the present discussion, deviations from nominal are treated as measurement biases or errors whose description can be accomplished by constructing pdfs that represent their statistical distributions. Knowledge of these distributions is acquired through measurement, tempered by certain a priori knowledge of their makeup and of the uncertainties surrounding the measurement process.

Whether the measurand is an element of a derived quantity (such as distance is an element of velocity) or stands alone as the quantity of interest, deviations of its true value from nominal are referred to herein as “error components.” Errors inherent in measurements of these components are labeled process errors.

From Eqs. (F.7) and (F.8), the process error is

$$e_p = b_{ms} + b_e + b_s + \varepsilon_{ms} + \varepsilon_e \, . \tag{F.10}$$

F.4.1.1 Development of the Measurement Results pdf

Let the variable x represent the deviation from nominal of a measured quantity (i.e., the error component of the quantity). Development of the pdf $f(y|x)$ for results produced by the measuring system begins by viewing the measurement result within the context of a given set of process errors. The pdf is

$$f(y\,|\,x, e_p) = f(y\,|\,b_{ms}, b_e, b_s, \varepsilon_{ms}, \varepsilon_e) \, . \tag{F.11}$$

F.4.1.2 Determining the Expectation Value for the Measurement Results pdf

The pdf $f(y|x)$ is found by averaging the error sources in Eq. (F.11) over their respective distributions.

[22] Examples of such nominal values are the length of a yardstick, the volume of a quart of milk, and the weight of a four-ounce sinker.

General Case — The general expression for performing this average is

$$
\begin{aligned}
f(y\,|\,x) &= \int_{\text{process errors}} f(e_p)f(y\,|\,x,e_p)de_p \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} db_{ms}db_e db_s d\varepsilon_{ms}d\varepsilon_e f(e_p)f(y\,|\,x,b_{ms},b_e,b_s,\varepsilon_{ms},\varepsilon_e).
\end{aligned}
\tag{F.12}
$$

s-Independent Sources — If the error sources are s-independent, then the joint pdf $f(y|x,e_p)$ is the product of the pdfs of the source distributions:

$$
f(e_p) = f(b_{ms})f(b_e)f(b_s)f(\varepsilon_{ms})f(\varepsilon_e)\,. \tag{F.13}
$$

With s-independent error sources, Eq. (F.12) can then be solved in a straightforward manner. The order of integration is usually unimportant. For example, we might first consider measurement uncertainty due to random fluctuations in the measuring environment. These fluctuations are accounted for by averaging Eq. (F.12) over the variable ε_e:

$$
f(y\,|\,x,b_{ms},b_e,b_s,\varepsilon_{ms}) = \int_{-\infty}^{\infty} d\varepsilon_e f(\varepsilon_e)f(y\,|\,x,b_{ms},b_e,b_s,\varepsilon_{ms},\varepsilon_e)\,.
$$

The other error sources are averaged in the same way.

F.4.2 The Perception Error Model

Once the measurement result y is obtained, it is perceived by the operator to have the value z. The distribution of z around y, described by the conditional pdf $f(z|y)$, can usually be determined by engineering analysis.

Determination of the pdf for Perceived Measurement Values — Using Eq. (F.12), the pdfs $f(z|y)$ and $f(y|x)$ can be used to determine the pdf for observed measurements of the value of the measurand:

$$
\begin{aligned}
f(z\,|\,x) &= \int_{-\infty}^{\infty} f(z\,|\,y)f(y\,|\,x)dy \\
&= \int_{-\infty}^{\infty} dy \int_{\text{process error}} de_p f(z\,|\,y)f(y\,|\,x,e_p).
\end{aligned}
\tag{F.14}
$$

Equation (F.14) describes a pdf for observed measurements taken on a given measurand value x. Prior to measurement, the available information on this value consists of knowing that the measurand attribute was drawn from a population of like attributes whose values are distributed according to some pdf $f(x)$. In many instances, sufficient a priori knowledge is available on this population to enable an approximate specification of the population's distribution prior to measurement. To illustrate, suppose the measuring situation is product acceptance testing. In this case, a priori knowledge of $f(x)$ can be obtained from design and manufacturing considerations and from product-testing-history data.

Armed with an a priori pdf $f(x)$, the expected distribution of observed measurements is given by

$$f(z) = \int_{-\infty}^{\infty} f(z \mid x) f(x) dx \, . \tag{F.15}$$

where $f(z|x)$ is given in Eq. (F.14).

F.4.3 Inferences Concerning Measurand Values

From a measurement or a set of measurements, we can infer what the most likely distribution of values for the measurand x might be. This is the distribution that could lead to obtaining the perceived values z from measurements of x. Of course, to be precise, the measurand's value is usually a fixed quantity, not a distribution of values. However, this quantity is unknown. In forming an estimate of its distribution, we are really trying to determine probabilities for incremental ranges or neighborhoods of values that contain the measurand value.

The pdf $f(x|z)$ for the distribution of values of x, given the observed measured values z, is obtained from the expression

$$f(x \mid z) = \frac{f(z \mid x) f(x)}{f(z)} \, . \tag{F.16}$$

The pdf $f(z|x)$ is given in Eq. (F.14) and the pdf $f(z)$ is computed using Eq. (F.15). The a priori pdf $f(x)$ is determined as described in the previous section. Equation (F.16) will be used in Section F.5 to determine confidence limits for x and to estimate the most probable value for x, given a perceived measurement z.

F.4.4 Example — Normally Distributed s-Independent Sources

For s-independent error sources, Eq. (F.13) is substituted into Eq. (F.12). If all error sources are normally distributed, performing the integration yields the result

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-(y-x)^2/2\sigma_p^2} \, . \tag{F.17}$$

where

$$\sigma_p^2 = \sigma_{b_{ms}}^2 + \sigma_{b_e}^2 + \sigma_{b_s}^2 + \sigma_{\varepsilon_{ms}}^2 + \sigma_{\varepsilon_e}^2 \, . \tag{F.18}$$

If errors of perception are normally distributed, as is the case with those that stem from random cognitive processes (such as parallax errors), the pdf $f(z|y)$ can be written

$$f(z \mid y) = \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon_0}} e^{-(z-y)^2/2\sigma_{\varepsilon_0}^2} \, , \tag{F.19}$$

where the variable ε_0 is the (random) perception or "observation" error. Substitution of Eqs. (F.19) and (F.17) in Eq. (F.14) yields

$$f(z\,|\,x)=\frac{1}{\sqrt{2\pi}\sigma_m}e^{-(z-x)^2/2\sigma_m^2}, \tag{F.20}$$

where

$$\sigma_m^2=\sigma_p^2+\sigma_{\varepsilon_0}^2\,. \tag{F.21}$$

For normally distributed measurand values, the a priori pdf $f(x)$ is (assuming zero-population bias)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma_x}e^{-x^2/2\sigma_x^2}\,. \tag{F.22}$$

Using this expression with Eq. (F.21) in Eq. (F.15) gives the expected distribution of measured values:

$$f(z)=\frac{1}{\sqrt{2\pi}\sigma_z}e^{-z^2/2\sigma_z^2}, \tag{F.23}$$

where

$$\sigma_z^2=\sigma_x^2+\sigma_m^2\,. \tag{F.24}$$

Combining Eqs. (F.23), (F.22) and (F.20) in Eq. (F.16) gives

$$\begin{aligned} f(x\,|\,z)&=\frac{\sigma_z}{\sqrt{2\pi}\sigma_m\sigma_x}e^{-(z-x)^2/2\sigma_m^2}e^{-x^2/2\sigma_x^2}e^{-z^2/2\sigma_z^2}\\ &=\frac{1}{\sqrt{2\pi}\sigma_{x|z}}e^{-(x-\beta z)^2/2\sigma_{x|z}^2}, \end{aligned} \tag{F.25}$$

where

$$\beta=\frac{1}{1+(\sigma_m/\sigma_x)^2} \tag{F.26}$$

and

$$\sigma_{x|z}=\sqrt{\beta}\sigma_m\,. \tag{F.27}$$

From Eqs. (F.17) through (F.21), it is obvious that the component pdfs obtained using the foregoing procedure could be calculated by recognizing that if the error sources are normally distributed, the component distributions are also normal with variances equal to the sums of the variances of the error sources. This is the familiar RSS result found in many treatments on uncertainty analysis. Note that the conditions for its validity are that error sources be both s-independent and normally distributed.

> For such situations, the statistical distribution construction procedure described above is pure overkill. The procedure becomes more relevant (practical) in cases where one or more error sources are not normally distributed.

F.4.5 Example — Mixed Error-Source Distributions

Consider for purposes of illustration, a case where all error sources except those for perception error are normally distributed. An example is where perception uncertainty is due to random fluctuations in the least-significant digit of a digital device readout. In using the device, the operator obtains a perceived value z. If there are k significant digits following the decimal, then the limits of uncertainty due to the least-significant digit can be expressed according to

$$y = z \pm \rho_k,$$

where $\rho_k = 5 \times 10^{-(k+1)}$.

The measuring system readout informs the operator that the measurement result is somewhere between z - ρ_k and z - ρ_k with uniform probability. The conditional distribution that applies to this uniformly distributed perception error is

$$f(z \mid y) = \begin{cases} \dfrac{1}{2\rho_k}, & y - \rho_k \le z \le y + \rho_k \\ 0, & otherwise. \end{cases} \tag{F.28}$$

Substitution of this pdf in Eq. (F.20) yields

$$\begin{aligned} f(z \mid x) &= \frac{1}{2\rho_k \sqrt{2\pi}\sigma_p} \int_{z-\rho_k}^{z+\rho_k} e^{-(y-x)^2/2\sigma_p^2} dy \\ &= \frac{1}{2\rho_k}\left[\Phi\left(\frac{z-x+\rho_k}{\sigma_p}\right) - \Phi\left(\frac{z-x-\rho_k}{\sigma_p}\right)\right]. \end{aligned} \tag{F.29}$$

where the variable σ_p is defined in Eq. (F.18). The function Φ is the Gaussian cumulative distribution function.

Rather than plugging Eq. (F.29) in Eq. (F.15) to obtain the pdf $f(z)$, it is more convenient to substitute Eq. (F.14) in Eq. (F.15) and perform the integration over first x and then y:

$$\begin{aligned} f(z) &= \int_{-\infty}^{\infty} f(z \mid x) f(x) dx \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, f(z \mid y) f(y \mid x) f(x) \\ &= \frac{1}{2\rho_k} \int_{-\rho_k}^{\rho_k} dy \int_{-\infty}^{\infty} dx\, f(y \mid x) f(x) \\ &= \frac{1}{2\rho_k}\left[\Phi\left(\frac{z+\rho_k}{\sigma_z}\right) - \Phi\left(\frac{z-\rho_k}{\sigma_z}\right)\right], \end{aligned} \tag{F.30}$$

where σ_z is now given by

$$\sigma_z = \sqrt{\sigma_p^2 + \sigma_x^2} \tag{F.31}$$

The construction of the pdf $f(x|z)$ follows the same procedure as that used for normally distributed components. Using Eqs. (F.22), (F.29), and (F.30) in Eq. (F.16), this pdf can be written

$$\begin{aligned} f(x\,|\,z) &= \frac{f(z\,|\,x)f(x)}{f(z)} \\ &= \left[\Phi\left(\frac{z-x+\rho_k}{\sigma_p}\right)-\Phi\left(\frac{z-x-\rho_k}{\sigma_p}\right)\right]\frac{1}{\varphi\left(z,\rho_k,\sigma_z\right)\sqrt{2\pi}\sigma_x}e^{-x^2/2\sigma_x^2}\,, \end{aligned} \tag{F.32}$$

where σ_z is defined in Eq. (F.31) and

$$\varphi\left(z,\rho_k,\sigma_z\right)=\Phi\left(\frac{z+\rho_k}{\sigma_z}\right)-\Phi\left(\frac{z-\rho_k}{\sigma_z}\right). \tag{F.33}$$

Comparing Eq. (F.32) with Eq. (F.25) shows that if even a single error source is nonnormal, the resultant pdf may be substantially different in character than if all sources are normally distributed.

F.5 Applications

F.5.1 Estimating Measurement Confidence Limits

Conventional methodologies calculate statistical confidence limits for measurements by inferring these limits from computed measurement variances. Alternatively, using the practical method, statistical confidence limits for observed measurements can be estimated directly using the pdf $f(z|x)$. For a $(1 - \alpha)\times$ 100% confidence level, the appropriate expressions are

$$\frac{\alpha}{2}=\int_{-\infty}^{L_1} f(z\,|\,x)dz \quad \text{(lower limit)}, \tag{F.34}$$

and

$$\frac{\alpha}{2}=\int_{L_2}^{\infty} f(z\,|\,x)dz\,. \quad \text{(upper limit)}. \tag{F.35}$$

F.5.2 Estimating Measurand Values

In making measurements, we are often primarily interested in ascertaining an estimate of the value of the measurand and in obtaining some confidence that this estimate is sufficiently accurate to suit our purposes. Extension of the foregoing methodology allows this objective.

In making this extension, we employ the pdf $f(x|z)$ to obtain a statistical expectation value for x, given a perceived measurement result z. The relevant expression is

$$\langle x\,|\,z\rangle=\int_{-\infty}^{\infty} x\,f(x\,|\,z)\,dx\,. \tag{F.36}$$

F.5.3 Estimating Confidence Limits for *x*

The conditional pdf $f(x|z)$ can be used to find upper and lower bounds for a neighborhood of measurand values that contains the value of the measurand with a specified level of confidence. If this level of confidence is $(1 - \alpha)\times$ 100%, then the confidence limits L_1 and L_2 for x are found by solving

$$\frac{\alpha}{2} = \int_{L_2}^{\infty} f(x\,|\,z)\,dx = \int_{-\infty}^{L_1} f(x\,|\,z)\,dx. \tag{F.37}$$

F.5.4 Estimating Measurement Decision Risk

Consumer and producer risk are two of the most powerful indicators of measurement decision risk. Consumer risk is defined as the probability that measurements of out-of-tolerance attributes will be perceived as being in-tolerance. Producer risk is defined as the probability that measurements of in-tolerance attributes will be perceived as being out-of-tolerance. Both variables are useful indicators of the quality or accuracy of a measuring process.

If the variable A denotes the acceptable (in-tolerance) range of attribute values, and its complement $\bar{A}$ denotes the corresponding range of out-of-tolerance values, then consumer risk (CR) and producer risk (PR) are calculated according to

$$\begin{aligned} CR &= P(z \in A, x \in \bar{A}) \\ &= P(z \in A) - P(z \in A, x \in A) \\ &= \int_A f(z)\,dz - \int_A dx \int_A dz\, f(z\,|\,x) f(x), \end{aligned} \tag{F.38}$$

and

$$\begin{aligned} PR &= P(z \in \bar{A}, x \in A) \\ &= P(x \in A) - P(z \in A, x \in A) \\ &= \int_A f(x)\,dx - \int_A dx \int_A dz\, f(z\,|\,x) f(x). \end{aligned} \tag{F.39}$$

F.5.5 Example — Normally Distributed s-Independent Sources

The pdfs for normally distributed s-independent sources will be employed in Eqs. (F.34) through (F.39) to estimate measurement confidence limits, measurand bias, confidence limits for this bias, and consumer and producer risks accompanying measurements.

F.5.5.1 Measurement Confidence Limits

Substitution of Eq. (F.20) in Eqs. (F.34) and (F.35) gives the (1 - α)× 100% confidence limits for observed measurements z :

$$L_1 = x - \sigma_m \Phi^{-1}(1 - \alpha/2),$$

and

$$L_2 = x + \sigma_m \Phi^{-1}(1 - \alpha/2),$$

or, alternatively,

$$L_1 = x - \sigma_m \Phi^{-1}(1 - \alpha/2) \le z \le x + \sigma_m \Phi^{-1}(1 - \alpha/2). \tag{F.40}$$

The operator Φ^{-1} is the inverse cumulative normal function, and the measurement standard deviation σ_m is defined in Eq. (F.21).

F.5.5.2 Measurand Bias Estimate

By substituting Eq. (F.25) into Eq. (F.36), the most likely value for the measurand, given the perceived measurement result z, turns out to be

$$\langle x\,|\,z \rangle = \beta z\,, \tag{F.41}$$

where β is given in Eq. F(26).

Note that, since $\beta > 1$ (unless $\sigma_m = 0$), the magnitude of the maximum-likelihood estimate of x is larger than the magnitude of z. This can be understood by recalling that the variable x is treated as a deviation from nominal, and noting that normally distributed measurements tend to regress toward nominal. With these considerations in mind, it can be anticipated that the maximum-likelihood estimate of the true deviation from nominal would be larger than the perceived or measured deviation from nominal.

It should be pointed out that the process of estimating a maximum-likelihood value for an attribute involves both measuring the attribute and making a priori statements about its distribution. If, in the development of Eq. (F.25), a nonzero mean value had been specified in the a priori distribution of x, the resultant maximum-likelihood value would have been centered around the nonzero mean value (i.e., away from nominal).

