NASA
REFERENCE
PUBLICATION
1364

March 1995

Metrology—Measurement Assurance Program Guidelines

NASA Metrology and Calibration Working Group

W. G. Eicke
J. P. Riley
K. J. Riley

Kennedy Space Center, Florida

This Publication is intended to provide technical and administrative guidance for developing, documenting, implementing, and maintaining Measurement Assurance Programs within and between National Aeronautics and Space Administration (NASA) field installations. This Publication strives to develop and maintain consistently cost effective, high quality value added programs for the Agency. It is not possible to address every measurement discipline and methodology in this guide so the reader must take this basic information and adapt it to their particular measurement requirements. Measurement assurance is continually evolving and will continue to do so. The reader is therefore urged to take advantage of new concepts, ideas and techniques to build on what is written in this guideline and elsewhere. Finally, this guideline would not have been possible without the pioneering work of others; especially J. M. Cameron, now retired from National Bureau of Standards, for his invaluable contributions, and C. Croarkin of National Institute of Standards and Technology for providing the metrology community with its first definitive guideline for establishing local measurement assurance programs. Where possible the authors have tried to adapt their material to meet NASA needs.

Figures

Tables

BIPM	Bureau of International Weights and Measures
CIPM	International Committee for Weights and Measures
CL	Central line (mean)
CRM	Certified reference material (see also RM and SRM)
EOP	End of period
GMAP	Group or regional measurement assurance program
GMAP-T	Group measurment assurance program transfer
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
LCL	Lower control limit
M&TE	Measurement and test equipment (also known as TME or T&ME)
MAP	Measurement assurance program
MAP-T	Measurement assurance program type transfer
MLS	Method of least squares
NBS	National Bureau of Standards (now NIST)
NIST	National Institute of Standards and Technology (formerly NBS)
OIML	Organization for Legal Metrology
RM	Reference material (see also CRM and SRM)
RMAP-T	Reverse measurement assurance program type transfer
RSS	Root sum of squares (square root of the sum-of-the-squares)
SPC	Statistical process control
SRM	Standard reference material (see also CRM and RM)
SSVR	Solid-state voltage reference
UCL	Upper control limit

1 Introduction

1.1 Purpose

The purpose of this publication is to provide guidance for the establishment and implementation of Measurement Assurance Programs (MAPs) which is defined as: *"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 used to eliminate sources of unacceptable uncertainty."*[1] Specific objectives are to:

- Ensure the quality of measurements made within NASA programs,
- Establish realistic measurement process uncertainties,
- Maintain continuous control over the measurement processes, and
- Ensure measurement compatibility among NASA facilities.

1.2 Applicability

This publication applies, to the extent practicable, to all NASA programs. It is applicable when referenced in systems contracts, major subcontracts, and may also be used for other contracts where measurements are an important part of the scope of work. In cases of conflict between the contractual document and the provisions of this publication, the contractual document shall take precedence. It is not the intent of this publication to impose additional requirements on existing contracts. The contractual metrology and calibration requirements should be determined for each project.

This publication references other NASA Handbooks and is consistent with them. Since measurement quality requirements are written at a high level and technical information is treated generically, it is recommended that functional requirements, performance specifications, and related requirements for each measurement activity be determined for each project.

1.3 Scope

This publication addresses measurement assurance program methods as applied within and among NASA installations and serves as a guide to:

- Control measurement processes at the local level (one facility),
- Conduct measurement assurance programs in which a number of field installations are joint participants, and
- Conduct measurement integrity (round robin) experiments in which a number of field installations participate to assess the overall quality of particular measurement processes at a point in time.

[1] NASA RP 1342, *Metrology—Calibration and Measurement Process Guideline*, p, 167, (June 1994).

1.4 Definitions

It is recognized that there are different definitions, connotations, and preferences for terms used in the statistical, instrumentation, aerospace and metrology communities. Terms used in this publication are defined in Appendix A, Definitions. Recognized definitions are used wherever possible. Occasionally, an important definition is given in the body of the document.

2 Applicable Quality Requirements

2.1 Introduction

The 5300.4 series of NASA Handbooks for Reliability and Quality Assurance Programs have provisions for the establishment and utilization of a documented metrology system to control measurement processes and to provide objective evidence of quality conformance. The intent of these provisions is to assure consistency and conformance to specifications and tolerances of equipment, systems, materials and processes procured and/or used by NASA, its international partners, contractors, subcontractors, and suppliers.

2.2 Applicable Quality Program Documents

Provisions and information relevant to measurement quality requirements, measurement processes, and calibrations are set forth in the following NASA publications.

NHB 5300.4(1B), "Quality Program Provisions for Aeronautical and Space Systems Contractors"

NHB 5300.4(1C) "Inspection System Provisions for Aeronautical and Space System Contractors"

NHB 5300.4(1D-2), "Safety, Reliability, Maintainability and Quality Provisions for the Space Shuttle Program".

NASA Reference Publication RP 1342, "Metrology—Calibration and Measurement Process Guidelines."

2.3 Measurement Quality Requirements

NASA RP 1342 states, "*The objective of the design and control of measurement processes is to manage the risks taken in making decisions based on measurement data.*" Recognizing that all measurements are only estimates of the "true" value, it is important to control the uncertainty of measuring processes to ensure that the risk of making an unsatisfactory decision is minimized. Certain fundamental concepts enumerated below are critical to establishing measurement quality.

- Measurement process quality must be consistent with the end user's measurement requirements and established accuracy ratios.
- The complete measurement process must be included in the evaluation of the measurement quality.
- Uncertainty is a property of the measurement process and must be stable and quantifiable. All sources of uncertainty, including standards, instruments, environment, operator, and sensors must be included in the estimate of total uncertainty.
- Uncertainties grow progressively through the chain of measurements.
- Uncertainties from earlier links in the measurement chain must be quantified and included in the final uncertainty
- The measurement uncertainty for a process usually grows with time and the resulting growth must be included in the uncertainty.

2.4 Measurement Accuracy

The NASA quality documents identified in Section 2.2 establish the following requirements for measurement processes:

- Combined random and systematic uncertainties in any article or material measurement process shall not exceed 10% of the tolerance of the article or material characteristic being measured.

- Combined random and systematic uncertainties in any calibration measurement process shall not exceed 25% of the tolerance of the parameter being measured.

Authorization for exception shall be requested from the procuring NASA installation in both cases. The reader should refer to Section 8 of NASA RP 1342, *Metrology—Calibration and Measurement Process Guidelines* and other applicable documents for waivers and exceptions to established NASA policy regarding accuracy ratios and other measurement requirements.

2.5 Traceability

Traceability is the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties[1]. All measurement standards must be traceable, in the context of the above definition, to standards maintained by the National Institute of Standards and Technology (NIST); other national laboratories recognized by NIST[2]; Recognized consensus standards; and locally established and maintained intrinsic standards.

2.6 Instrument Control Approach

The intent of the instrument control approach is to assure the adequacy of an end item by using a measuring system that will contribute a negligible uncertainty to the measurement result. Typically, instrument(s) used to measure a component are 1/10 of the allowable tolerance specified for the end item. Similarly, the uncertainty of the overall measuring process used to calibrate other instruments must be no greater than 1/4 of the tolerance of the instrument or standard being calibrated. Other Government agencies have adopted similar requirements and incorporated them into various documents used to control measurement quality (MIL STD 45662A for example). Additionally, the measuring systems must be traceable to appropriate higher level standards (see Section 2.5). The resulting measurement chain, starting at the national level and ending with final measurement, involves one or more measurement processes and standards. At each link in the chain, the total uncertainty of the measurement process must comply with the requirements of Section 2.4. Therefore as the length of the chain increases, requirements at the higher echelons become more stringent. Measurement control documents usually require that the various stages or levels of calibration be documented so that, in principle, any measurement can be traced to its source (traceability).

---

[1] ISO Publication, International Vocabulary of Basic and General Terms in Metrology, Definition 6.12, p. 47 (1993).

[2] NIST has entered into agreements with other nations to mutually recognize each others capabilities in specific areas. NIST should be consulted to verify that such a recognition exists for a particular quantity at the magnitude in question.

2.7 Process Control Approach

Improvements in measurement technology have led to the development of instrumentation whose performance could previously only be achieved in measurement standards laboratories. In many cases, the traditional hierarchy of calibrations, each performed at an established accuracy ratio has been compressed or eliminated entirely. The traditional approach becomes ineffective when the measurement uncertainties required to characterize the properties of some item, process, or material, approach or exceed the measurement capabilities of the highest levels of the national measurement system. An alternative methodology has evolved to determine objectively the capability of a calibration or measurement process to achieve an acceptable level of performance. It is a holistic approach that treats the standards, procedures, equipment, environments, operators, and other influences as interacting to define a process that produces measurement results as a product. Sound metrology and attention to all facets of the process, coupled with traditional statistical process control (SPC) techniques are used to establish and monitor the adequacy of the measurement process for its intended application. Measurement quality assurance can only be established when the higher level calibration process and the local surveillance are both in a state of continuous statistical control. Key elements to attaining measurement quality assurance are enumerated below.

- There must be an ongoing operating program to ensure that the local standards are calibrated after reasonable intervals.
- All uncertainties associated with higher level calibrations must be evaluated and quantified as a clear statement of the uncertainty which includes the uncertainty of the higher echelon calibration, and any influences that affect the standard.
- There must be well defined and stable measuring processes for both standards and regular calibration activities.
- At the local level, there must be a continuous surveillance process to monitor the local standards between higher level calibrations.
- The result of local surveillance must be a determination of the process uncertainties, and an uncertainty statement for the local process.
- Control procedures must be in place to ensure that the uncertainties of the process remain stable with time.
- When out-of-control conditions arise, procedures must be in place and followed to eliminate the out-of-control condition.
- The calibration laboratory must produce and disseminate a meaningful uncertainty statement to its clients.

3 Measurements and Calibrations

3.1 General

A measurement is the comparison of an unknown (measurand) to a known (standard) and the result is only an estimate of the "true"[1] value (M_{true}) of the measurand. The result (M_{obs}) is only complete when the uncertainty of the estimate (U) is specified such that:

$$M_{obs} - U \leq M_{true} \leq M_{obs} + U \qquad (3.1).$$

Unlike the measurement result which is current, the uncertainty is estimated from previous measurement process data and is determined by statistical and analytical means. As U has a significant statistical component there is a finite probability that M_{true} will not lie within the region described by Eq. (3.1). Since measurements are used to make decisions and the uncertainty of a measurement must have a negligible effect on the decision, it is critical the measurement uncertainties be realistic and carefully documented.

3.2 Traceability

A measurement result is expressed in terms of agreed upon units that are defined, maintained and disseminated by national laboratories (NIST and others). These serve as the common reference for expressing the magnitude of the quantity being measured. Global compatibility is realized through a chain of calibrations from the national laboratory to the final end use. This chain is traceability and provides the end user with the assurance that the calibrated standard or measurement and test equipment (M&TE) for a particular quantity is a representation of the national, international, or consensus standard. The traceability chain must (1) be unbroken, (2) provide the client with a value(s) assigned in terms of accepted units, and (3) have a statement of uncertainty. Achieving these objectives requires:

(1) Calibration or verification of all local standards in terms of higher level standards,
(2) A local surveillance process to ensure the integrity of the standards between higher level interactions, and
(3) A measuring process to serve clients that has a quantifiable uncertainty with respect to the appropriate standards.

A word about the last two. They are often thought of as being the same. Often they are not as the internal surveillance process may differ significantly from the process used to calibrate client's standards and M&TE.

3.2.1 Measurement Compatibility

When a measurement is made the result is expressed as the product of a pure number representing the magnitude of the quantity and the unit. That is, if the result of a measurement is 10 meters it is the product of the pure number 10 and the unit – 1 meter. Calibration is a tool of traceability that transfers the unit from one standard/instrument to another. The process is best illustrated by considering the following scenario. An *invariant* quantity is measured using two different instruments (or in terms of different standards).

[1] The true value is never known.

Representing the magnitude, a pure number, as $\{N\}_i$ and the unit as θ_i the relationship between the two measurements, in different units, is described by Eq. (3.2)

$$\{N\}_A \cdot \theta_A = \{N\}_B \cdot \theta_B = \text{CONSTANT} \tag{3.2}$$

rearranging

$$\frac{\{N\}_A}{\{N\}_B} = \frac{\theta_B}{\theta_A} \tag{3.3}$$

As long as either A or B is known then the other is determined. If neither is known then *only* the ratio is determined. Traceable calibrations accomplishes the former, verifiably propagating units from one level to the next.

3.2.2 Calibration Hierarchy

Traceability is accomplished through a hierarchy of calibrations starting at the national laboratory using suitable standards. In the United States NIST disseminates representations of the national standards and these services are described in the current *NIST Calibration Services Guide* (NIST SP 250). These and other services offered by calibration laboratories serve to disseminate the units at various levels to a large client base. Not all standards are of NIST origin. For example, there is no national standard for hardness. Rather this standard is realized by consensus through an agreed on methodology and reference materials (RM). A further discussion of this topic is contained in Section 5 of *Metrology-Calibration and Measurement Process Guidelines* (RP 1342). Locally, units are acquired by calibration of artifact standards at a higher echelon which provide the user with a value and an uncertainty.

> *A standard deriving its value(s) by calibration in terms of a higher-level standard is not subject to arbitrary change by the user. Therefore, unless there is evidence to the contrary it must be assumed that the standard remains within its stated uncertainty between calibrations.*

Evidence about the local unit comes from the internal surveillance process designed to detect changes in the values of the local standards. The local surveillance process only yields information about changes relative to the group as a whole. Higher level calibration and local surveillance are two distinct and independent processes, each producing its own uncertainties.

> The calibration process by an external activity determines the value and its uncertainty of the standard at the time and location of its calibration.

> Surveillance of the standards at the local level monitors their integrity between higher level calibrations.

3.2.3 Calibrations and Measurement Assurance Programs

An important distinction needs to be made between higher level calibrations and MAPs. Today, unfortunately MAPs are thought of as a calibration process between NIST and the local laboratory. This is

not true. A measurement assurance program characterizes the total measuring process and includes results from external higher level calibrations an other relevant data generated by the measuring process. A calibration, on the other hand, is an element of a MAP. The following sections will further amplify the distinction.

3.3 Calibrations

The calibration method has a significant impact on the overall uncertainty of the local unit. Today dissemination of the units is usually accomplished by either conventional or measurement assurance program-type (MAP) calibrations. A conventional calibration is one in which the local laboratory sends one or more of its standards to a higher level laboratory. The MAP type usually uses a traveling (transport) standard[1], usually under the control of the higher level laboratory. Unfortunately the term MAP has two meanings. NIST and others refer to this type of calibration service as a MAP which, strictly speaking, it is only an element of a MAP as defined in Appendix A and mentioned in Section 3.2.3. To distinguish between the two, the interlaboratory calibration process will hereafter be called a MAP transfer (MAP-T). The use of conventional calibrations sometime leads to a larger overall uncertainty. The advantage of a MAP-T is that it usually yields a lower uncertainty for the values of the reference standards at the local laboratory. The essential features and differences of each are listed in Table 3.1.

Until recently MAP-T calibrations have been thought of as exercises conducted between NIST and its clients. NIST is no longer the sole purveyor of this type of calibration as competent laboratories also provide similar services. Moreover, by careful planning and execution a conventional calibration can achieve uncertainties approaching those of a formal MAP-T.

Table 3.1
Comparison of conventional calibrations and MAP-T

Conventional Calibrations	MAP Type Calibrations
Values at the *calibrating laboratory.*	Values in the local *laboratory.*
Calibrated under the conditions of the calibrating laboratory.	Calibrated under the conditions in the local laboratory.
Uncertainty includes only the conditions at the calibrating laboratory.	Uncertainty of the calibration includes those conditions in the local laboratory.
Provides little or no information on the effects of transportation or time.	Uncertainty includes effects of transport and time.
	Can identify, eliminate or reduce certain constant local systematic errors.

[1] Since the late 1960's the term "transport standard" has been used. To maintain international consistency, this guide will use "traveling standard."

3.3.1 Conventional Calibrations

Conventional calibrations usually involve the local laboratory sending one or more of their standards to the higher level laboratory where they are calibrated, then returned to the local laboratory. For certain standards this type of service is adequate. Examples are; standards unaffected by transportation or time such as mass standards; lower accuracy standards that will not be significantly affected by the transport process, and so forth. Many standards such as standard platinum resistance thermometers (SPRT), voltage standards, capacitors, and some resistors may be adversely affected by the transport process. The effect may be permanent or require very long settling periods. If a conventional method is to be used then there are certain precautions that must be taken to minimize the transport and time related uncertainties.

(1) Use the same standard for repeated calibrations.

(2) Do not adjust the standard at any time during its life unless necessary. *Note:* If an adjustment or repair is made it is usually necessary to treat it as though it is a new standard.

(3) Use the remaining local standards to calibrate the traveling standard before and after external calibration. When satisfied that the effect of the journey had no significant effect recalibrate all standards using the method of Sec. 3.4.1.

(4) Careful control of the transportation process. Parameters that must be addressed are (a) packing standard, (b) method of transportation, and (c) if necessary, environmental control during transit.

(5) Obtain a detailed uncertainty statement from the calibrating facilities.

3.3.2 MAP Type Calibrations (MAP-T)

A MAP-T performs an *in situ* calibration of local standards using a traveling standard in conjunction with an established protocol. Originally developed by NIST to improve the dissemination of certain units they are, in one form or another, being adapted to provide a range of higher quality calibration services. A MAP-T can be operated in several ways and at various accuracy levels. Customarily the higher level laboratory carefully controls the transfer process and

(1) Provides a suitable traveling standard(s),

(2) Calibrates the traveling standard before and after transport,

(3) Schedules the experiment and oversees the transportation process,

(4) Often prescribes the measurements to be made by the client laboratory,

(5) Often prescribes the format of the data to be submmitted for analysis, and

(6) Analyzes the results and supplies the client with a report.

3.3.2.1 Current NIST MAP Services

Currently NIST offers MAP services in nine disciplines as listed in Table 3.2. The latest edition of SP 250 should be consulted for available services.

Table 3.2
Current NIST MAP services*

SERVICE	TEST NUMBER
DC resistance	51110M
DC voltage	53120M
Dose interpretation of ferrous-ferric dosimeters	48010M – 48011M
Laser power and energy	42120M – 42140M & 42150M
Mass	22180M
Platinum resistance thermometry	33370M – 33390M
Retroreflectance	38070M – 38074M
Transmittance	38080M
Watthour meters (electrical energy)	56210M

* As listed in the 1991 edition of SP 250

3.3.2.2 Reverse MAP Transfers

MAP type transfers have a corollary, the reverse measurement assurance program (RMAP-T). A RMAP-T is simply a MAP-T initiated and operated by the local laboratory rather than by a higher level laboratory. Any capable laboratory can establish a MAP-T with its clients or an RMAP-T with a higher level facility (i.e., NIST) using the same basic techniques. Minimum requirements are:

- A fully evaluated stable measuring process having a well-determined process uncertainty,
- A suitable traveling standard;
- A sound transport process;
- Detailed procedures for the whole experiment
- A suitable protocol to ensure that the process is in a state of continuous statistical control,
- Established operating procedures to deal with out-of-control situations, and
- Interaction among participants to ensure that the transportation process, administrative and technical matters associated with the overall experiment are under control.

3.3.3 Group (Regional) MAP Transfers

A modification to the traditional technique is the group or regional measurement assurance program (G-MAP). This group of laboratories usually, in a specific region, band together to obtain a MAP-T from a higher level laboratory (usually NIST). Rather than each member interacting with the higher laboratory they use a "hub and spoke" approach as illustrated in Figure 3.1. One laboratory acts as the pivot (hub) hosting each participant's traveling standard and that from the higher level laboratory. This method reduces the number of transfers and by tracking interlaboratory differences serves as a system check standard. This method is designed to serve a local region where private transportation can be used to move standards in a short period of time (hours) but can be adapted to situations where other modes of transportation are used.

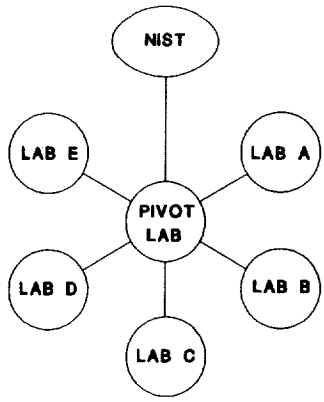

Figure 3.1 Pivot laboratory method for a GMAP. Although usual it is not necessary for the primary laboratory to be NIST. The participating laboratories surrounding the pivot laboratory are satellites to the pivot laboratory only for the experiment.

3.3.4 Other MAP Related Programs

A major variation in MAP-T is one in which the higher laboratory provides the client with a well characterized standard. The client then carries out the calibration using his own procedures, analyzes the results and makes his own decisions. The only services that the higher laboratory provide is the traveling standard (and value), transportation and usually some technical guidance. This method should produce results similar to that of the standard service. For situations where MAP-Ts are conducted between a calibration laboratory and users of M&TE this method is a very good compromise. For example, a laboratory that supports considerable M&TE might use a traveling standard to allow equipment in the field to be verified by either sending laboratory personnel or training the end user to perform the necessary calibrations.

3.3.4.1 NIST Services

The NIST offers several services whereby stable, well characterized artifact standards and/or instruments are provided to customers on a rental basis. The artifacts are characterized by NIST before and after being measured by the customer. The NIST data is then furnished to the customer. The customer uses the NIST results to establish the base for his in-house realization of the unit. Currently there two such rental standards available: the Luminous Intensity Rental Program (37015C) and the Photodiode Spectral Response Rental Package (39070C).

