2. Measurement Process Characterization

1. Characterization

1. Issues
2. Check standards

2. Control

1. Issues
2. Bias and long-term variability
3. Short-term variability

3. Calibration

1. Issues
2. Artifacts
3. Designs
4. Catalog of designs
5. Artifact control
6. Instruments
7. Instrument control

4. Gauge R & R studies

1. Issues
2. Design
3. Data collection
4. Variability
5. Bias
6. Uncertainty

5. Uncertainty analysis

1. Issues
2. Approach
3. Type A evaluations
4. Type B evaluations
5. Propagation of error
6. Error budget
7. Expanded uncertainties
8. Uncorrected bias

6. Case Studies

1. Gauge study
2. Check standard
3. Type A uncertainty
4. Type B uncertainty

Detailed table of contents

References for Chapter 2

2. Measurement Process Characterization - Detailed Table of Contents

1. Characterization [2.1.]
    1. What are the issues for characterization? [2.1.1.]
        1. Purpose [2.1.1.1.]
        2. Reference base [2.1.1.2.]
        3. Bias and Accuracy [2.1.1.3.]
        4. Variability [2.1.1.4.]
    2. What is a check standard? [2.1.2.]
        1. Assumptions [2.1.2.1.]
        2. Data collection [2.1.2.2.]
        3. Analysis [2.1.2.3.]

2. Statistical control of a measurement process [2.2.]
    1. What are the issues in controlling the measurement process? [2.2.1.]
    2. How are bias and variability controlled? [2.2.2.]
        1. Shewhart control chart [2.2.2.1.]
            1. EWMA control chart [2.2.2.1.1.]
        2. Data collection [2.2.2.2.]
        3. Monitoring bias and long-term variability [2.2.2.3.]
        4. Remedial actions [2.2.2.4.]
    3. How is short-term variability controlled? [2.2.3.]
        1. Control chart for standard deviations [2.2.3.1.]
        2. Data collection [2.2.3.2.]
        3. Monitoring short-term precision [2.2.3.3.]
        4. Remedial actions [2.2.3.4.]

3. Calibration [2.3.]
    1. Issues in calibration [2.3.1.]
        1. Reference base [2.3.1.1.]
        2. Reference standards [2.3.1.2.]
    2. What is artifact (single-point) calibration? [2.3.2.]
    3. What are calibration designs? [2.3.3.]
        1. Elimination of special types of bias [2.3.3.1.]
            1. Left-right (constant instrument) bias [2.3.3.1.1.]
            2. Bias caused by instrument drift [2.3.3.1.2.]
        2. Solutions to calibration designs [2.3.3.2.]
            1. General matrix solutions to calibration designs [2.3.3.2.1.]
        3. Uncertainties of calibrated values [2.3.3.3.]
            1. Type A evaluations for calibration designs [2.3.3.3.1.]
            2. Repeatability and level-2 standard deviations [2.3.3.3.2.]
            3. Combination of repeatability and level-2 standard deviations [2.3.3.3.3.]

4. Calculation of standard deviations for 1,1,1,1 design [2.3.3.3.4.]
5. Type B uncertainty [2.3.3.3.5.]
6. Expanded uncertainties [2.3.3.3.6.]

4. Catalog of calibration designs [2.3.4.]
   1. Mass weights [2.3.4.1.]
      1. Design for 1,1,1 [2.3.4.1.1.]
      2. Design for 1,1,1,1 [2.3.4.1.2.]
      3. Design for 1,1,1,1,1 [2.3.4.1.3.]
      4. Design for 1,1,1,1,1,1 [2.3.4.1.4.]
      5. Design for 2,1,1,1 [2.3.4.1.5.]
      6. Design for 2,2,1,1,1 [2.3.4.1.6.]
      7. Design for 2,2,2,1,1 [2.3.4.1.7.]
      8. Design for 5,2,2,1,1,1 [2.3.4.1.8.]
      9. Design for 5,2,2,1,1,1,1 [2.3.4.1.9.]
      10. Design for 5,3,2,1,1,1 [2.3.4.1.10.]
      11. Design for 5,3,2,1,1,1,1 [2.3.4.1.11.]
      12. Design for 5,3,2,2,1,1,1 [2.3.4.1.12.]
      13. Design for 5,4,4,3,2,2,1,1 [2.3.4.1.13.]
      14. Design for 5,5,2,2,1,1,1,1 [2.3.4.1.14.]
      15. Design for 5,5,3,2,1,1,1 [2.3.4.1.15.]
      16. Design for 1,1,1,1,1,1,1,1 weights [2.3.4.1.16.]
      17. Design for 3,2,1,1,1 weights [2.3.4.1.17.]
      18. Design for 10 and 20 pound weights [2.3.4.1.18.]
   2. Drift-elimination designs for gage blocks [2.3.4.2.]
      1. Doiron 3-6 Design [2.3.4.2.1.]
      2. Doiron 3-9 Design [2.3.4.2.2.]
      3. Doiron 4-8 Design [2.3.4.2.3.]
      4. Doiron 4-12 Design [2.3.4.2.4.]
      5. Doiron 5-10 Design [2.3.4.2.5.]
      6. Doiron 6-12 Design [2.3.4.2.6.]
      7. Doiron 7-14 Design [2.3.4.2.7.]
      8. Doiron 8-16 Design [2.3.4.2.8.]
      9. Doiron 9-18 Design [2.3.4.2.9.]
      10. Doiron 10-20 Design [2.3.4.2.10.]
      11. Doiron 11-22 Design [2.3.4.2.11.]
   3. Designs for electrical quantities [2.3.4.3.]
      1. Left-right balanced design for 3 standard cells [2.3.4.3.1.]
      2. Left-right balanced design for 4 standard cells [2.3.4.3.2.]
      3. Left-right balanced design for 5 standard cells [2.3.4.3.3.]
      4. Left-right balanced design for 6 standard cells [2.3.4.3.4.]
      5. Left-right balanced design for 4 references and 4 test items [2.3.4.3.5.]
      6. Design for 8 references and 8 test items [2.3.4.3.6.]
      7. Design for 4 reference zeners and 2 test zeners [2.3.4.3.7.]
      8. Design for 4 reference zeners and 3 test zeners [2.3.4.3.8.]
      9. Design for 3 references and 1 test resistor [2.3.4.3.9.]
      10. Design for 4 references and 1 test resistor [2.3.4.3.10.]
   4. Roundness measurements [2.3.4.4.]
      1. Single trace roundness design [2.3.4.4.1.]
      2. Multiple trace roundness designs [2.3.4.4.2.]
   5. Designs for angle blocks [2.3.4.5.]
      1. Design for 4 angle blocks [2.3.4.5.1.]
      2. Design for 5 angle blocks [2.3.4.5.2.]
      3. Design for 6 angle blocks [2.3.4.5.3.]
   6. Thermometers in a bath [2.3.4.6.]

7. Humidity standards [2.3.4.7.]
    1. Drift-elimination design for 2 reference weights and 3 cylinders [2.3.4.7.1.]
5. Control of artifact calibration [2.3.5.]
    1. Control of precision [2.3.5.1.]
        1. Example of control chart for precision [2.3.5.1.1.]
    2. Control of bias and long-term variability [2.3.5.2.]
        1. Example of Shewhart control chart for mass calibrations [2.3.5.2.1.]
        2. Example of EWMA control chart for mass calibrations [2.3.5.2.2.]
6. Instrument calibration over a regime [2.3.6.]
    1. Models for instrument calibration [2.3.6.1.]
    2. Data collection [2.3.6.2.]
    3. Assumptions for instrument calibration [2.3.6.3.]
    4. What can go wrong with the calibration procedure [2.3.6.4.]
        1. Example of day-to-day changes in calibration [2.3.6.4.1.]
    5. Data analysis and model validation [2.3.6.5.]
        1. Data on load cell #32066 [2.3.6.5.1.]
    6. Calibration of future measurements [2.3.6.6.]
    7. Uncertainties of calibrated values [2.3.6.7.]
        1. Uncertainty for quadratic calibration using propagation of error [2.3.6.7.1.]
        2. Uncertainty for linear calibration using check standards [2.3.6.7.2.]
        3. Comparison of check standard analysis and propagation of error [2.3.6.7.3.]
7. Instrument control for linear calibration [2.3.7.]
    1. Control chart for a linear calibration line [2.3.7.1.]