F.5.5.3 Measurand Confidence Limits

Upper and lower confidence limits for the measurand are obtained by substituting $f(x|z)$ from Eq. (F.25) in Eq. (F.37). The result is

$$\beta z - \sigma_{x|z}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \le x \le \beta z + \sigma_{x|z}\Phi^{-1}\left(1-\frac{\alpha}{2}\right). \tag{F.42}$$

F.5.5.4 Consumer/Producer Risk

To simplify the discussion, assume that the acceptance region for attribute deviations from nominal, represented by the variable x, is symmetrical about zero, i.e., that $A = [-L, L]$. From Eqs. (F.38) and (F.39), consumer risk and producer risk are given by

$$CR = P(z \in A) - P(z \in A, x \in A), \tag{F.43}$$

and

$$PR = P(x \in A) - P(z \in A, x \in A). \tag{F.44}$$

The component parts of these relations are easily calculated. From Eq. (F.23),

$$P(z \in A) = 2\Phi\left(\frac{L}{\sigma_z}\right) - 1, \tag{F.45}$$

where σ_z is defined in Eq. (F.24). From Eq. (F.20), the joint probability for both z and x lying within A is given by

$$\begin{aligned} P(z \in A, x \in A) &= \int_{-L}^{L} dz \int_{-L}^{L} dx\, f(z\,|\,x) f(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma_x}\left[\Phi\left(\frac{L+x}{\sigma_m}\right) + \Phi\left(\frac{L-x}{\sigma_m}\right) - 1\right], \end{aligned} \tag{F.46}$$

where σ_m is given in Eq. (F.21). Finally, using Eq. (F.22),

$$P(x \in A) = 2\Phi\left(\frac{L}{\sigma_x}\right) - 1. \tag{F.47}$$

Equations (F.45) and (F.46) are substituted into Eq. (F.43) to get an estimate of consumer risk. Equations (F.46) and (F.47) are substituted into Eq. (F.44) to get the corresponding producer risk.

F.5.6 Example — s-independent Error Sources with Mixed Distributions

In the example considered here for cases involving mixed distributions, perception errors are uniformly distributed, and errors from all other sources are normally distributed.

F.5.6.1 Measurement Confidence Limits

The same procedure is used to estimate confidence limits for both mixed distribution error sources and normally distributed error sources. For uniformly distributed errors of perception, the lower and upper confidence limits can be obtained from

$$\begin{aligned}\frac{\alpha}{2} &= \int_{-\infty}^{L_1} f(z\,|\,x)dz \\ &= \int_{-\infty}^{L_1} dz \int_{-\infty}^{\infty} dy\, f(z\,|\,y) f(y\,|\,x) \\ &= \int_{-\infty}^{L_1} dy\, f(y\,|\,x) \int_{y-\rho_k}^{y+\rho_k} dz\, f(z\,|\,y) + \int_{L_1-\rho_k}^{L_1+\rho_k} dy\, f(y\,|\,x) \int_{y-\rho_k}^{L_1} dz\, f(z\,|\,y) \\ &= \frac{1}{2\rho_k}\left\{ (L_1+\rho_k-x)\Phi\left(\frac{L_1+\rho_k-x}{\sigma_p}\right) - (L_1-\rho_k-x)\Phi\left(\frac{L_1-\rho_k-x}{\sigma_p}\right) \right. \\ &\qquad \left. + \frac{1}{\sqrt{2\pi}}\left[e^{-(L_1+\rho_k-x)^2/2\sigma_p^2} - e^{-(L_1-\rho_k-x)^2/2\sigma_p^2} \right] \right\},\end{aligned} \tag{F.48}$$

and

$$\begin{aligned}\frac{\alpha}{2} &= \int_{L_2}^{\infty} f(z\,|\,x)dz \\ &= \int_{L_2}^{\infty} dz \int_{-\infty}^{\infty} dy\, f(z\,|\,y) f(y\,|\,x) \\ &= \int_{L_2-\rho_K}^{L_2+\rho_K} dy\, f(y\,|\,x) \int_{L_2}^{y+\rho_k} dz\, f(z\,|\,y) + \int_{L_2+\rho_k}^{\infty} dy\, f(y\,|\,x) \int_{y-\rho_k}^{y+\rho_k} dz\, f(z\,|\,y) \\ &= \frac{1}{2\rho_k}\left\{ (L_2+\rho_k-x)\Phi\left(\frac{L_2+\rho_k-x}{\sigma_p}\right) - (L_2-\rho_k-x)\Phi\left(\frac{L_2-\rho_k-x}{\sigma_p}\right) \right. \\ &\qquad \left. + \frac{1}{\sqrt{2\pi}}\left[e^{-(L_2+\rho_k-x)^2/2\sigma_p^2} - e^{-(L_2-\rho_k-x)^2/2\sigma_p^2} \right] \right\}.\end{aligned} \tag{F.49}$$

Solving for L_1 and L_2 from Eqs. (F.48) and (F.49) requires numerical or graphical methods.

F.5.6.2 Measurand Bias Estimate

For the present example, the expectation value for the measurand is obtained from

$$
\begin{aligned}
\langle x|z\rangle &= \int_{-\infty}^{\infty} x\, f(x|z)\, dx \\
&= \int_{-\infty}^{\infty} x\, \frac{f(z|x) f(x)}{f(z)}\, dx \\
&= \frac{1}{f(z)} \int_{-\infty}^{\infty} x\, f(x)\, dx \int_{-\infty}^{\infty} f(z|y) f(y|z)\, dy \\
&= \frac{1}{f(z)} \int_{-\infty}^{\infty} dy\, f(z|y) \int_{-\infty}^{\infty} x\, f(y|x) f(x)\, dx\,.
\end{aligned}
$$

By using Eqs. (F.17), (F.22), (F.28), (F.30), and (F.31) and integrating,

$$
\langle x|z\rangle = \frac{2\rho_k \gamma}{\sqrt{2\pi}\sigma_z \varphi(z, \rho_k, \sigma_z)}\,, \tag{F.50}
$$

where

$$
\gamma = \frac{1}{1 + (\sigma_p / \sigma_x)^2}\,. \tag{F.51}
$$

F.5.6.3 Measurand Confidence Limits

Upper and lower confidence limits are calculated for this example by numerically or graphically solving the following expressions for L_1 and L_2

$$
\frac{\alpha}{2} = \frac{1}{\varphi(z, \rho_k \sigma_z)\sqrt{2\pi}\sigma_x} \int_{-\infty}^{L_1} \left[\Phi\left(\frac{z - x + \rho_k}{\sigma_p} \right) - \Phi\left(\frac{z - x - \rho_k}{\sigma_p} \right) \right] e^{-x^2/2\sigma_x^2}\, dx \tag{F.52}
$$

and

$$
\frac{\alpha}{2} = \frac{1}{\varphi(z, \rho_k \sigma_z)\sqrt{2\pi}\sigma_x} \int_{L_2}^{\infty} \left[\Phi\left(\frac{z - x + \rho_k}{\sigma_p} \right) - \Phi\left(\frac{z - x - \rho_k}{\sigma_p} \right) \right] e^{-x^2/2\sigma_x^2}\, dx\,. \tag{F.53}
$$

F.5.6.4 Consumer/Producer Risk

As with the example of normally distributed error sources, assume that the acceptance region A in Eqs. (F.38) and (F.39) is symmetrical about zero, i.e., $A = [-L, L]$. Using Eqs. (F.22), (F.29), and (F.30) yields the expressions

$$
P(z \in A) = 2\Phi\left(\frac{L}{\sigma_z} \right) - 1 \tag{F.54}
$$

$$
P(z \in A) = \frac{1}{2\rho_k} \int_{-L}^{L} \left[\Phi\left(\frac{z + \rho_k}{\sigma_z} \right) - \Phi\left(\frac{z - \rho_k}{\sigma_z} \right) \right] dz\,, \tag{F.55}
$$

and

$$
\begin{aligned}
P(z \in A, x \in A) = \frac{1/2\rho_k}{\sqrt{2\pi}\sigma_x} \int_{-L}^{L} dz \\
\times \int_{-L}^{L} dx \left[\Phi\left(\frac{z - x + \rho_k}{\sigma_p} \right) - \Phi\left(\frac{z - x - \rho_k}{\sigma_p} \right) \right] e^{-x^2/2\sigma_x^2}\,.
\end{aligned} \tag{F.56}
$$

Contrasting Eqs. (F.55) and (F.56) with Eqs. (F.45) and (F.46), respectively, shows that applying the assumption of normality to cases with mixed error component distributions may compromise the validity of measurement decision risk management.

F.6 Nomenclature

The following are terms and variables used in the discussion of the practical method. The definitions pertain to this discussion and do not necessarily reflect their general usage within given fields of study.

attribute - A measurable parameter or function.

confidence limits - Limits which are estimated to contain a given variable with a specified probability.

distribution - A mathematical expression describing the probabilities associated with obtaining specific values for a given attribute.

error component - If an attribute is a function of one or more variables, the deviation from nominal of a each variable is an error component .

error model - A mathematical expression describing the relationship of an error to its error components.

error source - A variable that influences the value of an error component.

expectation value - The most probable value of an attribute or variable.

measurement decision risk - The probability of an undesirable outcome resulting from a decision based on measurements.

measurement reliability - The probability that an attribute is in conformance with stated accuracy specifications.

population - All items exhibiting a given measurable property.

probability density function (pdf) - A mathematical expression describing the functional relationship between a specific value of an attribute or variable and the probability of obtaining that value.

statistical variance - The expectation value of the square of the deviation of a quantity from its mean value. A measure of the magnitude of the spread of values adopted by a variable.

s-independent - Statistical independence. Two variables are said to be s-independent if the values adopted by one have no influence on the values adopted by the other.

total error - The total deviation from nominal of the value of an attribute.

ε_{total}	-	Total error.
ε_i	-	The *ith* error component of the total error.
e_p	-	Measurement process error. Error due to the measuring system, environment and set-up.
e_{ms}	-	Error due to the measuring system.
e_e	-	Error due to the measuring environment.
e_s	-	Error due to the set-up and configuration of the measuring system.

b_{ms}	-	The part of measuring system error that remains fixed during a given measurement or set of measurements.
ε_{ms}	-	The part of measuring system error that varies randomly during a given measurement or set of measurements.
b_e	-	The part of measuring environment error that remains fixed during a given measurement or set of measurements.
ε_e	-	The part of measuring environment error that varies randomly during a given measurement or set of measurements.
b_s	-	Synonymous with e_s.
x	-	The true value of the deviation from nominal of an attribute being measured.
y	-	The value returned by the measuring system for a measurement of x.
z	-	The value of a measurement perceived or observed by the operator of the measuring system.
$f(y\|x)$	-	The pdf for obtaining a measured value y from a measurement of x.
$f(z\|y)$	-	The pdf for perceiving a measurement result z from a measured value y.
$f(z\|x)$	-	The pdf for a measurement result z being perceived from a measurement of x.
$f(x\|z)$	-	The pdf for an attribute having a value x given that its measurement is perceived to be z.
$f(x)$	-	The *a priori* pdf for attribute values prior to measurement.
$f(z)$	-	The pdf for perceived measurements taken on an attribute population.
L_1	-	Lower confidence limit.
L_2	-	Upper confidence limit.
$\langle x\|z\rangle$	-	The most probable value for an attribute being measured, given that its perceived measurement value is z.
CR	-	Consumer risk.
PR	-	Producer risk.
$P(z \in A)$	-	The probability that measurements of an attribute will be perceived to be in conformance with stated specifications.
$P(x \in A)$	-	The probability that an attribute is in conformance with specifications prior to measurement.
$P(z \in A, x \in A)$	-	The probability that an attribute is in conformance with specifications and is perceived to be in conformance with specifications.
$\Phi(\cdot)$	-	The cumulative normal distribution function.
$\Phi^{-1}(\cdot)$	-	The inverse of $\Phi(\cdot)$.
σ_p	-	The standard deviation for measurement process errors.
σ_{ε_0}	-	The standard deviation for errors of perception.
σ_m	-	The standard deviation for perceived measurement results.

σ_z - The standard deviation for perceived measurement results for measurements taken on an attribute population.

$\sigma_{x|z}$ - The standard deviation for the estimated distribution of true attribute values that is most likely to produce a perceived measurement result z.

ρ_k - One half the magnitude of the maximum range of perceived values that can contain a measurement result.

Appendix G DETERMINING UNCERTAINTY OF AN EXAMPLE DIGITAL TEMPERATURE MEASUREMENT SYSTEM

G.1 Introduction

The following example is fairly detailed in its identification of error sources and development of mathematical expressions. This is because the adage "garbage in garbage out" is especially relevant in error analysis. Small omissions or mistakes in identifying and specifying error components and in defining an error model can lead to significant departures from reality in the final analysis.

In the past, such departures were not always taken seriously, since the result of an error analysis ordinarily led either to highly conservative compensations or corrections in system design applications, or to excluded risk uncertainty statements intended to provide subjective "warm fuzzies" or similar effects of little concrete utility for measurement interpretation or evaluation.

With the advent of measurement-decision risk methods, this situation has changed. Measurement uncertainty has emerged as an essential element in the computation of risks involved in making erroneous decisions from measurement results.

In developing expressions for measurement uncertainty for use in measurement-decision risk analysis, it is evident that simply quantifying an overall system standard deviation is not sufficient. Instead, a mathematical expression of the statistical error distribution is required. The development of such distributions is described in Appendix F.

Once an attribute bias distribution is specified, it can be employed to determine confidence limits for bias values. In this way, a bias error is treated statistically as a random variable. This is justifiable on the grounds that the instrument was drawn randomly from a population of like instruments whose individual (and unknown) biases take on a distribution of values that can each be assigned a probability of occurrence.

It should be remarked that this practice is regarded by some as being too risky or speculative. Critics of bias distribution estimation usually prefer that the uncertainty limits bounding the attribute's bias be such that essentially *no* values can be found outside them. This approach is not recommended for the simple reason that it establishes bounds that would be applicable under highly unlikely circumstances, i.e., instances where biases are equal to extreme values. Moreover, if a set of limits can be said to satisfy this "excluded bias" requirement, then twice these limits also satisfies the requirement. Indeed, an infinite number of limits can be fixed that satisfy it. The choice of which to use is entirely subjective.

What results from excluded bias uncertainty limits is a "zero-information" condition. To be sure, the bias is likely to be contained within the limits, but the probability of this containment is unknown. This makes projections of risk or other variables by which measurement error can be managed all but impossible.

If a conservative set of bias uncertainty limits is desired, it is *far* more preferable to estimate the distribution and employ a high degree of confidence in specifying limits.

Methods for determining overall system standard deviations are provided in NIST Technical Note 1297 and in ISO/TAG4/WG3, *Guide to the Expression of Uncertainty in Measurement*, hereafter referred to as the "ISO Guide."

In the example, a simple system for converting a time-varying analog measurand value to a digital representation will be analyzed. Since a number of specialized disciplines are involved in the measurement, some detail will be given with regard to the physical and interpretive processes that define the Measurement System.

The foregoing steps will be followed in a more or less formal sequence in an example of a digital temperature measurement system. (This system was previously described in Section 4.) It should be mentioned that the sequence of steps need not be strictly followed. For instance, it may be preferred to develop an error model, based on the system model, prior to and as a means to identifying sources of error. Moreover, the development of a measurement process model may be done at any point. In all cases, however, the approach chosen should be rigorously followed. If not, glaring mistakes can result.

System Model ➟ System Equations ➟ System Error Model ➟ System Uncertainty Model

Later, the methodologies for developing system error models and uncertainty models will be described. These methodologies provide a framework by which measurement system errors and uncertainties can be identified, estimated, and analyzed.

G.2 Identifying the Measurement System Errors

The figure below shows a temperature measurement system. Following the prescription described in Section 4.4, the analysis of the measurement uncertainty of this system involves the development of a system error model. The development of this model will trace the measured value through the system stages and interfaces of the system, from the measurand input to the data processor output.

Identifying sources of measurement system error involves identifying and describing the physical processes that affect a measured value along the measurement path. First, one should draw a simple schematic of the system and then examine each of the system components in detail to identify error sources.

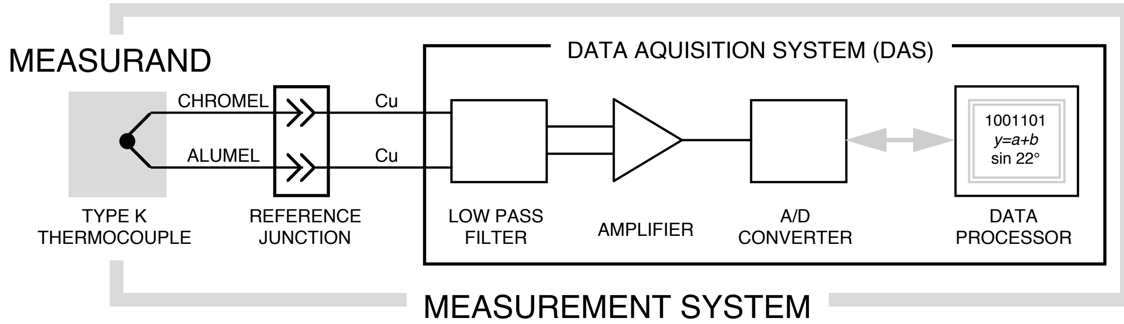


FIGURE G.1 — TEMPERATURE MEASUREMENT SYSTEM.
Differences in the thermoelectric properties of dissimilar conductors produce a voltage difference. This voltage difference is measured and expressed in terms of a temperature.