3.4 Using Calibration Results

At the local level the quality of the local standards is a function of the quality of the standard, the calibration including such parameters as measurements, transportation, time, etc., and how the resulting calibration data are used. Once the calibration is complete nothing can be done to improve the resulting data short of repeating the measurement.

A calibration should be considered as an experiment and planned to obtain the maximum information per measurement consistent with the purpose of the experiment.

Assuming that the guidelines for MAP-T or RMAP-T calibrations are followed, the first step is an evaluation of the data from the higher level laboratory and that obtained from the local before and after calibrations. Evidence of their quality lies in (1) the comparison of the current calibration with that from previous calibrations and (2) changes in the value of the traveling standard with respect to the remaining local standards. This evaluation should include all pertinent data including local surveillance data. A well conceived calibration produces three numbers, values for the traveling standard before and after the higher level calibration (T_{X1} and T_{X2}) and the result of the higher level calibration (T_S). Using Eq. (3.4) the difference between the two laboratories (ΔT) is calculated.

$$\Delta T = \frac{T_{X1} + T_{X2}}{2} - T_S \tag{3.4}$$

Eq. (3.4) assumes that there has been no significant drift in the traveling standard. If drift is significant then a suitable adjustment must be made. Assuming that there is a history on the traveling standard the change in the traveling standard ($T_{X1}-T_{X2}$) can be compared to data and formally tested using the student *t* test. (See Section 4.4.1). If it is determined that the difference is not statistically significant then one would accept the difference as valid and use the long-term uncertainty as the uncertainty of the current transfer. If statistically significant then (1) the experiment should be repeated or (2) use the result and assign a larger uncertainty based on the current transfer (calculate the standard deviation of the mean of the two values, $\sqrt{(T_{X1}-T_{X2})^2/2}$).

3.4.1 Adjusting Units

More than likely some adjustment to the local unit will be required. Whether or not to adjust after an external calibration varies with the circumstances. Traditionally, standards have been adjusted based on the last calibration.

> There is no physical adjustment to the standard. Rather an adjustment is made to the assigned numeric value so that the local unit is in agreement with the accepted one. It is a mathematical process based on Eq. (3.3)

Denoting the value assigned the traveling standard at the local laboratory as $[\{T(X)\}_{old} \cdot \theta(X)_{old}]$, that assigned by the calibrating laboratory as $[\{T(S)\} \cdot \theta(S)]$, the current value of the local standards as $[S(X)_{old} \cdot \theta(X)_{old}]$ the new value for the local standards is obtained using Eq. (3.5).

$$\{S(X)\}_{new} \cdot \theta(S) = [\{S(X)\}_{old} \cdot \theta(X)_{old}] \times \left(\frac{\{T(S)\} \cdot \theta(S)}{\{T(X)\}_{old} \cdot \theta(X)_{old}} \right) \tag{3.5}$$

This is, of course, identical to multiplying the magnitude of the value of the standard by $\{T(S)\}/\{T(X)\}$. This method is general and can be used to adjust the local unit regardless of the value of the traveling standard or the reference standard.

3.4.1.1 Example of Adjusting Standards

The local representation of the ohm is calibrated using a nominally 10 Ω traveling standard that was calibrated before and after the external calibration. The calibrating laboratory assigned the standard a value of 10.000 04 Ω(S). Locally, the value of the traveling standard was 10.000 08 Ω(X) in terms of the local standard having an accepted value of $R(ref)_{old}$ = 1.000 015 $\Omega(X)_{old}$. Using Eq. (3.5) the new value of the local standard is

$$R(ref)_{new} = R(ref)_{old} \times \left(\frac{R(S)}{R(X)} \right) = 1.000015\ \Omega(X)_{old} \times \left(\frac{10.000\ 04 \times \Omega(S)}{10.000\ 08 \times \Omega(X)_{old}} \right) = 1.000\ 011\ \Omega$$

As a "sanity check" substitute the new value for the reference and calculate the local value for the traveling standard. The result should be 10.000 04 Ω. Everything assigned using the local standards must be recalculated to reflect the adjustment. *Recall that this is not a change in the resistors, it is merely a reassignment of their values.*

3.4.2 Calibration Intervals

Calibration intervals depend on the standard, accuracy objectives, end-of-period (EOP) reliability and other factors. A extensive discussion of this topic can be found in NASA RP 1342 *Metrology-Calibration and Measurement Process Guidelines*. In general, establishment of intervals for standards is not quite as complex. Calibration intervals are basically driven by the uncertainty goals of the laboratory. Major factors that influence the establishment of realistic calibration intervals are the: (1) contractual or regulatory requirements, (2) final uncertainty goals, (3) the quality and age of the standards, (4) local environmental influences, (5) quality of the process used to transfer the unit to the local level, and (6) quality of the local measuring process. All must be carefully evaluated for each specific measurement area and procedures established and followed to meet the established uncertainty goals.

3.5 Local Surveillance

External calibrations are a necessary but not a sufficient condition for local measurement assurance. Between such calibrations there must be protocols in place to monitor the relative behavior of the standards and the day-to-day client related services to ensure they meet established end use requirements (satisfy the client). The precise techniques vary from standard to standard but there are three basic requirements that must be satisfied.

Requirement 1	Use the historical calibration data for the external calibration process to (1) assist in establishing realistic recalibration interval, and (2) estimate the long-term uncertainty of the standards.

Requirement 2	Monitor the differences between (or ratio of) local standards in the interval between higher level calibrations to (1) identify any abnormalities in performance, (2) estimate uncertainties associated with time and use, and (3) assign working values to them in accordance with the values assigned by the higher echelon.

Requirement 3	Use check standards to continuously monitor the process.

4.1 Control Charts

Control charts, first introduced by Shewhart in the 1920s for industrial process quality control, are now an important element for any measurement assurance program. Many magical properties are ascribed to the control chart but they only yield very specific information.

> Control charts detect variations in the process that are not random.

Control charts use process generated data to establish limits of expected variability for the process. A total process may require control of several independent variables such as assigned value of standards, check standards, etc. *Data falling within the assigned limits are deemed to be caused by random variations in the process and require no action. Those outside the limits are regarded as due to assignable causes and require action.* When all points fall within the limits the process is said to be in a "state of statistical control."

> Control charts do not detect constant systematic process errors.

Systematic errors or uncompensated effects that remain constant such as the incorrect value for a standard, incorrect constants, incorrect algorithms and improperly evaluated software, to name a few, are not detectable by control charts. Also, effects related to the external calibration are systematic to the process until the next calibration. Upon subsequent calibration, the latter should be detected .

> Control charts are valid only for a specific measuring process.

Since chart parameters are derived from the process, any process change or modification becomes a new process. This is a double edged sword. Failure to reevaluate a modified process can lead to serious problems. On the other hand, control charts serve to compare a new or modified measurement process with its predecessor. Such comparisons, often visual, help in evaluating the new process. No change in the monitored variable plus a variability reduction may be the basis for adopting the new or modified process. A shift in the variable means that there is a problem with one or both processes that must be explored if deemed significant.

> Control charts are not specification limits or tolerances.

They are helpful in establishing or reestablishing these parameters. Specifications and tolerances are externally imposed and determine the measurement process uncertainly requirements which are monitored by the control chart.

> Control charts provide data for establishing measurement uncertainty.

Data from the process control charts are essential for estimating the total measurement uncertainty. For example, control chart data from a check standard is a primary tool for estimating the local measuring

process uncertainty. Similarly, control charts for local standards is the basis for estimating their component of uncertainty for the process.

Croarkin, in NBS SP 676-II, *Measurement Assurance Programs Part II: Development and Implementation* summarized the role of the control chart in metrology as follows:

- The parameters for statistical control of the properties of the standard or measuring process are not imposed on the process but are a property of the process,
- If any measurement result for a check standard or other property monitored is outside the established limits the process is presumed to be out-of-control at the time of the measurement,
- The process precision is characterized by the standard deviation calculated from measurement results from the check standard or similar measurement, and
- In spite of the automation of control processes, visual inspection of the control charts is essential to understanding the process and detecting anomalies.

4.1.1 Control Chart Candidates

Any process or standard for which there is repetitive process data is a candidate for a control chart. Some important parameters that can be monitored using control charts are:

- Calibration history of local standards,
- Check standards,
- Internal surveillance of local standards,
- Standard deviations(ranges) associated with various portions of the process, and
- Instrument offsets, temperature, pressure and other influences that affect the process.

Measurement process control charts monitor quantities such as the value of a standard or the variability of the process. The former are called x-bar charts, the latter r- or s-charts (range or standard deviation). A single measurement process may require several charts to monitor various influences that affect the process.

4.1.2 Control Chart Basics

Control charts are based solely on process generated data and established procedures for their construction and have four main elements. First, a basic chart initially requires from 4 to 25 observations, Second, using that data a central line is calculated. For the x-bar chart, the central line is the mean; for the s-chart it is a function of the pooled standard deviation. Third, upper and lower control limits (UCL and LCL) are calculated based on the process data. Finally, new data is added, as acquired, to the chart and the state of control of each point determined by comparing it to the control limits.

4.1.3 X-Bar Charts

The x-bar charts monitor calibration data for standards, internal surveillance data, a check standard and influences. Depending on the parameter being monitored, they are based on either a sampling technique or single observations. The differences are the way the data is obtained and how the standard deviation used

to set the limits is calculated. Most situations in metrology deal with two basic components of variability called the within-day standard deviation (s_w) and the between-day standard deviation (s_b). The within-day standard deviation measures the variability of the process during the measurement interval that as a rule is short and is estimated from replicate measurements or the use of an experiment design that produces a standard deviation. For example, when calibrating a gage block the operator makes three observations. The standard deviation of the set estimates s_w. The between-day standard deviation measures the time variation of the overall process that includes the effect of variables such as temperature, humidity, and operator. It is calculated from repeated measurements over a time interval (days, weeks, or months). The within-day, between-day, and total standard deviation, s are related by Eq. (4.1).

$$s = \sqrt{s_w^2 + s_b^2} \tag{4.1}$$

For example, a group of standard cells calibrated using an experiment design yields an estimate of s_w that is usually 0.1 μV or less for a particular run. However, the standard deviation s of a number of runs made over a time interval, e.g., 10 days, is in the order of 0.2–0.3 μV. This latter figure is the root sum of squares value of s_w and s_b as defined by Eq. (4.1). The causes of s_b are real and include changes in the cell emf, effects of influences such as ambient temperature, temperature measurement errors, operator, and long-term instrument variability. If $s_w = 0.1$ μV and $s = 0.2$ μV, then s_b is calculated by solving Eq. (4.1) for s_b (0.17 μV). The presumption is that both the within-day and between-day components come from normally distributed populations.

4.1.3.1 Single Observation X-Bar Chart

The single observation x-bar chart is the one most commonly encountered in metrology. The data is usually a single observation (calibration of a standard) or the mean of a set that has a very small s_w in comparison to s_b, and is illustrated by the data of Table 4.1[1]. The data is a simulation of this type of chart. The four step procedure for creating and maintaining the chart is given below and illustrated in Figure 4.1.

Table 4.1
Example for single observation control chart

Time (days)	Obs'd value	Time (days)	Obs'd value	Time (days)	Obs'd value
0	9.7	112.4	12.4	217.3	9.9
10.6	10.4	123.2	10.3	236.6	10.0
14.8	11.8	134.1	9.3	243.1	11.3
22.4	10.0	141.0	12.2	249.8	8.5
41.1	10.1	142.6	8.7	264.5	10.6
56.0	9.5	161.5	10.8		
72.3	8.2	165.7	9.2		
80.8	11.5	169.4	9.6		
93.6	9.5	186.2	11.0		
94.9	10.5	198.9	10.2		

[1] The data for this table was created using random numbers and having a mean of 10 and a standard deviation of 1.

Step 1: Initial control chart parameters are calculated using from 4 to 25 observations. For this example 10 values were used. Using the first ten values of the table ($n = 10$), the mean ($\bar{x}$.) and standard deviation (s) are calculated using Eqs. (4.2) and (4.3) respectively.

Note – Control charts used in production situations generally produce copious amounts of data so a reasonable database is quickly acquired. This is not so in metrology. This example required about 3 months to obtain 10 data observations. Rather than wait that long, one should establish an interim control chart using four observations (or in some cases where data comes very slowly, months or years, with three observations). Had this been done for this example the mean and standard deviation would have been approximately 10.5 and 0.9 respectively. When the data base reaches the desired level then new limits can be calculated. In this example, the outcome using the 3σ control limits would not have changed.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 10.120 \tag{4.2}$$

$$s = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n} (x_i - \bar{x})^2} = 1.0326 \tag{4.3}$$

Step 2: Plot the 10 observed values as a function of time (t) and enter the mean line (central value) as shown in Figure 4.1.

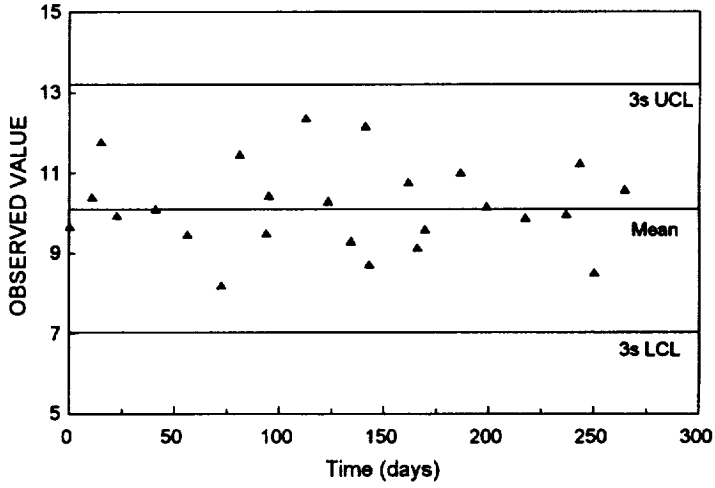

Figure 4.1 Control chart for single observation data. This example was created using normally distributed random numbers with a mean of 10 and a standard deviation of 1.

Step 3: Compute UCL and LCL using Eqs. (4.4) and (4.5) and enter them on the chart. The constant A 3 approximates the 3 sigma limits (See Section 4.1.4 for additional discussion on setting limits).

$$LCL = \bar{x} - A \cdot s = 10.12 - 3 \cdot 1.033 = 7.02 \quad (4.4)$$

$$UCL = \bar{x} + A \cdot s = 10.12 + 3 \cdot 1.033 = 13.22 \quad (4.5)$$

Step 4: As additional data is obtained, it is promptly added to the chart and inspected to ensure that the process is still in control. (Figure 4.1). If the initial data set is small, less than 10, it is wise to revise it later but not every time a new observation is acquired. If the process remains in control no further revisions are usually necessary.

4.1.3.2 Sample Based X-Bar Charts

Situations arise in which s_b is small compared to s_w. In such cases use s_w to construct the chart. The steps are the same as just discussed except computing the standard deviation. Instead of using Eq. (4.3) use either Eq. (4.6) or (4.7) to pool the individual within-day standard deviations used to set the limits. Otherwise, the technique is the same as that of Sec. 4.1.3.1. Table 4.2 and Figure 4.2 are examples of this technique. Again, the first ten runs serve to establish the limits. This example is based on a computer simulation of the dimensional measurement of the outside diameter standards (cylindrical plug gages) sized between 1 mm and 25 mm in 0.5 mm increments. The specified tolerance is ±0.0025 mm and the direct reading length measuring machine has a resolution and accuracy of 0.001 mm. The requirement for a conventional accuracy ratio of 4 to 1 cannot be satisfied by the available measuring machines. Accordingly, along with each set of test items, four check standards are also measured having nominal diameters of 1.5, 3, 9 and 20 mm. The check standards are periodically measured for diameter, and taper using a laser interferometer measuring machine and for roundness using a Talyrond. The check standard dimensional characteristics are known to 0.0001 mm. This information is not known to the operators. The final result is the mean of three measurements on each artifact. Note: One must be very careful about roundoff error. A safer way to handle the data would be to use the observed minus the nominal and expressed in parts in 10^{-6} or similar format.

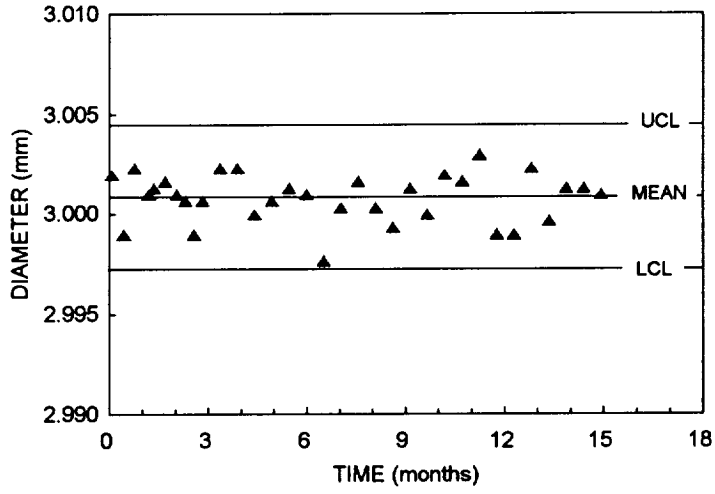

Figure 4.2 Simulated x-bar chart for a 3 mm plug gage check standard. Mean and control limits are based on the first 10 observations. Each point is the mean of 3 observations.

Table 4.2
Dimensional check standard measurements

	Date	Elapsed Time from (Months)	Obs 1	Obs 2	Obs 3	Mean	Std. Dev. of mean
			Diameter in (mm)				
1	01/03/91	0.066	3.003	2.999	3.003	3.0017	0.00133
2	01/14/91	0.427	2.999	2.999	2.999	2.9990	0.00000
3	01/24/91	0.756	3.002	3.003	3.002	3.0023	0.00033
4	02/06/91	1.183	3.000	3.003	3.000	3.0010	0.00100
5	02/11/91	1.347	3.002	3.000	3.002	3.0013	0.00067
6	02/21/91	1.676	3.004	2.997	3.004	3.0017	0.00233
7	03/04/91	2.037	3.003	2.997	3.003	3.0010	0.00200
8	03/12/91	2.300	3.002	2.998	3.002	3.0007	0.00133
9	03/20/91	2.563	2.999	2.999	2.999	2.9990	0.00000
10	0.001177	2.825	3.001	3.000	3.001	3.0007	0.00033
11	04/13/91	3.351	3.003	3.001	3.003	3.0023	0.00067
12	04/29/91	3.877	3.002	3.003	3.002	3.0023	0.00033
13	05/15/91	4.402	3.000	3.000	3.000	3.0000	0.00000
14	05/31/91	4.928	3.000	3.002	3.000	3.0007	0.00067
15	06/16/91	5.454	3.003	2.998	3.003	3.0013	0.00167
16	07/02/91	5.979	3.001	3.001	3.001	3.0010	0.00000
17	07/18/91	6.505	2.997	2.999	2.997	2.9977	0.00067
18	08/03/91	7.031	3.001	2.999	3.001	3.0003	0.00067
19	08/19/91	7.556	3.002	3.001	3.002	3.0017	0.00033
20	09/04/91	8.082	3.000	3.001	3.000	3.0003	0.00033
21	09/20/91	8.608	2.999	3.000	2.999	2.9993	0.00033
22	10/06/91	9.133	3.002	3.000	3.002	3.0013	0.00067
23	10/22/91	9.659	2.999	3.002	2.999	3.0000	0.00100
24	11/07/91	10.185	3.004	2.998	3.004	3.0020	0.00200
25	11/23/91	10.710	3.003	2.999	3.003	3.0017	0.00133
26	12/09/91	11.236	3.004	3.001	3.004	3.0030	0.00100
27	12/25/91	11.762	2.997	3.003	2.997	2.9990	0.00200
28	01/10/92	12.287	2.997	3.003	2.997	2.9990	0.00200
29	01/26/92	12.813	3.003	3.001	3.003	3.0023	0.00067
30	02/11/92	13.339	2.999	3.001	2.999	2.9997	0.00067
31	02/27/92	13.864	3.003	2.998	3.003	3.0013	0.00167
32	03/14/92	14.390	3.001	3.002	3.001	3.0013	0.00033
33	03/30/92	14.916	3.000	3.003	3.000	3.0010	0.00100

4.1.4 Setting X-Bar Control Limits

The traditional control chart practice is to set A to 3 that corresponds to the 3-sigma control limits. If sigma is known the probability that any observed value will be out of control is about 1 chance in a 1000. Alternately, Croarkin[1] proposed using the Student-t distribution to establish control limits. For small samples the limits can be very large at, say, the 99% confidence level. In this example, $t_{0.99}$=9.92 for 2 degrees of freedom. This rigorous method will not quickly detect a problem, so the use of A=3 is preferred. Recall that the function of a control chart is to set off an alarm to warn the operator of potential problems. The 3-sigma limit is well suited for that purpose but it has a price. The 3-sigma limit is not good at detecting trouble when it exists. To alleviate this problem, some establish warning limits at 2-sigma. Such limits say that there *may* be a problem in the making which it is probably not serious now but should not be ignored. For routine control this practice is not recommended but can be used when refining a process of looking for subtle effects. Finally, control charts should be examined for possible trends, small process shifts and other changes that may be harbingers of future problems. If one observation is out of control, or eight (or nine) successive points are above or below the central line the process should be investigated for possible assignable causes.