4. Gauge R & R studies [2.4.]
    1. What are the important issues? [2.4.1.]
    2. Design considerations [2.4.2.]
    3. Data collection for time-related sources of variability [2.4.3.]
        1. Simple design [2.4.3.1.]
        2. 2-level nested design [2.4.3.2.]
        3. 3-level nested design [2.4.3.3.]
    4. Analysis of variability [2.4.4.]
        1. Analysis of repeatability [2.4.4.1.]
        2. Analysis of reproducibility [2.4.4.2.]
        3. Analysis of stability [2.4.4.3.]
            1. Example of calculations [2.4.4.4.4.]
    5. Analysis of bias [2.4.5.]
        1. Resolution [2.4.5.1.]
        2. Linearity of the gauge [2.4.5.2.]
        3. Drift [2.4.5.3.]
        4. Differences among gauges [2.4.5.4.]
        5. Geometry/configuration differences [2.4.5.5.]
        6. Remedial actions and strategies [2.4.5.6.]
    6. Quantifying uncertainties from a gauge study [2.4.6.]

5. Uncertainty analysis [2.5.]
    1. Issues [2.5.1.]
    2. Approach [2.5.2.]
        1. Steps [2.5.2.1.]
    3. Type A evaluations [2.5.3.]
        1. Type A evaluations of random components [2.5.3.1.]
            1. Type A evaluations of time-dependent effects [2.5.3.1.1.]
            2. Measurement configuration within the laboratory [2.5.3.1.2.]
        2. Material inhomogeneity [2.5.3.2.]

1. Data collection and analysis [2.5.3.2.1.]
3. Type A evaluations of bias [2.5.3.3.]
1. Inconsistent bias [2.5.3.3.1.]
2. Consistent bias [2.5.3.3.2.]
3. Bias with sparse data [2.5.3.3.3.]
4. Type B evaluations [2.5.4.]
1. Standard deviations from assumed distributions [2.5.4.1.]
5. Propagation of error considerations [2.5.5.]
1. Formulas for functions of one variable [2.5.5.1.]
2. Formulas for functions of two variables [2.5.5.2.]
3. Propagation of error for many variables [2.5.5.3.]
6. Uncertainty budgets and sensitivity coefficients [2.5.6.]
1. Sensitivity coefficients for measurements on the test item [2.5.6.1.]
2. Sensitivity coefficients for measurements on a check standard [2.5.6.2.]
3. Sensitivity coefficients for measurements from a 2-level design [2.5.6.3.]
4. Sensitivity coefficients for measurements from a 3-level design [2.5.6.4.]
5. Example of uncertainty budget [2.5.6.5.]
7. Standard and expanded uncertainties [2.5.7.]
1. Degrees of freedom [2.5.7.1.]
8. Treatment of uncorrected bias [2.5.8.]
1. Computation of revised uncertainty [2.5.8.1.]

6. Case studies [2.6.]
1. Gauge study of resistivity probes [2.6.1.]
1. Background and data [2.6.1.1.]
1. Database of resistivity measurements [2.6.1.1.1.]
2. Analysis and interpretation [2.6.1.2.]
3. Repeatability standard deviations [2.6.1.3.]
4. Effects of days and long-term stability [2.6.1.4.]
5. Differences among 5 probes [2.6.1.5.]
6. Run gauge study example using Dataplot™ [2.6.1.6.]
7. Dataplot™ macros [2.6.1.7.]
2. Check standard for resistivity measurements [2.6.2.]
1. Background and data [2.6.2.1.]
1. Database for resistivity check standard [2.6.2.1.1.]
2. Analysis and interpretation [2.6.2.2.]
1. Repeatability and level-2 standard deviations [2.6.2.2.1.]
3. Control chart for probe precision [2.6.2.3.]
4. Control chart for bias and long-term variability [2.6.2.4.]
5. Run check standard example yourself [2.6.2.5.]
6. Dataplot™ macros [2.6.2.6.]
3. Evaluation of type A uncertainty [2.6.3.]
1. Background and data [2.6.3.1.]
1. Database of resistivity measurements [2.6.3.1.1.]
2. Measurements on wiring configurations [2.6.3.1.2.]
2. Analysis and interpretation [2.6.3.2.]
1. Difference between 2 wiring configurations [2.6.3.2.1.]
3. Run the type A uncertainty analysis using Dataplot™ [2.6.3.3.]
4. Dataplot™ macros [2.6.3.4.]
4. Evaluation of type B uncertainty and propagation of error [2.6.4.]

7. References [2.7.]

NIST
SEMATECH

2. Measurement Process Characterization

2.1. Characterization

The primary goal of this section is to lay the groundwork for understanding the measurement process in terms of the errors that affect the process.

What are the issues for characterization?

1. Purpose
2. Reference base
3. Bias and Accuracy
4. Variability

What is a check standard?

1. Assumptions
2. Data collection
3. Analysis

2.1.1. What are the issues for characterization?

'Goodness' of measurements

A measurement process can be thought of as a well-run production process in which measurements are the output. The 'goodness' of measurements is the issue, and goodness is characterized in terms of the errors that affect the measurements.

Bias, variability and uncertainty

The goodness of measurements is quantified in terms of

- Bias
- Short-term variability or instrument precision
- Day-to-day or long-term variability
- Uncertainty

Requires ongoing statistical control program

The continuation of goodness is guaranteed by a statistical control program that controls both

- Short-term variability or instrument precision
- Long-term variability which controls bias and day-to-day variability of the process

Scope is limited to ongoing processes

The techniques in this chapter are intended primarily for ongoing processes. One-time tests and special tests or destructive tests are difficult to characterize. Examples of ongoing processes are:

- Calibration where similar test items are measured on a regular basis
- Certification where materials are characterized on a regular basis
- Production where the metrology (tool) errors may be significant
- Special studies where data can be collected over the life of the study

Application to production processes

The material in this chapter is pertinent to the study of production processes for which the size of the metrology (tool) error may be an important consideration. More specific guidance on assessing metrology errors can be found in the section on gauge studies.

2.1.1. What are the issues for characterization?

2.1.1.1. Purpose

Purpose is to understand and quantify the effect of error on reported values

The purpose of characterization is to develop an understanding of the sources of error in the measurement process and how they affect specific measurement results. This section provides the background for:

- identifying sources of error in the measurement process
- understanding and quantifying errors in the measurement process
- codifying the effects of these errors on a specific reported value in a statement of uncertainty

Important concepts

Characterization relies upon the understanding of certain underlying concepts of measurement systems; namely,

- reference base (authority) for the measurement
- bias
- variability
- check standard

Reported value is a generic term that identifies the result that is transmitted to the customer

The reported value is the measurement result for a particular test item. It can be:

- a single measurement
- an average of several measurements
- a least-squares prediction from a model
- a combination of several measurement results that are related by a physical model

2.1.1.2. Reference base

Ultimate authority

The most critical element of any measurement process is the relationship between a single measurement and the reference base for the unit of measurement. The reference base is the ultimate source of authority for the measurement unit.

For fundamental units

Reference bases for fundamental units of measurement (length, mass, temperature, voltage, and time) and some derived units (such as pressure, force, flow rate, etc.) are maintained by national and regional standards laboratories. Consensus values from interlaboratory tests or instrumentation/standards as maintained in specific environments may serve as reference bases for other units of measurement.