G.2.1. Sensing (Thermocouple)

Temperature differences between the ends of conductors give rise to voltage gradients. Because of differences in thermoelectric properties, different conductors exhibit different voltage gradients for the same temperature difference. This is the case for chromel and alumel. A given temperature difference across a chromel lead produces a different voltage gradient than the same temperature difference across an alumel lead.

Sensitivity — Chromel and alumel leads that are connected from a measurand to a reference junction produces a voltage between the leads at the reference junction. To convert the voltage to a temperature requires knowledge of the thermocouple's sensitivity to temperature differences. An error in the assumed value for this sensitivity, expressed in terms of μV/°C, produces an error in the sensed value of the temperature of the measurand.

Errors are possible from other sources as well. These include the following:

Hysteresis — Hysteresis is the resistance of a response to a change in stimulus. If the measurand temperature is time varying, any lack of response of the thermocouple to rapid temperature changes is a source of error.

Measurand Fluctuations — If the measured value is a quantity that will be communicated for use in some practical application, random fluctuations that cause deviations from this reported value are a source of error. Randomly occurring differences in measurand value should not be confused with any time-varying aspect of a measurand, such as its signal frequency. Measurand fluctuations are unknown and undesirable phenomena that randomly alter measurement results and may introduce errors in reported measurement values.

Nonlinearity — The potential developed across the thermocouple leads follows a defined functional relationship to the measurand temperature. This relationship is embodied in a mathematical model of temperature versus difference potential. Given the use of the model, any departure between the assumed relationship and the actual temperature constitutes an error.

For example, let

$$\Delta T = T_M - T_R$$
$$V_{0A} = V_0 - V_A$$
$$V_{0C} = V_0 - V_C \ .$$

The voltage differences are given in terms of ΔT by

$$V_{0A} = a_0 + a_1 \Delta T + a_2 (\Delta T)^2 + a_3 (\Delta T)^3 + \cdots$$
$$V_{0C} = b_0 + b_1 \Delta T + b_2 (\Delta T)^2 + b_3 (\Delta T)^3 + \cdots$$

from which the voltage difference $\Delta V = V_C$ - V_A is expressed as

$$\Delta V = (a_0 - b_0) + (a_1 - b_1)\Delta T + (a_2 - b_2)(\Delta T)^2 + (a_3 - b_3)(\Delta T)^3 + \cdots \ .$$

Differences between actual values of the coefficients in this expression and their assumed values give rise to nonlinearity error.

Noise — Since the thermocouple leads are conductors, externally applied electromagnetic fields may introduce stray emfs. Such "noise" comprises an error. Noise is usually random in character.

Thermally generated noise is also possible. If the bandwidth of the signal being measured is B Hz, the ambient temperature is T, and the resistance of a given lead is R, the thermal noise level in the lead is equal to $k_B BRT$, where k_B is Boltzmann's constant. For the present example, thermal noise can be considered negligible.

Junction Temperature — Although the reference junction is an ice bath, impurities in the bath may cause the temperature to differ slightly from 0°C. In addition, the temperature may not be precisely uniform over the physical extent of the bath, differing from location to location by small amounts.

G.2.2. Interfacing (Reference Junction—Low-Pass Filter)

The potential difference at the reference junction output terminals is transmitted through copper wires and applied across the input terminals of a low-pass filter. The copper wires and the filter terminals comprise an interface between the reference junction and the data acquisition system. The sources of error are

Interface Loss — The voltage applied across the terminals of the low-pass filter suffers a drop due to the resistance of the connecting leads from the reference junction and of the low-pass filter contacts.

Noise — Electromagnetic noise is a factor for the connecting leads, while both the connecting leads and the low-pass filter terminals are subject to thermal noise.

Crosstalk — Leakage currents between input filter terminals may alter the potential difference across the terminals.

G.2.3. Filtering (Low-Pass Filter)

The potential difference that survives the reference junction–low-pass filter interface is altered by the low-pass filter. The filter attenuates noise that may be present and provides a "cleaned-up" potential difference to an amplifier. However, some noise gets through. Also, the filter attenuates the signal somewhat and itself generates a small noise component. The sources of error in the low-pass filter interface are the following:

Interface Loss — Although the filter is intended to attenuate unwanted noise, some signal attenuation also occurs.

Nonlinearity — The response of a filter over the range from its cutoff frequency f_c to its terminating frequency f_n is usually considered to be linear. Departures from this assumed linearity constitute errors.

Noise — Not all the input noise will be filtered out. The noise that remains will be attenuated by an amount that depends on the roll-off characteristics of the filter. These characteristics are usually assumed to be linear and are expressed in terms of dB per octave. Thermal noise is also generated within the filter itself.

G.2.4. Interfacing (Low-Pass Filter—Amplifier)

The potential difference output by the low-pass filter is fed to the amplifier across an interface comprised of the leads from the low-pass filter and the input terminals of the amplifier. The sources of error are

Interface Loss — The voltage at the amplifier terminals suffers a drop due to the resistance of the connecting leads from the low-pass filter and of the input terminal contacts.

Noise — Electromagnetic noise is a factor for the connecting leads, while both the connecting leads and the amplifier terminals are subject to thermal noise.

Crosstalk — Leakage currents between input amplifier terminals may cause a decrease in the potential difference across the terminals.

G.2.5. Amplification (Amplifier)

The amplifier amplifies the potential difference (and any noise received from the low-pass filter) and outputs the result to an A/D converter. Several sources of error are present:

Gain — Gain is the ratio of the amplifier output signal voltage to the input signal voltage. Gain errors are those that lead to a uniform shift in expected amplifier output versus actual output. Gain errors are composed of inherent (bias) errors and temperature-induced (precision and bias) errors.

Gain Stability — If the amplifier voltage gain is represented by G_v, its input resistance by R, and its feedback resistance by R', then oscillations are possible when

$$\frac{RG_V}{R+R'} = \pi .$$

These oscillations appear as an instability in the amplifier gain.

Normal Mode — Normal mode voltages are differences in zero potential that occur when amplifier input (signal) lines are not balanced. Normal mode voltages are essentially random in character.

Common Mode — Common-mode voltage consists of unwanted voltages in the measurement system that are common to both amplifier input terminals. They produce a shift in the zero baseline of the signal to be amplified.

Common Mode Rejection Ratio (CMRR) — The CMRR is the ratio of the amplifier signal voltage gain to the common-mode voltage gain. CMRR is often useful in estimating errors in amplifier output.

Offset — Offset voltages and currents are applied to the amplifier input terminals to compensate for systematically unbalanced input stages.

The various parameters involved in offset compensation are the following:

> **Input Bias Current** — A current supplied to compensate for unequal bias currents in input stages. It is equal to one-half the sum of the currents entering the separate input terminals.
>
> **Input Offset Current** — The difference between the separate currents entering the input terminals.
>
> **Input Offset Current Drift** — The ratio of the change of input offset current to a change in temperature.
>
> **Input Offset Voltage** — The voltage applied to achieve a zero amplifier output when the input signal is zero.
>
> **Input Offset Voltage Drift** — The ratio of the change of input offset voltage to a change in temperature.
>
> **Output Offset Voltage** — The voltage across the amplifier output terminals when the input terminals are grounded.
>
> **Power Supply Rejection Ratio** — The ratio of the change in input offset voltage to the corresponding change in a given power supply voltage, with all other power supply voltages held fixed.
>
> **Slew Rate** — The maximum time rate of change of the amplifier output voltage under large-signal (usually square-wave) conditions. Slew rate usually applies to the slower of the leading-edge and trailing-edge responses.

Nonlinearity — As with other components, actual amplifier response may depart from the assumed output-versus-input curve. Unlike gain errors, which are uniform differences between expected output-versus-input, nonlinearity errors are point-by-point differences in actual-versus-expected response over the range of input signal levels and frequencies. Nonlinearity error consists of the disagreement between the characteristic signature of an amplifier's response and its expected characteristic.

Noise — Noise generated within the amplifier that enters the signal path causes errors in amplifier output.

G.2.6. Interfacing (Amplifier—A/D Converter)

The amplified potential difference is applied across the A/D converter input terminals. The interface between the amplifier and the A/D converter is prone to the following error sources:

Interface Loss — The voltage at the A/D converter terminals suffers a drop due to the resistance of the connecting leads from the amplifier.

Noise — Electromagnetic noise is a factor for the connecting leads, while both the connecting leads and the A/D converter terminals are subject to thermal noise.

Crosstalk — Leakage between the input and the A/D converter may cause a decrease in the potential difference across the terminals.

G.2.7. Sampling (A/D Converter)

The potential difference applied to the A/D converter terminals is sampled. Samples are taken in windows (apertures) of time of finite duration. Several sources of error accompany the sampling process. (Refer to Section 4.4.3 for a detailed treatment of this subject.)

Sampling Rate — The input signal is sampled at a finite rate. Because of this, an incomplete representation of the waveform is available for analog-to-digital conversion. The sampled points that are converted to binary code for processing purposes must be eventually reconverted back to some form of analog or quasi-analog representation for information.

Aperture Time — A finite amount of time δt is required to sample the signal voltage V. During this time, the signal value changes by an amount δV.

Hysteresis — In sampling the signal, the sampling circuit must be able to respond to and recover from signal changes. If the rise times and recovery times of the sampling circuit are not negligible in comparison with the sampling aperture time, hysteresis errors occur.

Aliasing — An alias is an artifact of the sampling process masquerading as a signal component. As stated in Section 4.4.3, it is important to remember that once A/D conversion is completed, there is no way to know from the sampled data whether aliasing has occurred. Once sampled, there is no way to correct the data for alias-induced errors.

Digital Filtering — The output from the A/D converter contains coded amplitude variations that may represent alias frequencies. At this point, the signal has been digitized and the filtering process must take place in the digital domain.

The elimination of alias frequencies by digital filtering is not a free ride, however. The process introduces some error. Fortunately for the frequencies involved in the present example, these errors are negligible and will not be covered here. For cases where these errors are significant, the reader is encouraged to survey the literature on anti-aliasing filters.[23]

---

[23] See, for example, Himelblau, et al., *Handbook for Dynamic Data Acquisition and Analysis*, IES (Institute of Environmental Sciences) Design, Test, and Evaluation Division Recommended Practice 012.1; IES-RP-DTE012.1, 1994.

G.2.8. Sensing (A/D Converter)

In digitizing the analog potential difference, the sampled potential difference is applied across a network of analog components. These components are a set of analog sensing elements. The network outputs a coded pulse consisting of ones and zeros. The location of these ones and zeros is a function of the input signal level and the response of the network to this signal level.

Errors that may be present in sensing (responding to) the input signal level and converting this level into digital code are the following:

Gain — One type of A/D converter employs a ladder network of resistors. The configuration of the network is such that different signal levels cause different discrete responses. A major factor affecting the accuracy of these responses is the error in the value of the resistors in the network. This is because the voltage drop (negative gain) across each component resistor is a function of the signal level and the component's DC resistance.

Noise — As expected, stray voltages are sensed along with the signal voltage and contribute to the voltage level applied to the network. In addition, thermal fluctuations in components cause fluctuations in voltage drops.

G.2.9. Quantizing (A/D Converter)

The potential drop (or lack of a potential drop) across each element of the A/D converter sensing network produces either a “1” or “0”. This response constitutes a “bit” in the binary code that represents the sampled value. The position of the bit in the code is determined by which network element originated it.

Even if no errors were present in sampling and sensing the input signal, errors would still be introduced by the discrete nature of the encoding process. Suppose, for example, that the full-scale signal level (dynamic range) of the A/D converter is A volts. If n bits are used in the encoding process, then a voltage V can be resolved into 2^n discrete steps, each of size $A/2^n$. The error in the voltage V is thus

$$\varepsilon(V) = V - m\frac{A}{2^n},$$

where m is some integer determined by the sensing function of the A/D converter. As will be discussed later, the uncertainty associated with each step is one-half the value of the magnitude of the step. Consequently, the uncertainty inherent in quantizing a voltage A is $(1/2)(A/2^n)$, or $A/2^{n+1}$. This is embodied in the expression

$$V_{quantized} = V_{sensed} \pm \frac{A}{2^{n+1}}.$$

G.2.10. Data Reduction and Analysis (Data Processor)

The quantized output from the A/D converter is input to a data processor. Since the output is digital, the interface between the A/D converter and the data processor will be assumed not to constitute an error source. The data processor converts the binary coded number to a value and applies any correction factors that may be appropriate.

Two of the principal sources of error in this process are *correction-factor error* and *data-reduction error*.

Correction-Factor Error — The correction factor applied to the digitally encoded voltage difference attempts to correct for losses that occur between the reference junction and the data processor. Uncertainties in estimating these losses may lead to errors in the correction factors.

Data-Reduction Error — In converting the corrected value for the voltage difference into a temperature difference, the data processor attempts to solve the equation

$$\Delta V = (a_1 - b_1)\Delta T + (a_2 - b_2)(\Delta T)^2 + (a_3 - b_3)(\Delta T)^3 + \cdots .$$

In arriving at the solution, the series is truncated at some polynomial order. This truncation leads to a discrepancy between the solved-for temperature difference and the actual temperature difference.

For example, suppose that the series is truncated to second order. Then the data processor solution for the temperature difference becomes

$$\Delta T = -\frac{1}{2}\left(\frac{a_1 - b_1}{a_2 - b_2}\right) + \sqrt{\frac{1}{4}\left(\frac{a_1 - b_1}{a_2 - b_2}\right)^2 + \frac{\hat{V}_C - \hat{V}_A}{a_2 - b_2}} + \mathrm{O}(3),$$

where the quantities $\hat{V}_C$ and $\hat{V}_A$ are corrected values for V_C and V_A, and $\mathcal{O}(3)$ represents the error due to neglecting third order and higher terms.

G.2.11. Decoding (Data Processor)

The output of the data processor is a corrected result that is displayed as a decimal number. The following error source is relevant in developing and displaying this number.

Binary to Decimal Conversion — Suppose that the digital "resolution" of the binary encoded signal is $A/2^n$. Suppose further that the full-scale value Data Processor readout is S and that m digits are displayed. Then the resolution of the decimal display of the Data Processor is $S/10^m$. Another way of saying this is that the input to the Data Processor is a multiple of steps of size

$$h_b = \frac{A}{2^n},$$

while the decimal encoded display is presented in steps of size

$$h_d = \frac{S}{10^m}.$$

What this means is that a binary encoding of a voltage V into a representation $V' = 2^x h_b$ will be translated into a decimal representation $V'' = 10^y h_d$, where x and y are integers. The quantization error that results from expressing an analog value first as a binary coded value and second as a decimal coded value is the sum of these two errors:

$$\text{Quantization error} = \pm (h_b \pm h_d)/2 = \pm \frac{A}{2^{n+1}} \pm \frac{S/2}{10^m} \quad .$$

G.3 Identifying the Measurement Process Errors

Measurement process errors arise from the measurement procedure, measuring environment, measurement system operation, and the perception and interpretation of measurement results. These errors can be broadly grouped in the following categories:

Measuring Parameter Precision Error — This error is due to random changes in the measurement system output with the input held fixed. It is observed during random sampling in which successive sampled measurements differ randomly with respect to sign and magnitude.

Measurand Precision Error — This error is due to short-term random variations in the measurand that occur during the taking of a measurement sample. Note that it is necessary to have a basic understanding of the measurand so that random variations are not mistakenly interpreted as errors — i.e., the variations may be a dynamic characteristic of the phenomenon being measured or measurand anomalies.

Precision Error — This is the combined precision error due to measuring-parameter and measurand fluctuations. This error has a category in its own right in that random measuring-parameter and measurand errors are often not distinguishable as separate entities. In many cases, what is observed or estimated is instead their combined effect.

Ancillary Error — Ancillary error is due to errors or instabilities in such ancillary equipment as power supplies, and secondary monitoring devices. For example, if temperature corrections are applied to measured values, then the error in a given temperature measurement constitutes an ancillary error.

Operator Error — Operator error occurs as a result of a discrepancy between the measured value provided by a measuring system and the perception of this value.

G.3.1. Precision Error

Precision error cannot be estimated directly. Instead, the error is acknowledged and the resultant of uncertainty based on a sample of measurements is computed.

In cases where samples of data are not available, yet an estimate of precision uncertainty is needed, it may suffice to infer the uncertainty from estimated limits that are assumed to bound the error with some degree of confidence.