4.1.5 Standard Deviation Charts

Standard deviation control charts for the single observation case cannot be constructed but can and should be for the sample case (Sec. 4.1.3.2). Control of s_w monitors the performance of day to day measurements and their use should be coordinated with the x-bar chart. The technique for constructing a s-chart is similar to that for the x-bar. The example below is based on the standard deviations in the last column of Table 4.2, that is, the standard deviation of the mean of the three observations.

4.1.5.1 Standard Deviation Charts Using Pooled Standard Deviation

Standard deviation control charts are constructed in the same general way as x-bar charts as illustrated below. But before proceeding, a word about the LCL for s. Control charts detect out-of-control conditions therefore, points outside the LCL require the same attention as those exceeding the UCL because it suggests assignable cause. Often such a condition suggests more than one process. Many points exceeding the LCL also suggest that with modification s can be permanently reduced. Using the data of Table 4.2 and referring to Figure 4.3 an s chart is constructed in the following manner.

Step 1: Using an initial set of ten observations estimate σ by pooling the standard deviations (s_p) in the last column of Table 4.2 by the RMS method to estimate sigma (σ).

$$\sigma \cong s_p = \sqrt{\frac{\upsilon_1 s_1^2 + \upsilon_2 s_2^2 + \ldots \upsilon_n s_n^2}{\upsilon_1 + \upsilon_2 + \ldots \upsilon_n}} = 0.00121 \text{ mm} \qquad (4.6).$$

When all data sets have the same numbers of degrees of freedom then Eq. (4.6) simplifies to

[1] Croarkin, C., *Measurement Assurance Programs Part II: Development and Implementation*, NBS SP 676-II, pp. 95, (April 1984).

$$s_p = \sqrt{\frac{s_1^2 + s_2^2 + \ldots s_n^2}{n}} = 0.00121 \text{ mm} \quad (4.7).$$

Step 2: The central line (CL) is obtained from Table B.1 for 3 sigma and 3 observations. Unlike x-bar charts the center line is not s_p. Instead it is based on the χ^2 distribution such that 50% of the observed standard deviations will be above the CL and 50% below.

$$CL = B_{CL}\, s_p = 0.833 \cdot 0.00121 = 0.00101 \text{ mm} \quad (4.8)$$

Step 3: Establish the lower and upper control limits using Eqs. (4.9) and (4.10) and Table B.1. The limits are not symmetrical about the CL because the nature of the χ^2 distribution. For small sample sizes the lower limit is zero (5 for 3σ and 2 for 2σ).

$$LCL = B_{LCL}\, s_P = 0 \cdot 0.00121 = 0 \text{ mm} \quad (4.9)$$

$$UCL = B_{UCL}\, s_P = 2.76 \cdot 0.00121 = 0.00334 \text{ mm} \quad (4.10)$$

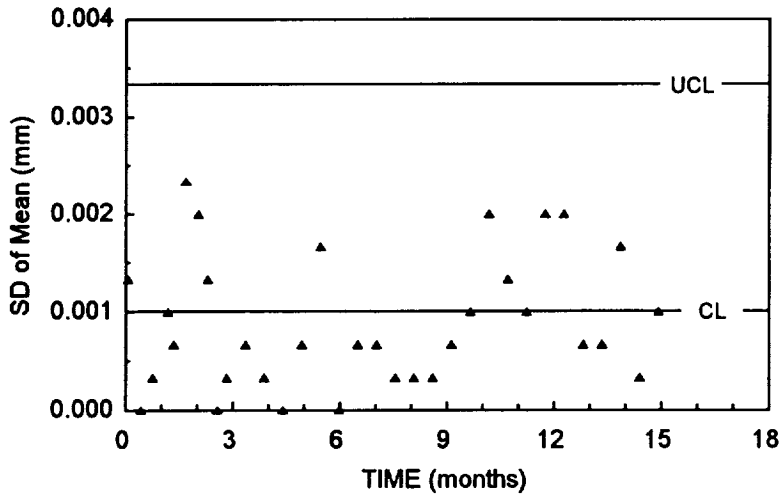

Figure 4.3 Simulated *s* chart for a 3 mm plug gage check standard. The CL and UCL are based on the first 10 observations. The number of observations is too small to establish a LCL.

Step 4: Maintain the chart in conjunction with the x-bar chart.

The average standard deviation ($\bar{s}$) can also be used to establish control limits for *s* in conjunction with tables found in statistical process control texts and manuals. Today with the availability of the spreadsheet and computer tools the method just discussed is preferred because s_p is needed to calculate uncertainty.

4.1.6 Control Charts with Drift

Many artifact standards do not remain constant with time. If a standard shows predictable behavior (usually a linear drift) then another method can be employed to establish a control chart. Using an empirical model and the Method of Least Squares (MLS)[1] one can create a control chart, predict future values, and estimate the drift rate of a standard.

> There is usually no physical basis for using the model, therefore, the model must be tested every time new data is added to the database.

Linear fits of data can be carried out by hand, using special least square programs or using a spreadsheet with the latter offering the most convenience. Most spreadsheets have a single command to fit data to a variety of functions. The command structure of the latter requires only three inputs; an array containing the values of x (usually time), one containing the observations to be fitted to the x's, and the location of the output. The output gives the coefficients, their standard deviation, and the standard deviation of a single observation. Unlike conventional control charts, this type must be updated every time new data is added. The procedure for constructing a control chart is illustrated using the data of Table 4.3 (and plotted in Figure 4.4). The table summarizes the results of periodic NIST calibrations of the 10 V output of a client's solid-state voltage reference (SSVR).

Table 4.3
Calibration data for a 10 v solid-state voltage reference

Time (mo)	Deviation from Nominal (ΔE) μV
0.99	2.00
5.00	3.40
9.57	3.80
13.94	3.70
18.08	4.70
23.31	6.20
28.01	7.70
39.98	8.62

Step 1. Inspection of the data shows that a linear model of the form described by Eq.(4.11) approximates the standard's behavior.

[1] The Method of Least Squares is also known as regression or for the linear case "linear regression." When using a spreadsheet the command is usually "regression."

$$y = \beta_0 + \beta_1(x - x_0) \tag{4.11}$$

where β_0 is the intercept and β_1 the slope. Substituting the SSVR voltage change (ΔE) for y and time interval ($t-t_0$) for x Eq. (4.11) becomes

$$\Delta E = \beta_0 + \beta_1(t - t_0) \tag{4.12}$$

Step 2: Using the MLS calculate the intercept (β_0), slope (β_1), the standard deviation of a single observation (s_y) and the standard deviation of the slope ($s_{\beta 1}$) as summarized below.

$n =$ 8
$\beta_0 =$ 2.003 µV
$\beta_1 =$ 0.173 µV/mo
$s_y =$ 0.558 µV
$s_{\beta 1} =$ 0.016 µV/mo

Step 3: Using the above coefficients in Eq. (4.12) calculate the predicted line and draw it on the chart.

Step 4: Using Eq. (4.13) calculate s_{pred}. The control limits (LCL and UCL) are constructed as $3 \pm s_{pred}$ about the predicted values. Unlike the previous control charts the control limits are not constant because the standard deviation (s_{pred}) is time dependent. The parameters of Eq. (4.13) are n, the number of observations, t, the time and $\bar{t}$, the mean of the n observed times.

$$s_{pred} = s_y \sqrt{1 + \frac{1}{n} + (t-\bar{t})^2 \left(\frac{s_{\beta 1}}{s_y} \right)^2} \tag{4.13}$$

A word about the equation. For values near $\bar{t}$, s_{pred} approaches s_y but as $(t-\bar{t})$ increases so does s_{pred}. Eventually the $(t-\bar{t})$ term becomes dominant especially when extrapolating for a large time interval.

Step 5: When a new data set is obtained, verify its control status and if in control *update the chart by repeating Steps 1-4.* An out-of-control condition must be dealt with on a case by case basis.

An alternate method is simply to use the current s_y, a constant to set the limits (Step 4). The lines are not shown in the figure but would be inside the limits and nearly tangent at (at $t=\bar{t}$, $s_{pred} = s_y \sqrt{(1 + 1/n)}$). Additionally, they would be parallel to the prediction line. This method may sound the alarm a bit more often but it does simplify the calculations. If used and a point is out of control the exact method could be used for a final confirmation. How well the empirical model fits the data can be estimated by examining the chart. In this instance there may be some very small unidentified cyclic process taking place. Because the model is empirical it often fails when used over an extended time interval When this happens dropping earlier points will often correct the problem. If not, another model should be developed.

Such control limits are valid only for the next observation and must be updated to ensure the validity of the control process.

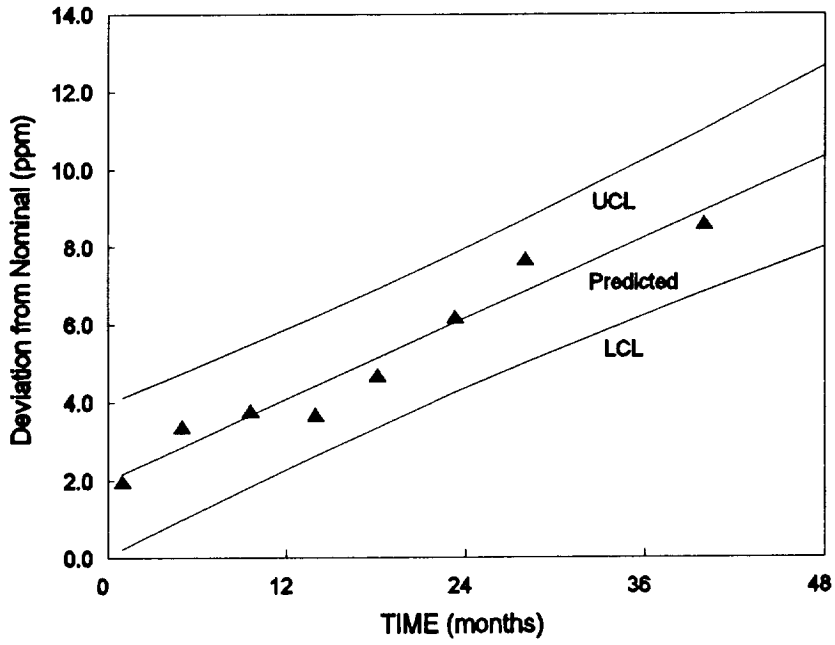

Figure 4.4 Control chart for a standard with empirically predictable drift. Data based on NIST calibration of a SSVR. Control limits are based on Eq. (4.13). If the simple limits discussed below are used they will be parallel to the predicted line and inside those shown.

4.1.7 Predicting Future Values

This type of control chart has another and perhaps more important function: predicting future values for the standard. Conventionally, the last calibrated value is used to assign a value for use until the next calibration. When drift is predictable as in Figure 4.4 it is better to predict future values than to simply use the last value. If the model has worked in the past, there is no reason to think it will not be valid until the next calibration. Therefore the best value for the standard at any time before the next calibration is the one predicted by the Eq. (4.12). Additionally, the uncertainty at the time of use (s_{calc}) can be estimated using Eq. (4.14).

$$s_{calc} = \sqrt{\frac{1}{n} + (t-\bar{t})^2 \left(\frac{s_{\beta 1}}{s_y} \right)^2} \qquad (4.14)$$

This equation is very similar to that used for control, differing in that the unity term is missing and is used to estimate the uncertainty of a calculated value. Note that the s_{calc} is time dependent that results in a moving uncertainty statement. One conservative way to stabilize the uncertainty is to calculate the uncertainty at the time of the next scheduled calibration.

Data from the control processes is an important element for setting calibration intervals.

4.2 Expressing Measurement Uncertainty

Many methods exist for expressing measurement uncertainty and most of them yield different results given the same input data. This lack of agreement has often brought on confusion, and sometimes acrimony between parties. As tolerances tighten, accuracy ratios get smaller, and economies become dependent on other nations, there needs to be a uniform method for expressing measurement uncertainty. The International Committee for Weights and Measures (CIPM), in 1978, recognized this problem and instituted a study to bring about such uniformity. The study, completed in 1980 was the basis for a uniform method for expressing the uncertainty of physical measurements. Starting with the 1980 CIPM recommendations, the BIPM, IEC, ISO, OIML, and other international organizations developed the *Guide to the Expression of Uncertainty in Measurement* (issued in 1993) which serves to harmonize expressions of uncertainty for calibrations, basic research, the certification of standard reference materials, instruments, and other measurements.

4.2.1 Conventional Expressions of Uncertainty

Measurement uncertainty is expressed in many ways that lead to widely differing results given the same starting data. Three are in common use in metrology which, for this discussion, are called the "linear," "quadrature" and "hybrid" methods. Each has its adherents and detractors for combining systematic errors (symbol B) and random errors (symbol s)[1].

Linear method: The linear method assumes that all errors are additive in one direction and represents the worst case scenario. It combines random and systematic errors by first adding the magnitude (without regard to sign) of each type separately then adding the sums of the two types as illustrated by Eqs (4.15), (4.16) and (4.17).

$$\text{systematic errors} \qquad B = B_1 + B_2 + \ldots B_n \tag{4.15}$$

$$\text{random errors} \qquad s = s_1 + s_2 + \ldots s_3 \tag{4.16}$$

$$\text{uncertainty} \qquad U = B + t_p s \tag{4.17}$$

The multiplier t_p is the Student's-t distribution (see section 4.4.1) for the appropriate degrees of freedom for the random error that establishes the confidence interval for random uncertainty. Systematic errors are estimated from scientific judgment, manufacturer's specifications or other sources and is rarely stated at a confidence level.

---

[1] The terms and symbols used in this paragraph should not be confused with those to follow. In the context of this section B and s refer to systematic and random error respectively.

Quadrature method: The quadrature method assumes that each class of errors are independent and are combined by the root-sum-of-squares (RSS) method. Individual errors are estimated as above but are combined using Eqs (4.18), (4.19), and (4.20). This method yields the smallest uncertainty of the three.

$$\text{systematic error} \qquad B = \sqrt{B_1^2 + B_2^2 + \ldots B_n^2} \tag{4.18}$$

$$\text{random error} \qquad s = \sqrt{s_1^2 + s_2^2 + \ldots s_n^2)} \tag{4.19}$$

$$\text{uncertainty} \qquad U = \sqrt{B^2 + (t_p s)^2} \tag{4.20}$$

Hybrid method: The hybrid method combines random and systematic uncertainties using Eqs (4.18) and (4.19) then combines the two classes using the linear one as shown in Eq. (4.21).

$$U = B + t_p s \tag{4.21}$$

Uncertainties estimated by the first two can differ significantly (sometimes as much as a factor of 2) and often lead to different decisions given the same input data. The last method gives results somewhere in between the other two. How systematic errors are assessed is not defined nor are other important factors that affect the final uncertainty.

4.3 CIPM Method (NIST Interpretation)

NIST developed and published its interpretation of the CIPM method as presented in the ISO guide as Technical Note 1297, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results* in 1993 and revised in 1994, referred to hereafter as the *Guidelines*. Use of the latest *Guidelines* is recommended. The CIPM method addresses four major components: (1) estimation of the standard uncertainty (estimated standard deviation) for each contributing uncertainty component; (2) determination of the combined standard uncertainty; (3) calculation of the expanded uncertainty; and (4) reporting the uncertainty of the measurement result. A summary of the points addressed in the *Guidelines* for the CIPM method follows and the latest complete text is incorporated into this document by reference. Additionally, the National Conference of Standards Laboratories (NCSL) has also issued NCSL RP-12, *Determining and Reporting Measurement Uncertainty,* which is based on the CIPM method.

1. Components of uncertainty are grouped into two categories depending on the method used to estimate their numerical value (**Type A** and **Type B**)

   Note: Adopting entirely new terms is intended to eliminate the confusion and controversy over the terms random error, systematic error and bias. Other new terms introduced in the Guide are also highlighted in bold as they are introduced.

2. Type A components are denoted by u_i and are evaluated by statistical methods; in particular, they are evaluated by the calculation of the familiar statistical standard deviation s_i based on the experimental data. Each Type A component, denoted by u_i, is called the **standard uncertainty**.

   $$u_i = s_i$$

3. Type B components, denoted by u_j, are those which cannot be evaluated by statistical means and are evaluated by other means. Type B components are also evaluated as standard deviations and are also called **standard uncertainties** but their evaluation uses a different methodology. Type B standard uncertainties "may be considered as an approximation to the corresponding [Type A] standard deviation; it is equal to the positive square root of u_j^2 and which may be considered an approximation to the corresponding [Type A] variance and is obtained from an assumed probability distribution based on all available information." (NIST TN 1297 paragraph 2.6) Type B standard uncertainties are evaluated based on scientific judgment. NIST TN 1297 lists several methods for quantifying this type of uncertainty component.

4. Type A and B do not always correspond to the terms "random" and "systematic" or "bias." The type is use independent. For example: if the uncertainty of a calibration contains only Type A components, the resulting uncertainty is always Type A no matter how the result of the calibration is used (i.e., it never becomes Type B in the manner that a "random" component can become a "systematic" component).

5. The individual standard uncertainties are combined using the *law of propagation of uncertainty*, usually called the "root-sum-of-the-squares" (square root of the sum-of-the-squares) or RSS method, to form the **combined standard uncertainty** which is denoted by the symbol u_c. Although specifically not stated in the NIST TN-1297 or elsewhere, individual Type A and Type B standard uncertainties are often combined separately, these are combined to yield the final standard uncertainty. That is:

$$u_i = \sqrt{u_{i,1}^2 + u_{i,2}^2 \ldots + u_{i,n}^2} \tag{4.22}$$

and

$$u_j = \sqrt{u_{j,1}^2 + u_{j,2}^2 \ldots + u_{j,n}^2} \tag{4.23}$$

which are then combined to yield the overall uncertainty,

$$u_c = \sqrt{u_i^2 + u_j^2} \tag{4.24}$$

Of course this is identical to combining both Type A and Type B components at one time, but this method serves to highlight the magnitude of the two categories.

6. Estimation of uncertainties assumes that corrections for all determinable or significant systematic effects have been made. Practically speaking this is not always the case, especially when an instrument is verified to ensure that it is within certain specifications.

7. For many situations the resulting uncertainty can be assumed to characterize an approximately normal (Gaussian) distribution.

8. The terms "confidence interval" and "confidence level" are not used because of their very specific statistical definition. Instead for the latter, the terms **coverage probability** or **level of confidence** (p)

are used to avoid ambiguity. The *Guidelines* notes, that for those cases where the normal distribution condition exists or are assumed, the probability that an observed value will lie in the interval $y \pm u_c$ is approximately 68 percent.

9. Many situations arise where it is necessary to state an uncertainty to a specified level of confidence. This uncertainty is known as the **expanded uncertainty** (symbol U) and is calculated by multiplying the combined standard uncertainty by a **coverage factor** (k):

$$U = k u_c \tag{4.25}$$

10. Unless otherwise justified NIST takes k to be 2[1]. Where the normal distribution situation applies, this corresponds to a level of confidence of approximately 95%. Because of the uncertainties in determining the precise probability distributions of the various components, a more precise statement of the level of confidence is not practical.

11. CIPM does not specify how to establish the relationship between k and p but some possible methods are presented in the *Guidelines*.

12. **All information required to reconstruct or dissect the reported uncertainty must be provided. This would include but not be limited to the following:**

    - **Report U with coverage factor used;**
    - **List all components of standard uncertainty and their type (A or B);**
    - **Describe how each component was evaluated;**
    - **Describe how k was obtained if not taken equal to 2; and**
    - **If stated, describe how and on what basis the level of confidence for U or u_c was obtained.**

4.3.1 Using the CIPM Method

Table 4.4 is one such example based on a typical calibration of standard cells. Estimating measurement uncertainty depends on an in-depth understanding of the measuring process and a suitable mathematical model that includes all of the parameters that affect the final result. The individual sources of uncertainty are listed and their type given and have been normalized (reported in ppm). The numbers to the left are for reference only. Briefly each was arrived at as shown in Table 4.5.

[1] NIST adopted 2 instead of 3 to be consistent with international practice.

Table 4.4
Uncertainty analysis for standard cells using the CIPM method

	Source of Uncertainty	1 std. dev. estimate (ppm)	Type
1	NIST calibration uncertainty	0.065	A&B*
2	Temperature measurement error at NIST	0.100	B
3	Transportation effects	0.200	B
4	Changes of unit with time	0.333	B
5	Local cell calibration	0.07 0	A
6	Local temperature measurements	0.08 0	B
7	Measurement system	0.07 0	A
	Total Estimated uncertainty	**0.426**	

* For convenience they have been combined (SP 250-24, p.21)

Table 4.5
Sources of uncertainty for Table 4.4

Uncertainty Source	How Estimated
NIST calibration (1&2)	SP 250-24 (p.21) plus an allowance for errors in measuring the client's enclosure temperature (±0.002 °C).
Transportation (3)	Based on NIST standard cell MAP data.
Changes with time (4)	NIST general information on expected standard cell changes with time is the range of ±1 ppm per year based on repeated calibration data at NIST. Lacking any other information one can assume that this figure to represent the 3σ limit or 1σ = 0.33 ppm./yr.[1] It does *not* suggest any further information about the model or models applicable to the reported number. Therefore one would use this number until data becomes available on *their* standards. As historical data is acquired this figure can be estimated more precisely for the particular standards configuration.
Local calibration (5)	Based on data generated by the local measuring process for the local standards (between-day variability) or from check standards.
Local temperature measurements (6)	Includes resolution and other temperature measurement uncertainties for both the standard and unit under test.
Measurement system (7)	Based on data generated by the measuring system (within-day variability)

[1] Although the data is based on a single cell behavior groups of cells tend to behave in a similar manner because they usually come from the same lot.