For comparison purposes

A reference base, for comparison purposes, may be based on an agreement among participating laboratories or organizations and derived from

- measurements made with a standard test method
- measurements derived from an interlaboratory test

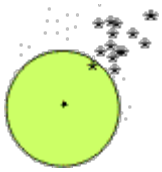
2.1.1.3. Bias and Accuracy

Definition of Accuracy and Bias

Accuracy is a qualitative term referring to whether there is agreement between a measurement made on an object and its true (target or reference) value. Bias is a quantitative term describing the difference between the average of measurements made on the same object and its true value. In particular, for a measurement laboratory, bias is the difference (generally unknown) between a laboratory's average value (over time) for a test item and the average that would be achieved by the reference laboratory if it undertook the same measurements on the same test item.

Depiction of bias and unbiased measurements

Unbiased measurements relative to the target



Biased measurements relative to the target

Identification of bias

Bias in a measurement process can be identified by:

1. Calibration of standards and/or instruments by a reference laboratory, where a value is assigned to the client's standard based on comparisons with the reference laboratory's standards.
2. Check standards , where violations of the control limits on a control chart for the check standard suggest that re-calibration of standards or instruments is needed.
3. Measurement assurance programs, where artifacts from a reference laboratory or other qualified agency are sent to a client and measured in the client's environment as a 'blind' sample.
4. Interlaboratory comparisons, where reference standards or materials are circulated among several laboratories.

Reduction of

Bias can be eliminated or reduced by calibration of

bias

standards and/or instruments. Because of costs and time constraints, the majority of calibrations are performed by secondary or tertiary laboratories and are related to the reference base via a chain of intercomparisons that start at the reference laboratory.

Bias can also be reduced by corrections to in-house measurements based on comparisons with artifacts or instruments circulated for that purpose (reference materials).

Caution

Errors that contribute to bias can be present even where all equipment and standards are properly calibrated and under control. Temperature probably has the most potential for introducing this type of bias into the measurements. For example, a constant heat source will introduce serious errors in dimensional measurements of metal objects. Temperature affects chemical and electrical measurements as well.

Generally speaking, errors of this type can be identified only by those who are thoroughly familiar with the measurement technology. The reader is advised to consult the technical literature and experts in the field for guidance.

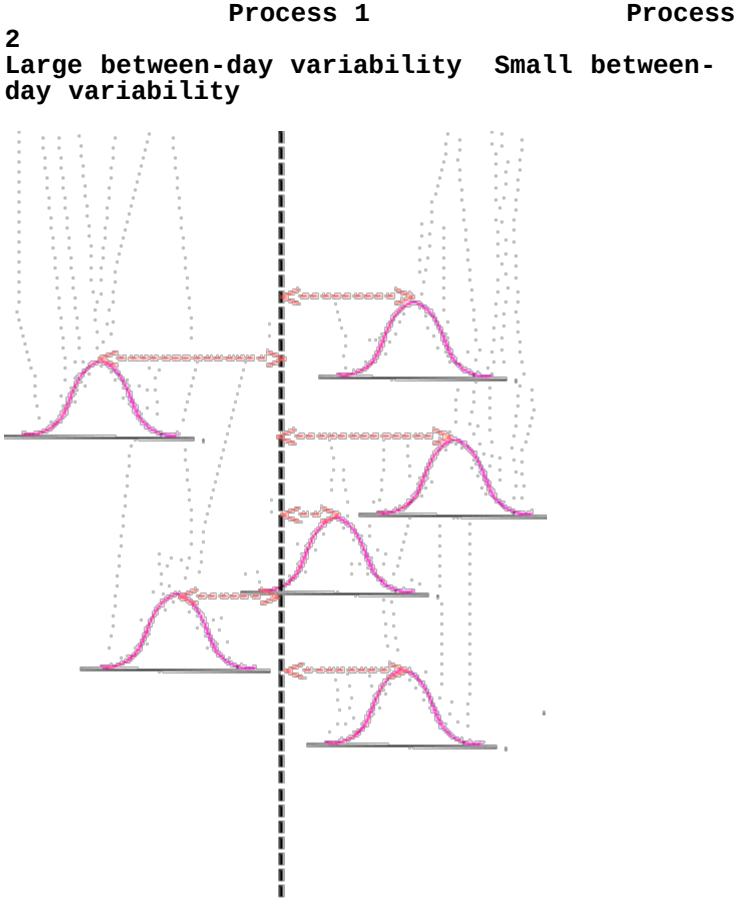
2.1.1.4. Variability

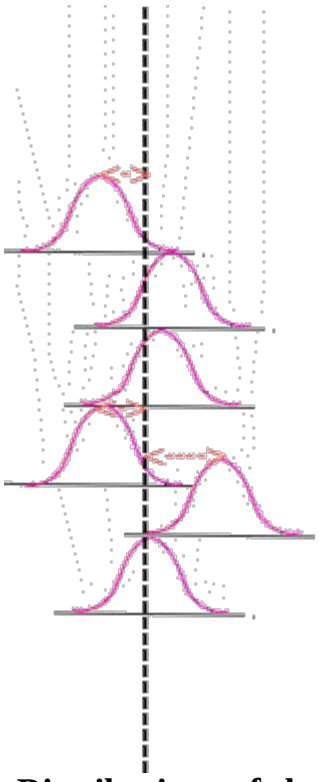
Sources of time-dependent variability

Variability is the tendency of the measurement process to produce slightly different measurements on the same test item, where conditions of measurement are either stable or vary over time, temperature, operators, etc. In this chapter we consider two sources of time-dependent variability:

- Short-term variability ascribed to the precision of the instrument
- Long-term variability related to changes in environment and handling techniques

Depiction of two measurement processes with the same short-term variability over six days where process 1 has large between-day variability and process 2 has negligible between-day variability





Distributions of short-term measurements over 6 days where distances from the centerlines illustrate between-day variability

Short-term variability

Short-term errors affect the precision of the instrument. Even very precise instruments exhibit small changes caused by random errors. It is useful to think in terms of measurements performed with a single instrument over minutes or hours; this is to be understood, normally, as the time that it takes to complete a measurement sequence.

Terminology

Four terms are in common usage to describe short-term phenomena. They are interchangeable.

1. precision
2. repeatability
3. within-time variability
4. short-term variability

Precision is quantified by a standard deviation

The measure of precision is a standard deviation. Good precision implies a small standard deviation. This standard deviation is called the short-term standard deviation of the process or the repeatability standard deviation.

Caution -- long-term variability may be dominant

With very precise instrumentation, it is not unusual to find that the variability exhibited by the measurement process from day-to-day often exceeds the precision of the instrument because of small changes in environmental conditions and handling techniques which cannot be

controlled or corrected in the measurement process. The measurement process is not completely characterized until this source of variability is quantified.

Terminology

Three terms are in common usage to describe long-term phenomena. They are interchangeable.

1. day-to-day variability
2. long-term variability
3. reproducibility

Caution -- regarding term 'reproducibility'

The term 'reproducibility' is given very specific definitions in some national and international standards. However, the definitions are not always in agreement. Therefore, it is used here only in a generic sense to indicate variability across days.

Definitions in this Handbook

We adopt precise definitions and provide data collection and analysis techniques in the sections on check standards and measurement control for estimating:

- Level-1 standard deviation for short-term variability
- Level-2 standard deviation for day-to-day variability

In the section on gauge studies, the concept of variability is extended to include very long-term measurement variability:

- Level-1 standard deviation for short-term variability
- Level-2 standard deviation for day-to-day variability
- Level-3 standard deviation for very long-term variability

We refer to the standard deviations associated with these three kinds of uncertainty as "Level 1, 2, and 3 standard deviations", respectively.

Long-term variability is quantified by a standard deviation

The measure of long-term variability is the standard deviation of measurements taken over several days, weeks or months.

The simplest method for doing this assessment is by analysis of a check standard database. The measurements on the check standards are structured to cover a long time interval and to capture all sources of variation in the measurement process.

2.1.2. What is a check standard?

A check standard is useful for gathering data on the process

Check standard methodology is a tool for collecting data on the measurement process to expose errors that afflict the process over time. Time-dependent sources of error are evaluated and quantified from the database of check standard measurements. It is a device for controlling the bias and long-term variability of the process once a baseline for these quantities has been established from historical data on the check standard.