G.3.2. Ancillary Error

Amplifier — Suppose that amplifier gain is dependent on temperature according to the equation

$$p_{51} = p_{51}^0 + \kappa(T - T_0) \, ,$$

where κ is a coefficient whose units are volts/°C, T is the ambient temperature, and T_0 is the nominal or calibration temperature for the amplifier. Then the error in amplifier gain, $\varepsilon(p_{51})$ should be written

$$\begin{aligned}\varepsilon(p_{51}) &= \varepsilon\left[p_{51}^0 + \kappa(T - T_0)\right] \\ &= \varepsilon(p_{51}^0) + \kappa\varepsilon(T) + (T - T_0)\varepsilon(\kappa).\end{aligned}$$

The terms $\kappa\varepsilon(T) + (T - T_0)\varepsilon(\kappa)$ are process error terms. The term $\varepsilon(T)$ arises from errors in measuring or estimating the value of the ambient temperature used in the equation to compute amplifier gain. The term $\varepsilon(\kappa)$ arises from errors in estimating the temperature coefficient κ. This last term can often be ignored.

Noise — The error in the outputs of several of the system stages includes a component due to noise. Since noise is dependent on temperature, estimating its value involves knowing ambient and operating temperatures. Errors in these ancillary measurements of temperature appear as process errors.

G.3.3. Operator Error

In a system employing an analog display, operator error may arise from parallax in lining up a meter needle relative to marked values or in interpolating "between the lines" in nonlinearly scaled displays. Since the system in this example provides a digital readout, operator error will be taken to be zero.

G.4 Methodology for Developing a Measurement System Error Model

In this treatment, systems are considered as collections of stages whose responses are functions of inputs from other stages and of parameters that characterize the stage and the measuring environment.

Representing the output of the ith stage of a system by Y_i and the input by $\mathbf{X}_i$, the equation for each stage is

$$Y_i = Y_i(\mathbf{X}_i, \mathbf{p}_i),$$

where the vector $\mathbf{p}$ is the ith stage's parameter vector. (Note that, for a series system, $\mathbf{X}_i = Y_{i-1}$.)

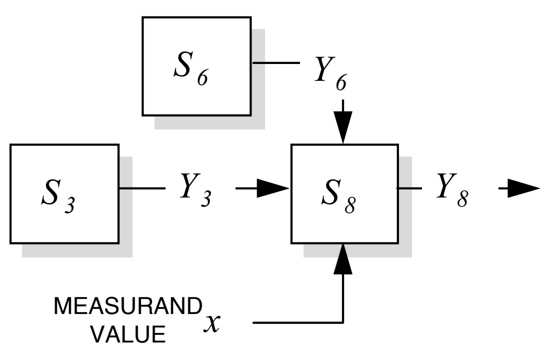


FIGURE G.2 — THE MEASUREMENT MODEL.

The output of the 8th stage is a function of the parameters of the stage and of the input vector $\mathbf{X} = (x, Y_3, Y_6)$.

The output of the measurement system, denoted $y(\mathbf{Y}|x)$, differs from the measurand by an error:

$$\varepsilon(\mathbf{Y} \mid x) = y(\mathbf{Y} \mid x) - x$$

This error is a function of the individual responses of the measurement system and of the errors in these responses. This functional relationship is developed using a Taylor series expansion. For systems whose component errors are small relative to the outputs of the stages, the expansion can be terminated at first order in the error components.

In most cases, the output of the system will be the output of the nth stage. For these systems, the measurement error is given by (the variable Y_0 is the measurand value x)

$$\varepsilon(\mathbf{Y}|x) = \sum_{i=1}^{q_n - 1} \left(\frac{\partial Y_n}{\partial X_i} \right) \varepsilon_i + \sum_{j=1}^{m_n} \left(\frac{\partial Y_n}{\partial p_{nj}} \right) \varepsilon(p_{nj}) = \sum_{i=0}^{n-1} \left(\frac{\partial Y_n}{\partial Y_i} \right) \varepsilon_i + \sum_{j=1}^{m_n} \left(\frac{\partial Y_n}{\partial p_{nj}} \right) \varepsilon(p_{nj}),$$

where q_n is the number of inputs to the nth stage, and where each error component ε_i is expressed in terms of the errors of other system responses and of the errors of its characterizing parameters:

$$\varepsilon_i = \sum_{k \neq i} \left(\frac{\partial Y_i}{\partial Y_k} \right) \varepsilon_k + \sum_{j=1}^{m_i} \left(\frac{\partial Y_i}{\partial p_{ij}} \right) \varepsilon(p_{ij}).$$

The quantity m_i is the number of components of the parameter vector for the ith stage and p_{ij} is the jth component.

This method of establishing system errors will be illustrated in an example. In the example, an error model will be developed from which the computation of measurement uncertainty can be made. The overall system uncertainty will be expressed in terms of the uncertainties of component uncertainties derived form component errors.

G.5 Developing the System Error Model

Referring to the previously discussed hypothetical temperature measurement system, we can construct the following system block diagram:

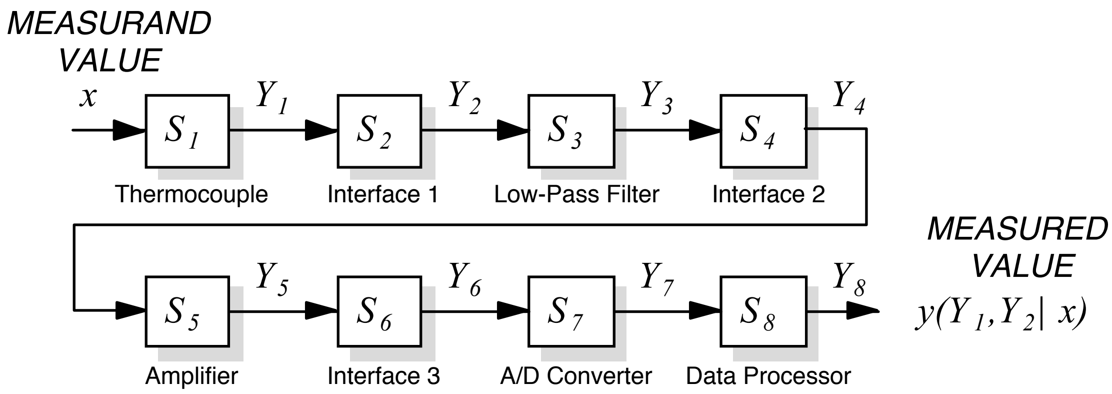


FIGURE G.3 — THE TEMPERATURE MEASUREMENT SYSTEM MODEL.

Thermocouple Output (Y_1) — The relevant parameters are

p_{11}	Sensitivity (temperature to voltage)
p_{12}	Thermocouple/Reference Junction hysteresis
p_{13}	Thermocouple non-linearity
p_{14}	Noise
p_{15}	Junction temperature deviation

Assuming that hysteresis and nonlinearity can be expressed in terms of percentage of measurand value, the output is given by

$$\begin{aligned} Y_1 &= (p_{11} + p_{13})(1 + p_{12})(x + p_{15}) + p_{14} \\ &= Y_1^0 + \varepsilon_1 . \end{aligned}$$

Using the general error equations, the error in Y_1 is given by

$$\begin{aligned} \varepsilon_1 &= \varepsilon(\mathbf{p}_1 | x) \\ &= \sum_{j=1}^{4} \frac{\partial Y_1}{\partial p_{1j}} \varepsilon(p_{1j}) + \frac{\partial Y_1}{\partial x} \varepsilon(x) \\ &= (x + p_{15})\{(1 + p_{12})[\varepsilon(p_{11}) + \varepsilon(p_{13})] + (p_{11} + p_{13})\varepsilon(p_{12})\} \\ &\quad + (p_{11} + p_{13})(1 + p_{12})[\varepsilon(p_{15}) + \varepsilon(x)] + \varepsilon(p_{14}) . \end{aligned}$$

If measurand fluctuations are not a factor, then $\varepsilon(x)$ can be set to zero and

$$\begin{aligned} \varepsilon_1 = {} & (x + p_{15})\{(1 + p_{12})[\varepsilon(p_{11}) + \varepsilon(p_{13})] \\ & + (p_{11} + p_{13})\varepsilon(p_{12})\} + (p_{11} + p_{13})(1 + p_{12})\varepsilon(p_{15}) + \varepsilon(p_{14}) . \end{aligned}$$

It may be that some simplification can be made at this point. For example, suppose that

$$p_{13} << p_{11} \, , \; p_{12} << 1 \text{ and } p_{15} << x .$$

Then

$$\varepsilon_1 = x[\varepsilon(p_{11}) + p_{11}\varepsilon(p_{12}) + \varepsilon(p_{13})] + \varepsilon(p_{14}) + p_{11}\varepsilon(p_{15}) .$$

Interface 1 — Thermocouple to Filter (Y_2) — The relevant parameters are

p_{21} Interface loss factor
p_{22} Crosstalk
p_{23} Noise

The input to the Low Pass Filter from Interface 1 is

$$Y_2 = (1 + p_{22})(1 + p_{21})Y_1 + p_{23}\,.$$

Using the general model described earlier, the error in Y_2 is found to be

$$\begin{aligned}\varepsilon_2 &= \varepsilon(\mathbf{p}_2, Y_1|x) \\ &= \frac{\partial Y_2}{\partial Y_1}\varepsilon_1 + \sum_{j=1}^{3}\frac{\partial Y_2}{\partial p_{2j}}\varepsilon(p_{2j}) \\ &\cong (1 + p_{22})(1 + p_{12})\varepsilon_1 + Y_1^0[(1 + p_{22})\varepsilon(p_{21}) + (1 + p_{21})\varepsilon(p_{22})] + \varepsilon(p_{23})\,.\end{aligned}$$

At this point, we seek to simplify the analysis, as we did in the previous step, retaining only terms considered to be significant. For instance, suppose that the interface loss and the crosstalk parameters *are* small relative to unity. If so, the above expression becomes, to first order in error terms,

$$\varepsilon_2 \cong \varepsilon_1 + Y_1^0[\varepsilon(p_{21}) + \varepsilon(p_{22})] + \varepsilon(p_{23})\,.$$

Substituting for ε_1 obtained in the previous step yields

$$\varepsilon_2 \cong x[\varepsilon(p_{11}) + p_{11}\varepsilon(p_{12}) + \varepsilon(p_{13})] + \varepsilon(p_{14}) + Y_1^0[\varepsilon(p_{21}) + \varepsilon(p_{22})] + \varepsilon(p_{23})\,.$$

At this point, we observe that substituting error terms from previous steps can lead to equations at subsequent stages that become extremely complicated. This argues that the general expressions for error developed earlier may be appropriate for a computer-based error model, but can be cumbersome if doing analyses by hand. The extent of manual processing and mathematical bookkeeping becomes quickly prohibitive. For this reason, in what follows, we will not substitute error expressions from previous stages in writing the outputs of successive stages.

Low Pass Filter (Y_3) — The parameters are

p_{31} Filter signal attenuation
p_{32} Filter noise
p_{33} Cut-off frequency, f_c
p_{34} Maximum frequency output, f_n

The output of the filter is given by

$$Y_3 = \begin{cases} (1+p_{31})Y_2 & , f \le f_c \\ (1+p_{31})Y_2 + \dfrac{p_{32}-(1+p_{31})Y_2}{p_{34}-p_{33}}(f - p_{33}) & , f_c \le f \le f_n \\ p_{32} & , f \ge f_n \,. \end{cases}$$

Where the variable f is the input frequency. Applying the usual expressions gives

$$\varepsilon_3 = \begin{cases} \dfrac{\partial Y_3}{\partial Y_2}\varepsilon_2 + \dfrac{\partial Y_3}{\partial p_{31}}\varepsilon(p_{31}) & , f \le f_c \\ \dfrac{\partial Y_3}{\partial Y_2}\varepsilon_2 + \displaystyle\sum_{i=1}^{4}\dfrac{\partial Y_3}{\partial p_{3i}}\varepsilon(p_{3i}) & , f_c \le f \le f_n \\ \varepsilon(p_{32}) & , f \ge f_n \,, \end{cases}$$

which becomes, to first order,

$$\varepsilon_3 \cong \begin{cases} (1+p_{31})\varepsilon_2 + Y_2^0\varepsilon(p_{31}) & , f \le f_c \\ \dfrac{f_n - f}{f_n - f_c}[(1+p_{31})\varepsilon_2 + Y_2^0\varepsilon(p_{31})] + \dfrac{f - f_c}{f_n - f_c}\varepsilon(p_{32}) - \dfrac{p_{32}-(1+p_{31})Y_2^0}{(f_n - f_c)^2}[(f_n - f)\varepsilon(f_c) + (f - f_c)\varepsilon(f_n)], & f_c \le f \le f_n \\ \varepsilon(p_{32}) & , f \ge f_n \,, \end{cases}$$

where the parameters p_{33} and p_{34} have been replaced by f_c and f_n, respectively. If errors in these frequencies can be ignored, which is usually the case, then the above result can be greatly simplified. In addition, if the filter attenuation $p_{31} << 1$, further simplification is possible. The final expression, accurate to first order is given by

$$\varepsilon_3 \cong \begin{cases} \varepsilon_2 + Y_2^0\varepsilon(p_{31}) & , f \le f_c \\ \dfrac{f_n - f}{f_n - f_c}[\varepsilon_2 + Y_2^0\varepsilon(p_{31})] + \dfrac{f - f_c}{f_n - f_c}\varepsilon(p_{32}), & f_c \le f \le f_n \\ \varepsilon(p_{32}) & , f \ge f_n \,. \end{cases}$$

Note that the parameter p_{31} is important in describing the roll-off of the filter. It constitutes an error source in that errors in its value introduce a departure of the roll-off from the assumed or nominal value. In the present discussion, this departure can be thought of as a non-linearity error in that it represents a discrepancy between assumed filter performance and actual filter performance. Strictly speaking, non-linearity error would also include error due to a departure of the filter roll-off curve from the assumed straight line. Ordinarily, such errors are thought to be small enough to ignore.

Interface 2 — Filter to Amplifier (Y_4) — The parameters are

p_{41}	Interface loss factor
p_{42}	Crosstalk
p_{43}	Noise

The input to the Amplifier from Interface 2 is

$$Y_4 = (1 + p_{41})(1 + p_{42})Y_3 + p_{43} .$$

The error in this input is

$$\begin{aligned}\varepsilon_4 &= \frac{\partial Y_4}{\partial Y_3}\varepsilon_3 + \sum_{i=1}^{3} \frac{\partial Y_4}{\partial p_{4i}}\varepsilon(p_{4i}) \\ &= (1 + p_{41})(1 + p_{42})\varepsilon_3 + Y_3^0[(1 + p_{42})\varepsilon(p_{41}) + (1 + p_{41})\varepsilon(p_{42})] + \varepsilon(p_{43}) .\end{aligned}$$

Assuming that $p_{41} << 1$, and $p_{42} << 1$, permits us to write

$$\varepsilon_4 \cong \varepsilon_3 + Y_3^0[\varepsilon(p_{41}) + \varepsilon(p_{42})] + \varepsilon(p_{43}) .$$

Amplifier (Y_5) — The parameters are

p_{51}	Amplifier gain	p_{55}	Non-linearity
p_{52}	Gain instability	p_{56}	Common mode voltage
p_{53}	Normal mode voltage	p_{57}	Noise
p_{54}	Offset		

The output is given by

$$Y_5 = (p_{51} + p_{52} + p_{55})(Y_4 + p_{53} + p_{54} + p_{56}) + p_{57} .$$

The error in the amplifier output is

$$\begin{aligned}\varepsilon_5 &= \frac{\partial Y_5}{\partial Y_4}\varepsilon_4 + \sum_{i=1}^{7} \frac{\partial Y_5}{\partial p_{5i}}\varepsilon(p_{5i}) \\ &= (p_{51} + p_{52} + p_{55})[\varepsilon_4 + \varepsilon(p_{53}) + \varepsilon(p_{54}) + \varepsilon(p_{56})] \\ &\quad + (Y_4^0 + p_{53} + p_{54} + p_{56})[\varepsilon(p_{51}) + \varepsilon(p_{52}) + \varepsilon(p_{55})] + \varepsilon(p_{57}) .\end{aligned}$$

Assuming that $p_{52} << p_{51}$, $p_{55} << p_{51}$, $p_{53} << Y_4^0$, $p_{54} << Y_4^0$, and $p_{56} << Y_4^0$, the error in the amplifier output can be approximated by

$$\varepsilon_5 \cong p_{51}[\varepsilon_4 + \varepsilon(p_{53}) + \varepsilon(p_{54}) + \varepsilon(p_{56})] + Y_4^0[\varepsilon(p_{51}) + \varepsilon(p_{52}) + \varepsilon(p_{55})] + \varepsilon(p_{57}) .$$

Interface 3 — Amplifier to A/D Converter (Y_6) — The parameters for Interface 3 are

p_{61}	Interface loss factor
p_{62}	Crosstalk
p_{63}	Noise

The input to the A/D Converter from Interface 3 is

$$Y_6 = (1 + p_{61})(1 + p_{62})Y_5 + p_{63} .$$

The expression for the error in this input is

$$\begin{aligned}\varepsilon_7 &= \frac{\partial Y_6}{\partial Y_5}\varepsilon_5 + \sum_{i=1}^{3}\frac{\partial Y_6}{\partial p_{6i}}\varepsilon(p_{6i}) \\ &= (1 + p_{61})(1 + p_{62})\varepsilon_5 + Y_5^0[(1 + p_{62})\varepsilon(p_{61}) + (1 + p_{61})\varepsilon(p_{62})] + \varepsilon(p_{63}) .\end{aligned}$$