4.4 Other Statistical Tools

Control charts are the single most valuable tool for monitoring standards and measuring processes, however it is sometimes necessary to use other statistical tools to assist in the decision making process. Although there are many statistical tools to aid in the analysis of measurement results, the t test, the F test and tests for outliers are the three most often used. Users should consult references on the topics before using these tests; although they are easy to use there are dangers that must be understood.

4.4.1 The *t* Test

The t test is used to test for differences between observed values such as two calibrations at different laboratories or the difference between a current set of measurements and a known value. It statistically tests the hypothesis that the two values come from the same population. Eq. (4.26) is the general form of the relationship where the $\bar{X}$'s are the observed values, the s's the standard deviations of a single observation and the n's the number of observations (See Handbook. 91 or other texts on statistics for modifications for other situations).

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4.26}$$

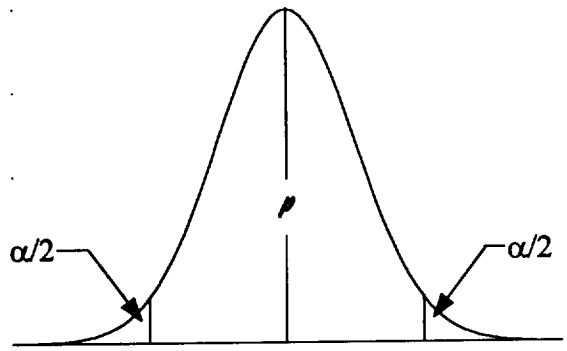

Figure 4.5 The normal distribution curve showing the relationship between p and α.

The calculated t is compared to tabulated values for the Student-t distribution such as those in Table B.2. The data tabulated in the table is based to the two tailed Student t distribution rather the one tailed usually tabulated (See Table A-4 NBS Handbook 91). The Student t distribution can be used in two ways, (1) to establish limits, or confidence intervals, and (2) to test for significance and each has its own terms and symbols. Figure 4.5 shows the relationship between the two. The symbol α is usually used when testing for significance and is (1-p). When establishing control limits or confidence intervals, etc., the symbol p is

usually used. Table B.2 is tabulated as p. The test for differences between means is conducted as outlined in the five steps listed below.

(1) The hypothesis to be tested is that the two means are from the same population.

(2) Establish the level of significance, α, for the test. Typically α=0.05 or 0.01 are the levels of choice.

(3) Calculate the statistic t using Eq. (4.26) or a suitable modification.

(4) Using Table B.2, select the appropriate value of p (p=1-α) and look up the value for the appropriate degrees of freedom.

(5) If the calculated t exceeds the tabulated value then the hypothesis is rejected and it is concluded that there is a difference between the two means. The test says the probability of the difference exceeding the tabulated t is α so there is always a chance that a difference will be claimed when none exists.

Example: A traveling voltage standard used to conduct a RMAP is measured twice by the initiating laboratory and once by the higher level laboratory and the results are tabulated below. Does the difference exceed that resulting from the uncertainty of the experiment?

Results reduced by 1.010 000 V and expressed in microvolts.

LAB 1	Before:	$\bar{X}_{1,1}$ = 8 157.782 μV	$s_{1,1}$ = 0.0670 μV	$n_{1,1}$ = 21
	After:	$\bar{X}_{1,2}$ = 8157.766 μV	$s_{1,2}$ = 0.0270 μV	$n_{1,2}$ = 11
LAB 2		$\bar{X}_2$ = 8158.067 μV	s_2 = 0.0810 μV	n_2 = 16

The observed before and after differences are consistent with previous experiments.

Step 1: Establish the hypothesis – there is no difference between the two as-maintained units.

Step 2: Select a probability α – in this case 0.05. *Note* – The choice of α is a matter of choice but 1-α is usually 95 or 99 percent with 95 % being the most widely used.

Step 3: Pool the before and after standard deviations and use Eq. (4.26) to calculate t.

$$t = \frac{|8157.774 - 8158.067|}{\sqrt{\frac{0.0565^2}{32} + \frac{0.081^2}{16}}} = \frac{0.293}{0.0226} = 12.96$$

Step 4: Referring to Table B.2 for 45 degrees of freedom t_α = 2.02 (α = 0.05).

Step 5: Since t exceeds the critical t the hypothesis is deemed false and it is concluded that there is a difference between the two quantities. This test is independent of which laboratory is the

calibrating one, therefore it serves equally well for MAP and RMAP transfers. There are other cases such as the $s = \sigma$ or one value is known (See Section 7.5.5 for the latter) that are covered in statistic texts and handbooks (See Chapter 2 of NBS Handbook 91, *Experimental Statistics*)[1].

4.4.2 Testing Equality of Variances

A second test, the F test is used to compare variances such as those in the example of the previous section. The statistic F is calculated by Eq. (4.27)

$$F = \frac{s_1^2}{s_2^2} \tag{4.27}$$

where s_1 and s_2 are the standard deviations to be compared. Since the purpose of the test is to detect a difference the smaller of the two is placed in the denominator. Values of F are found in Tables B.3 though B.5. For more detailed tables consult NBS Handbook 91 or other statistical texts. To test the data of the previous example to determine if the two laboratories have the same variability calculate the variance for each laboratory (s^2) and calculate $F = 2.03$ using Eq. (4.27). Referring to Table B.3 for α=0.05 $F_{15,30} = 2.01$, thus it is concluded that the two laboratories do not have the same process standard deviation at the 95% level.

4.4.3 Outliers

Frequently a question arises about observations that are removed from a cluster of data. Any observations that appears to be an outlier should be examined and a decision made whether to retain or remove it from the data set. Chapter 17 of NBS Handbook 91, *Experimental Statistics* presents detailed methods for a variety of cases for rejecting apparently aberrant observations. The task is straightforward when there are many degrees of freedom. The reader is warned, however, that extreme caution must be exercised for small data sets; especially for sets containing only three or four observations.

[1] Table B.2 differs from Table A-4 of Handbook 91 in that the former is for both tails of the distribution and the latter one-tail. The p of Table B.2 corresponds the $p/2$ in Table A-4.

5.1 General

Measurement uncertainty requirements are customer driven while measurement assurance monitors process capability and verifies whether or not the customer's requirements are satisfied. Measurement assurance uses historical data to predict expected future behavior.

An established, documented, continuous measurement assurance program in statistical control and having a documented uncertainty is considered as objective evidence that a calibration or measurement process is meeting the established uncertainty requirements.

5.2 Measurement Process Control

The control chart is the primary tool to demonstrate statistical control and estimate process uncertainty. Figure 5.1 is a simplified block diagram of a typical measuring process showing the various inputs (calibrated standards, influences, etc.) and a single output, a measurement result shown in the block "workload." There are two paths — measurement data and uncertainty. Measurement data is the current

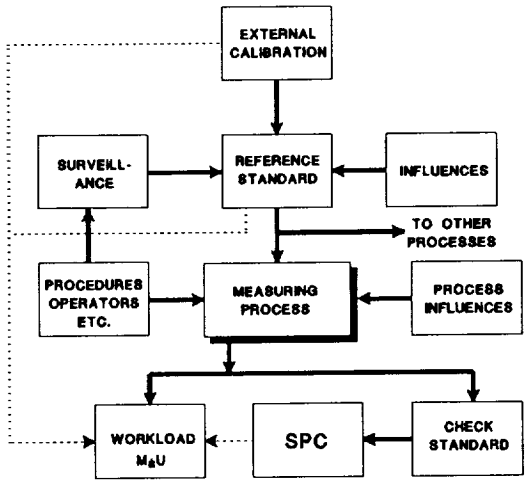

Figure 5.1 Typical calibration process. Block diagram of a calibration process from the calibration of the reference standards to the final workload. The process can be broken down into several subprocesses. The outputs of each can serve more than one client process. Solid lines are flow of measurement data; dotted the flow of uncertainty information.

result whereas the uncertainty is based on analysis of historical data from the same process. Measurement assurance dictates that parameters affecting process quality be independently monitored. One parent process (primary standard, etc.) may serve more than one child process — a SPRT may be the primary reference (parent) for calibrating other PRTs, liquid in glass thermometers, and thermocouples. No action is needed as long as the process remains in a state of statistical control. Out of control situations must be promptly addressed to determine:

- The cause of the condition and if possible actions necessary to remedy the condition;
- If necessary, a new applicable uncertainty for the process; and
- Whether or not the increased uncertainty for the out of control condition exceeds the customer imposed limits.

There will be cases where an out of control condition signals a real change in the process that requires reestablishment of control at a new level. Unlike many manufacturing processes that can be brought back to their original conditions, a measurement process may require the establishment of new operating parameters as the result of out of control conditions.

5.2.1 Measurement Assurance Documentation

Organizations using measurement assurance methods to control calibration or measurement processes should document each system and process. The documentation should include all necessary instructions, test and measurement procedures, data and analysis procedures, equipment description and set up, environment, operator, check standards, and established process control limits. Continuous contemporary evidence of process control is demonstrated by the control charts and ancillary data maintained for important measurement system parameters and influences that affect the process. All out of control conditions and corrective actions taken to restore the process to a state of statistical control should be documented. To the greatest extent possible, the measurement process, data logging, data reduction, and data analysis should be automated to improve data quality, permit more sophisticated data reduction and analysis, reduce the manpower needs, permit the use of less skilled personnel, and provide real-time results. A successful MAP requires an understanding of the physical principles underlying and affecting the measurement process; the standards employed; the role of the operators and other personnel involved in the process; the measuring apparatus and methodology; and the data reduction, analysis and interpretation for the total measuring process. A minimum MAP requires that:

- Local standards must be periodically calibrated using MAP transfer techniques or equivalent;
- The calibration must include an uncertainty for the assigned values;
- The uncertainties due to influences such as transportation must be quantified,
- There must be a continuous surveillance of the local standards between external calibrations which includes SPC and other techniques to detect anomalous behavior,
- Out-of-control conditions must be promptly investigated and corrective action taken, and
- There must be a documented current uncertainty for the process output.

In the context of standards, scaling from the local standard to higher and lower values must also have their own MAP embodying the above six points

5.3 External Calibrations

All laboratory primary standards require periodic external calibration, even so their uncertainties are often a significant part of the total uncertainty. Components are (1) calibration uncertainty, (2) effects of transportation related uncertainties, and (3) time and use related uncertainties. Ideally, the last two should be individually evaluated and combined with the calibration uncertainty to provide an overall uncertainty for the quantity represented by the standard(s). A sound estimate of each requires analysis of historical information.

5.3.1 All Standards Externally Calibrated

A laboratory sending out all standards for calibration gains only limited information about their behavior. Although satisfactory for certain standards this method provides only two pieces of information — *the value of the standard and its uncertainty at the calibrating facility*. It does not include uncertainties arising from transportation influences or changes between calibrations. When historical information exists for a standard, control chart like techniques can be used to monitor its long-term behavior and estimate an overall local uncertainty for the standard (excluding internal systematic effects). Lacking other information data from external calibrations of the type just discussed only yields a single Type A uncertainty that includes the three components of Section 5.3. The step by step technique is given in Section 5.3.1.1 which is based on NIST calibrations of a group of standard cells.

5.3.1.1 Example (All Standards Externally Calibrated)

The data of Table 5.1 (also plotted in Figure 5.2) is for the mean (corrected to 30 °C) of four standard cells in a temperature controlled enclosure. The enclosure is the laboratory's sole standard and was calibrated at approximately 24 month intervals.

Table 5.1
Calibration history for the mean of four standard cells

Time (months)	Temp (°C)	Mean* (μV)	No. of Obs. (n)	Mean** (μV)	Std. Dev.** (μV)	L
0	30.001	8130.60				
17.9	30.004	8130.50				
45.2	30.005	8131.70	3	8130.93	0.666	
83.4	30.003	8132.20	4	8131.25	0.835	0.634
120.3	30.001	8132.07	5	8131.41	0.811	0.327
169.0	29.999	8130.05	6	8131.19	0.914	0.561
185.0	29.998	8130.51	7	8131.09	0.873	0.247

* Reduced by 1 010 000 μV
** Mean and std. dev. of all data to this point

Step 1: Starting with data (3 points minimum) and other information about the standards, develop a model describing their expected behavior. The model for this analysis assumes that the (1) standards do not change with time and (2) the mean of the cells in the enclosure is constant with time. If the process remains in control the model is considered as valid. If not in control then action is required. Data for this example is for t = 120.3 mo.

Step 2: Using the previous calibrations calculate the mean ($\bar{E}$) and standard deviation (s) for the previous n =4 calibrations (see columns 5 and 6).

$$\bar{E} = \frac{1}{n}\sum_{i=1}^{n} E_i = \frac{1}{4}\sum(8130.60+8130.50+8131.70+8132.20) = 8131.25\ \mu V \qquad (5.1)$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(E_i - \bar{E})^2}{(n-1)}} = \sqrt{\frac{\sum_{i=1}^{4}(E_i - 8131.25)^2}{3}} = 0.835\ \mu V \qquad (5.2)$$

Step 3: For the fifth calibration, calculate L using Eq. (5.3) where E_{curr} is the current calibrated (5 th) value and the vertical lines indicate the absolute value of the contained expression.

$$L = \frac{|E_{curr} - \bar{E}|}{3s} = \frac{|8132.07 - 8131.25|}{(3)(0.835)} = 0.327 \qquad (5.3)$$

Step 4a: If $L \leq 1$ the current value is in control with respect to the previous calibrations, then calculate a new mean and standard deviation using Eqs. (5.1) and (5.2) and include the latest calibration. This figure is now the new Type A estimate of u_i for the current calibrated value. Use the mean supplied by the calibrating laboratory.

Step 4b: If $L > 1$ the current value is out of control and action is required. Actions could range from doing nothing to repeating the calibration. Some possible solutions are: investigate other models such as a linear one; use the standard deviation of the current and previous calibration only; or drop some earlier data from the analysis. It is not unusual for artifact standards to show unexpected changes over time. When this happens, it must be factored into the overall uncertainty usually by reevaluating the process and estimating an uncertainty that reflects the standard's variability.

Remember the purpose of this process is to ensure that future calibration uncertainties are defensible.

Step 5: Prepare a current control chart for this portion of the process as illustrated in Figure 5.2 up to and including the current one (t=120.3 mo) and examine the data for trends or other possible anomalies.

Step 6: Repeat steps (1) through (4) each time an external calibration is completed.

This process evaluates only the uppermost box of Figure 5.1 ("External Standards") and does not include allowances for local surveillance or use. As a "reality check" the derived uncertainty should be compared with general information from NIST, the manufacturer, or other sources. For this example, NIST reports that the expected change with time of standard cells is about ±0.4 μV/yr (1σ), a figure that is in reasonable agreement with that obtained for a calibration interval of approximately 24 months. For a stable process, the standard deviation will tend to stabilize to a value representative of the process σ.

(1) The uncertainty from 4a. or 4b is the new uncertainty for the assigned value of the reference standard until the next calibration, and is all inclusive for the calibration process.

(2) All other uncertainties affected by this component must be updated.

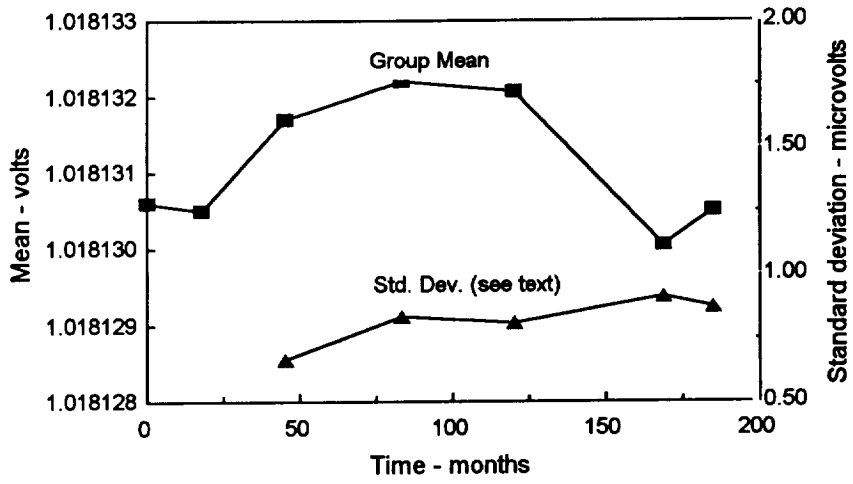

Figure 5.2 Control chart for calibration data. Calibration data for the mean (corrected to 30.000 °C) of a group of four standard cells periodically calibrated by NIST. This is the laboratory's only primary voltage standard.

5.3.2 Using Traveling Standards

Whereas the previous process directly assigns a value to the local standards, MAP type transfers act as a transfer agent. Rather than directly monitoring the local unit, the measured difference is used. If the local laboratory assigns a value T_X to a traveling standard and the higher echelon laboratory assigns a value T_S, then the difference $T_X - T_S$ is a measure of the difference between the two as-maintained units which can be used to make the difference zero to within experimental uncertainty. This difference can then be used to estimate the long-term uncertainty of the unit. Using the same basic technique as for Section 5.3.1.1, the overall process uncertainty is assessed as illustrated by the example of Section 5.3.2.1.

5.3.2.1 Example (Calibration Using Traveling Standards)

Nine years of MAP-T data between NBS[1] and a client laboratory are summarized in Table 5.2 and Figure 5.3. All transfers were conducted using the NBS volt transfer program protocol that measures the difference ($E_{LAB} - E_{NBS}$) between the client assignment (E_{LAB}) and the NBS assignment (E_{NBS}) using a four standard cell enclosure traveling standard. All results are the mean of the four cells[2]. After each transfer the laboratory adjusted its unit so that $E_{LAB} - E_{NBS} = 0$ using Eq. (3.6) and the new value is assumed constant until the next calibration. Using the procedure of Example 5.1 at t=63.4 months the following results were calculated:

---

[1] Since the data was obtained before the name change NBS is used instead of NIST.

[2] In practice each cell would be analyzed in the same manner as discussed in the example.

$$\bar{E} = -0.090\ \mu V$$

$$s = +0.271\ \mu V$$

$$E(63.4) = +0.230\ \mu V$$

$$L = \frac{|0.230 - (-0.090)|}{(3)(0.271)} = 0.39$$

This and all other transfers are in control. The overall Type A uncertainty (u_i) is between 0.2-0.3 μV, which agrees with expectations. Visual inspection of Figure 5.3 reveals that the model, although adequate, is not perfect. Clearly, except for the last two points the observed difference is increasing slowly with time. One could use a linear model but it would fail at the next to the last point. Furthermore it is likely that the correct model is one of increasing difference until 60-72 months followed by a decreasing one. The cause is not determined but not unexpected given the vagaries of standard cells.

Always chose the simplest model that meets the prescribed measurement requirements.

Table 5.2
History of a laboratory NBS volt MAP with standard cells

Time (months)*	$E_{LAB} - E_{NBS}$ (μV)	No. of Transfers	Cumulative average (μV)**	Std. Dev. (μV)	L
6.3	-0.50	1			
13.9	-0.18	2			
22.7	-0.06	3	-0.247	0.227	
29.7	-0.05	4	-0.198	0.210	0.29
36.1	0.01	5	-0.156	0.204	0.33
43.2	-0.24	6	-0.170	0.186	0.14
59.2	0.39	7	-0.090	0.271	1.00
63.4	0.23	8	-0.050	0.275	0.39
74.8	0.28	9	-0.013	0.280	0.40
89.5	-0.07	10	-0.019	0.265	0.07
113.7	-0.11	11	-0.027	0.253	0.11

* Zero time referred to January 1

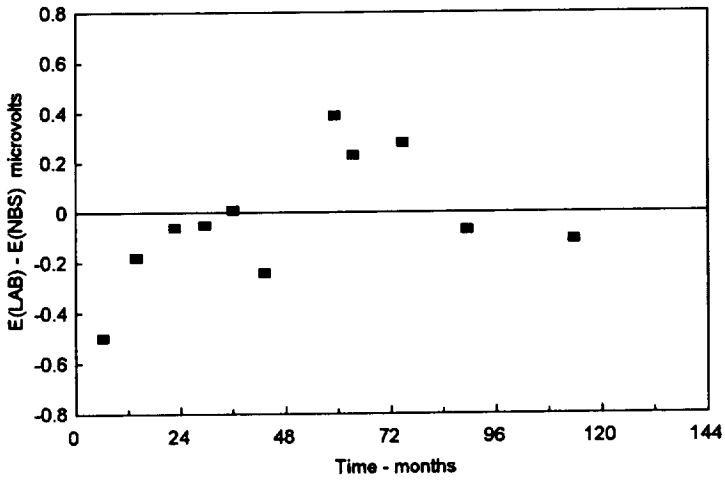

Figure 5.3 A MAP transfer history. The interlaboratory difference, (E_{LAB}-E_{NBS}) as a function of time.

5.3.3 Intrinsic Standards

The nature of an intrinsic standard suggests the lack of need for external calibration. Although the physical constant or phenomenon used is invariant, the related measuring process may introduce serious measurement uncertainties in the final result. The measurement system may be flawed and introduce errors, usually systematic, into the final result and the system can fail in subtle ways during use. They are addressed by (1) verifying the system at the time of installation , (2) establishing a rigorous operating protocol, (3) continuously monitoring the system using suitable check standards and (4) through round robin type experiments. Finally, remember that any changes in the measuring system constitute a new measuring process that must be operationally verified. Such tests can probably identify problems before they become serious. It is essential that laboratory personnel not be lulled into a state of overconfidence about the infallibility of intrinsic standards.

5.4 Internal Surveillance

External calibrations produce sparse data while internal surveillance can produce a continuous data stream. Control charts should be established for each attribute affecting the process uncertainty. Control charts should be maintained on the processes used to monitor the local standards, scaling, routine calibration services and for any influence affecting the final uncertainty.