Think in terms of data

A check standard can be an artifact or defined quantity

The check standard should be thought of in terms of a database of measurements. It can be defined as an artifact or as a characteristic of the measurement process whose value can be replicated from measurements taken over the life of the process. Examples are:

- measurements on a stable artifact
- differences between values of two reference standards as estimated from a calibration experiment
- values of a process characteristic, such as a bias term, which is estimated from measurements on reference standards and/or test items.

An artifact check standard must be close in material content and geometry to the test items that are measured in the workload. If possible, it should be one of the test items from the workload. Obviously, it should be a stable artifact and should be available to the measurement process at all times.

Solves the difficulty of sampling the process

Measurement processes are similar to production processes in that they are continual and are expected to produce identical results (within acceptable limits) over time, instruments, operators, and environmental conditions. However, it is difficult to sample the output of the measurement process because, normally, test items change with each measurement sequence.

Surrogate for unseen measurements

Measurements on the check standard, spaced over time at regular intervals, act as surrogates for measurements that could be made on test items if sufficient time and resources were available.

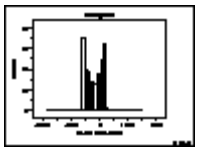
2.1.2.1. Assumptions

Case study:

Resistivity check standard

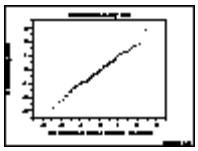
Before applying the quality control procedures recommended in this chapter to check standard data, basic assumptions should be examined. The basic assumptions underlying the quality control procedures are:

1. The data come from a single statistical distribution.
2. The distribution is a normal distribution.
3. The errors are uncorrelated over time.

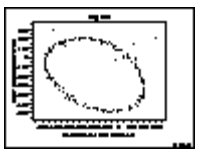


An easy method for checking the assumption of a single normal distribution is to construct a histogram of the check standard data. The histogram should follow a bell-shaped pattern with a single hump. Types of anomalies that indicate a problem with the measurement system are:

1. a double hump indicating that errors are being drawn from two or more distributions;
2. long tails indicating outliers in the process;
3. flat pattern or one with humps at either end indicating that the measurement process in not in control or not properly specified.



Another graphical method for testing the normality assumption is a probability plot. The points are expected to fall approximately on a straight line if the data come from a normal distribution. Outliers, or data from other distributions, will produce an S-shaped curve.



A graphical method for testing for correlation among measurements is a time-lag plot. Correlation will frequently not be a problem if measurements are properly structured over time. Correlation problems generally occur when measurements are taken so close together in time that the instrument cannot properly recover from one measurement to the next. Correlations over time are usually present but are often negligible.

2.1.2.2. Data collection

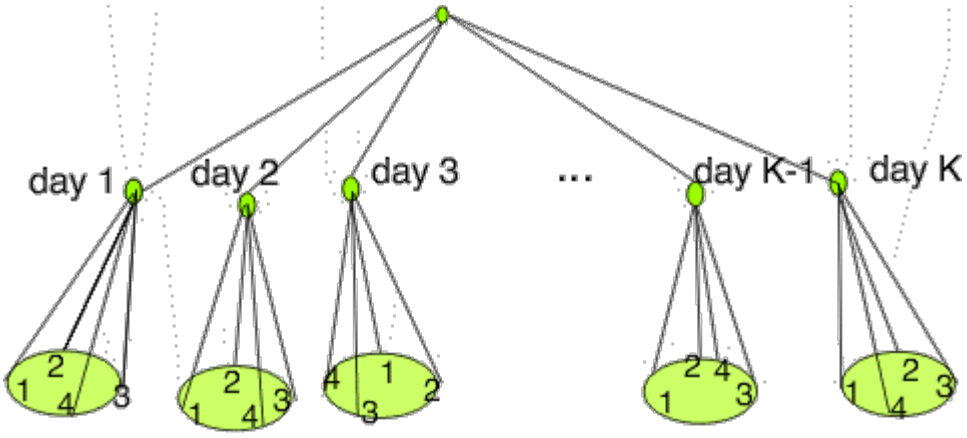
Schedule for making measurements

A schedule for making check standard measurements over time (once a day, twice a week, or whatever is appropriate for sampling all conditions of measurement) should be set up and adhered to. The check standard measurements should be structured in the same way as values reported on the test items. For example, if the reported values are averages of two repetitions made within 5 minutes of each other, the check standard values should be averages of the two measurements made in the same manner.

Exception

One exception to this rule is that there should be at least $J = 2$ repetitions per day. Without this redundancy, there is no way to check on the short-term precision of the measurement system.

Depiction of schedule for making check standard measurements with four repetitions per day over K days on the surface of a silicon wafer with the repetitions randomized at various positions on the wafer



K days - 4 repetitions

2-level design for measurement process

Case study: Resistivity check standard for measurements on silicon wafers

The values for the check standard should be recorded along with pertinent environmental readings and identifications for all other significant factors. The best way to record this information is in one file with one line or row (on a spreadsheet) of information in fixed fields for each check standard measurement. A list of typical entries follows.

1. Identification for check standard
2. Date
3. Identification for the measurement design (if applicable)
4. Identification for the instrument

5. Check standard value
6. Short-term standard deviation from J repetitions
7. Degrees of freedom
8. Operator identification
9. Environmental readings (if pertinent)

2.1.2.3. Analysis

Short-term or level-1 standard deviations from J repetitions

An analysis of the check standard data is the basis for quantifying random errors in the measurement process -- particularly time-dependent errors.

Given that we have a database of check standard measurements as described in data collection where

$$Y_{kj}\left(k=1,\cdots K;j=1,\cdots J\right)$$

represents the jth repetition on the kth day, the mean for the kth day is

$$\bar{Y}_{k\bullet}=\frac{1}{J}\sum_{j=1}^{J}Y_{kj}$$

and the short-term (level-1) standard deviation with $v = J - 1$ degrees of freedom is

$$s_k=\sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left(Y_{kj}-\bar{Y}_{k\bullet}\right)^2}.$$

Drawback of short-term standard deviations

An individual short-term standard deviation will not be a reliable estimate of precision if the degrees of freedom is less than ten, but the individual estimates can be pooled over the K days to obtain a more reliable estimate. The pooled level-1 standard deviation estimate with $v = K(J - 1)$ degrees of freedom is

$$s_1=\sqrt{\frac{1}{K}\sum_{k=1}^{K}s_k^2}.$$

This standard deviation can be interpreted as quantifying the basic precision of the instrumentation used in the measurement process.

Process (level-2)

The level-2 standard deviation of the check standard is appropriate for representing the process variability. It is

standard deviation

computed with $v = K - 1$ degrees of freedom as:

$$s_{chkstd} = s_2 = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{Y}_{k\bullet} - \bar{Y}_{\bullet\bullet}\right)^2}$$

where

$$\bar{Y}_{\bullet\bullet} = \frac{1}{K}\sum_{k=1}^{K}\bar{Y}_{k\bullet}$$

is the grand mean of the KJ check standard measurements.

Use in quality control

The check standard data and standard deviations that are described in this section are used for controlling two aspects of a measurement process:

1. Control of short-term variability
2. Control of bias and long-term variability

Case study: Resistivity check standard

For an example, see the case study for resistivity where several check standards were measured $J = 6$ times per day over several days.

2. Measurement Process Characterization

2.2. Statistical control of a measurement process

The purpose of this section is to outline the steps that can be taken to exercise statistical control over the measurement process and demonstrate the validity of the uncertainty statement. Measurement processes can change both with respect to bias and variability. A change in instrument precision may be readily noted as measurements are being recorded, but changes in bias or long-term variability are difficult to catch when the process is looking at a multitude of artifacts over time.

What are the issues for control of a measurement process?

1. Purpose
2. Assumptions
3. Role of the check standard

How are bias and long-term variability controlled?