Making the usual assumptions $p_{61} << 1$, and $p_{62} << 1$, yields

$$\varepsilon_7 = \varepsilon_5 + Y_5^0[\varepsilon(p_{61}) + \varepsilon(p_{62})] + \varepsilon(p_{63}) .$$

A/D Converter (Y_7) — The parameters are

p_{71}	Analog loss	p_{72}	Aperture time error $(\delta V/V)$
p_{73}	Sampling rate error $(\delta V/V)$	p_{74}	Quantization error $(\delta V/V)$
p_{75}	Linearity error $(\Delta V/V)$	p_{76}	Noise level

The output is given by

$$Y_7 = (1 + p_{75})(1 + p_{73})(1 + p_{72})(1 + p_{71})Y_6 + p_{74} + p_{76} ,$$

for which

$$\begin{aligned}\varepsilon_7 &= \frac{\partial Y_7}{\partial Y_6}\varepsilon_6 + \sum_{i=1}^{5}\frac{\partial Y_7}{\partial p_{7i}}\varepsilon(p_{7i}) \\ &\cong (1 + p_{75})(1 + p_{73})(1 + p_{72})(1 + p_{71})\varepsilon_6 + Y_6^0[(1 + p_{75})(1 + p_{73})(1 + p_{72})\varepsilon(p_{71}) \\ &\quad + (1 + p_{75})(1 + p_{73})(1 + p_{71})\varepsilon(p_{72}) + (1 + p_{75})(1 + p_{72})(1 + p_{71})\varepsilon(p_{73}) \\ &\quad + (1 + p_{73})(1 + p_{72})(1 + p_{71})\varepsilon(p_{75})] + \varepsilon(p_{74}) + \varepsilon(p_{76}) .\end{aligned}$$

With the usual assumptions regarding relative magnitudes of parameters (except for p_{71}) , the A/D Converter error can be written

$$\varepsilon_7 \cong (1 + p_{71})\varepsilon_6 + Y_6^0\{\varepsilon(p_{71}) + (1 + p_{71})[\varepsilon(p_{72}) + \varepsilon(p_{73}) + \varepsilon(p_{75})]\} + \varepsilon(p_{74}) + \varepsilon(p_{76}) .$$

Data Processor (Y_8) — The parameters are

p_{81}	Voltage to temperature conversion factor
p_{82}	Resolution
p_{83}	Correction factor (applied as a compensation for losses and gains in the signal path)

The output of the measurement system is given by

$$Y_8 = p_{83} p_{81} Y_7 + p_{82} \text{ ,}$$

with error

$$\begin{aligned}\varepsilon_8 &= \frac{\partial Y_8}{\partial Y_7}\varepsilon_7 + \sum_{i=1}^{3}\frac{\partial Y_8}{\partial p_{8i}}\varepsilon(p_{8i}) \\ &= p_{83}p_{81}\varepsilon_7 + Y_7^0[p_{83}\varepsilon(p_{81}) + p_{81}\varepsilon(p_{83})] + \varepsilon(p_{82})\end{aligned}$$

The correction factor error is composed of two parts. First is the error due to any discrepancy between the computed and actual signal path gains and losses. Second is the error due to the fact that the correction is applied digitally and is subject to quantization error. Thus

$$\varepsilon(p_{83}) = \varepsilon_{p_{83}}^{analytical} + \varepsilon_{p_{83}}^{quantization} \text{ .}$$

One step remains to complete the development of the system error equation. By rearranging terms in the expression for ε_8, we can write

$$\varepsilon_8 \cong p_{81}p_{83}\varepsilon_7 + \left[\frac{\varepsilon(p_{81})}{p_{81}} + \frac{\varepsilon(p_{83})}{p_{83}}\right]p_{81}p_{83}Y_7^0 + \varepsilon(p_{82}) \text{ .}$$

The term $p_{81}p_{83}Y_7$ is the signal processor's estimate of the measurand value x. If the measuring system is accurate to first order in errors, and the signal frequency is less than f_c, then we can write

$$\varepsilon_8 \cong p_{81}p_{83}\varepsilon_7 + x\left[\frac{\varepsilon(p_{81})}{p_{81}} + \frac{\varepsilon(p_{83})}{p_{83}}\right] + \varepsilon(p_{82}) \text{ .}$$

G.6 Methodology for Developing a Measurement System Uncertainty Model

The system error model forms the backbone of the system uncertainty model. Estimating the measurement uncertainty of a system involves building the uncertainty on the system error model framework, term by term. The sources of error become sources of uncertainty. The contribution of each source to the total uncertainty is governed by the coefficient of the source in the model. These coefficients are obtained directly from the system equations.

For example, consider the measurement of the velocity v of a body. The measurement is decomposed into measurements of distance d and time t. The system equation is

$$v = \frac{d}{t} \text{ .}$$

Using the methodology described above for constructing the error model yields

$$\varepsilon(v) = \frac{\partial v}{\partial d}\varepsilon(d) + \frac{\partial v}{\partial t}\varepsilon(t)$$
$$= \left[\frac{\varepsilon(d)}{d} - \frac{\varepsilon(t)}{t}\right] v\,.$$

Developing the uncertainty model from the error model involves first writing an expression for the *statistical variance* of the error in *v*. It will be worthwhile to pause here and discuss some of the properties of variances.

G.6.1 Statistical Variance

In general, the variance in the sum of two quantities *x and y* is given by

$$\mathrm{var}(x + y) = \mathrm{var}(x) + \mathrm{var}(y) + 2\,\mathrm{cov}(x, y),$$

where the term cov(*x*,*y*) is the "covariance" of *x* and *y*. The nature and computation of the covariance is discussed in detail in the ISO and NIST guidelines and will not be covered here. This is for two reasons. First, developing an understanding of the basic approach to uncertainty analysis, which is the intent of this discussion, will not be overly enhanced by delving into what can turn into an involved and difficult subject.[24] Second, many, if not most, error sources exhibit a property called "statistical independence." Uncertainties in statistically independent sources do not influence one another with the result that their covariance vanishes. So, in most cases, the covariance is zero anyway. This allows us to concentrate almost exclusively on the variance.[25]

There is a simple rule that governs variances that is extremely useful in developing uncertainty estimates. This rule states that, if *a* and *b* are constants (or, if you will, "coefficients"), and if *x* and *y* are statistically independent variables, then

$$\mathrm{var}(ax + by) = a^2\,\mathrm{var}(x) + b^2\,\mathrm{var}(y)\,.$$

Applying this rule to the error model of the velocity measurement example above gives

$$\mathrm{var}(\varepsilon(v)) = \left(\frac{\partial v}{\partial d}\right)^2 \mathrm{var}(\varepsilon(d)) + \left(\frac{\partial v}{\partial t}\right)^2 \mathrm{var}(\varepsilon(t))$$
$$= (v/d)^2\,\mathrm{var}(\varepsilon(d)) + (v/t)^2\,\mathrm{var}(\varepsilon(t))\,.$$

The variance in the error of a quantity is just the variance in the quantity itself. Thus, for a component of error *x*,

$$\mathrm{var}(\varepsilon(x)) = \mathrm{var}(x)$$
$$\equiv \sigma_x^2\,.$$

---

24 The subject involves the concept of expectation value. The expectation value of a variable is obtained by integrating or summing the product of the variable and its probability density function over all values accessible to the variable. The expectation value for a variable *x* is written $E(x)$. The covariance of two variables *x* and *y* is written $E\{[x - E(x)][y - E(y)]\}$.

25 The variance of a variable *x* is the expectation value $E\{[x - E(x)]^2\}$.

With this in mind, the variance of the velocity measurement is written

$$
\begin{aligned}
\sigma_v^2 &= \left(\frac{\partial v}{\partial d}\right)^2 \sigma_d^2 + \left(\frac{\partial v}{\partial t}\right)^2 \sigma_t^2 \\
&= (v/d)^2\sigma_d^2 + (v/t)^2\sigma_t^2 \text{ , or} \\
\sigma_v^2 &= (1/t)^2\sigma_d^2 + (d/t^2)^2\sigma_t^2 \text{ .}
\end{aligned}
$$

This simple example contains the seeds of uncertainty analysis in general. Using the expression for the output of the *i*th system stage given earlier a general expression for the variance in this output can be constructed. If the errors of input stages are statistically independent of one another, then this expression can be written

$$
\sigma_i^2 = \sum_{k \neq i} \left(\frac{\partial Y_i}{\partial Y_k}\right)^2 \sigma_i^2 + \sum_{j=1}^{m_i} \left(\frac{\partial Y_i}{\partial p_{ij}}\right)^2 \sigma_{p_{ij}}^2 \text{ .}
$$

This variance provides the form of the general uncertainty model for a system with statistically independent error sources.

G.6.2 Relationship of Standard Deviation to System Uncertainty

The square root of the variance of a quantity is called the *standard deviation*. The standard deviation is an important parameter in defining the way that a quantity is statistically distributed, i.e., the way in which the values of the quantity are related to their probabilities of occurrence. In particular, the standard deviation is a measure of the spread of the values of the quantity around some reference point, such as a mean value, mode value or median value.

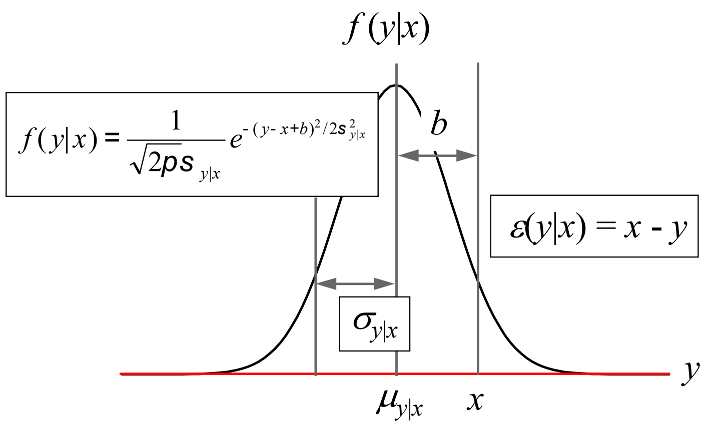


FIGURE G.4 — THE MEASUREMENT DISTRIBUTION.
The quantity b is the average bias of the measurement system.

In general, the larger the standard deviation, the greater the spread. This means that, with large standard deviations, values of a quantity tend not to be "localized," i.e., the confidence with which they are known tends to be low. Equating the word "confidence" with the less precise but more comfortable word "certainty," we see that the standard deviation for a quantity is related to its *un*certainty. In fact, in the ISO and NIST references, the standard deviation of a quantity is *equated* to its uncertainty. This means that we can write

$$\sigma_i^2 = \sum_{k \neq i} \left(\frac{\partial Y_i}{\partial Y_k} \right)^2 \sigma_i^2 + \sum_{j=1}^{m_i} \left(\frac{\partial Y_i}{\partial p_{ij}} \right)^2 \sigma_{p_{ij}}^2 .$$

G.7 Evaluating the Measurement Uncertainty

G.7.1 Thermocouple

From the expression for thermocouple error, the uncertainty in the thermocouple output is given by

$$\sigma_1 = \sqrt{x^2(\sigma_{p_{11}}^2 + p_{11}^2 \sigma_{p_{12}}^2 + \sigma_{p_{13}}^2) + \sigma_{p_{14}}^2 + p_{11}^2 \sigma_{p_{15}}^2} ,$$

where

$\sigma_{p_{11}}$ = $T \rightarrow V$ translation uncertainty

$\sigma_{p_{12}}$ = junction hysteresis uncertainty

$\sigma_{p_{13}}$ = thermocouple nonlinearity uncertainty

$\sigma_{p_{14}}$ = noise uncertainty

$\sigma_{p_{15}}$ = junction temperature uncertainty .

G.7.2 Interface 1 (Reference Junction—Low-Pass Filter)

The uncertainty in the signal passed input to the Low Pass Filter is given by

$$\sigma_2 = \sqrt{\sigma_1^2 + (Y_1^0)^2(\sigma_{p_{21}}^2 + \sigma_{p_{22}}^2) + \sigma_{p_{23}}^2} ,$$

where

$\sigma_{p_{21}}$ = interface loss uncertainty

$\sigma_{p_{22}}$ = crosstalk uncertainty

$\sigma_{p_{23}}$ = interface noise uncertainty,

$Y_1^0 = p_{11} x$.

G.7.3 Low-Pass Filter

The uncertainty in the output of the Low Pass Filter is given by

$$\sigma_3 = \begin{cases} \sqrt{\sigma_2^2 + (Y_2^0)^2 \sigma_{p_{31}}^2} & , f \le f_c \\ \sqrt{\left(\frac{f_n - f}{f_n - f_c} \right)^2 [\sigma_2^2 + (Y_2^0)^2 \sigma_{p_{31}}^2] + \left(\frac{f - f_c}{f_n - f_c} \right)^2 \sigma_{p_{32}}^2} & , f_c \le f \le f_n \\ \sigma_{p_{32}} & , f \ge f_n , \end{cases}$$

where

$\sigma_{p_{31}}$ = filter attenuation uncertainty

$\sigma_{p_{32}}$ = filter noise uncertainty

$Y_2^0 = Y_1^0 = p_{11} x$.

G.7.4 Interface 2 (Low-Pass Filter—Amplifier)

The uncertainty in the signal input to the amplifier is given by

$$\sigma_4 = \sqrt{\sigma_3^2 + (Y_3^0)^2(\sigma_{p_{41}}^2 + \sigma_{p_{42}}^2) + \sigma_{p_{43}}^2}\ ,$$

where

$$\sigma_{p_{41}} = \text{ interface loss uncertainty}$$
$$\sigma_{p_{42}} = \text{ crosstalk uncertainty}$$
$$\sigma_{p_{43}} = \text{ interface noise uncertainty.}$$

The variable Y_3^0 is given by

$$Y_3^0 = \begin{cases} p_{11}x\,, & f \le f_c \\ p_{11}x - \dfrac{p_{11}x}{f_n - f_c}(f - f_c)\,, & f_c \le f \le f_n \\ 0\,, & f \ge f_n\,. \end{cases}$$

G.7.5 Amplifier

The uncertainty in the amplifier output is

$$\sigma_5 \cong \sqrt{p_{51}^2(\sigma_4^2 + \sigma_{p_{53}}^2 + \sigma_{p_{54}}^2 + \sigma_{p_{56}}^2) + (Y_4^0)^2(\sigma_{p_{51}}^2 + \sigma_{p_{52}}^2 + \sigma_{p_{55}}^2) + \sigma_{p_{57}}^2}\ ,$$

where

$\sigma_{p_{51}}$ = amplifier gain uncertainty (includes process uncertainty)
$\sigma_{p_{55}}$ = amplifier non-linearity uncertainty
$\sigma_{p_{52}}$ = gain instability uncertainty
$\sigma_{p_{56}}$ = common mode voltage uncertainty
$\sigma_{p_{53}}$ = normal mode voltage uncertainty
$\sigma_{p_{57}}$ = amplifier noise uncertainty
$\sigma_{p_{54}}$ = offset uncertainty
Y_4^0 = $p_{51}Y_3^0$.

Recalling the discussion on process error, the uncertainty in the amplifier gain is given by

$$\sigma_{p_{51}} = \sqrt{(\sigma_{p_{51}}^0)^2 + \kappa^2\sigma_T^2 + (T - T_0)^2\sigma_\kappa^2}\ .$$

G.7.6 Interface 3 (Amplifier—A/D Converter)

The uncertainty of the input to the A/D converter is

$$\sigma_7 = \sqrt{\sigma_5^2 + (Y_5^0)^2(\sigma_{p_{61}}^2 + \sigma_{p_{62}}^2) + \sigma_{p_{63}}^2}\ ,$$

where

$\sigma_{p_{61}}$ = interface loss uncertainty

$\sigma_{p_{62}}$ = A / D input crosstalk uncertainty

$\sigma_{p_{63}}$ = interface noise uncertainty

$Y_5^0 = Y_4^0 = p_{51} Y_3^0$.

G.7.7 Sampling (A/D Converter)

Several sources of uncertainty are inherent in converting the analog voltage input to a digital representation. These include analog loss uncertainty, sampling rate uncertainty, aperture time uncertainty, quantization uncertainty, and noise uncertainty. Uncertainties for analog loss and noise can be obtained in a straightforward way from specifications for the A/D converter stage. Uncertainties due to sampling rate, aperture time, and quantization are more elusive and may require some extra computation.

The expression for sampling uncertainty is

$$\sigma_7 \cong \sqrt{[(1+(p_{71})^2]\sigma_6^2 + (Y_6^0)^2\{\sigma_{p_{71}}^2 + [1+(p_{71})^2](\sigma_{p_{72}}^2 + \sigma_{p_{73}}^2 + \sigma_{p_{74}}^2 + \sigma_{p_{75}}^2)\} + \sigma_{p_{76}}^2} \ .$$

where

$\sigma_{p_{71}}$ = D / A converter analog loss uncertainty

$\sigma_{p_{72}}$ = aperture time uncertainty

$\sigma_{p_{73}}$ = sampling rate uncertainty

$\sigma_{p_{74}}$ = quantization uncertainty

$\sigma_{p_{75}}$ = linearity uncertainty

$\sigma_{p_{76}}$ = noise uncertainty

$Y_6^0 = Y_5^0 = p_{51} Y_3^0$.