5.4.1 Process Parameters

Every measurement process is affected by influences such as temperature, instrument gain, or operator which may or may not have a significant impact on the quality of the final result. This interaction must first be identified, then steps taken to minimize its impact on the final result. For every measuring process one should prepare an exhaustive list of influences that can introduce errors into the process.

5.4.1.1 Interactions

Influences such as temperature, pressure, instrument offset or gain affect the measurement by increasing the uncertainty due to influence variations, introducing a bias, or a combination of both. Tests can be carried out to ascertain if an influence affects the (1) observed result or (2) the result after corrections are made for the influence. The first is especially important for a new measuring process. The presence or absence is most easily graphically detected by plotting the measurand as a function of the influence as illustrated in Figure 5.4. The mass of a 200 g mass standard is plotted as a function of temperature after making the buoyancy corrections. Visual inspection shows no correlation between the two. There are formal statistical tests, however, usually inspection of the plot suffices.

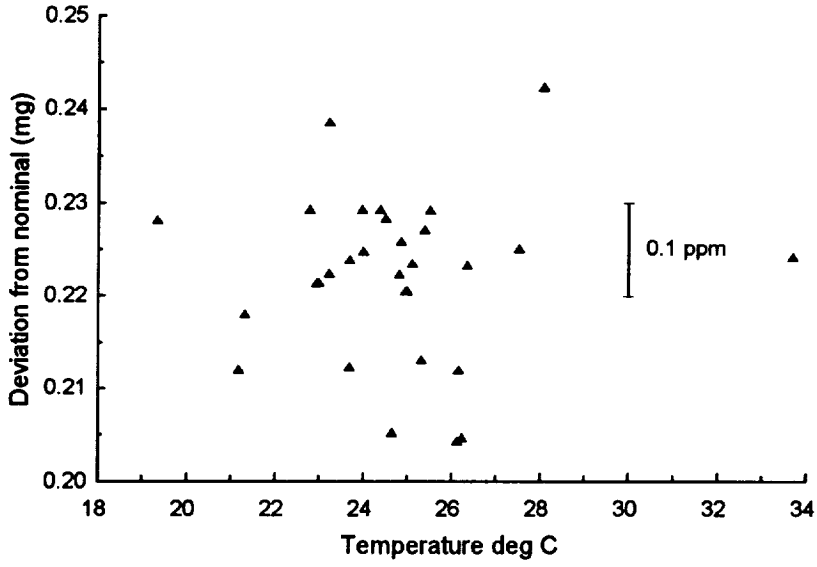

Figure 5.4 Plot of the mass of a 200 g standard as a function of temperature. The randomness of the data indicate that the final result is unaffected by temperature variations.

5.4.1.2 Monitoring Influences

Influences such as instrument gain or instrument offset may introduce a bias into the measurement and are best eliminated by experiment design or direct measurement. The former is preferable, as it usually requires less effort and yields a better result. For example, low-level voltage measurements are sensitive to spurious emfs in the measuring circuit that may be either constant or time varying. If X is the quantity being measured and Δ a small constant spurious component of the observed value (M), the two can be estimated by taking two measurements – one with the instrument connected in the usual fashion (M_{NORM}) the other reversed (M_{REV})[1]. As depicted below, X is the *average difference* between the two observations and Δ the *average* of the two.

$$M_{NORM} = X + \Delta \qquad \text{and} \qquad M_{REV} = -X + \Delta$$

[1] See NBS TN 430 for a further discussion of this topic.

from the measurement result.

$$\Delta = 1/2(M_{NORM} + M_{REV}) \qquad \text{and} \qquad X = 1/2(M_{NORM} - M_{REV})$$

Example: The observed difference between two standard cells connected in series-opposition is 5.78 μV and -5.34 μV for (M_{NORM}) and (M_{REV}) respectively. Using the two relationships above Δ= 0.20 μV and X=5.56 μV respectively.

While a well designed process routinely estimates and eliminates Δ as a part of the process with minimal extra effort, Δ should still be monitored as unexpected changes can be a harbinger of problems. Figure 5.5 is a control chart for the left-right component[1] in a standard cell measuring system that is routinely generated by the measuring process. The offset is about 0.2 ppm and if not eliminated it would be a major Type B uncertainty.

Monitor any parameter or influence that can affect the overall quality of the measurement.

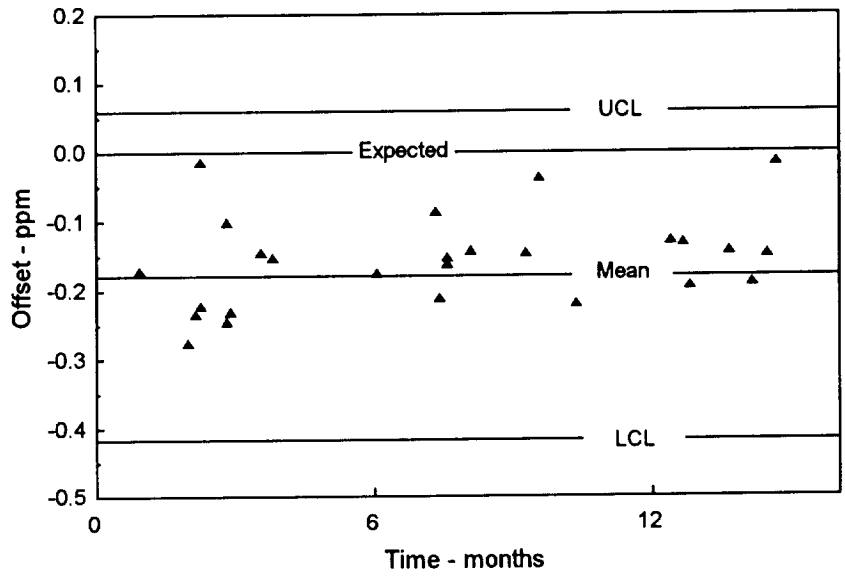

Figure 5.5 Left-right effect for a standard cell calibration system. The data was from the output of designs routinely run when calibrating standards.

5.4.2 Standards

Between calibrations there must be a surveillance procedure in place to monitor the local standards. How the surveillance is carried out depends on the type and number of standards available. Internal surveillance:

[1] When making low-level voltage measurements there is always a possibility that there are small spurious emfs that can contaminate the measurement.

- Provides information about possible changes in the standards between external calibrations,
- Provides information about the overall behavior of the standards and measuring process, and
- Determines the combined within and between day uncertainty.

5.4.2.1 Multiple Standards

Groups of standards are often used to maintain the local unit to reduce the effect of individual variations. As a rule the mean of the group is assumed to remain constant[1]. Some examples are, groups of standard cells or resistors. Their individual behavior is always tied to the accepted group mean. Very simply, every time the standards are intercompared the difference of each from the mean (Δ_i) is determined. The sum of the differences for the standards will always be zero[2] due to fact that the mean is externally assigned. For example if three standards are intercompared the result of a single intercomparison would yield the following where M is the accepted group mean.

$$X_1 = \Delta_1 + M$$
$$X_2 = \Delta_2 + M$$
$$X_3 = \Delta_3 + M$$

Because of the constraint, any change in one standard will affect the calculated values of the others and the magnitude of the shift of the other standards depends on the number of standards. Figure 5.6 is a 12 month history of the apparent difference of one cell from the mean of the four (assumed constant with time). Because of the constraint, the linear drift of about –0.4 ppm/yr with respect to the mean must be offset by the drift rate of the remainder. The net drift rate must be zero.

(1) Each chart provides control over individual standards.

(2) The charts provide the user with the current value for a standard with respect to the assigned mean.

(3) The computed standard deviation is a part of the overall uncertainty.

Under no circumstances make any unilateral adjustments to the assigned mean based on surveillance data. *Note*: When one of more standards must be removed because they are "bad actors" a new mean is calculated so that it remains consistent with the originally assigned mean

[1] Other models are used such as linear drift with time based on external calibration data. In such cases the mean is updated.

[2] This applies only to those items that are a part of the group of standards. Often other items may also be calibrated at the same time but their differences are not included. There will also be differences from the mean but their sum will not be zero.

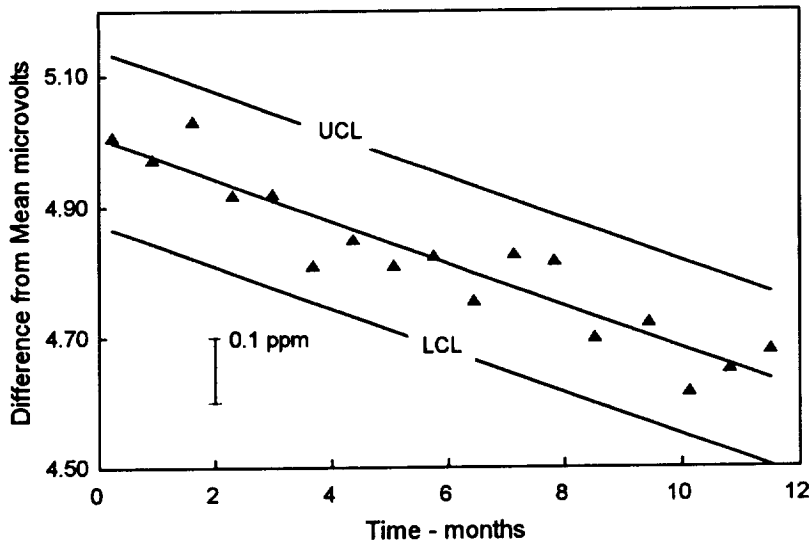

Figure 5.6 Control chart of the difference from the mean of one cell of a group of cells. The accepted value of the group is externally constrained usually to a constant.

5.5 Check Standards

A check standard demonstrates the state of control of a measuring process and provides essential information needed to estimate the assignable workload uncertainty. As depicted in Figure 5.7 the check standard monitors the process output. Check standards can and cannot do certain things.

- They can detect unexpected or abnormal behavior,
- They cannot pinpoint the cause,
- They do not monitor the external calibration,
- They provide a database of information about the process, and
- They provide uncertainty information assignable to the calibration of the workload.

Additionally, check standards should be used to monitor critical elements of the process. The following are important considerations in the selection and use of a check standard.

(1) The check standard should monitor the process output, have characteristics similar to the workload, be dedicated, and remain under the control of the laboratory. It must be emphasized that it must be clearly understood precisely what is being monitored by the check standard. Some possible check standards are:

(a) Differences between the observed values of two reference standards at least one of which had a value assigned by a higher echelon.

Examples: A check standard can be created for two SPRTs by monitoring the difference in observed temperature at some specified temperature or the difference between two standard cells.

(b) A separate artifact in an experiment design used to calibrate several standards or instruments at one time.

Example: Standard cells, mass standards, gage blocks, etc., are often calibrated using redundant experiment design that includes a check standard. Mass calibrations often add check standards at several levels (100, 10, 1, etc.). The number of check standards must be sufficient to monitor the overall process without using them at every level.

(c) Measurements made on an artifact using a direct reading instrument.

Examples: Making selected measurements of a check standard DVM to monitor a calibrator.

(d) Calibration of an artifact using a ratio technique.

Example: Using a check standard to monitor scaling by a ratio method such as a resistance bridge from one value to another, i.e., 1 to 100 Ω.

(2) The check standard should be integrated into the normal operating procedure so it duplicates the normal operating mode of the process. Where a calibration process such as mass or gage blocks covers a range, check standards should be incorporated at different levels, i.e., 0.1 g, 1 g, 10 g, or 100 g.

(3) For multirange instruments, it is not necessary to make measurements at every point; instead points should be chosen to evaluate various process functions (usually full scale). For example, when using a calibrator to monitor a DVM calibration, measurements should be made on each range calibrated.

(4) Control charts (x-bar and s if possible) should be maintained on each check standard. If the results from the check standard are in control then the overall process is deemed to be operating properly. If not in control, action is required.

(5) If the appropriate check standard is used and its measurement representative of the normal use, the check standard variability estimates the measuring process variability including the variability of the check standard. If the check standard represents the workload, its variability represents the workload variability. *It does not evaluate (1) items calibrated by that process that differ in characteristics or performance from the check standard or (2) any portions of the process that depends on an externally assigned value derived either locally or from a higher echelon* (see Figure 5.1).

(6) Before making use of a check standard, experiments should be conducted to ascertain the efficacy of a particular one and to establish provisional control limits and a central value.

5.5.1 Guide for Establishing a Check Standard

Establishing a check standard to monitor a measuring process should be viewed as an experiment designed to evaluate the potential check standard. Remember it is a decision making tool – if the check standard is out of control the process is presumed to be out of control. As a rule the check standard should be as good as the highest accuracy item measured by the process. If the workload for a calibrator is only 5 1/2 digits or less, the check standard need only be a 5 1/2 digit instrument. Influences that affect the check standard such as environmental variations, or operators, should be evaluated by varying each and observing their effect on the check standard result (see Section 5.4.1.1). If the check standard represents the workload, the variability will also include the effects of these influences on the test item. When establishing the initial

conditions, outliers can be a problem because the initial database is usually small and can therefore introduce a bias in the control limits. These biases can be evaluated by visual inspection of a time-attribute plot. Outliers that are eliminated still should be retained in the overall database for future reference. Analysis of the outliers is helpful in determining an assignable cause. Control charts (and limits) for a check standard should be started when 5 to 10 points have been acquired and then updated later at about 20 points. They should be continuously reviewed for trends, small shifts and other anomalous behavior.

5.5.2 Using Check Standards

Check standards are monitored using control chart techniques as illustrated in Figure 5.7, which is one of several used to monitor the process. Here, a 100 g check standard monitors the process at this level and is constructed using the technique of Section 4.1.3.1. The limits are based on the first five observations which would normally be updated later. The initial values, m=0.985 mg and s =0.0120 mg are used to set the central value and limits. The Type A uncertainty (u_i), *excluding the uncertainty of the reference standard,* is the process standard deviation (s) or 0.0120 mg. Check standards can and often drift with time.

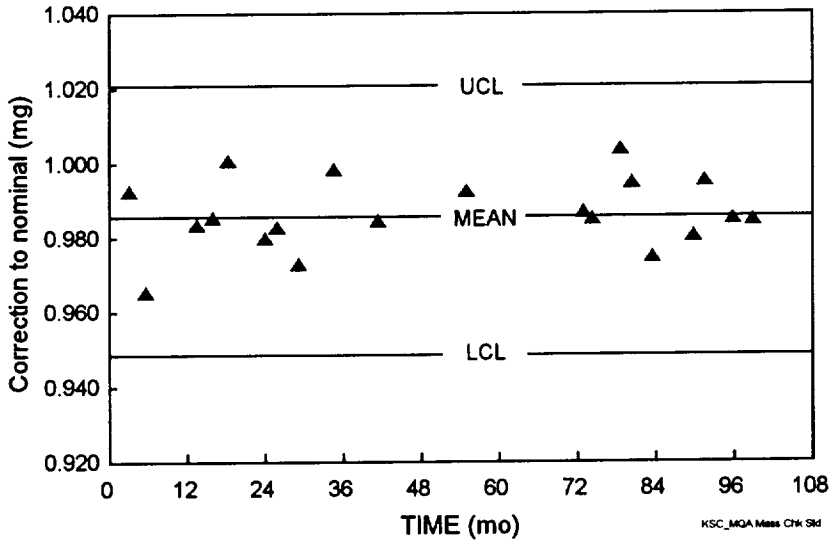

Figure 5.7 Control Chart for a mass check standard. Correction to nominal for a 100 g check standard integrated into the design used to calibrated weight sets. Control limits and central value are based on the first five observations

The plot in Figure 4.4 could very well be a check standard for monitoring voltage calibrations. If that were the case the s_{pred} calculated by Eq. (4.14) is a Type A estimate of the process uncertainty u_i (u_i = s_{pred} = 0.056 ppm). When combined with the uncertainty of the reference standards, it estimates the uncertainty (u_c) of the process as shown below.

Uncertainty Source		Type
Maintenance of the local unit =	0.33 ppm	Type A & Type B combined
Check Standard (u_i) =	0.056 ppm	Type A
Combined standard uncertainty (u_c) =	0.335 ppm	
Expanded uncertainty $2U$ =	0.670 ppm	

6 Group Measurement Assurance Programs

6.1 General

A Group Measurement Assurance Program (GMAP) provides an accepted methodology for maintaining measurement consistency among participating laboratories and traceability to national standards. A GMAP usually (1) yields the lowest uncertainty with respect to national standards, (2) identifies local measurement problems and (3) ensures the best interlaboratory agreement among its participants. Successful GMAPs require that each participating installation be fully committed to the program. The major elements of a NASA GMAP are a local measuring process meeting the measurement assurance criteria of the section; a sound MAP transfer procedure; sound procedures to conduct the measurements; periodic audits via round robins; and commitment.

6.1.1 Identifying a Potential Group MAP

When a measurement area has been identified for a NASA GMAP the need and degree of interest of potential participants must be determined. A GMAP may require a significant investment in resources. It is, therefore, important that the surveyor advise each potential participant of what is required both with respect to equipment and time. Generally, most GMAPs require a significant startup investment in time which drops to a level commensurate with current operational practices. Measurement processes which are or which can be semiautomated or automated are most suitable for GMAPs. Although desirable it not essential that all participants have equivalent measurement capabilities since each laboratory maintains a capability necessary to meet its own program requirements. Still, additional resources may be required to bring participants to a desired capability level.

6.1.2 Selecting Group MAP Candidates

The selection and prioritization of NASA GMAP's is accomplished by the NASA Metrology and Calibration Working Group. Accordingly, priority is given to projects with the broadest base for participation, where national standards are inadequate; where timely opportunity for joint participation with other Government agencies exists; where the maximum benefit can be realized from the investment; and where measurement requirements push the state of the art. The prioritization process is described in the *Working Group Operating Procedures*.

6.1.3 Confidentiality Guidelines

Reports of NASA GMAPs results, including round robins, are generally published so that the identity of individual participants is indicated by a code. This convention will be followed when results are published in a public forum or when participants other than NASA and NASA contractors are involved. Internal NASA round robin reports will identify participant laboratory data unless otherwise agreed upon.

6.1.4 Participation

Participation in a NASA GMAP is open to all NASA installations, the Jet Propulsion Laboratory, and their respective mission support contractors. The general requirements for participation are:

- Designation of an individual(s) at the installation who will accept administrative responsibility for the G-MAP;

- Designation of a technical contact at each installation who will either perform the required measurements or who has direct technical responsibility for the measurement system and process;
- A commitment to make the required measurements within the established time period and provide the results to the GMAP Coordinator in the format requested;
- The willingness to periodically act as a pivot laboratory (see Section 6.3.2);
- A commitment to making and recording the results of the in-house measurements necessary to maintain measurement process control on a continuous basis; and
- Incurring nominal transportation and other related operational expenses;

6.2 Operational Requirements and Responsibilities

A successful GMAP is a long term endeavor therefore it must have a structure to ensure continuity and a "corporate memory." The structure should be simple, have clear lines of responsibility and open communications among participating parties. The first two should be documented as an operating manual. There should be a lead installation having overall operational responsibility for the program.

6.2.1 Lead Organization and Structure

Primary administrative and technical responsibility is delegated to a lead organization that is a member of the group. The lead organization is responsible for establishing objectives, planning, budgeting, design, development, scheduling, implementation, follow-up, the appointment of Group Coordinator, and status reporting for the NASA GMAP.

6.2.1.1 Lead Organization

The lead organization will generally assign responsibility to a Group Coordinator who becomes the point of contact for all related activities. Group Coordinator duties include:

- Collaborating with other participants and NIST, to develop and implement baseline experiments to assess each installation's capability;
- Ascertaining group equipment needs such as traveling standards, special shipping containers, etc.;
- Developing a preliminary procedure for implementing the transfers;
- Preparing a final procedure after review by participants;
- Identifying transportation problems and developing alternatives to ensure safe and timely transportation of traveling standards;
- Establishing workable schedules including coordinating MAP transfers between NIST and the group;
- Initiating and continuously monitoring progress of transfers;
- Ensuring that each installation always has an administrative and technical contact;
- Preparing or overseeing the preparation of an operating manual;
- Ensuring that each participant remains on schedule and adjusting the schedule for unforeseen incidents;
- Handling all data for the transfer;
- Maintaining a permanent database for all important results from transfers and other experiments;

- Providing reports and other relevant material to each installation in a timely manner; and
- Presenting results annually to the NASA Metrology and Calibration Working Group.

6.2.1.2 Participating Installations

Participating installations are responsible for designating a Local Coordinator for each NASA GMAP with duties to:

- Be aware of the GMAP operating procedures and policy;
- Have in place a documented continuous measurement assurance program;
- Have a documented current measurement process uncertainty;
- Be prepared to receive the traveling standard and protect it from damage or deterioration;
- Confirm to the sender the arrival of the traveling standard and its condition;
- Immediately advise the Group Coordinator and other affected parties of any problems;
- Be prepared to make the required measurements in a timely manner and promptly forward the data to the designated individual or installation;
- Arrange transportation to the next recipient and confirm arrangements with the next recipient;
- Promptly advise the coordinator and others affected of any unexpected delays.