1. Shewhart control chart
2. Exponentially weighted moving average control chart
3. Data collection and analysis
4. Control procedure
5. Remedial actions & strategies

How is short-term variability controlled?

1. Control chart for standard deviations
2. Data collection and analysis
3. Control procedure
4. Remedial actions and strategies

2.2.1. What are the issues in controlling the measurement process?

Purpose is to guarantee the 'goodness' of measurement results

The purpose of statistical control is to guarantee the 'goodness' of measurement results within predictable limits and to validate the statement of uncertainty of the measurement result.

Statistical control methods can be used to test the measurement process for change with respect to bias and variability from its historical levels. However, if the measurement process is improperly specified or calibrated, then the control procedures can only guarantee comparability among measurements.

Assumption of normality is not stringent

The assumptions that relate to measurement processes apply to statistical control; namely that the errors of measurement are uncorrelated over time and come from a population with a single distribution. The tests for control depend on the assumption that the underlying distribution is normal (Gaussian), but the test procedures are robust to slight departures from normality. Practically speaking, all that is required is that the distribution of measurements be bell-shaped and symmetric.

Check standard is mechanism for controlling the process

Measurements on a check standard provide the mechanism for controlling the measurement process.

Measurements on the check standard should produce identical results except for the effect of random errors, and tests for control are basically tests of whether or not the random errors from the process continue to be drawn from the same statistical distribution as the historical data on the check standard.

Changes that can be monitored and tested with the check standard database are:

1. Changes in bias and long-term variability
2. Changes in instrument precision or short-term variability

2.2.2. How are bias and variability controlled?

Bias and variability are controlled by monitoring measurements on a check standard over time

Bias and long-term variability are controlled by monitoring measurements on a check standard over time. A change in the measurement on the check standard that persists at a constant level over several measurement sequences indicates possible:

1. Change or damage to the reference standards
2. Change or damage to the check standard artifact
3. Procedural change that vitiates the assumptions of the measurement process

A change in the variability of the measurements on the check standard can be due to one of many causes such as:

1. Loss of environmental controls
2. Change in handling techniques
3. Severe degradation in instrumentation.

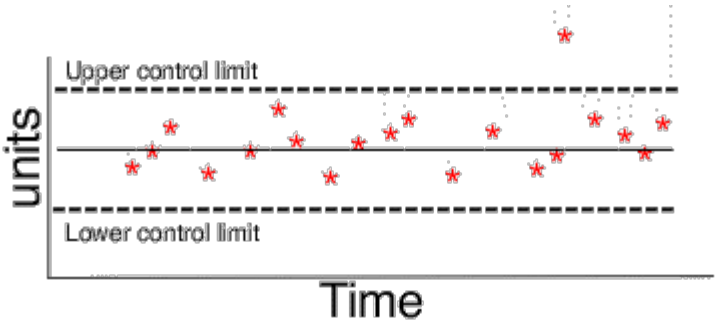
The control procedure monitors the progress of measurements on the check standard over time and signals when a significant change occurs. There are two control chart procedures that are suitable for this purpose.

Shewhart Chart is easy to implement

The Shewhart control chart has the advantage of being intuitive and easy to implement. It is characterized by a center line and symmetric upper and lower control limits. The chart is good for detecting large changes but not for quickly detecting small changes (of the order of one-half to one standard deviation) in the process.

Depiction of Shewhart control chart

In the simplistic illustration of a Shewhart control chart shown below, the measurements are within the control limits with the exception of one measurement which exceeds the upper control limit.

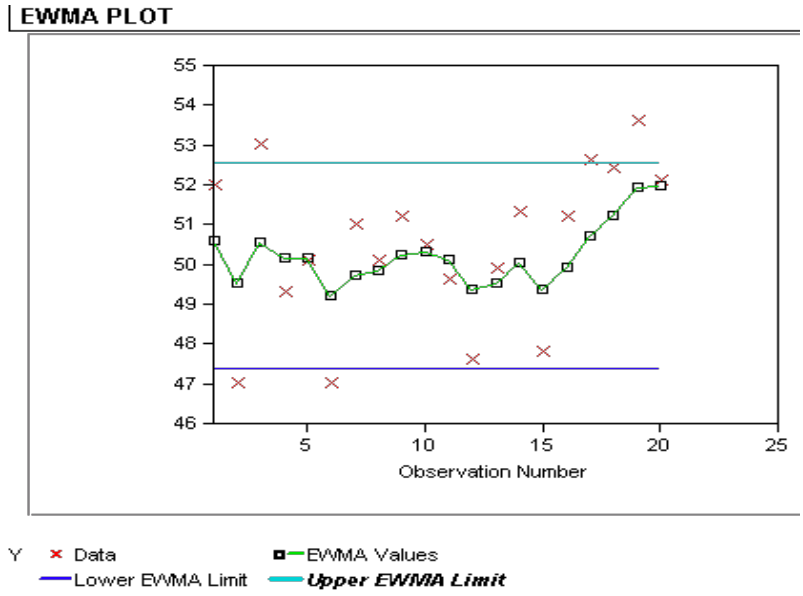


EWMA Chart is better for detecting small changes

The EWMA control chart (exponentially weighted moving average) is more difficult to implement but should be considered if the goal is quick detection of small changes. The decision process for the EWMA chart is based on an exponentially decreasing (over time) function of prior measurements on the check standard while the decision process for the Shewhart chart is based on the current measurement only.

Example of EWMA Chart

In the EWMA control chart below, the red dots represent the measurements. Control is exercised via the exponentially weighted moving average (shown as the curved line) which, in this case, is approaching its upper control limit.



Artifacts for process control must be stable and available

The check standard artifacts for controlling the bias or long-term variability of the process must be of the same type and geometry as items that are measured in the workload. The artifacts must be stable and available to the measurement process on a continuing basis. Usually, one artifact is sufficient. It can be:

Case study: Resistivity

1. An individual item drawn at random from the workload
2. A specific item reserved by the laboratory for the purpose.

Topic covered in this section>

The topics covered in this section include:

1. Shewhart control chart methodology

2. EWMA control chart methodology
3. Data collection & analysis
4. Monitoring
5. Remedies and strategies for dealing with out-of-control signals.

2.2.2.1. Shewhart control chart

Example of Shewhart control chart for mass calibrations

The Shewhart control chart has a baseline and upper and lower limits, shown as dashed lines, that are symmetric about the baseline. Measurements are plotted on the chart versus a time line. Measurements that are outside the limits are considered to be out of control.

Baseline is the average from historical data

The baseline for the control chart is the accepted value, an average of the historical check standard values. A minimum of 100 check standard values is required to establish an accepted value.

Caution - control limits are computed from the process standard deviation -- not from rational subsets

The upper (UCL) and lower (LCL) control limits are:

$$UCL = \text{Accepted value} + k * \text{process standard deviation}$$

$$LCL = \text{Accepted value} - k * \text{process standard deviation}$$

where the process standard deviation is the standard deviation computed from the check standard database.

Individual measurements cannot be assessed using the standard deviation from short-term repetitions

This procedure is an *individual observations* control chart. The previously described control charts depended on *rational subsets*, which use the standard deviations computed from the rational subsets to calculate the control limits. For a measurement process, the subgroups would consist of short-term repetitions which can characterize the precision of the instrument but not the long-term variability of the process. In measurement science, the interest is in assessing individual measurements (or averages of short-term repetitions). Thus, the standard deviation over time is the appropriate measure of variability.

Choice of k depends on number of measurements we are willing to

To achieve tight control of the measurement process, set

$$k = 2$$

in which case approximately 5% of the measurements from a process that is in control will produce out-of-control

reject signals. This assumes that there is a sufficiently large number of degrees of freedom (>100) for estimating the process standard deviation.

To flag only those measurements that are egregiously out of control, set

$$k = 3$$

in which case approximately 1% of the measurements from an in-control process will produce out-of-control signals.