G.7.8 System Uncertainty

The uncertainty of the output of the data processor is the uncertainty in the measurement system. If we can make the accuracy claims that were made in the discussion on data reduction error, then this uncertainty can be written

$$\sigma_{system} \cong \sqrt{(p_{81}p_{83})^2\sigma_7^2 + x^2\left(\frac{\sigma_{p_{81}}^2}{p_{81}^2} + \frac{\sigma_{p_{83}}^2}{p_{83}^2}\right) + \sigma_{p_{82}}^2} \ .$$

where

$\sigma_{p_{81}}$ = data processor voltage to temperature conversion unceraintty

$\sigma_{p_{82}}$ = resolution uncertainty

$\sigma_{p_{83}}$ = error correction factor uncertainty

x = measurand value $\cong p_{81} p_{83} Y_7$.

G.8 Establishing the Standard Deviations for Uncertainty Components

Standard deviations will no be sought for the outputs of the various stages of the system.

G.8.1 Thermocouple

Suppose that the temperature to be measured varies sinusoidally in time over 20°C to 100°C with a frequency of from 0 to 10 Hz. Under these conditions, the normal Type K thermocouple sensitivity is 22.8 ±0.2 μV/°C. Because of the low frequency, we can ignore hysteresis. This leaves nonlinearity, noise, and junction temperature uncertainty.

In the temperature range of interest, for the differences in temperature under consideration, nonlinearity is negligible. In addition, since the resistance of the chromel and alumel leads is less than 1Ω, and since the temperature is near room temperature, the noise signal for a bandwidth of 10 Hz is on the order of 10^{-21} V — clearly negligible. As for noise induced by stray electromagnetic signals, at an upper frequency of 10 Hz, the noise generated by these is also negligible.

Using a typical specification of ±0.25°C for junction temperature and ±0.1°C for bath uniformity, the reference junction error limits are $\pm\sqrt{(0.25)^2+(0.10)^2}$ °C, or about ±0.27°C.

In summary, the thermocouple parameters are

Sensitivity (p_{11}):	22.8 $\pm$ 0.2 μV/°C
Junction Hysteresis (p_{12}):	$\simeq 0$
Non linearity (p_{13}):	$\simeq 0$
Noise (p_{14}):	$\simeq 0$
Junction Temperature (p_{15}):	0 °C $\pm$ 0.27 °C.

Assume that the ± error limits in these specifications are stated without an accompanying statistical confidence limit, as is often the case with specifications from equipment manufacturers. Without such a confidence limit or other supporting statistics, estimates of uncertainty obtained from these limits are heuristic in nature. Such estimates are referred to in the ISO and NIST guidelines as *Type B* estimates.[26]

> It should not be assumed that evaluations based on repeated observations are necessarily superior to evaluations obtained by other means. Type A evaluations of standard uncertainty are not necessarily more reliable than Type B evaluations, and in many practical measurement situations, the components obtained from Type B evaluations may be better known than the components obtained from Type A evaluations.

Obtaining Type B uncertainty estimates from the above data involves estimating what the probabilities of error containment are for the ± limits and making some assumptions as to how errors are distributed within these limits. For this example, we will assume that the ±0.2 μV/°C sensitivity limits bound sensitivity errors with approximately 99% probability and that the junction temperature limits of ±0.27°C bound errors from this source with 99.73% probability.

[26] As described in the ISO/TAG4/WG3 "Guide to the Expression of Uncertainty in Measurement," Type A estimates are those that are evaluated by applying statistical methods to a series of repeated observations—a posteriori. Type B estimates are other evaluations—subjective and otherwise—a priori.

We will also assume that sensitivity and junction temperature errors are normally distributed with zero mean within their respective limits. From statistical tables, the normal deviates of 2.576 and 3.000 are found for 99% and 99.73% significance levels. This means that ±0.2 μV/°C corresponds to 2.576 standard deviations from the mean for sensitivity errors, and that ±0.27°C corresponds to 3.000 standard deviations from the mean for junction temperature errors. The respective standard deviations are thus

$$\sigma_{p_{11}} = \sigma_{sensitivity} \cong \frac{0.2\mu V/^\circ C}{2.576} \cong 0.078\mu V/^\circ C, \text{ and}$$

$$\sigma_{p_{15}} = \sigma_{Junction} \cong \frac{0.27\,^\circ C}{3.000} \cong 0.09\,^\circ C.$$

From the expression for σ_1:

$$\sigma_1 = \sqrt{x^2(\sigma_{p_{11}}^2 + p_{11}^2\sigma_{p_{12}}^2 + \sigma_{p_{13}}^2) + \sigma_{p_{14}}^2 + p_{11}^2\sigma_{p_{15}}^2},$$

we see that

$$\sigma_1 = \sqrt{x^2(0.078\mu V/^\circ C)^2 + (22.8\mu V/^\circ C)^2(0.09^\circ C)^2}$$
$$\cong \sqrt{x^2(0.006/^\circ C^2) + 4.21}\ \mu V.$$

From this result, it is apparent that the maximum uncertainty occurs at the upper end of the temperature range (x = 100 °C). Inserting this number in the expression for σ_1 gives

$$\sigma_1 \cong 8.07\,\mu V.$$

Obviously, the dominant term is the σ_{11} term.

G.8.2 Interface 1 (Reference Junction—Low-Pass Filter)

The parameters for interface 1 are

p_{21} Interface loss factor
p_{22} Crosstalk
p_{23} Noise,

and the uncertainty is

$$\sigma_2 \cong \sqrt{\sigma_1^2 + x^2(\sigma_{p_{21}}^2 + \sigma_{p_{22}}^2) + \sigma_{p_{23}}^2}.$$

Ordinarily, the standard deviations would be estimated from heuristic data as was done in estimating thermocouple uncertainty. However, with the temperatures and frequencies under consideration in this example, these standard deviations can be considered negligible relative σ_1. Accordingly,

$$\sigma_2 \cong \sigma_1 \cong 8.07\,\mu V\ .$$

G.8.3 Low-Pass Filter

Reiterating from earlier, the uncertainty in the output of the Low Pass Filter is given by

$$\sigma_3 = \begin{cases} \sqrt{\sigma_2^2 + (Y_2^0)^2 \sigma_{p_{31}}^2} & , f \le f_c \\ \sqrt{\left(\dfrac{f_n - f}{f_n - f_c}\right)^2 [\sigma_2^2 + (Y_2^0)^2 \sigma_{p_{31}}^2] + \left(\dfrac{f - f_c}{f_n - f_c}\right)^2 \sigma_{p_{32}}^2} & , f_c \le f \le f_n \\ \sigma_{p_{32}} & , f \ge f_n\ , \end{cases}$$

where

$$\sigma_{p_{31}} = \text{ filter attenuation uncertainty}$$
$$\sigma_{p_{32}} = \text{ filter noise uncertainty}$$
$$Y_2^0 = Y_1^0 = p_{11}x\ .$$

Assume that $f_c >> f$ for the 0 to 10 Hz range of interest in this example. Then

$$\sigma_3 = \sqrt{\sigma_2^2 + (Y_2^0)^2 \sigma_{p_{31}}^2}$$
$$\cong \sqrt{\sigma_2^2 + x^2 p_{11}^2 \sigma_{p_{31}}^2}\ .$$

Suppose that we have a specification for the non linearity of the filter of $\pm$ 0.15% of input signal level. From calls to the manufacturer, we determine that these are 95% confidence limits for the filter at the upper end of our frequency range (i.e., 10 Hz). In other words, the limits $\pm$ 1.5% bound errors in filter linearity with 95% probability. We again assume a normal distribution with zero mean for these errors and consult a table of normal deviates, where we find that 95% confidence corresponds to about 1.960 standard deviations from the mean. The standard deviation for filter linearity errors is thus

$$\sigma_{filter} = \sigma_{p_{31}} = \frac{0.0015}{1.960} \cong 7.7 \times 10^{-4}\ .$$

Consequently, the uncertainty in the filter output is

$$\sigma_3 = \sqrt{\sigma_2^2 + (Y_2^0)^2 \sigma_{p_{31}}^2}$$
$$\cong \sqrt{(8.07)^2 + (100)^2 (22.80)^2 (5.86 \times 10^{-7})}\ \mu V$$
$$= 8.26\,\mu V\ \text{(cumulative uncertainty)}\ .$$

G.8.4 Interface 2 (Low-Pass Filter—Amplifier)

The uncertainty in the signal input to the amplifier is given by

$$\sigma_4 \cong \sqrt{\sigma_3^2 + p_{11}^2 x^2 (\sigma_{p_{41}}^2 + \sigma_{p_{42}}^2) + \sigma_{p_{43}}^2}\ ,$$

where

$$\sigma_{p_{41}} = \text{ interface loss uncertainty}$$
$$\sigma_{p_{42}} = \text{ crosstalk uncertainty}$$
$$\sigma_{p_{43}} = \text{ interface noise uncertainty.}$$

We assume interface loss to be negligible, as is crosstalk and noise. Consequently,

$$\sigma_4 \cong \sigma_3 \cong 8.26\,\mu V \text{ (cumulative uncertainty)}.$$

G.8.5 Amplifier

The uncertainty in the amplifier output is

$$\sigma_5 \cong \sqrt{p_{51}^2(\sigma_4^2 + \sigma_{p_{53}}^2 + \sigma_{p_{54}}^2 + \sigma_{p_{56}}^2) + (Y_4^0)^2(\sigma_{p_{51}}^2 + \sigma_{p_{52}}^2 + \sigma_{p_{55}}^2) + \sigma_{p_{57}}^2}\,,$$

where

$\sigma_{p_{51}}$ = amplifier gain uncertainty (includes process uncertainty)
$\sigma_{p_{55}}$ = amplifier non-linearity uncertainty
$\sigma_{p_{52}}$ = gain instability uncertainty
$\sigma_{p_{56}}$ = common mode voltage uncertainty
$\sigma_{p_{53}}$ = normal mode voltage uncertainty
$\sigma_{p_{57}}$ = amplifier noise uncertainty
$\sigma_{p_{54}}$ = offset uncertainty
Y_4^0 = $p_{51}Y_3^0$,

and where the uncertainty in the amplifier gain is given by

$$\sigma_{p_{51}} = \sqrt{(\sigma_{p_{51}}^0)^2 + \kappa^2\sigma_T^2 + (T - T_0)^2\sigma_\kappa^2}\,.$$

Assume that we have the following specifications:

p_{51} (amplifier gain)	=	20 dB $\pm$ 0.5%
p_{55} (amplifier non-linearity error)	=	$\pm$ 0.02%
p_{52} (gain instability error)	=	$\pm$ 0.25%
p_{56} (common mode rejection error)	=	$\pm$ 0.002% of common mode input[27]
common mode voltage	=	10μV (maximum)
p_{53} (normal mode voltage error)	=	0
p_{57} (amplifier noise level)	=	$\pm$ 2.5 μV
p_{54} (offset error)	=	$\pm$ 3.2 μV
κ (thermal gain coefficient)	=	2% / °C

The amplifier manufacturer has assured us that these specification are made with 95% confidence, corresponding to 1.960 standard deviations from the mean. Hence, noting that a 20 dB amplitude gain represents a factor of 10 increase, the uncertainties are

---

[27]Based on a common mode rejection ratio of 120 dB.

$\sigma^0_{p_{51}} \cong (.005)(10)/1.96 \cong .026$ $\quad\sigma_{p_{55}} = 1.02 \times 10^{-4}$

$\sigma_{p_{52}} = 0.001$ $\quad\sigma_{p_{56}} = 2 \times 10^{-4} \mu V$

$\sigma_{p_{53}} = 0 \ \mu V$ $\quad\sigma_{p_{57}} = 1.28 \ \mu V$

$\sigma_{p_{54}} = 1.63 \ \mu V$.

With regard to ancillary uncertainty, the ambient temperature is measured by a thermometer with the specifications $T \pm 0.1$ °C $\pm$ .5% of reading. At 100 °C, this translates to

$$\begin{aligned}\text{Temperature} &= T \pm \sqrt{(0.1)^2 + (.005 \times 100)^2} \ ^\circ C \\ &\cong T \pm 0.51 ^\circ C .\end{aligned}$$

If the error limits are stated with 95% confidence, then this specification corresponds to

$$\sigma_T \cong \frac{0.51}{1.96} {}^\circ C \cong 0.26 {}^\circ C .$$

For this example, we will assume that $\sigma_\kappa \cong 0$. Then

$$\begin{aligned}\sigma_{p_{51}} &= \sqrt{(0.026)^2 + (0.02)^2 (0.26)^2} \\ &\cong 0.027 ,\end{aligned}$$

i.e., the ancillary contribution to amplifier output uncertainty is small, but not negligible.

The total output uncertainty for this stage is

$$\begin{aligned}\sigma_5 &\cong \sqrt{(10)^2[\sigma_4^2 + (1.63\mu V)^2 + (2 \times 10^{-4} \mu V)^2] + (Y_4^0)^2[(0.027)^2 + (0.001)^2 + (1.02 \times 10^{-4})^2] + (1.28\mu V)^2} \\ &\cong \sqrt{100\sigma_4^2 + 267.3(\mu V)^2 + (0.027)^2 (Y_4^0)^2} .\end{aligned}$$

To a good approximation $Y_4^0 \cong p_{11} x$. Recalling that $p_{11} = 22.8$ μV/°C, and that we are using the maximum value of $x = 100$ °C, gives $Y_4^0 \cong 2280 \mu V$. Substituting this value and the value $\sigma_4 \cong 8.26 \mu V$, gives

$$\begin{aligned}\sigma_5 &\cong \sqrt{100(8.26)^2 + 265.7 + (0.027)^2 (2280)^2} \ \mu V \\ &\cong 104.3 \mu V \ \text{(cumulative uncertainty)} .\end{aligned}$$

G.8.6 Interface 3 (Amplifier—A/D Converter)

The uncertainty of the input to the A/D converter is

$$\sigma_6 = \sqrt{\sigma_5^2 + (Y_5^0)^2 (\sigma_{p_{61}}^2 + \sigma_{p_{62}}^2) + \sigma_{p_{63}}^2} ,$$

where

$\sigma_{p_{61}}$ = interface loss uncertainty
$\sigma_{p_{62}}$ = A / D input crosstalk uncertainty
$\sigma_{p_{63}}$ = interface noise uncertainty
$Y_5^0 = Y_4^0 = p_{51}Y_3^0 \cong p_{51}p_{11}x$.

Assume the following for the interface:

p_{61} = -1% $\pm$ 0.1% of input signal (at 95% confidence)
$p_{62} \cong 0$
$p_{63} \cong 0$,

Thus we get for the interface loss uncertainty term:

$$\sigma_{p_{61}} = \frac{0.001}{1.960} \cong 5.1 \times 10^{-4} .$$

Since $Y_5^0 = p_{51}p_{11}x = (10)(2280)\ \mu V = 22{,}800\ \mu V$, note for future reference that $Y_6 \cong 0.99Y_5 \cong$ 22,570 μV. With these results, and using $\sigma_5 = 104.3\ \mu V$, we get

$$\sigma_6 = \sqrt{(104.3)^2 + (22{,}800)^2(5.1\times10^{-4})^2}\ \mu V \cong 104.9\ \mu V \ \text{(cumulative uncertainty)}.$$

G.8.7 Sampling (A/D Converter)

The expression for sampling uncertainty is

$$\sigma_7 \cong \sqrt{[(1+(p_{71})^2]\sigma_6^2 + (Y_6^0)^2\{\sigma_{p_{71}}^2 + [1+(p_{71})^2](\sigma_{p_{72}}^2 + \sigma_{p_{73}}^2 + \sigma_{p_{75}}^2)\} + \sigma_{p_{74}}^2 + \sigma_{p_{76}}^2} .$$

where

$\sigma_{p_{71}}$ = D / A converter analog loss uncertainty
$\sigma_{p_{72}}$ = aperture time uncertainty
$\sigma_{p_{73}}$ = sampling rate uncertainty
$\sigma_{p_{74}}$ = quantization uncertainty
$\sigma_{p_{75}}$ = linearity uncertainty
$\sigma_{p_{76}}$ = noise uncertainty
$Y_6^0 = Y_5^0 = p_{51}Y_3^0$.

Assume the following specifications.

p_{71} = -0.5% $\pm$ .05% (at 99% confidence)
p_{72} = (see below)
p_{73} = (see below)
p_{74} = (see below)
$p_{75} \cong \pm$ 0.1% (at 99.73% confidence)
$p_{76} \cong 0\ \mu V$.

The parameters p_{72}, p_{73} and p_{74} are not, in themselves, of interest. Their uncertainties, which are of interest, can be computed directly. Suppose that the D/A converter specifications include

aperture time = 1 mSec
sampling rate = 200 Hz
number of bits = 14

A/D full scale = 100 μV
signal frequency = 10 Hz.