6.3 Group MAP Structure

There are many ways to conduct a GMAP, but the one most commonly used is known as a pivot laboratory or "hub and spoke" method as illustrated in Figure 3.1. The basic operating principle is quite simple. Each laboratory calibrates its traveling standard and sends it to the pivot laboratory. At the same time the pivot laboratory arranges for a NIST traveling standard to be at its laboratory. All standards are then compared using a prescribed method. From this data, each installation receives a calibration report. NIST will sometimes manage the data reduction and report issuance for the whole group however this is more expensive and time consuming. The transfer between the pivot laboratory and the other participants introduces an additional uncertainty and is about the same as the transfer uncertainty between NIST and the pivot laboratory. This added uncertainty of all participants, except the pivot laboratory, is about 1.4 times the pivot laboratories. By rotating the pivot laboratory, all laboratories are equalized over the long-term. Occasionally the capability of one or more installations cannot support the uncertainty requirements. These installations should not become a pivot laboratory. Often after the GMAP is well established and all participants have sound MAPs in place the interaction with NIST can be reduced while still using the GMAP approach. Such a program, although a true GMAP, is often though of as a round robin (see Section 7). In fact, the distinction between the formal GMAP of this section and the measurement integrity experiments of the next section sometimes become blurred. This is not a problem as long as the participants understand that the objective of both is traceability.

When a traveling standard is very stable with time, a single traveling standard can be circulated around the loop with the pivot laboratory coordinating and periodically verifying the standard's performance. No matter what structure is used it must be experimentally verified before becoming operational. Another form of GMAP is when a major laboratory takes on the role of providing MAP transfers to client laboratories. In this instance the laboratory becomes a surrogate NIST and will operate a program paralleling NIST MAP transfers. This type is often employed for clients not requiring the lowest possible uncertainty.

6.3.1 Preliminary Evaluations

Before initiating a GMAP preliminary experiments should be conducted to identify and correct any local measurement problems. These can be internal experiments designed to identify the presence of systematic errors, round robins, or other techniques. Such evaluations include training personnel to use new procedures for making measurements if one is to be introduced. Generally the Group Coordinator will monitor this phase of the program.

6.3.2 Pivot Laboratory Duties

The quality of a transfer depends to a great extent on the performance of the pivot laboratory. Although measurements last a relatively short time they will be intensive. Additionally, the pivot laboratory in collaboration with the Group Coordinator will:

- Schedule the experiment in consultation with NIST;
- Advise each participant of the schedule;
- Promptly acknowledge receipt and condition of the traveling standard;
- Carry out all measurements;
- Process or have processed all data;
- Send each participant a copy of his data only;
- Inform the Group Coordinator of any problems;
- Arrange return of all traveling standards at the completion of the measurement; and
- Issue or forward all reports.

6.3.3 Participants Duties

The participant's primary duty is to ensure that what they do does not interfere with or impede the schedule established by the Group Coordinator. Each participant's duties include:

- Calibrating the traveling standards according to the agreed schedule;
- Arranging prepaid shipping following agreed on procedures including verification of the exact shipping address;
- Packing following established procedures;
- Coordinating the exact shipping schedule with the pivot laboratory;
- Immediately advising the pivot laboratory of the mode of transportation, expected delivery, bill of lading number, and all other pertinent information;
- Coordinating return with the pivot laboratory;
- Promptly recalibrating the transport standard after its return; and
- Immediately sending all pertinent data to pivot laboratory or other entity doing the final processing.

6.3.4 NIST and NASA Group MAPs

GMAPs usually directly involve NIST. Additionally, NIST has a great deal of experience conducting MAP transfers and can provide guidance in selecting traveling standards, establishing operating procedures, designing measurement protocols, data processing and analysis, and other important matters. The experiment is very simple; it measures the difference between two calibrating facilities (M_{LAB}– M_{NIST}). This data is in turn used to make the differences between the two units zero (θ_{LAB}= θ_{NIST}) to within experimental error. Since the whole process effectively calibrates the local standards process at the output terminals, constant systematic effects can often be eliminated or significantly reduced.

6.3.5 Group MAP Logistics and Techniques

All MAP transfers have several key elements; the traveling standard; the transportation process; the measurement processes; data reduction and analysis; and reporting. Although each must be tailored to the specific disciplines there are some properties common to each.

6.3.6 Traveling Standards

Traveling standards must be robust and predictable during a MAP transfer. It is wise to select traveling standards known to have suitable performance characteristics. Consultation with NIST and others who have had experience conducting MAP transfers and round robins is recommended. A traveling standard consists not only of the standard proper but includes its shipping container, battery if necessary, and any instrumentation needed to monitor it while in transit[1]. Carefully document the transportation process so that it can be repeated for consistency, and need not be reinvented each time. When a traveling standard is not in use, it can serve as a check standard which will ensure that it is routinely monitored.

It is incumbent on the initiators of the transport process to inform the recipient of all information necessary for proper handling of the traveling standard; any data to be recorded upon receipt and before departure; and precautions to be taken before placing it in normal operation. This is best done by sending detailed instructions before shipment, including any data sheets to be completed and a reminder in the packing case. Remember these experiments are carried out infrequently and participants do not always recall or properly document the previous experiments. New traveling standards can be evaluated by round trip shipping using the worst of the expected transportation systems and comparing their behavior with similar known traveling standards. When packing standards, follow the manufacturer's and NIST's recommendation regarding protection from influences such as shock, vibration, temperature, humidity, etc. and if necessary, monitor critical influences.

6.3.7 Transportation

Once the traveling standard enters the transportation process all control is lost and asking for non-standard special handling is usually to no avail.

> Do not attempt or expect commercial carriers to adapt their system to your needs; you must adapt to their system. The only way to control the transportation process totally is by hand carrying.

[1] Transit is defined as the time shipper packs the standard to the time that it is unpacked and put into its normal operating mode.

The traveling standard must withstand expected abuses and the process including its selection, packing, and method of transportation must be taken into account. Some general guidelines for successfully managing the transportation phase of a GMAP follow.

(1) Obtain suitable reusable packing containers for the traveling standard that ensures its safe transit. NIST, the standard's manufacturer, or shipping container manufacturers are excellent sources of information. The latter often provides information about expected conditions within the freight system.

(2) Work with a shipping specialist to avoid problems from hazardous material, unsafe practices, damage due to improper packing, transit delays, and other problems arising during the transit process. Unless the standards are hand carried, they are at the mercy of a monolithic transportation system. Assurance by those marketing the service does not guarantee proper treatment.

(3) Mark the shipping container "FRAGILE SCIENTIFIC INSTRUMENTS". If necessary equip the shipping container with suitable sensors (temperature, shock, etc.) to monitor conditions during transit.

(4) Determine the best way to ship the traveling standard to its destination. Options include overnight air shippers, direct arrangement with the airlines, small package freight companies and finally, hand carry. Discuss the problem with their representatives. *Remember the transportation process begins when the traveling standard leaves the laboratory and does not end until it reaches the destination laboratory.*

(5) Obtain the recipient's exact shipping address. If possible have it shipped directly to the laboratory rather than a shipping room. *Most problems can be traced to local shipping rooms.*

(6) Coordinate the shipping schedule with the recipient and advise them of the final arrangements. The shipping laboratory should provide the recipient with (1) the carrier, (2) the exact travel mode (3) estimated time of arrival, (4) any shipper provided identification numbers, and (5) the exact delivery address.

(7) Confirm that the traveling standard is in transit.

(8) Instruct the recipient to confirm receipt and its condition upon arrival to the sender.

(9) Provide the recipient with any special handling instructions before shipment with a copy accompanying the shipment.

6.3.8 Measurement Protocols

NIST usually specifies one or more acceptable measurement protocols to ensure that the process is fully evaluated. When conducting GMAP transfers between laboratories, similar protocols should be adopted to reduce the possibility of introducing biases into the final result. All measurements should be made with the laboratory's instrumentation configured in its normal operating mode unless the protocol directs otherwise. The protocol should specify:

- Description of the artifacts;
- Normalization period;
- Preliminary check measurements;

- Test equipment setup and any special services;
- The measurements to be made including any specific sequence;
- What environmental influences should be measured and, if critical, how and where they are to be measured;
- The minimum number of data sets and the time interval over which they should be made;
- What if any preliminary measurements are required;
- Data reporting instructions; and
- Algorithms and instructions for data reduction and software if available.

6.3.9 Automation and Data Reduction

The data reduction and analysis software for a NASA GMAP is an integral part of the program and should be well documented. Software for use by participants should run on various platforms and reduce raw data to the form specified by the GMAP protocol using dedicated programs written in BASIC, FORTRAN, C, etc. or using a spreadsheet. In either case the output should be in computer readable form (floppy disk, etc.). Again look to NIST for guidance and assistance as they may have already developed software packages. Finally, automate as much of the process including data collection as is practical. As a rule, automated systems yield lower uncertainties and reduce calibration costs. Additional software is also required to carry out the final data reduction, usually by the Group Coordinator. *Ideally, once the data is in a computer compatible format it should never be touched by human hands.*

6.3.10 Reports

When a GMAP transfer is complete, NIST or the Group Coordinator will issue a report; its content will depend on how the MAP is constructed. A report may be issued to each participant about their standards or it may only provide the group with the calibration of the pivot laboratory's standards. In the latter case it is the responsibility of the Group coordinator and the pivot laboratory to prepare reports for the other participants. Copies of all reports should be retained by the Group Coordinator.

6.3.11 Database Management

The long-term success or failure of a GMAP depends on maintaining a continuous long-term database for the activity. The Lead Installation or Group Coordinator should maintain records for all transfers and adjustments made by each participant as well as a general log to facilitate future trouble shooting and calibration interval adjustment. Each participant should maintain records of its own that should be far more detailed those of the Group Coordinator. Among the important information to be conserved is the transportation phase of the experiments. *Remember GMAPs may be continuous but GMAP-Ts are usually only conducted every one to two years.*

6.3.12 Communications

Timely and accurate communications are critical to the success of any GMAP. Basic information is best communicated by telephone or FAX. Data is best exchanged electronically to speed up the process and reduce errors due to transcription. Today direct exchange using microcomputers equipped with a modem and suitable communications software is easy and can be done at moderate cost. It has the advantage that it can be transmitted in the desired format. The goal should be – *data untouched by human hands.*

6.4 NASA Group MAP Program Descriptions and Procedures

Each NASA GMAP will have a program description and a procedure which will be incorporated into this publication as an appendix that will be revised as GMAPs are developed and/or modified. Distribution of revisions will be made only to participating installations and NIST. General distribution of procedure appendix revisions will be made only when the publication as a whole is revised.

6.4.1 Local Process Descriptions and Procedures

Each participant in a NASA GMAP will maintain a description of the measuring process, standards, and operating procedure as a local portion of the Guideline appendix pertaining to the particular GMAP and should retain all previous ones for possible future reference. This information will be kept current as a part of their internal documentation.

6.5 Group MAP Example

Several laboratories execute a GMAP-T of the type illustrated in Figure 3.1 which includes NIST. For this example only the pivot laboratory and one satellite laboratory, X, will be considered as the experiment is the same for all satellite laboratories. All traveling standards will be at the pivot laboratory simultaneously. The role of each party is listed below.

(1) **NIST:** The traveling standard is calibrated by NIST before and after transit. If the transfer is satisfactory then the mean of the two is taken as the value at the pivot laboratory (E_{NIST}).

Note: This phase can also be conducted as an RMAP with the same result as discussed in Section 3.4. In such a situation there will be only one calibration by NIST and two at the pivot laboratory.

(2) **Pivot Laboratory:** The pivot laboratory:

(a) monitors all traveling standards until they are stable;

(b) measures the NIST traveling ($E_{NIST,PIVOT}$) and assigns a value in terms of the pivot laboratories standards ($S_{PIVOT,OLD}$); and

(c) measures the difference between the NIST traveling standard and the satellite traveling standards ($\Delta E_{X,\,NIST}$).

Note: It is prudent to also assign a value to each traveling standard in terms of the pivot laboratory's local standards as a check on the overall GMAP-T.

(3) **Satellite Laboratories:** Each measures their traveling standard before and after transport in terms of their local standards ($S_{X,OLD}$). If the transfer process is satisfactory then the mean is used as the value (E_X) at the pivot laboratory.

Note: This phase of the process is basically a RMAP-T between the satellite and pivot laboratories.

There are, of course, possible variations so long as the basic data outlined above results. The data for each part is summarized below and the calculations based on Sections 3.4 and 3.4.1 follow. The subscripts for the various units are used solely to avoid ambiguity.

A	Assigned value of the NIST traveling standard at NIST	E_{NIST}	10.000 045 V_{NIST}
B	Assigned value of the NIST traveling standard at the Pivot Lab in terms of the local unit.	$E_{NIST,PIVOT}$	10.000 075 $V_{NIST,PIVOT}$
C	Current assigned mean of the Pivot Lab reference standards	$S_{PIVOT,OLD}$	10.000 085 $V_{PIVOT,OLD}$
D	Measured difference between the Lab X and NIST traveling standards at the Pivot Lab	$\Delta E_{X,NIST} = E_{X,PIVOT} - E_{NIST,PIVOT}$	-0. 000 0035 V_{NIST}
E	Assigned value of the Lab X traveling standard at Lab X	E_X	10.000 055 $V_{X,OLD}$
F	Current assigned mean of Lab X reference standards.	$S_{X,OLD}$	10.000 070 $V_{X,OLD}$

Using data elements A, B, and C and Eq. (3.5) the adjusted value for the pivot laboratory standards is

$$S_{PIVOT,NEW} = 10.000085V_{PIVOT,OLD} \times \frac{10.000045V_{NIST}}{10.000075V_{PIVOT,OLD}} = 10.000055V_{NIST}$$

The adjustment for each satellite laboratory is made in a similar fashion except that an extra step is required as shown below using data elements A, D, E, and F.

$$E_{X,PIVOT} = E_{NIST} + \Delta E_{X,NIST} = 10.000\,045\ V_{NIST} + (-0.000\,035\ V_{NIST}) = 10.000\,010\ V_{NIST}$$

and the new value for the standards is

$$S_{X,NEW} = 10.000\,070\ V_{X,OLD} \times \frac{10.000\,010\ V_{NIST}}{10.000\,055\ V_{X,OLD}} = 10.000\,025\ V_{NIST}$$

Occasionally, a transfer will not be satisfactory and must be repeated. If it is carried out promptly then the satellite laboratory can conduct a regular RMAP-T with the pivot laboratory with only a very small inflation of the uncertainty. In this case $S_{PIVOT,NEW}$ is used to calculate the value of the traveling standard.

7.1 General

Measurement integrity experiments or round robins do not of themselves constitute objective evidence of continuous measurement or calibration process control. Instead each is a snapshot, auditing a laboratory's current measurement capability at a point in time. Well-designed round robins:

- Provide independent verification of the bias and precision of the measurement process;
- Are an effective method of surveying participants' measurement capability; and
- Serve as an assessment tool to determine the readiness of a group of laboratories to participate in a G-MAP.

A round robin usually audits a process in its normal operating mode. That is, it looks at the process output in a mode that is similar to the normal workload. Changes, intentional or unintentional, are evaluated by repeated round robins.

7.2 Identifying Requirements

Before initiating a NASA round robin, the degree of interest of individual field installations in assessing/upgrading the measurement capability should be evaluated. Usually the installation having a proprietary interest in a proposed project will seek designation as the Lead Center for the proposed round robin. Preliminary estimates of installation capability can be obtained from the *NASA Metrology Laboratory Measurement Capabilities Document* followed by a more detailed survey to identify the need, objectives, and potential participants. Generally, programmatic or institutional requirements are the drivers for establishing and subsequently upgrading measurement capabilities. Decisions about the initiation of a round robin are made by the Metrology and Calibration Working Group based on the results of the survey. Since measurement capabilities are developed and maintained to satisfy identified programmatic requirements, all installations may not have the same measurement capabilities.

7.2.1 Setting Priorities

Establishment of priorities for a proposed measurement integrity experiment is accomplished by the NASA Metrology and Calibration Working Group as described in the *Working Group Operating Procedure*. Round robins are to some degree labor and equipment intensive, thus consuming installation resources. Therefore, projects with the broadest base for potential participation and the maximum potential benefit for the required investment should be favored.

7.2.2 Participation

Participation in a NASA measurement integrity experiment is open to all NASA field installations, the Jet Propulsion Laboratory, and their respective mission support contractors. The general requirements for participation are a commitment to the experiment and the willingness to incur modest expenses associated with transportation and other matters.

7.2.3 Lead Center Responsibilities

The duties of the Lead Center responsible for establishing a NASA round robin include establishing objectives, devising a plan, preparing a budget, designing the experiment, developing a schedule, implementation of the experiment, follow-up, and status reporting. These responsibilities are generally assigned to the Interlaboratory Coordinator, designated by the Lead Center, who becomes the internal and external point of contact for all related activities.

7.2.4 Participating Installations

Participating field installations are responsible for designating a Local Coordinator for each NASA round robin. This individual will assure the installation:

- Is aware of the experiment's operating procedures and policy;
- Has in place a documented measurement process to be used to measure the traveling standard;
- Has a documented current measurement process uncertainty;
- Is prepared to receive the traveling standard and protect it from damage or deterioration;
- Will confirm to the sender arrival of the traveling standard and its condition;
- Will immediately advise the Interlaboratory Coordinator and other affected parties of any transportation or technical problems;
- Is prepared to and make the required measurements in a timely manner and promptly forward the data to the designated recipient;
- Will arrange transportation to the next recipient and confirm arrangements with the next recipient;
- Will promptly advise the coordinator and others affected of any unexpected delays.

7.3 Types of Measurement Integrity Experiments

Measurement integrity experiments or round robins fall into two broad classes, those using artifact standards and those using reference materials (RM). The former evaluates a calibration process of the type depicted in Figure 5.1 while the latter usually evaluates a measurement process that determines a material's property or composition.

7.3.1 Artifact Measurement Integrity Experiments

As the name implies an artifact measurement integrity experiment uses one or more artifacts which are circulated among the participants under the assumption that they do not change (or are predictable) over the course of the experiment. A lead laboratory usually provides a baseline for the data and may remeasure the traveling standards periodically during the course of the experiment to ensure its integrity. Because measurements are sequential these experiments may take a considerable time.

7.3.2 Reference Material Measurement Integrity Experiments

Historically the earliest measurement integrity experiments evaluated analytical procedures used to test a variety of materials. Even now many processes depend on the use of a reference material of known composition to calibrate a process. NIST has developed a series of carefully characterized Standard Reference Materials (SRM) to meet a variety of need. Their composition or property and uncertainty has been experimentally determined which makes them exceptionally well suited for use as round robin

artifacts. SRMs offered for sale by NIST are described in *NIST SP 260, Standards Reference Materials Catalog* which is revised and published biennially. Many RMs are suitable for use as round robin artifacts for metrology applications, such as hardness, surface finish, density, particle size, and thermometric fixed points. Some will be destroyed during the experiment while others such as hardness or surface finish SRM may be merely circulated.

7.4 Logistics and Operating Procedures

The basic logistics and operating procedures are similar for both types of round robins with the overall success depending on the dedication of the Interlaboratory Coordinator, Local Coordinators, and personnel directly involved in making the measurements, handling data, and arranging transportation.

7.4.1 Responsibilities of the Lead Center

The Lead Center will generally assign responsibility to an Interlaboratory Coordinator who becomes the point of contact for all related activities. Interlaboratory Coordinator duties include:

- Collaborating with other participants and NIST (if involved), to develop a realistic schedule for the experiment;
- Reviewing equipment needs such as traveling standards, special shipping containers, etc.;
- Preparing a final procedure and issuing it well before the start of the experiment;
- Identifying transportation problems and developing alternatives to ensure safe and timely transportation of traveling standards;
- Continuously monitoring the experiment's progress;
- Ensuring that each installation always has an administrative and technical contact;
- Ensuring that each participant stays on schedule and adjusting the schedule for unforeseen incidents during a transfer;
- Handling all data for the transfer;
- Maintaining a database for all important results from transfers and other experiments;
- Promptly provides reports and other relevant information to those involved; and
- Annually presenting results to the NASA Metrology and Calibration Working Group.

7.4.2 Participants Duties

Each participating installation is responsible for designating a Local Coordinator whose duties include:

- Being aware of the round robin operating procedures and policies;
- Having a documented measurement process;
- Advising the Interlaboratory Coordinator before starting the experiment of any changes in the measuring process since the last round robin;
- Being prepared to receive the traveling standard and protect it from damage or deterioration;
- Confirming to the sender the arrival of the traveling standard and its condition;
- Immediately advising the Interlaboratory Coordinator and other affected parties of any transportation problems;

- Being prepared to and make the required measurements in a timely manner and promptly forward the data to the Interlaboratory Coordinator;
- Arranging transportation to the next recipient and confirming arrangements with the next recipient;
- Promptly advising the coordinator and others affected of any unexpected delays.

7.4.3 Confidentiality Guidelines

Reports on the results of a round robin are generally published so that the identity of individual participants is indicated by a code. This convention will be followed when results are published in a public forum or when participants other than NASA and NASA contractors participate. Unless otherwise agreed upon internal NASA round robin reports will identify participant laboratory data.

7.4.4 Software

Software includes both manual and computer based methods for reducing raw data to produce a final result and is an integral part of the program that should be well documented. Each laboratory should be capable of calculating the requested results from its own efforts. Additionally, the Interlaboratory Coordinator must have the necessary software to properly analyze all supplied data. All software must be properly documented. When computer software is modified, copies of old programs should be retained.

7.4.5 Procedures

Overall procedure development is the responsibility of the Interlaboratory Coordinator. Procedures must contain sufficient detail to assure that the objectives of the round robin can be attained. It is beneficial to circulate a draft of the procedure to prospective participants for review and comment well before the experiment begins. This will allow time for incorporating any revisions. The procedure should contain the following information:

- A description of the round robin artifacts and if necessary operating instructions and precautions for use;
- Any special electrical or environmental requirements;
- Shipping packing and handling instructions for the artifact or RM;
- Any preliminary checks or measurements needed to ensure proper operation;
- The normalization period;
- Data recording requirements, instructions, and any special data sheets;
- Instructions, including forms, software, etc. for reporting the results; and
- Data reporting instructions including deadlines.