2.2.2.1.1. EWMA control chart

Small changes only become obvious over time

Because it takes time for the patterns in the data to emerge, a permanent shift in the process may not immediately cause individual violations of the control limits on a Shewhart control chart. The Shewhart control chart is not powerful for detecting small changes, say of the order of 1 - 1/2 standard deviations. The EWMA (exponentially weighted moving average) control chart is better suited to this purpose.

Example of EWMA control chart for mass calibrations

The exponentially weighted moving average (EWMA) is a statistic for monitoring the process that averages the data in a way that gives less and less weight to data as they are further removed in time from the current measurement. The data

$$Y_1, Y_2, \ldots, Y_t$$

are the check standard measurements ordered in time. The EWMA statistic at time t is computed recursively from individual data points, with the first EWMA statistic, $EWMA_1$, being the arithmetic average of historical data.

$$EWMA_{t+1} = \lambda Y_t + (1-\lambda) EWMA_t$$

Control mechanism for EWMA

The EWMA control chart can be made sensitive to small changes or a gradual drift in the process by the choice of the weighting factor, λ. A weighting factor of 0.2 - 0.3 is usually suggested for this purpose (Hunter), and 0.15 is also a popular choice.

Limits for the control chart

The target or center line for the control chart is the average of historical data. The upper (UCL) and lower (LCL) limits are

$$UCL = EWMA_1 + ks\sqrt{\frac{\lambda}{(2-\lambda)}}$$

$$LCL = EWMA_1 - ks\sqrt{\frac{\lambda}{(2-\lambda)}}$$

where *s* times the radical expression is a good approximation to the standard deviation of the EWMA statistic and the factor *k* is chosen in the same way as for the Shewhart control chart -- generally to be 2 or 3.

Procedure for implementing the EWMA control chart

The implementation of the EWMA control chart is the same as for any other type of control procedure. The procedure is built on the assumption that the "good" historical data are representative of the in-control process, with future data from the same process tested for agreement with the historical data. To start the procedure, a target (average) and process standard deviation are computed from historical check standard data. Then the procedure enters the monitoring stage with the EWMA statistics computed and tested against the control limits. The EWMA statistics are weighted averages, and thus their standard deviations are smaller than the standard deviations of the raw data and the corresponding control limits are narrower than the control limits for the Shewhart individual observations chart.

2.2.2.2. Data collection

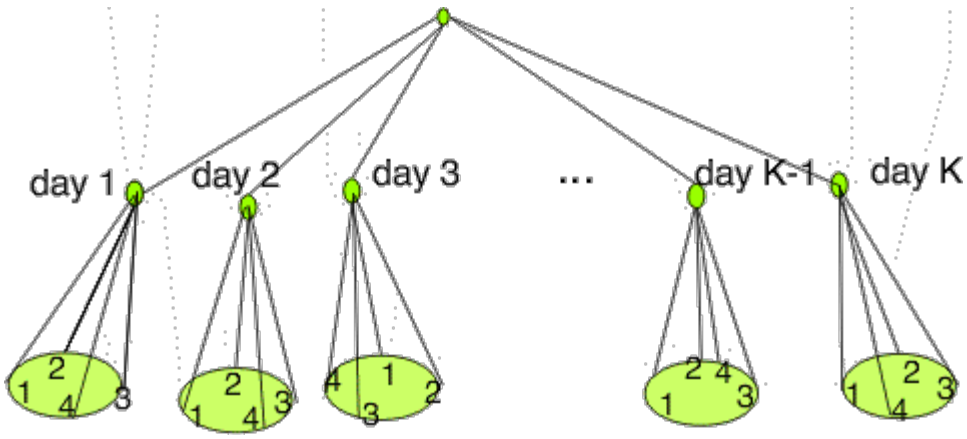
Measurements should cover a sufficiently long time period to cover all environmental conditions

A schedule should be set up for making measurements on the artifact (check standard) chosen for control purposes. The measurements are structured to sample all environmental conditions in the laboratory and all other sources of influence on the measurement result, such as operators and instruments.

For high-precision processes where the uncertainty of the result must be guaranteed, a measurement on the check standard should be included with every measurement sequence, if possible, and at least once a day.

For each occasion, *J* measurements are made on the check standard. If there is no interest in controlling the short-term variability or precision of the instrument, then one measurement is sufficient. However, a dual purpose is served by making two or three measurements that track both the bias and the short-term variability of the process with the same database.

Depiction of check standard measurements with J = 4 repetitions per day on the surface of a silicon wafer over K days where the repetitions are randomized over position on the wafer



K days - 4 repetitions

2-level design for measurements on a check standard

Notation

For *J* measurements on each of *K* days, the measurements are denoted by

$$Y_{kj}\,(k=1,\cdots,K;\ j=1,\cdots J)$$

The check standard value is

The check standard value for the *k*th day is

defined as an average of short-term repetitions

$$\bar{Y}_{k\bullet} = \frac{1}{J}\sum_{j=1}^{J} Y_{kj}$$

Accepted value of check standard

The accepted value, or baseline for the control chart, is

$$\bar{Y}_{\bullet\bullet} = \frac{1}{K}\sum_{k=1}^{K} \bar{Y}_{k\bullet}$$

Process standard deviation

The process standard deviation is

$$s_2 = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{Y}_{k\bullet} - \bar{Y}_{\bullet\bullet}\right)^2}$$

Caution

Check standard measurements should be structured in the same way as values reported on the test items. For example, if the reported values are averages of two measurements made within 5 minutes of each other, the check standard values should be averages of the two measurements made in the same manner.

Database

Case study: Resistivity

Averages and short-term standard deviations computed from *J* repetitions should be recorded in a file along with identifications for all significant factors. The best way to record this information is to use one file with one line (row in a spreadsheet) of information in fixed fields for each group. A list of typical entries follows:

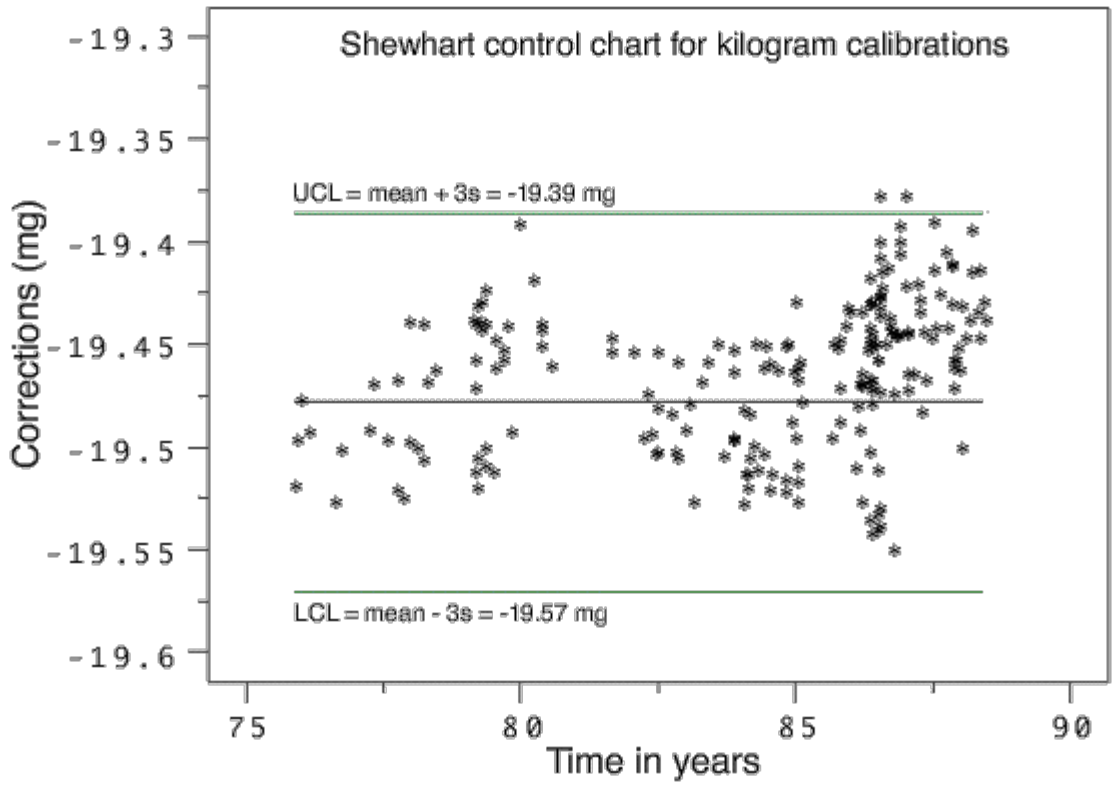
1. Month
2. Day
3. Year
4. Check standard identification
5. Identification for the measurement design (if applicable)
6. Instrument identification
7. Check standard value
8. Repeatability (short-term) standard deviation from *J* repetitions
9. Degrees of freedom
10. Operator identification
11. Environmental readings (if pertinent)