Using these data with the methods of Appendix G yields

Aperture time: $\sigma_{p_{72}} = \dfrac{2\pi(10)(1\times10^{-6})}{2\sqrt{6}} = 1.28\times10^{-5}$

Sampling rate: $\sigma_{p_{73}} \cong 1.26\times10^{-4}$

Quantization: $\sigma_{p_{74}} = (100/2^{14+1})/\sqrt{3}\,\mu V = 1.76\times10^{-3}\,\mu V.$

Also, for p_{71} and p_{75}, we have

$$\sigma_{p_{71}} = \frac{(0.0005)}{2.5758} \cong 1.94\times10^{-4}\ ,$$

and

$$\sigma_{p_{75}} = \frac{(0.001)}{3.000} \cong 3.33\times10^{-4}\ .$$

Putting the numbers together yields the output of the D/A Converter as

$$Y_7 = (0.995)Y_6 \cong 22{,}460\,\mu V,\ \text{with}$$

$$\sigma_7 \cong \sqrt{(104.9)^2 + (22{,}570)^2[(1.94)^2 + (.128)^2 + (1.26)^2 + (3.33)^2]\times10^{-8} + (0.00176)^2}\ \mu V$$
$$\cong 105.2\,\mu V\ \text{(cumulative uncertainty)}.$$

G.8.8 System Uncertainty

As discussed earlier, the uncertainty in the output of the system is given by

$$\sigma_{system} \cong \sqrt{(p_{81}p_{83})^2\sigma_7^2 + x^2\left[(\sigma_{p_{81}}/p_{81})^2 + (\sigma_{p_{83}}/p_{83})^2\right] + \sigma_{p_{82}}^2}\ .$$

where

$\sigma_{p_{81}}$ = data processor voltage to temperature conversion uncerainty
$\sigma_{p_{82}}$ = system resolution uncertainty
$\sigma_{p_{83}}$ = error correction factor uncertainty
x = measurand value $\cong p_{81}p_{83}Y_7$.

The voltage to temperature conversion process is approximately the reciprocal of the temperature to voltage conversion process encountered earlier. Thus

$$p_{81} \cong 1/22.8\ ^\circ C/\mu V = 0.044\ ^\circ C/\mu V,\ \text{and}$$

$$\sigma_{p_{81}} \cong \frac{0.5/22.8}{2.576}\times(0.044)\ ^\circ C/\mu V = 3.75\times10^{-4}\ ^\circ C/\mu V.$$

Suppose that the decimal output is given on a 100 °C scale to three significant digits. Then the decimal resolution is

$$p_{82} = \pm\frac{100\,^\circ C}{1000} = \pm 0.10\,^\circ C,$$

and the resolution uncertainty is

$$\sigma_{p_{82}} = 0.10/\sqrt{3}\,^\circ C = 0.058\,^\circ C.$$

The error correction factor is obtained by attempting to compensate or correct for gains and losses that occur in the measurement system. These are summarized as

Description	**Loss**	**Error**	**Uncertainty**
Interface1	0	0	0
Low Pass Filter	0	0	0
Amplifier	1000% (gain)	$\pm$0.5%	0.027
Interface 3	1%	$\pm$0.1%	5.1×10^{-4}
A/D Converter	0.5%	$\pm$0.05%	1.94×10^{-4}

From these data, the value of p_{83} is estimated at

$$\begin{aligned} p_{83} &= \left(\frac{1}{p_{51}}\right)\left(\frac{1}{1+p_{61}}\right)\left(\frac{1}{1+p_{71}}\right) \\ &\cong \left(\frac{1}{10}\right)\left(\frac{1}{0.99}\right)\left(\frac{1}{0.995}\right) = 0.102. \end{aligned}$$

The analog error in this term is determined using the Taylor series method:

$$\varepsilon(p_{83}) = -p_{83}\left[\frac{\varepsilon(p_{51})}{p_{51}} + \frac{\varepsilon(p_{61})}{1+p_{61}} + \frac{\varepsilon(p_{71})}{1+p_{71}}\right].$$

Recall that the correction will be in the digital rather than analog domain. Thus a quantization error component must be added to this analog error. Using the result

$$\text{Quantization uncertainty} = (100/2^{14+1})/\sqrt{3}\,\mu V = 1.76\times10^{-4}\ \mu V$$

obtained in the analysis of A/D conversion uncertainty gives

$$\begin{aligned} \sigma_{p_{83}} &= \sqrt{p_{83}^2\{(\sigma_{p_{51}}/p_{51})^2 + [\sigma_{p_{61}}/(1+p_{61})]^2 + [\sigma_{p_{71}}/(1+p_{71})]^2\} + (\sigma_{p_{83}}^{quantization})^2} \\ &= \sqrt{p_{83}^2[(.027/10)^2 + (5.1\times10^{-4}/0.99)^2 + (1.94\times10^{-4}/0.995)^2] + (1.76\times10^{-4}\mu V)^2} \\ &\cong p_{83}\sqrt{(0.003)^2 + (1.76\times10^{-4}\mu V/p_{83})^2} \\ &\cong 0.003 p_{83}. \end{aligned}$$

Combining these results gives the total system uncertainty as[28]

$$\sigma_{system} \cong \sqrt{(0.044\times 0.102)^2(105.2)^2 + (100)^2\left[(3.75\times 10^{-4}/0.044)^2 + (0.003)^2\right] + (0.058)^2}\ {}^\circ C$$
$$\cong 1.02\ {}^\circ C \text{ (cumulative uncertainty)}.$$

If it were desired to tolerance the system at 100 °C, using the NIST convention of multiplying the uncertainty by a factor of 2, we would write

$$T = T_{output} \pm 2.04\ {}^\circ C, \text{ with approximately 95\% confidence.}$$

Note — The length of time that this tolerance is applicable depends on the stability of the various parameters of the system from which component uncertainties were computed.

G.9 Estimating the Process Uncertainty

The contribution of process uncertainties due to environmental error sources have been included in the system uncertainty estimates developed above. The contribution of random sampling uncertainties in system output will now be discussed.

As the ISO and NIST guidelines show, random sampling uncertainty can be estimated by taking a sample of measurements and computing a sample standard deviation. Suppose that *n* values of the output of the system are obtained by measurement of a fixed temperature. If each sampled value is denoted y_i, $i = 1,2,\cdots,n$, then the mean of these values is

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i ,$$

and the sample standard variance is given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 .$$

The square root of the sample variance is the sample standard deviation. This quantity can be used to represent the uncertainty due to short-term random fluctuations in the output of the system. This uncertainty does not exactly characterize the random uncertainty of the system output but, rather, applies to random uncertainties in single measurements made using the system.

An estimate that better serves the purpose of representing the characteristic random output of the system is the sampling standard deviation given by

[28] Interestingly, the dominant term in the square root is the second term, which is driven by the uncertainty of the system following the thermocouple. The first term, which is driven primarily by the uncertainty of the thermocouple and has been dominant up to this point, is now subordinate.

$$s_{system} = s / \sqrt{n} \ .$$

G.10 Estimating the Total Uncertainty

The total uncertainty is obtained by combining system and random uncertainty components:

$$\sigma_{total} \cong \sqrt{\sigma_{system}^2 + s_{sample}^2} \ .$$

The equality is only approximate, since the sampling standard deviation is only an estimate of the "true" random uncertainty in the output of the measuring system.

Before leaving this example, it will be worthwhile to make a few observations about the system and random parts of the total uncertainty. The random uncertainty s_{sample} is a quantity that depends on the stability of the system to short-term environmental and other stresses and on the vagaries of the measurement process.

The quantity σ_{system}, on the other hand, represents the uncertainty in the bias of the system. The various component uncertainties that were used in determining this quantity were estimated using error limits based on specifications. In each case, the error containment probability of the error limits was employed. As will be discussed in Chapter 6, these containment probabilities may change with time. Likewise, the bias uncertainty of the system may be time varying. For this reason, it is always a good practice to write

$$\sigma_{total}(t) \cong \sqrt{\sigma_{system}^2(t) + s_{sample}^2} \ .$$

where the variable t indicates the time-dependence of the system bias uncertainty.

Appendix H THE INTERNATIONAL SYSTEM OF UNITS (SI)

H.1 The SI

The SI or modern metric system is a universally accepted system of units. It was adopted in 1960 by the General Conference of Weights and Measures (CGPM) to harmonize physical measurement throughout the world. It is a dynamic system that is continually evolving to meet measurement needs. The SI defines classes of units; establishes names, symbols, and prefixes (multipliers) for the units; and addresses other matters important to ensuring measurement accord. Also, NIST, the International Standards Organization (**IS0**), and the American National Standards Institute (**ANSI**) have published detailed information about the system and use. Although nearly universal, there are small variations between nations. These differences are mostly in the spelling of certain units and other minor matters. Both the NIST and ANSI documents are the United States' interpretation of the SI.

H.2 SI Units

Three classes of units were established, *base units, supplemental units,* and *derived units.* The system is coherent; that is, all units derived from base units have the implied multiplier of one (1).

H.2.1 Base Units

Seven base units were chosen by convention and are regarded as dimensionally independent. Each, except the kilogram, is defined in terms of a physical phenomenon or constants of nature. For example; the meter is the length of the path traveled by light during an interval of 1/299 792 458 of a second. The interval is the reciprocal of the speed of light in vacuum. The kilogram is a carefully preserved artifact residing at the International Bureau of Weights and Measures (BIPM). Also, it is the only unit that includes a prefix, "kilo," in its name. All other units are derived in terms of these seven (and two supplementary units discussed later.) Table H.1 lists the base units. The term "quantity" used in the heading of this and other tables means measurable attribute of phenomena or matter. For each quantity in Table H.1, there is an SI unit name and symbol.

TABLE H.1 SI Base Units

TABLE H.1
SI Base Units

Quantity	***Name***	***Symbol***
amount of substance	*mole*	mol
electric current	*ampere*	A
length	*meter*	m
luminous intensity	*candela*	cd
mass	*kilogram*	kg
thermodynamic temperature	*kelvin*	K
time	*second*	s

H.2.2 Supplementary Units

CGPM adopted two supplementary units, the SI unit of plain angle and the SI unit of solid angle. Plane angle is generally expressed as the ratio between two lengths and solid angle the ratio between an area and the square of length. Both are dimensionless derived quantities. Table H.2 gives the particulars on both.

TABLE H.2 SI Supplementary Units

TABLE H.2
SI Supplementary Units

Quantity	Name	Symbol	Expression in Terms of SI Base Units
plane angle	radian	rad	$m \cdot m^{-1} = 1$
solid angle	steradian	sr	$m^2 \cdot m^{-2} = 1$

H.2.3 Derived Units

Derived units are expressed algebraically in terms of base units by the mathematical symbols of multiplication and division. Because the system is coherent, the product or quotient of any two quantities is the unit of the resulting quantity. Table H.3 gives several examples of derived units expressed exclusively in base units.

TABLE H.3 Examples of SI-Derived Units Expressed in Base Units

TABLE H.3
Examples of SI-Derived Units Expressed in Base Units

Quantity	Name	Symbol
area	square meter	m^2
volume	cubic meter	m^3
speed, velocity	meter per second	m/s
acceleration	meter per second squared	m/s^2
wave number	reciprocal meter	m^{-1}
density, mass density	kilogram per cubic meter	kg/m^3
specific volume	cubic meter per kilogram	m^3/kg
current density	ampere per square meter	A/m^2
magnetic field strength	ampere per meter	A/m
concentration (of amount of substance)	mole per cubic meter	mol/m^3
luminance	candela per square meter	cd/m^2

Certain derived units have been given special names and symbols were established. They may themselves be used to express other derived units. In Table H.4 the name, symbol, and expression in terms of other units and the base units are given for each.

TABLE H.4 Derived Units with Special Names

TABLE H.4
Derived Units with Special Names

Quantity	Name	Symbol	Expression in Terms of Other Units	Expression in Terms of SI Base Units
frequency	hertz	Hz		s^{-1}
force	newton	N		$m \cdot kg \cdot s^{-2}$
pressure, stress	pascal	Pa	N/m^2	$m^{-1} \cdot kg \cdot s^{-2}$
energy, work, quantity of heat	joule	J	$N \cdot m$	$m^2 \cdot kg \cdot s^{-2}$
power, radiant flux	watt	W	J/s	$m^2 \cdot kg \cdot s^{-3}$
electric charge, quantity of electricity	coulomb	C		$s \cdot A$
electric potential, potential difference, electromotive force	volt	V	W/A	$m^{-2} \cdot kg \cdot s^{-3} \cdot A^{-1}$
capacitance	farad	F	C/V	$m^{-2} \cdot kg^{-1} \cdot s^4 \cdot A^2$
electric resistance	ohm	ž	V/A	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-2}$
electric conductance	siemens	S	A/V	$m^{-2} \cdot kg^{-1} \cdot s^3 \cdot A^2$
magnetic flux	weber	Wb	$V \cdot s$	$m^2 \cdot kg \cdot s^{-2} \cdot A^{-1}$
magnetic flux density	tesla	T	Wb/m^2	$kg \cdot s^{-2} \cdot A^{-1}$
inductance	henry	H	Wb/A	$m^2 \cdot kg \cdot s^{-2} \cdot A^{-2}$
Celsius temperature ✶	degree Celsius	°C		K
luminous flux	lumen	lm		$cd \cdot sr$ ✸
illuminance	lux	lx	lm/m^2	$m^{-2} \cdot cd \cdot sr$ ✸
activity (of a radionuclide)	becquerel	Bq		s^{-1}
absorbed dose, specific energy imparted, kerma, absorbed dose index	gray	Gy	J/kg	$m^2 \cdot s^{-2}$
dose equivalent, dose equivalent index	sievert	Sv	J/kg	$m^2 \cdot s^{-2}$

✶ Besides the thermodynamic temperature (symbol *T*) expressed in kelvins (see Table H.1), use is also made of the Celsius temperature (symbol *t*) defined by the equation $t = T - T_0$, where T_0 = 273.15K by definition. To express Celsius temperature, the unit "degree Celsius" which is equal to the unit "kelvin" is used; here "degree Celsius" is a special name used for "kelvin." An interval of difference of Celsius temperature can, however, be expressed in kelvins and in degrees Celsius.

✸ In photometry, the symbol *sr* is maintained in expressions for units.

Table H.5 gives some examples of derived units expressed by special names.

TABLE H.5 Example of SI-Derived Units Expressed by Special Names

TABLE H.5
Examples of SI-Derived Units Expressed by Special Names †

Quantity	*Name*	*Symbol*	*Expression in Terms of Other Units*
dynamic viscosity	*pascal second*	Pa • s	$m^{-1} \cdot kg \cdot s^{-1}$
moment of force	*newton meter*	N • m	$m^{2} \cdot kg \cdot s^{-2}$
surface tension	*newton per meter*	N/m	$kg \cdot s^{-2}$
heat flux density, irradiance	*watt per square meter*	W/m^{2}	$kg \cdot s^{-3}$
heat capacity, entropy	*joule per kelvin*	J/K	$m^{2} \cdot kg \cdot s^{-2} \cdot K^{-1}$
specific heat capacity, specific entropy	*joule per kilogram kelvin*	J/(kg • K)	$m^{2} \cdot s^{-2} \cdot K^{-1}$
specific energy	*joule per kilogram*	J/kg	$m^{2} \cdot s^{-2}$
thermal conductivity	*watt per meter kelvin*	W/(m • K)	$m \cdot kg \cdot s^{-3} \cdot K^{-1}$
energy density	*joule per cubic meter*	J/m^{3}	$m^{-1} \cdot kg \cdot s^{-2}$
electric field strength	*volt per meter*	V/m	$m \cdot kg \cdot s^{-3} \cdot A^{-1}$
electric charge density	*coulomb per cubic meter*	C/m^{3}	$m^{-3} \cdot s \cdot A$
electric flux density	*coulomb per square meter*	C/m^{2}	$m^{-2} \cdot s \cdot A$
permittivity	*farad per meter*	F/m	$m^{-3} \cdot kg^{-1} \cdot s^{4} \cdot A^{2}$
permeability	*henry per meter*	H/m	$m \cdot kg \cdot s^{-2} \cdot A^{-2}$
molar energy	*joule per mole*	J/mol	$m^{2} \cdot kg \cdot s^{-2} \cdot mol^{-1}$
molar entropy, molar heat capacity	*joule per mole kelvin*	J/(mol •K)	$m^{2} \cdot kg \cdot s^{-2} \cdot K^{-1} \cdot mol^{-1}$
exposure (x and γ)	*coulomb per kilogram*	C/kg	$kg^{-1} \cdot s \cdot A$
absorbed dose rate	*gray per second*	Gy/s	$m^{2} \cdot s^{-3}$

† See ANSI Std. 268-1982, Table 4, for more derived units.

A unit name may correspond to several different quantities. In the previous tables, there are several examples. The joule per kelvin (J / K) is the SI unit for the quantity heat capacity and for the quantity entropy (Table H.5). The name of the unit is not sufficient to define the quantity measured. Specifically, measuring instruments should indicate not only the unit but also the measure quantity concerned.