7.4.6 Transportation

Transportation requirements are the same as those used for a Group MAP described in Section 6.4.2.

7.5 Multi-Artifact Measurement Integrity Experiments (Youden Charts)

Measurement integrity experiments are best conducted using two artifacts[1] in conjunction with the Youden method. Originally developed by W. J. Youden in the 1950's as a graphical way of diagnosing interlaboratory test results, the method has been adapted to meet the needs of the metrology community. It is predicated on the simple hypothesis: *laboratories measuring the same material or standard should obtain the same result to within the experimental uncertainty.* Youden charts:

- Are easily constructed and require a minimum of computation;
- Are graphical to facilitate presentation and interpretation;
- Clearly show laboratory bias; and
- Provide quantitative information about the participants.

7.5.1 The Youden Chart

The Youden chart is a graphical procedure based on measurements made on two artifacts at several laboratories. The resulting data from the measurements is plotted using one artifact as the x-axis and the other as the y-axis (the choice is usually unimportant). The results are then visually examined for possible effects. If the data is randomly scattered then it is presumed that there is no interlaboratory effect. On the other hand, a trend along a 45° reference line (with respect to the x-axis) indicates a between-laboratory bias or offset. This process is best understood by example. Table 7.1 contains data simulating a Youden-type round-robin in which artifacts were measured at fifteen laboratories. The data in the first two columns was constructed using randomly distributed numbers from a population having a $\mu = 10$ and $\sigma = 1$. To simulate laboratory bias, a second set of random numbers (μ=0 and σ=3) was generated and added to the data of columns 1 and 2 as shown in columns 4 and 5. Finally, the difference between the two samples is given in column 3 of the table. This difference is the same for both the no-bias and bias cases. If $Z_{i,j}$ is the i th artifact (i = 1 or 2) at the j th laboratory (for the example 1-15) the expected value of $Z_{i,j}$ [$E(Z_{i,j})$] of an artifact is given by Eq. (7.1);

$$E(Z_{i,j}) = K + \epsilon_i + \tau_j \qquad (7.1)$$

where ϵ_i is the random error of the laboratory's measurements and τ_j the laboratory's offset and K the "true" value of the artifact (in this example 10). Two Youden charts are created from the data in the table, one with random uncertainties only (Figure 7.1), the other with random and offset uncertainties (Figure 7.2). The step-wise procedure for creating the Youden chart is given below using the data of Table 7.1.

7.5.1.1 Creating a Youden Chart

Step 1: Plot the results from each laboratory by assigning one sample to the x-axis (Sample 1) and the other to the y-axis (Sample 2). The chart is easier to make, understand, and interpret if the scales of the two axes are equal as shown in Figure 7.1.

---

[1] As used in this section the term "artifact" includes standards, instruments, reference materials and any other articles used to carry out a round robin. Any artifact so used must have predictable behavior.

Table 7.1
Data for Sample Youden Charts

Random Error Only		Difference	Random Error plus Laboratory Bias	
Sample No. 1	Sample No. 2		Sample No. 1	Sample No. 2
9.34	8.78	0.56	12.55	11.99
8.46	9.79	-1.33	5.56	6.89
9.92	11.08	-1.16	10.60	11.76
8.81	9.74	-0.93	13.95	14.88
9.17	10.81	-1.64	13.84	15.48
10.08	10.36	-0.28	6.53	6.81
9.60	9.46	0.14	11.26	11.12
8.78	12.16	-3.38	14.30	17.68
8.69	10.53	-1.84	7.87	9.71
9.40	9.63	-0.23	10.25	10.48
8.96	9.53	-0.57	10.67	11.24
8.43	8.27	0.16	10.39	10.23
9.89	8.93	0.96	8.66	7.70
9.40	9.59	-0.19	7.72	7.91
11.00	10.27	0.73	14.83	14.10
	Youden Std. Dev.=	0.81		

Step 2: Determine the median for each sample (the value for which half the points are greater than that value and half less). Youden chose the median because (1) it is less sensitive to outliers and (2) ideally, one fourth of the observations should be in each quadrant. Today, some Youden charts use the mean or the accepted value if known. As a rule this choice usually does not affect the basic interpretation of the chart.

Step 3: Enter the medians to form four quadrants as shown. It is not necessary that the medians intersect at the center of the chart.

Step 4: If the two axes have the same scale draw a 45° reference line with respect to the x-axis. If the scales are unequal the line should have a mathematical slope of 1.

Step 5: Calculate the Youden standard deviation (s_Y) of the differences (d_i) where d_i is the difference between the x and y observations for each laboratory using the relationship (column 3 of Table 7.1). (see *Graphical Diagnosis of Interlaboratory Test Results* for a discussion of the origin of the equation)

$$s_Y = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - n\bar{d}^2}{2(n-1)}} \qquad (7.2)$$

Note the 2 in the denominator. Before calculating s_Y outliers must first be dealt with, either by inspection or formally, so as to not distort s_Y. In this example the one point of Figure 7.1 lying outside the circle was not considered as an outlier. *Any deletion of data from the calculations must be made with great care, documented, and the point(s) should be retained on the chart.*

Step 6: Draw a circle with the center at the intersection of the two medians and a diameter of $3s_Y$ (Youden also suggests as an option $2.5s_Y$). This circle serves as a pseudo-control limit to help in analyzing the chart.

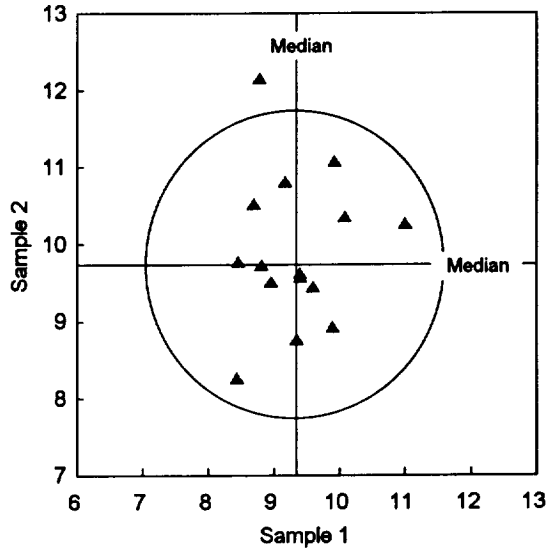

Figure 7.1 Youden chart with only random uncertainty.

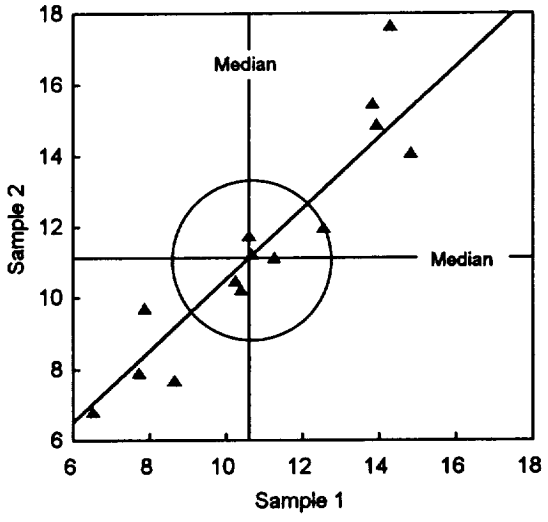

Figure 7.2 Sample Youden chart with laboratory bias.

7.5.2 Interpreting the Youden Chart

Once the chart has been constructed, visual inspection is the primary tool for its interpretation based on the guideline below:

(1) The random within-laboratory component of uncertainty is indicated by the scatter about the reference line and quantified using Eq. (7.2). The distribution of the results in the four quadrants provides information about the randomness of the data. Clearly the data in Figure 7.1 are random appearing in approximately equal numbers in each quadrant and falling within the $3s_Y$ limit with the exception of one point. It is important to note that the Youden standard deviation s_Y is only a measure of the group's random measurement uncertainty.

(2) Laboratory bias is indicated by the stringing of the points along the 45° reference line as shown in Figure 7.2. The limit circle (the same $3s_Y$ for both Figures 7.1 and 7.2) serves to show how each laboratory behaves with respect to the within-laboratory precision. In this example the bias is 3 times the random uncertainty and the bias component is obvious.

(3) Although the median serves to determine randomness or a lack thereof, other possible quantities may also play a major role in interpreting the results.

   (a) The artifacts may have a known value in which case all biases are reckoned relative to those values. *Example:* A round robin conducted by NIST.
   (b) Occasionally the mean of the group of laboratories is taken as the reference value. Then the reference line passes through that point and the uncertainty circle is centered on that point.

(4) The overall uncertainty from this analysis is a rough estimate of the combined capability of all participants. That is, the standard deviation of all data is a measure of the group capability to make the particular measurement. An estimate of the group's capability can be made by calculating the standard deviation of the results from each artifact and pooling the two. For this example, the standard deviations for the no bias and bias cases are 0.8 and 3 respectively. Thus, for the bias case, at the 2σ level, the uncertainty of a calibration performed at a laboratory selected at random would be about ±6 units with respect to the median of the group.

(5) These results are the property of the system and must be further examined as they affect the required performance of the system. Since most measurement integrity exercises are conducted to improve the system results, bias and other anomalous results must be investigated and corrected.

7.5.3 Youden Chart Enhancements

The value of the Youden chart can be further enhanced by linear least squares fit of the data and testing the resulting slope. If the slope is statistically different from that predicted then there is reason to believe that the model is not correct. There are two scenarios, no bias and bias, and the statistic t is calculated differently for each. To test for no bias the expected slope should be zero, therefore the difference between the calculated slope and zero is calculated. The bias case uses one instead of zero. When bias is present the slope may not be one because the measurement system may be sensitive to artifact value (artifacts having widely differing values). The spreadsheet is an easy way to make the calculation using the "Regression" command found on many. This function calculates the intercept, slope, standard deviation of a single observation, and the standard deviation of the slope as summarized in Table 7.2. Visual inspection of the table indicates that the two slopes agree with the expected outcomes, which can

be formalized using the *t* test. The statistic *t* is calculated based on Section 4.4.1 and is given for the two cases. The critical values of *t* were obtained from Table B.2 (Appendix B) for 13 degrees of freedom. One would conclude, for both cases, that the slopes are not statistically significant. One further note. The least square calculation assumes no error in the x's, which is not the case, but it is still a reasonable approximation.

Table 7.2
Least squares results for Table 7.1

Parameter	No Bias	Bias Present
Intercept (β_0)	8.63	0.149
$s_Y = \frac{s_{obs}}{\sqrt{2}}$ (note 1)	0.716	0.836
Slope (β_1)	0.139	1.043
Standard deviation of slope ($s_{\beta 0}$)	0.387	0.108
$t = \frac{\lvert 0 - \beta_1 \rvert}{s_{\beta 0}}$	0.359	—
$t = \frac{\lvert 1 - \beta_1 \rvert}{s_{\beta 0}}$	—	0.398
t_p (d.f. = 13), p=95%	2.16	2.16

Note 1: Corresponds to the Youden std. dev. calculated above.

7.5.4 Youden Chart Example - Rockwell Hardness

A round robin conducted among a number of measurement laboratories can assess the reproducibility of the Rockwell C Hardness Scale as maintained by participating laboratories using a variety of commercial hardness test machines and procedures. Accordingly, uniformity of hardness measurements is established by following commonly accepted measurement standards that specify characteristics for testing machines, indenters, and hardness test blocks which are generally accepted and used by industry. The circulating artifacts chosen for the round robin were a pair of Rockwell C test blocks (RC25.8 and RC60). Before circulation, the hardness value stamped on each block by the manufacturer was removed. The reference document for all participants was ASTM Designation: E18-67, *Rockwell Hardness and Rockwell Superficial Hardness of Metallic Materials*. A hardness test machine measures hardness by determining the depth of penetration of an indentor of specified geometry into the test specimen under fixed conditions as detailed in E18-67. The procedure calls for five hardness measurements to be performed on each specimen. The round robin package was circulated among nine participating laboratories and all measurement results were forwarded to the Interlaboratory Coordinator. Table 7.3 summarizes relevant data from the experiment by laboratory and specimen. Each laboratory measured

each specimen 5 times and the mean and standard deviation of the mean are recorded in the table. The results are plotted in Figure 7.3 along with the medians (62.25 and 25.78), reference 45° line and the $3s$ circle to complete the Youden chart. Using Eq. (7.2) the $s_Y = 0.45$ units ($3s_Y = 1.35$). Visual inspection of Figure 7.3 and the table of data reveals a great deal about interlaboratory agreement and precision.

(1) There is clearly a laboratory bias even if the points lying outside the limit circle are eliminated. Note that the outliers are still close to the reference line. Two of the points outside the circle are from the same laboratory and were very close together.

(2) The dispersion about the 45° line of about 0.45 units; approximately twice the internal precision of each laboratory's measurements (pooled std. dev. for all measurement is about 0.26).

Table 7.3
Rockwell Hardness Round Robin Results

Lab Code	Test Block 1 (Y)		Test Block 2 (X)	
	Hardness Mean	Standard Deviation	Hardness Mean	Standard Deviation
DD	22.1	0.09	59.2	0.28
CC	25.4	0.66	60.5	0.10
FF	25.5	0.10	61.3	0.27
HH	25.7	0.12	61.5	0.10
BB	25.8	0.09	62.2	0.30
EE	25.8	0.00	62.3	0.12
MFR	25.9	0.10	62.8	0.25
AB	26.4	0.43	63.1	0.49
AA	28.2	0.21	63.8	0.12
AA	28.3	0.17	64.0	0.30

7.5.4.1 Reviewing the Results

When evaluated in terms of other information, pooled standard deviation, the mean of each set of data, manufacturer's test results and specifications, etc., more is learned about the process. First, the means and medians are quite close in both cases indicating that the bias is fairly normally distributed. Second, the reproducibility of the measurements within a laboratory is about the same for each specimen which indicates that the method and apparatus yield similar results at both ends of the scale. Third, the agreement with the manufacturer's value is reasonable in light of the scatter. One unexpected result deals with the stated uniformity of the test blocks. The RC 25.8 and RC 60 blocks had RC 1.0 and RC 0.5 uniformity specifications respectively. These results indicate that the internal uniformity of the blocks is more nearly equal. It is probable that the specifications also include block-to-block uniformity. The results and conclusions from the round robin require action to bring about better interlaboratory agreement.

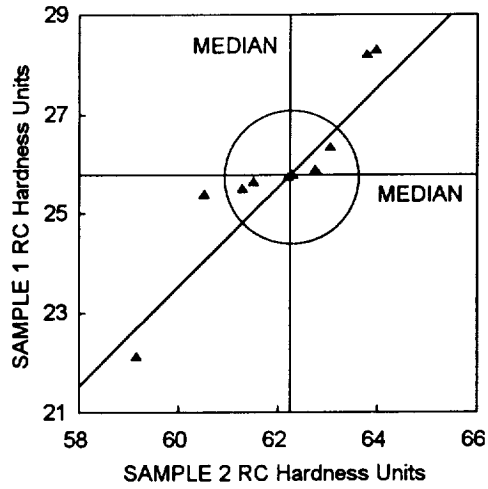

Figure 7.3 NASA hardness round robin. Two RMs circulated among 10 laboratories.

The utility of a round robin experiment is enhanced by preplanning to gather the most information from the measurements made by participants. Rarely does the first experiment in a discipline come out perfectly but subsequent ones can be enhanced by a careful critique of the previous ones. Since the process depends on machines and procedures the experiment can be modified to investigate specific potential error sources. One in particular, the indentor, can be evaluated by adding an indentor to the package and making an extra set of measurements to better determine the source of the disagreement.

7.5.5 Artifact Round Robins (Voltage)

The capability of laboratories maintaining the volt using groups of standard cells in temperature controlled enclosures or solid-state voltage references (SSVR) can be assessed by circulating a pair of SSVRs[2]. Because of the length of time required and the need for nearly state-of-the-art measurements the experiment required modification. Specifically:

(1) Allow one week, minimum for settling and data taking;

(2) A measurement scheme based on well-established principles was provided to assess each laboratory's potential capability;

(3) A low-thermal switch to facilitate Item (2) was included in the package;

(4) Establish a pivot laboratory (Lab E in the example) to measure the standards before, at the midpoint, and the end of the experiment to monitor the traveling standard's drift. For a large number of laboratories more pivot laboratory visits are required.

---

[2] Standard cell enclosures could be used but they will require more time, at least two weeks per laboratory.

(5) If possible calibrate the standards in terms of the U.S. volt and adjust the results for any drift caused by the length of the experiment.

An example of such an experiment is presented in Table 7.4 and Figure 7.4. The chart is constructed in the manner of Section 7.5.1.1 with only minor variations. The circle was omitted because it was so small (0.12 ppm). Second, a regression line was calculated along with the 45° reference line using the techniques of Section 7.1.6. The regression analysis estimate of the intercept, slope, and standard deviation of a single point are 0.091, 1.014, and 0.053 ppm respectively. Third, the SSVRs were initially adjusted so that the difference with respect to the U.S. volt was initially zero with an uncertainty of 0.3 ppm (this is not necessary nor it always desirable). Finally, each laboratory was asked to provide its estimated process uncertainty. Analysis of the Youden plot and associated data show:

Table 7.4
Data from an 11 Laboratory SSVR Round Robin

Lab Code	SSVR-A Diff. from nominal (ppm)	SSVR-B Diff. from nominal (ppm)	Lab Claimed Uncertainty (ppm)
A	-2.49	-2.39	3
B	-1.38	-1.27	5
C	-0.92	-0.83	1.8
D*	-0.53	-0.45	0.44
E-2	-0.10	-0.03	0.3
E-3	-0.09	-0.11	0.3
E-1	-0.01	0.01	0.3
G	0.27	0.33	1
H	0.31	0.43	2.3
J*	0.81	0.97	0.8
K	1.46	1.59	NA
L*	2.49	2.68	1.5
M*	3.48	3.60	2

(1) The precision of the experiment is more than an order of magnitude greater than the between laboratory differences which shows that all laboratories have a sound basic system or can establish one. All laboratories have comparable precision. The standard deviation of the fit (0.053 ppm) is a good indicator of the overall ability to intercompare standards and is typical.

(2) The calculated and expected slope (1) can be tested for significance using the technique of Section 4.4.1 as shown below. In this case, the expected slope, 1, is known.

$$t = \frac{|1 - \beta_1|}{s_{\beta 1}} = \frac{|1 - 1.01415|}{0.00974} = 1.453 \qquad (7.3)$$

At the 95 % significance level $t_{\alpha} = 2.20$ thus concluding that the slope does not differ statistically from the expected value of 1.

(3) Five of the 11 laboratories exceeded the 1 ppm volt potential maintenance capability.

(4) Four of the 11 laboratories exceeded their own claimed uncertainty.

As a minimum all laboratories exceeding 1 ppm need to review their operation and in several instances the claimed uncertainty does not match the observed. Although the experiment is directly traceable to NIST, an adjustment based on the experiment could be dangerous unless the causes of the offsets are identified and corrected. One point not mentioned are the sources of the individual calibrations. Not all laboratories obtained their unit directly from NIST therefore, this avenue needs to be explored as possible sources of the offsets.

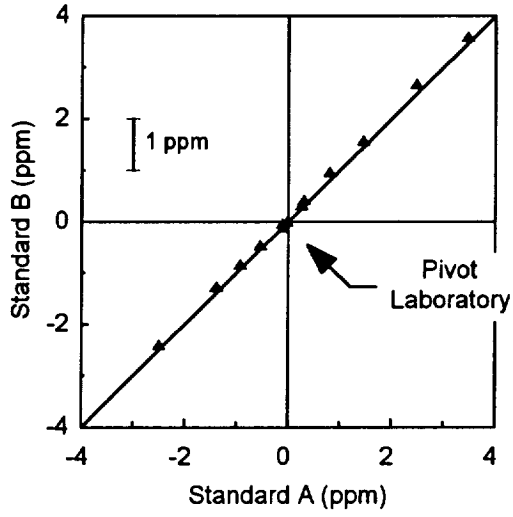

Figure 7.4 SSVR round robin using 10 V SSVRs. The standards were initially adjusted to 10.000 000 V and periodically returned to the pivot laboratory during the experiment.

7.5.6 Youden Chart Using Only One Standard

There are times when multiple transport standards may not be available so another strategy needs to be developed using a single standard. One is a variation of the Youden chart. Rather than circulate a pair of standards, one is circulated twice using a single pivot laboratory. This method works best when there is little or no drift of local standards or the traveling standard. The pattern and timing must be carefully controlled to eliminate possible drifts of the standards being monitored. As a minimum, the pivot laboratory must see the standard at the beginning, middle and end of the experiment. If many laboratories are involved, additional visits to the pivot laboratory are in order. Timing is also important because it is assumed that during the course of the round robin all local and traveling standards drift

linearly. The success of this method depends on the stability of the pivot laboratory's standards and the time required for the experiment. To eliminate this effect, a symmetrical pattern is employed as illustrated below.

> Starting at the pivot laboratory (P) the traveling standard follows the route below. What is most important is that the time interval between the departure from one laboratory and the next is approximately the same for each segment. As the path lengthens then more visits to the pivot laboratory are in order
>
> $$P_1 \rightarrow\rightarrow A_1 \rightarrow\rightarrow B_1 \rightarrow\rightarrow C_1 \rightarrow\rightarrow P_2 \rightarrow\rightarrow P_3 \rightarrow\rightarrow C_2 \rightarrow\rightarrow B_2 \rightarrow\rightarrow A_2 \rightarrow\rightarrow P_4$$
>
> If the time intervals are nearly equal and the drift linear, the averages of the values at each laboratory will be free of the drift. Another alternative is to estimate the drift from pivot laboratory data and make a correction to the remaining results.