2.2.2.3. Monitoring bias and long-term variability

Monitoring stage

Once the baseline and control limits for the control chart have been determined from historical data, and any bad observations removed and the control limits recomputed, the measurement process enters the monitoring stage. A Shewhart control chart and EWMA control chart for monitoring a mass calibration process are shown below. For the purpose of comparing the two techniques, the two control charts are based on the same data where the baseline and control limits are computed from the data taken prior to 1985. The monitoring stage begins at the start of 1985. Similarly, the control limits for both charts are 3-standard deviation limits. The check standard data and analysis are explained more fully in another section.

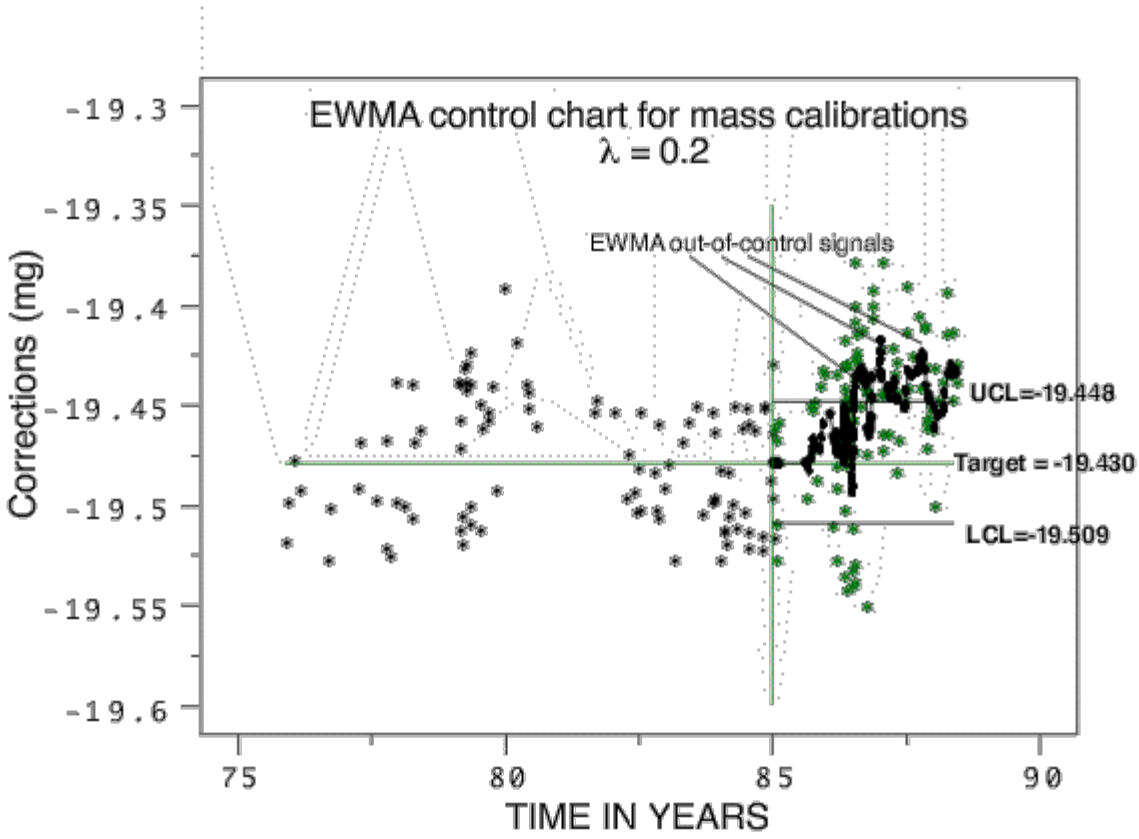
Shewhart control chart of measurements of kilogram check standard showing outliers and a shift in the process that occurred after 1985



EWMA chart for measurements on kilogram check standard

In the EWMA control chart below, the control data after 1985 are shown in green, and the EWMA statistics are shown as black dots superimposed on the raw data. The EWMA statistics, and not the raw data, are of interest in looking for out-of-control signals. Because the EWMA statistic is a weighted average, it has a smaller standard deviation than a single control measurement, and, therefore, the EWMA control limits are narrower than the limits for the Shewhart control chart shown above.

showing multiple violations of the control limits for the EWMA statistics



Measurements that exceed the control limits require action

The control strategy is based on the predictability of future measurements from historical data. Each new check standard measurement is plotted on the control chart in real time. These values are expected to fall within the control limits if the process has not changed. Measurements that exceed the control limits are probably out-of-control and require remedial action. Possible causes of out-of-control signals need to be understood when developing strategies for dealing with outliers.

Signs of significant trends or shifts

The control chart should be viewed in its entirety on a regular basis] to identify drift or shift in the process. In the Shewhart control chart shown above, only a few points exceed the control limits. The small, but significant, shift in the process that occurred after 1985 can only be identified by examining the plot of control measurements over time. A re-analysis of the kilogram check standard data shows that the control limits for the Shewhart control chart should be updated based on the the data after 1985. In the EWMA control chart, multiple violations of the control limits occur after 1986. In the calibration environment, the incidence of several violations should alert the control engineer that a shift in the process has occurred, possibly because of damage or change in the value of a reference standard, and the process requires review.

2.2.2.4. Remedial actions

Consider possible causes for out-of-control signals and take corrective long-term actions

There are many possible causes of out-of-control signals.

A. Causes that do not warrant corrective action for the process (but which do require that the current measurement be discarded) are:

1. Chance failure where the process is actually in-control
2. Glitch in setting up or operating the measurement process
3. Error in recording of data

B. Changes in bias can be due to:

1. Damage to artifacts
2. Degradation in artifacts (wear or build-up of dirt and mineral deposits)

C. Changes in long-term variability can be due to:

1. Degradation in the instrumentation
2. Changes in environmental conditions
3. Effect of a new or inexperienced operator

4-step strategy for short-term

An immediate strategy for dealing with out-of-control signals associated with high precision measurement processes should be pursued as follows:

Repeat measurements

1. Repeat the measurement sequence to establish whether or not the out-of-control signal was simply a chance occurrence, glitch, or whether it flagged a permanent change or trend in the process.

Discard measurements on test items

2. With high precision processes, for which a check standard is measured along with the test items, new values should be assigned to the test items based on new measurement data.

Check for

3. Examine the patterns of recent data. If the process is

drift

gradually drifting out of control because of degradation in instrumentation or artifacts, then:

- Instruments may need to be repaired
- Reference artifacts may need to be recalibrated.

Reevaluate

4. Reestablish the process value and control limits from more recent data if the measurement process cannot be brought back into control.

2.2.3. How is short-term variability controlled?

Emphasis on instruments

Short-term variability or instrument precision is controlled by monitoring standard deviations from repeated measurements on the instrument(s) of interest. The database can come from measurements on a single artifact or a representative set of artifacts.