H.2.4 Other Units

Certain units are not part of the SI but are important and widely used. The International Conference of Weights and Measures (**CIPM**) recognized the need for these units because of their importance. The units in this category accepted for use in the United States with the SI are listed in Table H.6. The combination of units of this table with SI units to form compound units should be restricted to special cases in order not to lose the advantage of coherence. Examples of combining the units of Table H.6 with SI units are ampere hour (A•h), kilowatt hour (kW•h), and kilometer per hour (km/h). The corresponding coherent SI units are coulomb (C), joule (J), and meter per second (m/s), respectively.

TABLE H.6 Units in Use with the SI

TABLE H.6
Units in Use With the SI

Name	**Symbol**	**Value in SI Unit**
minute (time)	min	*1 min = 60 s*
hour	h	*1 h = 60 min = 3,600 s*
day	d	*1 d = 24 h = 86,400 s*
degree (angle)	°	*1° = (š/180) rad*
minute (angle) ✶	'	*1' = (1/60)° = (š/10,800) rad*
second (angle) ✶	"	*1" = (1/60)' = (š/648,000) rad*
liter ✷	L	*1 L = 1 dm^3 = 10^{-3}m^3*
metric ton ✹	t	*1 t = 10^3 kg*
electron volt ❋	eV	*1 eV = 1.602 19 x 10^{-19} J, approx.*
unified atomic mass unit	u	*1 u = 1.660 57 x 10^{-27} kg, approx.*

✶ Use discouraged except for special fields such as cartography.

✷ Both L and l are internationally accepted symbols for liter. Because "l" can be confused with the numeral "1," the symbol "L" is recommended for the United States. ANSI/IEEE Std 268-1982 states: "The use of this unit is restricted to volumetric, capacity, dry measure, and measure of fluids (both liquids and gases). No prefix other than milli- or micro- should be used with liter."

✹ In many countries, this unit is called "tonne."

❋ The values of these units expressed in terms of the SI units must be obtained by experiment, and therefore are not known exactly. The electronvolt is the kinetic energy acquired by an electron passing through a potential difference of 1 volt in vacuum. The unified atomic mass is equal to (1/12) of the mass of the atom of the nuclide ^{12}C.

H.3 Units in Temporary Use

In those fields where their use is well established, the units in Table H.7 are acceptable, subject to future review. These units should not be introduced where they are not presently in use.

TABLE H.7 Units in Temporary Use With SI

TABLE H.7
Units in Temporary Use With SI

angström	are *(unit of land area)*	bar
barn	curie	gal *(unit of acceleration)*
knot	nautical mile	rad *(unit of absorbed dose)*
rem *(unit of dose equivalent)*	roentgen	

H.4 Obsolete Units

The 1990 Federal Register notice lists several units no longer accepted for use in the United States. They are myriameter, stere, millier, tonneau, quintal, myriagram, and kilo (for kilogram). Also, CIPM has recommended that several units in common use be avoided. Table 12 of NBS SP-330 (1986) lists a number in temporary use. Last, the CIPM recognizes the centimeter-gram-second system of units and the special names but urges that they no longer be used.

H.5 Rules for Writing and Using Symbols

The general principles for writing unit symbols were adopted by the CGPM:

(1) Roman (upright) type, generally, lower case, is used for the unit symbol. If, however, the name of the unit is derived from a proper name, the first letter of the symbol is in upper case.

(2) Unit symbols are unaltered in the plural.

(3) Unit symbols are not followed by a period.

To insure uniformity in the use of SI unit symbols, ISO has made certain recommendations. They are:

(a) The product of two or more units may be shown in any of the following ways:[29]

$$\mathrm{N} \bullet \mathrm{m} \text{ or } \mathrm{Nm}$$

(b) A solidus (oblique stroke, /), a horizontal line, or negative exponent may be used to express a derived unit formed from two others by division:

---

[29] From footnote on page 9 of NBS SP330 (1986). "See American National Standard ANSI/IEEE Std 260-1978, which states that in USA practice only the raised dot of these three ways is used."

$$m/s \text{ or } m \bullet s^{-1}$$

(c) The solidus must not be repeated on the same line unless ambiguity is avoided by parentheses. In complicated cases, negative exponents or parentheses should be used:

$$m/s^2 \text{ or } m \bullet s^{-2} \text{ but not } m/s/s$$

$$m \bullet kg/(s^3 \bullet A) \text{ or } m \bullet kg \bullet s^{-3} \bullet A^{-1} \text{ but not } m \bullet kg/s^3/A.$$

H.6 SI Prefixes

CGPM adopted a series of prefixes and symbols of prefixes for names and symbols of the decimal multiples and submultiples of SI units. They are given in Table H.8.

TABLE H.8 SI Prefixes

TABLE H.8
SI Prefixes

Factor	*Prefix*	*Symbol*	*Factor*	*Prefix*	*Symbol*
10^{18}	*exa*	E	10^{-1}	*deci*	d
10^{15}	*peta*	P	10^{-2}	*centi*	c
10^{12}	*tera*	T	10^{-3}	*milli*	m
10^{9}	*giga*	G	10^{-6}	*micro*	μ
10^{6}	*mega*	M	10^{-9}	*nano*	n
10^{3}	*kilo*	k	10^{-12}	*pico*	p
10^{2}	*hecto*	h	10^{-15}	*femto*	f
10^{1}	*deka* †	da	10^{-18}	*atto*	a

† The spelling “deca” is used extensively outside the United States,

In accord with the general principles adopted by the ISO, the CIPM recommends certain rules for using the SI prefixes. They are

(1) Prefix symbols are printed in Roman (upright) type without spacing between the prefix and the unit symbol.

(2) The grouping formed by a prefix symbol attached to the unit symbol is a new inseparable symbol that can be raised to a positive or negative power and that can be combined with other unit symbols to form compound unit symbols:

$$1\,cm^3 = (10^{-2}\,m)^3 = 10^{-6}\,m^3$$

$$1\,V/cm = (1/V)/(10^{-2}\,m) = 10^2\,V/m.$$

(3) Compound prefixes formed by the juxtaposition of two or more SI prefixes are not to be used:

1nm but not 1mμm.

(4) A prefix should never be used alone.

(5) Errors in calculations can be avoided by replacing the prefixes with powers of 10.

Definitive discussions of prefix rules and the use of exponents are found in ANSI/IEEE Std 268-1982, NBS Special Publication 330 (1986 edition), and ISO Standard Handbook 2 (1982 edition). All three of these documents are revised occasionally and the most recent versions take precedence.

H.7 Conversion to Metric

It will be necessary to convert many units from those in current use in the United States to metric. Such conversion can be carried out using Eq. (5.2) and a knowledge of the relationship between the two θ_A / θ_B . Be careful, as serious errors often happen when making conversions. ANSI/IEEE Std 268-1982 lists many conversion factors to obtain the SI units (but not the reverse). The standard also provides rules for conversion and rounding. There are two facets of this problem—the conversion proper and handling any associated tolerance.

Bibliography

Abdel-Hameed, M.S., Cinlar, E., and Quinn, J., ed. *Reliability Theory and Models*, Academic Press, Orlando, FL, 1984.

Abernathy, R.B., et al., *Measurement Uncertainty Handbook*, Research Triangle Park, Instrument Society of America, 1980.

Abernethy, R.B., and Benedict, R.P., *Measurement Uncertainty: A Standard Methodology,* ISA Transactions, Vol. 24, No. 1, 1985.

Abernethy, R.B., and Ringhiser, B., *The History and Statistical Development of the New ASME-SAE-AIAA-ISO Measurement Uncertainty Methodology*, AIAA-85-1403, 1985.

Accuracy Transfer/Software Corrections in ATE Systems, Navy METCAL Program Draft Technical Note, U.S. Navy Metrology Engineering Department, Naval Warfare Assessment Center, Corona, CA, 29 September 1989.

ANSI/NCSL Z540-1, *General Requirements for Calibration Laboratories and Measuring and Test Equipment*, Final Draft, October 29, 1993.

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

Bartlett, M., *An Introduction to Stochastic Processes*, Cambridge University Press, New York, 1955.

Belanger, B.C., *Traceability: An Evolving Concept*, ASTM Standardization News, American Society for Testing and Materials, 1980, Vol. 8, No. 1, Pages 22–28.

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

Bishop, Y., Feinberg, S. and Holland, P., *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, 1975.

Brownlee, K., *Statistical Theory and Methodology in Science and Engineering*, 2nd Ed., John Wiley & Sons, New York, 1965.

Castrup, H., “Calibration Requirements Analysis System,” *Proceedings of the 1989 NCSL Workshop and Symposium*, Denver, CO, July 1989.

Castrup, H., “Navy Analytical Metrology RD&E,” *Navy Metrology Research & Development Program Conference Report*, Dept. of the Navy, Metrology Engineering Center, NWS, Seal Beach, CA, Corona Annex, March 1988.

Castrup, H., *Evaluation of Customer and Manufacturer Risk v. Acceptance Test In-Tolerance Level*, TRW Technical Report No. 99900-7871-RU-00, April 1978.

Castrup, H., *Intercomparison of Standards: General Case, SAI Comsystems Technical Report*, Department of the Navy Contract N00123-83-D-0015, Delivery Order RM03, 1984.

Draper, N., and Smith, H., *Applied Regression Analysis*, John Wiley & Sons, Inc., New York, 1966.

Eicke, W.G., and Auxier, L.M., *Regional Maintenance of the Volt Using NBS Volt Transfer Techniques*, Reprinted in NBS SP 705 (1985); Original, IEEE Trans. Inst. & Meas., Vol. IM-23, No. 4, (Dec. 1974).

Eisenhart, C., "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," J. Res. NBS, Vol. 67C, No. 2, April–June 1963.

Ferling, J. and Caldwell, D., "Analytic Modeling for Electronic Test Equipment Adjustment Policies," Presentation, *1989 NCSL Workshop and Symposium*, Denver, CO, July 1989.

Ferling, J., "Calibration Intervals for Multi-Function Test Instruments, A Proposed Policy," *Proceedings of the 1987 Measurement Science Conference*, Irvine, CA, January 1987.

Ferling, J., "The Role of Accuracy Ratios in Test and Measurement Processes," *Proceedings of the 1984 Measurement Science Conference*, Long Beach, CA, January 19–20.

Ferling, J., *Calibration Interval Analysis Model*, SAI Comsystems Technical Report, Prepared for the U.S. Navy Metrology Engineering Center, Contract N00123-87-C-0589, April 1979.

Glassman, J., "Intervals by Exception," *Proceedings of the 1970 NCSL Workshop and Symposium*, Gaithersburg, MD, October 1970.

Goldman, D.T., and Bell, R.J. (Editors), *The International System of Units (SI)*, NBS SP 330 1986 Edition.

Grenander, U., and Rosenblatt, M., *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, New York, 1963.

Hannan, E., *Time Series Analysis*, Methuen, London, 1960.

Hartley. H., "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares," *Technometrics*, Vol. 3, No. 2, p. 269, 1961.

Hayes, J., *Factors Affecting Measuring Reliability*, U.S. Naval Ordnance Laboratory, TM No. 63-106, October 1955.

Hayes, J., "Calibration and Maintenance of Test and Measuring Equipment," *Encyclopedia of Applied Physics*, Vol. 3, VCH Publishers, Inc., 1992.

IEEE, American National Standard-Metric Practice, ANSI/IEEE Std 268-1982.

Instrument Intercomparison: A General Methodology, Analytical Metrology Note AMN 86-1, U.S. Navy Metrology Engineering Department, NWS Seal Beach, CA, Corona Annex, 1 January 1986.

International Organization for Standardization (ISO), *Guide to the Expression of Uncertainty in Measurement*, ISO/TAG4/WG3, 1st Ed., June 1992.

International Organization for Standardization (ISO), *Units of Measurement*, ISO Handbook 2 (1982), 2nd Edition.

Jackson, D., *Analytical Methods to be Used in the Computer Software for the Manometer Audit System*, SAIC Technical Report TR-830016-4M112/006-01, Computer Software Specification, Department of the Navy Contract N00123-83-D-0016, Delivery Order 4M112, 8 October 1985.

Jackson, D., Ferling, J. and Castrup, H., "Concept Analysis for Serial Number Based Calibration Intervals," *Proceedings of the 1986 Measurement Science Conference*, Irvine, CA, January 1986.

Jackson, D.H., "Reliability Analysis Methods for Calibration Intervals: Analysis of Type III Censored Data," *Proceedings of the 1987 NCSL Workshop and Symposium*, Denver, CO, 1987.

Jennrich, R. and Sampson, P., "Application of Stepwise Regression to Non-Linear Estimation," *Technometrics*, Vol. 10, No. 1, p. 63, 1968.

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

Ku, H., "Expressions of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value," *Precision Measurements and Calibration, Statistical Concepts and Procedures*, NBS Special Pub. 300, Vol. 1, Ed. by H. Ku, February 1969.

Ku, H., "Statistical Concepts in Metrology," *Handbook of Industrial Metrology*, American Society of Tool and Manufacturing Engineers, Prentice-Hall, Inc., New York, 1967.

Kuskey, K., and Kraft, R., *An Analysis and Evaluation of Management and Information Systems for the Calibration and Repair of the Navy's Electronic Test Equipment*, DDI Technical Report TR 81-9-177, December 1981.

Kuskey, K., *Cost-Benefit Model for Analysis of Calibration-System Designs in the Case of Random-Walk Equipment Behavior*, SAI Comsystems Technical Report, Prepared for the U.S. Navy Metrology Engineering Center, Contract N00123-76-C-0589, January 1979.

Larsen, J.C., "The Navy's New Manufacturer Initial Calibration Interval System," *Proceedings of the 1988 NCSL Workshop and Symposium*, Washington, DC, August 1988.

Mann, N., Schafer, R., and Singpurwalla, N., *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, Inc., New York, 1974.

Manufacturer Initial Calibration Interval System (MICIS), Navy Metrology and Calibration Program, Metrology Engineering Center, Engineering Circular EC-40, November 1987.

Meeker, W., and Nelson, W., "Weibull Percentile Estimates and Confidence Limits from Singly Censored Data by Maximum-likelihood," *IEEE Transactions*. Rel, Vol. R-25, No. 1, April 1976.

Morris, M.D., "A Sequential Experimental Design for Estimating a Scale Parameter from Quantal Life Testing Data," *TECHNOMETRICS*, Vol. 29, No. 2, May 1987.

NBS Handbook 91, *Experimental Statistics*, M. G. Natrella, U. S. Government Printing Office, Washington, DC, 1963; reprinted October 1966 with corrections.

NBS Special Publication 673, *Economic Model of Calibration Improvements for Automatic Test Equipment*, Weber, S. and Hillstrom A., U.S. Department of Commerce, April 1984.

NBS Special Publication 676-I, *Measurement Assurance Programs Part I: General Introduction*; Brian Belanger, U.S. Department of Commerce, May, 1984 — Library of Congress Catalog Card No. 84-601030.

NBS Special Publication 676-II, *Measurement Assurance Programs Part II: Development and Implementation*; M. Carroll Croarkin, U.S. Department of Commerce, April, 1985 — Library of Congress Catalog Card No. 84-601030.

NIST Technical Note 1297, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*; Barry N. Taylor and Chris E. Kuyatt, U.S. Department of Commerce, January, 1993.

Parzen, E., *Stochastic Processes*, Holden-Day, Inc., San Francisco, 1962.

Proschan, F., "Confidence and Tolerance Intervals for the Normal Distribution," *Proc. Annual American Statistical Association*, Boston, December 1951.

Riley, J.P., Ten Volt Round Robin Using Solid State Standards, *Proc. 1990 Measurement Science Conference*, February 8–9, 1990.

Shooman, M.L., *Probabilistic Reliability: An Engineering Approach*, McGraw-Hill, New York, 1968.

'Student,' (William S. Gosset), *Biometrica*, 6, 1908.

System for Tradeoff Analysis and Reporting, User Manual, Version 3.0, Department of the Navy, Metrology Engineering Center, September 1990.

Taylor, J., *Fundamentals of Measurement Error*, Neff Instrument Corporation, Monrovia, CA, 1988.

Wise, J.H., *Aliased Data: Description and Avoidance Methods*, Measurement Technology Center Technical Note, JPL D-8816, (internal document), Jet Propulsion Laboratory, Pasadena, CA, January 27, 1992.

Wold, H., "Forecasting by the Chain Principle," *Time Series Analysis*, ed. by M. Rosenblatt, pp 475-477, John Wiley & Sons, Inc., New York, 1963.

Wold, H., *A Study in the Analysis of Stationary Time Series*, 2nd Ed., Upsala, Sweden, 1954.

Youden, W.J., *Graphical Diagnosis of Interlaboratory Test Results*, Reprinted in - NBS SP300, Vol. 1, pp. 122-137 (1969); Original published - Ind. Qual. Control, Vol. XV, No. 11, (1959).

Youden, W., "Uncertainties in Calibration," *Proc. 1962 International Conference on Precision Electromagnetic Measurements*, August 1962.

Youden, W.J., "Uncertainties in Calibrations," *IRE Transactions*. Instr. I–11, Nos. 3&4, pp 133–138, December 1962.

Index