An example of this type of experiment involved eight[3] laboratories and a single traveling standard. A pivot laboratory was selected which supplied and calibrated a single SSVR to serve as the traveling standard. The selected pattern was:

Pass 1: P →→ A →→ B →→ P →→ C →→ D →→ P →→ E →→ F →→ P →→ G →→ P
Pass 2: P →→ G →→ P →→ F →→ E →→ P →→ D →→ C →→ P →→ B →→ A →→ P

Each stop requires about one week during which time each laboratory measured the standard several times. Each laboratory measured the traveling standard twice and the pivot laboratory measured the standard ten times. The analysis proceeds as follows.

Step 1. Using the pivot laboratory data calculate the time dependence of the SSVR with respect to the pivot laboratory's reference standards either graphically or using the least square method of Section 7.1.6. In this example, the time dependence of the SSVR with respect to the pivot laboratory ($\Delta E_{TRN,P}$) was found to be

$$\Delta E_{TRN,P} = -6.30376 - 0.005769\,(t - t_0)$$

where t and t_0 are the time and starting time respectively.

Step 2. Adjust each participant's results, correcting for drift using the information of Step 1 (which includes pivot laboratory drift).

Step 3: Plot the results from Run 1 against Run 2. The balance of the chart is constructed as in the example of Section 7.5.1 except that the mean is used instead of the median. The finished chart is shown in Figure 7.5.

Interpretation is similar to the previous examples but several points are worth noting. First, the $3s$ reference contains only one point and the dispersion along the 45 ° line is much larger. Second, the chart is relative. That is, there is no reference to a recent external calibration. Finally, the dispersion about the line is relatively large indicating (1) there may be local measuring problems, (2) the traveling standard may have a large uncertainty, or (3) there were changes in the local standards. Without further

---

[3] One of the laboratories had problems with a transfer and is therefore not include in the example.

information no problem can be clearly identified as a possible cause. Finally, the results clearly indicate that all processes should be studied and future experiments conducted.

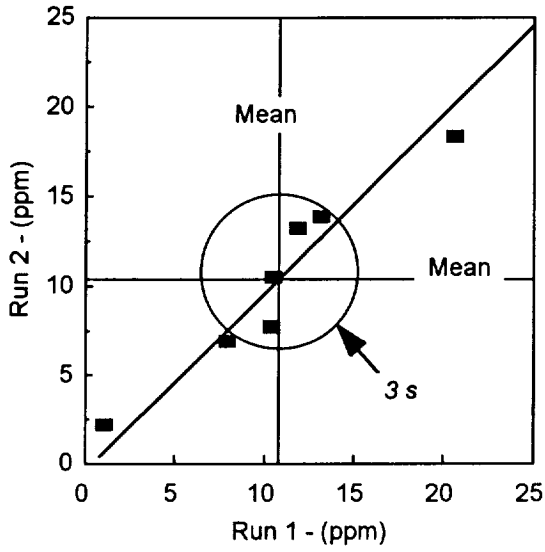

Figure 7.5 Youden plot for a single 10 V SSVR round robin. Data was acquired by circulating a single traveling standard twice.

7.6 Limited Standards Round Robins

Often it is not feasible to conduct a repeat experiment as was done for the previous example so another course of action must be taken. Generally, a single artifact is circulated in the same manner as before with each laboratory providing measurement data and uncertainty information which is then used to analyze the experiment. Using the data from only one traveling standard (A) of Section 7.5.5 proceed as follows.

Step 1. Plot the values of Table 7.4 as shown in Figure 7.6. The x-axis is laboratory designation and the y-axis the observed deviation from nominal.

Step 2. For each value, calculate the error bar (E_{obs}±U).

Step 3. Add an error bar for the "true" value which, in this case, is 0.3 ppm with respect to NIST to complete the chart.

Note that the conclusions are very similar to the Youden example but convey less information. This method is better suited to those processes which have stable standards such as mass or gage blocks.

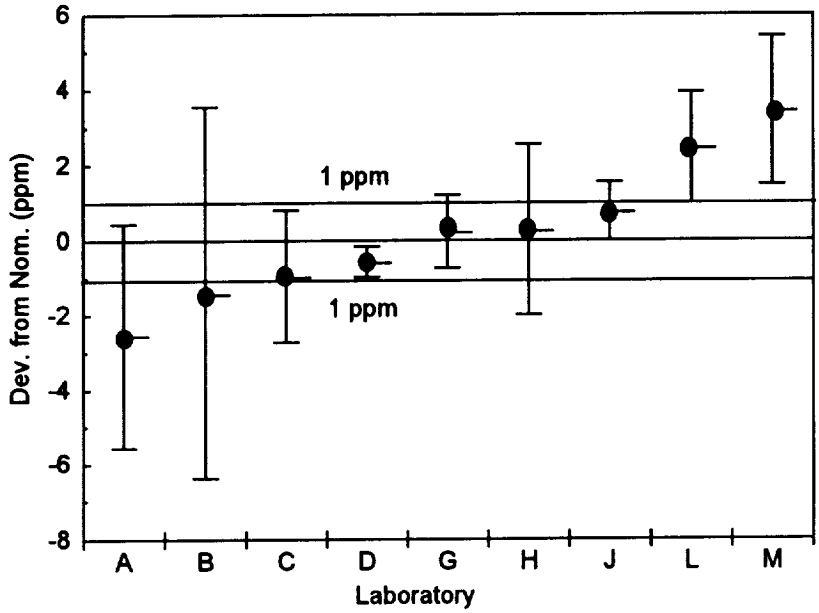

Figure 7.6 Interlaboratory experiment using a single 10 V SSVR. Using the data for a single SSVR of Table 7.3. The 1 ppm limit is based on realizable long-term capability.

7.7 Interlaboratory Agreement Summary

Interlaboratory agreement experiments, especially Youden charts, are powerful tools to identify aberrant measuring process behavior but, rarely pinpoint the causes by themselves. In addition to the basic data of the previous examples other important ancillary data should also be recorded and forwarded to the Interlaboratory Coordinator for analysis and inclusion. Once a problem condition has been identified, clues to the possible causes can often be found by careful examination of the results of the experiment in conjunction with other data. Some possible areas warranting investigation are listed in Table 7.5.

7.7.1 Group Uncertainty

The objective of calibration is to ensure that a calibration made in terms of one standard agrees with that made in terms of another to within the combined uncertainty of the two. Interlaboratory tests quantify this process for a defined group. The dispersion of the results when viewed in light of expected performance indicate those facilities having problems. Additionally, interlaboratory tests quantify the group's capability to make measurements meeting the prescribed tolerance or specification. For example, if a specification specifies a hardness measurement to ±0.75 units, clearly capability does not exist within the group of Section 7.5.4. In fact, the relative group capability, as measured by the standard deviation is 1.5 units which translates to group uncertainty U of 3. Similarly the group capability for voltage is only about 3.5 ppm (2σ). In other words, if a 6 1/2 digit DVM was calibrated at two different laboratories the worst case differences (2σ) could be as large as 7 ppm. Whether or not the capability is acceptable must be weighed in terms of the externally imposed measurement requirements.

Table 7.5
Possible Problems Identified Through Round Robins

Dispersion along the Reference Line (Bias or offset)	**Dispersion about the Reference Line** (random)
• Standards out of calibration Changed with time • Bias in the local measuring process Sensor error Indicating equipment errors Connection errors • Bias error introduced during data reduction Errors in making corrections Incorrect algorithm Software errors • External influences Temperature, etc. • Operator • Traveling standard has not settled down Transportation effects Local influences • Possible shift in the traveling standards	• Large within-day uncertainties Noisy measuring process • Large between-day uncertainties Day-to-day variations of standards Day-to-day instrument variations • Variability of the transport standard Local influences • Variations of external influences Local influences • Operator

8 Bibliography

ANSI/NCSL, *Calibration Laboratories and Measuring and Test Equipment – General Requirements.* ANSI/NCSL Z540-1-1994 (July 27, 1904).

AT&T Technologies, *Statistical Quality Control Handbook*, (available through AT&T Technologies, PO Box 19901, Indianapolis IN 46219).

Belanger, B. C., *Measurement Assurance Programs:* Part I, NBS Special Publication (SP) 676-I, May 1984.

Belecki, N. B., Dziuba, R. F., Field, B. F., Taylor, B. N., *Guidelines for Implementing the New Representations of the Volt and the Ohm Effective January 1, 1990*, NIST Technical Note TN 1263, (June 1989).

Croarkin, C., *Measurement Assurance Programs Part II: Development and Implementation*, NBS SP676-II, (April 1984).

Davidson, G. M., *Regional Measurement Assurance Programs Past and Future*, 1980 ASQC Technical Conference Transactions - Atlanta GA.

Dixon, W. J., Massey, F. J., Introduction to Statistics, 2 nd Edition, Mc Graw-Hill Book Co,, Inc. New York (1957).

Eicke, W, G., Cameron, J. M., *Designs for the Surveillance of Small Groups of Standard Cells*, Reprinted in NBS SP705 pp. 2893-311 (1985); originally NBS TN430 (1967).

Eicke. W. G., Auxier, L. M., *Regional Maintenance of the Volt Using NBS Volt Transfer Techniques*, Reprinted in NBS SP705 pp. 327-331 (1985); originally published in IEEE Trans. Inst. & Meas., Vol. IM-23, No. 4, (December 1974).

Grant, E. L., Leavenworth, R. S., *Statistical Quality Control*, 6th Ed. McGraw-Hill, Inc. (1988).

International Organization for Standards (ISO), *Units of Measurement*, ISO Standards Handbook 2, 2nd Ed., 1982

International Organization for Standards (ISO), *Guide To the Expression of Uncertainty in Measurement*, ISO/TAG4/WG3, 1st Ed., 101 pages, (1993).

Juran, J. M, Editor, *Juran's Quality Control Handbook*, 4th edition, McGraw-Hill Book Co. Chapter 24, (1988).

Lapin, L. L., *Probability and Statistics for Engineers*, PWS-KENT Pub. Co.(1990).

NASA Metrology and Working Group, *Working Group Operating Procedure.*

NASA Reference Publication RP 1342, *Metrology-Calibration and Measurement Process Guidelines* (1994).

NASA Metrology Laboratory Measurement Capabilities Document.

Natrella, M. G., *Experimental Statistics*, NBS Handbook 91, (1963).

NCSL Glossary Committee, *NCSL Glossary of Metrology-Related Terms*, NCSL (August 1994).

NCSL RP-12, *Determining and Reporting Measurement Uncertainties* (1994).

NIST SP260, *Standards Reference Materials Catalog*. This document is revised on a regular basis and the latest version should always be consulted.

Riley, J. P., *Ten Volt Round Robin Using Solid State Standards;* Proceedings of the 1990 Measurement Science Conference, (1990).

Simmons, J. D., Editor, *NIST Calibration Services Guide 1991 Edition*, NIST SP 250, October 1991. This document is revised on a regular basis and the latest version should always be consulted.

Snedecor, G. W., Cochran, W. G., *Statistical Methods*, 7th Ed., Iowa State University Press, (1980).

Taylor, B. N., Kuyatt, C. E. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297, January 1993.

Taylor, J., *Fundamentals of Measurement Error*, Neff Instrument Corp., Monrovia CA, 1988.

Youden W. J., *Statistical Methods for Chemists*; John Wiley & Sons (1951).

Youden, W. J., *Expermentation and Measurement*, NIST Spec. Pub. SP 672 (May 1994), Originally published as a VISTA of SCIENCE book in 196 2 for high school students but is an excellent introduction to measurements for all ages.

Youden, W. J., *Graphical Diagnosis of Interlaboratory Test Results,* Reprinted in NBS SP300 Vol. 1, pp. 122-137; originally published in Ind. Qual. Control, Vol. XV, No. 11, May 1959.

Youden, W. J., *The Sample, The Procedure, The Laboratory*, Reprinted in NBS SP300 Vol. 1, pp. 138-145; originally published in Report for Analytical Chemists.

Zucker, A., Eicke, W. G., *Regional Measurement Assurance using Solid-State References*, Proceedings of the 1989 Measurement Science Conference, Jan 27-28, 1989, Annaheim CA..

Appendix A
Definitions

Unless other wise noted all definitions were extracted from Appendix A of RP1342 Metrology – *Calibration and Measurement Process Guidelines* (June 1994) along with the following:

> "**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 true value of the measurand. *Notes – 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)

base unit – A unit of measurement of a base quantity in a given system of quantities. (VIM)

bias error – The inherent bias (off-set) of a measurement process or (of) one of its components. (See also 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. *NOTE* – (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)

certified reference material (CRM) – A reference material, one or more of whose property values are certified by a technically valid producer, accompanied by or traceable to a certificate or other documentation that is issued by a certifying body. *Note* – NIST issues Standard Reference Materials (SRM) which are in effect CRM.

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. *Note* – (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)

consensus standard – A standard not traceable to national standards but has an agreed on method for realization of the quantity. *Example:* The Rockwell Hardness Scale that depends on specifying a procedure and an apparatus meeting certain specifications.

confidence interval – An interval about the result of a measurement or computation within which the true 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 true value of a measurement.

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 (conventional) true 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 can not 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 can not be known exactly the correction factor is subject to uncertainty. (VIM)

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 which 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.

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 gage 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 measurements to determine the values of influence quantities in order to make corresponding corrections. *Example*: (a) measurement of a length using a graduated rule, (b) measurement of a mass using an equal-arm balance. (VIM)

drift – The slow variation with time of a metrological characteristic of a measuring instrument. (VIM)

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 effect the result of a measurement. *Note* – The sensor does not measure an environmental variable: it measures an *observable*. (Also known as influence quantities)

error – The difference between the result of a measurement and the true 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 specialty 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)

indicating (measuring) instrument – A measuring instrument that displays the value of a measurand or a related value. *Examples*: (a) analog voltmeter: (b) digital voltmeter, (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) 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 semi-digital 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) has to 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) (see also **environmental variables**)

instrument constant – The coefficient by which a direct indication must be multiplied to obtain the indication of a measuring instrument. *NOTE* – (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)

international standard – A standard recognized by an international agreement to serve internationally as the basis for fixing the value of all other of the quantity concerned. (VIM)

intrinsic error (of a measuring instrument) – Errors inherent in a measuring instrument. *Example*: non-linearity, gain accuracy, noise, offset, 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 Non-Linearity).

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 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 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 of 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 gage block; (c) 100 ohm standard resistor; (d) saturated Weston standard cell, (e) standard ammeter; (d) 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 electroacoustic 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)

non-linearity – 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. This is classified as a bias error and is expressed in percent of full scale.

normalization period – The time required for a standard to return to its normal operating mode after being subjected to an external influence. *Examples* – Time required for a standard cell to return to its value after a temperature excursion; time for an oscillator to return to its frequency after power is turned on.
Note – Normalization period is also referred to as stabilization time or settling time.

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)

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)

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 of whose property values are sufficiently homogenous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials. *Note* – A reference material may be in the form of a pure or mixed gas, solid or liquid. *Examples:* Water for the calibration of viscometers, sapphire as a heat capacity calibrant, and solutions used for calibration in chemical analysis. (See also certified reference material and standard reference material). (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 (conventional) true value of the measurand. (VIM)

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 under changing conditions such as: (a) method of measurement; (b) observer; (c) measuring instrument; (d) location; (e) conditions of use; (f) time. (VIM) (See precision).

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; (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)

secondary standard – A standard whose value is fixed by comparison with a primary standard. (VIM)

SI units – The coherent system of units adopted and recommended by the General Conference on Weights and Measures (CGPM). (VIM)

standard deviation – For a series of n measurements of the same measurand, the quantity s characterizing the dispersion of the results and given by the formula:

$$s = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}{n-1}$$

x_i being the result of the ith measurement and $\bar{x}$ being the arithmetic mean of the n results considered.

Notes – (1) The experimental standard deviation should not be confused with the population standard deviation of a population of size N and of mean n, given by the formula:

$$\sigma = \frac{\sqrt{\sum_{i=1}^{n} (x_i - m)^2}}{N}$$

(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 m of the overall population. The expression $s/\sqrt{n}$ is called the experimental standard deviation of the mean. (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. *Notes:* (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 – Property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties. (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 intercompare end standards. (VIM)

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) – An estimate characterizing the range of values within which the true value of a measurand lies. *Note* – Uncertainty of measurement comprises, in general, many components. Some of these components may be estimated on the basis of the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations. Estimates of other components can only be 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 all functional requirements have been met. Includes all sub-system and system tests done at the functional level.

working standard – A standard which, usually calibrated against a reference standard, is used routinely to calibrate or check material measures or measuring instruments. (VIM)

Appendix B
Statistical Tables

All tables were generated by Quattro Pro for Windows®[1] using the appropriate statistical functions and have been verified by random checks with the tables in NBS Handbook 91, *Experimental Statistics*

® Quattro Pro is the registered Trademark of the Borland Corporation.

Table B.1

Control limits for the standard deviation

No. of Obs	3-sigma limits*		2-sigma limits**		Central Line (CL)
	Lower Limit B_L	Upper Limit B_U	Lower Limit B_L	Upper Limit B_U	
2	0	3.48	0.000	2.241	0.674
3	0	2.76	0.159	1.921	0.833
4	0	2.43	0.268	1.765	0.888
5	0	2.24	0.348	1.669	0.916
6	0.178	2.10	0.408	1.602	0.933
7	0.223	2.00	0.454	1.552	0.944
8	0.263	1.93	0.491	1.512	0.952
9	0.298	1.87	0.522	1.480	0.958
10	0.329	1.82	0.548	1.454	0.963
11	0.356	1.77	0.570	1.431	0.967
12	0.380	1.74	0.589	1.412	0.970
13	0.401	1.70	0.606	1.395	0.972
14	0.421	1.68	0.621	1.379	0.974
15	0.439	1.65	0.634	1.366	0.976
16	0.455	1.63	0.646	1.354	0.978
17	0.470	1.61	0.657	1.343	0.979
18	0.484	1.59	0.667	1.333	0.980
19	0.497	1.57	0.676	1.323	0.981
20	0.508	1.56	0.685	1.315	0.982
22	0.530	1.53	0.700	1.300	0.984
24	0.549	1.50	0.713	1.287	0.985
26	0.565	1.48	0.724	1.275	0.987
28	0.580	1.46	0.735	1.265	0.988
30	0.594	1.45	0.744	1.256	0.988

* 3-sigma and 2-sigma are the terms commonly used in SPC. The actual limits are α=0.001 and α=0.05 respectively which approximate those named limits.

Table B.2

Values of $t_p(\nu)$ from the t-distribution for degrees of freedom ν that defines the interval $-t_p(\nu)$ to $+t_p(\nu)$ that encompasses the fraction p of the distribution

Degrees of freedom ν	Fraction of p in percent					
	68.27[a]	90	95	95.45[a]	99	99.73[a]
1	1.84	6.31	12.71	13.97	63.66	235.78
2	1.32	2.92	4.30	4.53	9.92	19.21
3	1.20	2.35	3.18	3.31	5.84	9.22
4	1.14	2.13	2.78	2.87	4.60	6.62
5	1.11	2.02	2.57	2.65	4.03	5.51
6	1.09	1.94	2.45	2.52	3.71	4.90
7	1.08	1.89	2.36	2.43	3.50	4.53
8	1.07	1.86	2.31	2.37	3.36	4.28
9	1.06	1.83	2.26	2.32	3.25	4.09
10	1.05	1.81	2.23	2.28	3.17	3.96
11	1.05	1.80	2.20	2.25	3.11	3.85
12	1.04	1.78	2.18	2.23	3.05	3.76
13	1.04	1.77	2.16	2.21	3.01	3.69
14	1.04	1.76	2.14	2.20	2.98	3.64
15	1.03	1.75	2.13	2.18	2.95	3.59
16	1.03	1.75	2.12	2.17	2.92	3.54
17	1.03	1.74	2.11	2.16	2.90	3.51
18	1.03	1.73	2.10	2.15	2.88	3.48
19	1.03	1.73	2.09	2.14	2.86	3.45
20	1.03	1.72	2.09	2.13	2.85	3.42
22	1.02	1.72	2.07	2.12	2.82	3.38
24	1.02	1.71	2.06	2.11	2.80	3.34
26	1.02	1.71	2.06	2.10	2.78	3.32
28	1.02	1.70	2.05	2.09	2.76	3.29
30	1.02	1.70	2.04	2.09	2.75	3.27
50	1.01	1.68	2.01	2.05	2.68	3.16
100	1.005	1.660	1.984	2.025	2.626	3.077
∞	1.000	1.645	1.960	2.000	2.576	3.000

[a] For p= 68.27, 95.45, and 99.73 corresponds approximately k = 1, 2, and 3 respectively

Table B.3
Percentiles of the F Distribution
$F_{.95}(\nu_1 \nu_2)$

Degrees of freedom in Denominator ν_2	Degrees of Freedom in the Numerator ν_1												
	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.78
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.77	1.68
∞	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.68	1.58

Table B.4
Percentiles of the F Distribution
$F_{.99}(\nu_1, \nu_2)$

Degrees of freedom in Denominator ν_2	Degrees of Freedom in the Numerator ν_1												
	1	2	3	4	5	6	7	8	9	10	12	15	20
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.56	2.42	2.27
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.37	2.22	2.07
∞	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.20	2.06	1.90

END DATE MAY 5, 1993