Artifacts - Case study: Resistivity

The artifacts must be of the same type and geometry as items that are measured in the workload, such as:

1. Items from the workload
2. A single check standard chosen for this purpose
3. A collection of artifacts set aside for this specific purpose

Concepts covered in this section

The concepts that are covered in this section include:

1. Control chart methodology for standard deviations
2. Data collection and analysis
3. Monitoring
4. Remedies and strategies for dealing with out-of-control signals

2.2.3.1. Control chart for standard deviations

Degradation of instrument or anomalous behavior on one occasion

Changes in the precision of the instrument, particularly anomalies and degradation, must be addressed. Changes in precision can be detected by a statistical control procedure based on the *F*-distribution where the short-term standard deviations are plotted on the control chart.

The base line for this type of control chart is the pooled standard deviation, s_1, as defined in Data collection and analysis.

Example of control chart for a mass balance

Only the upper control limit, *UCL*, is of interest for detecting degradation in the instrument. As long as the short-term standard deviations fall within the upper control limit established from historical data, there is reason for confidence that the precision of the instrument has not degraded (i.e., common cause variations).

The control limit is based on the F-distribution

The control limit is

$$UCL = s_1\sqrt{F_{\alpha, J-1, K(J-1)}}$$

where the quantity under the radical is the upper α critical value from the *F* table with degrees of freedom *(J - 1)* and *K(J - 1)*. The numerator degrees of freedom, *v1 = (J -1)*, are associated with the standard deviation computed from the current measurements, and the denominator degrees of freedom, *v2 = K(J -1)*, correspond to the pooled standard deviation of the historical data. The probability α is chosen to be small, say 0.05.

The justification for this control limit, as opposed to the more conventional standard deviation control limit, is that we are essentially performing the following hypothesis test:

$H_0: \sigma_1 = \sigma_2$
$H_a: \sigma_2 > \sigma_1$

where σ_1 is the population value for the s_1 defined above and σ_2 is the population value for the standard deviation of the current values being tested. Generally, s_1 is based on

sufficient historical data that it is reasonable to make the assumption that σ_1 is a "known" value.

The upper control limit above is then derived based on the standard *F* test for equal standard deviations. Justification and details of this derivation are given in Cameron and Hailes (1974).

Sample Code

Sample code for computing the *F* value for the case where $\alpha = 0.05$, $J = 6$, and $K = 6$, is available for both Dataplot and R.

2.2.3.2. Data collection

Case study: Resistivity

A schedule should be set up for making measurements with a single instrument (once a day, twice a week, or whatever is appropriate for sampling all conditions of measurement).

Short-term standard deviations

The measurements are denoted

$$Y_{kj}\left(k=1,\cdots,K;j=1,\cdots J\right)$$

where there are J measurements on each of K occasions. The average for the kth occasion is:

$$\bar{Y}_{k\bullet}=\frac{1}{J}\sum_{j=1}^{J}Y_{kj}$$

The short-term (repeatability) standard deviation for the kth occasion is:

$$s_1=\sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left(Y_{kj}-\bar{Y}_{k\bullet}\right)^2}$$

with *(J-1)* degrees of freedom.

Pooled standard deviation

The repeatability standard deviations are pooled over the K occasions to obtain an estimate with $K(J-1)$ degrees of freedom of the level-1 standard deviation

$$s_1=\sqrt{\frac{1}{K}\sum_{k=1}^{K}s_{1k}^2}$$

Note: The same notation is used for the repeatability standard deviation whether it is based on one set of measurements or pooled over several sets.

Database

The individual short-term standard deviations along with identifications for all significant factors are recorded in a file. The best way to record this information is by using one file with one line (row in a spreadsheet) of information in fixed

fields for each group. A list of typical entries follows.

1. Identification of test item or check standard
2. Date
3. Short-term standard deviation
4. Degrees of freedom
5. Instrument
6. Operator

2.2.3.3. Monitoring short-term precision

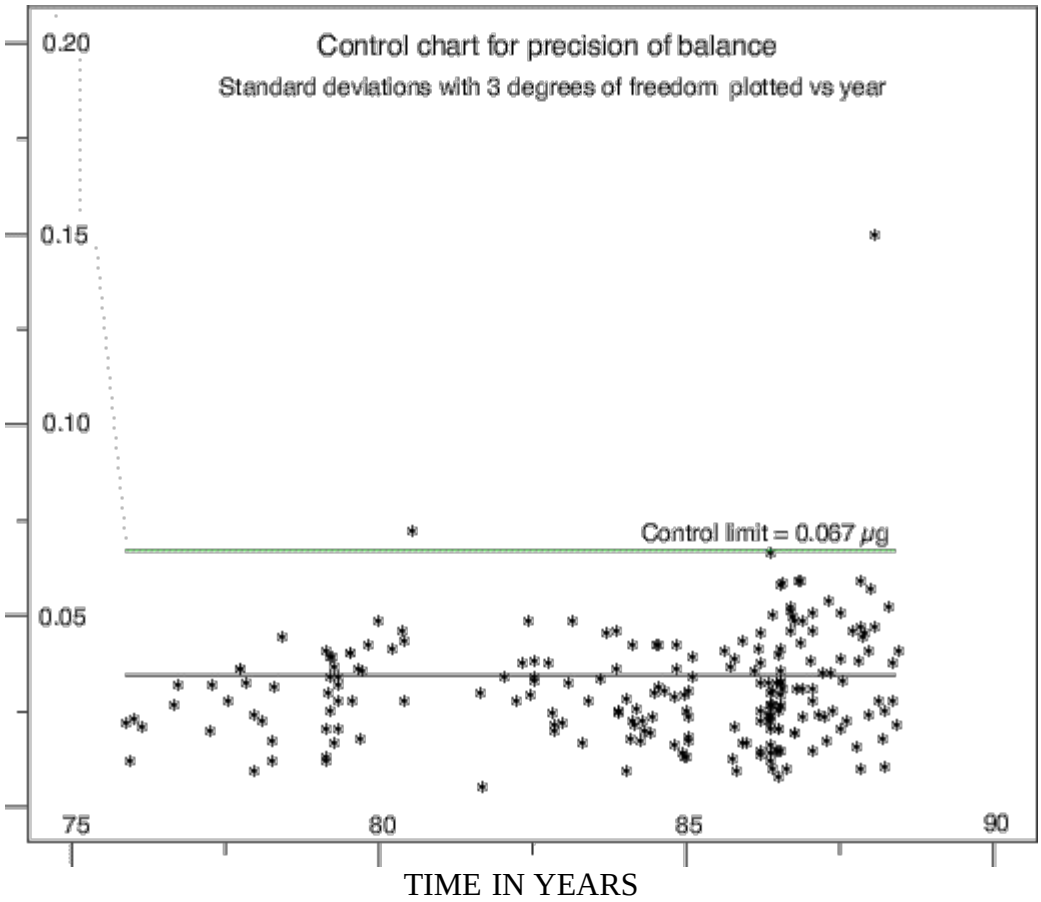
Monitoring future precision

Once the base line and control limit for the control chart have been determined from historical data, the measurement process enters the monitoring stage. In the control chart shown below, the control limit is based on the data taken prior to 1985.

Each new standard deviation is monitored on the control chart

Each new short-term standard deviation based on *J* measurements is plotted on the control chart; points that exceed the control limits probably indicate lack of statistical control. Drift over time indicates degradation of the instrument. Points out of control require remedial action, and possible causes of out of control signals need to be understood when developing strategies for dealing with outliers.

Control chart for precision for a mass balance from historical standard deviations for the balance with 3 degrees of freedom each. The control chart identifies two outliers and slight degradation over time in the precision of the balance



Monitoring where the number of

There is no requirement that future standard deviations be based on *J*, the number of measurements in the historical database. However, a change in the number of measurements leads to a change in the test for control, and it may

measurements are different from J

not be convenient to draw a control chart where the control limits are changing with each new measurement sequence.

For a new standard deviation based on *J'* measurements, the precision of the instrument is in control if

$$s_{new} < s_1 \sqrt{F_{\alpha, J'-1, K(J-1)}}$$

.

Notice that the numerator degrees of freedom, *v1 = J'- 1*, changes but the denominator degrees of freedom, *v2 = K(J - 1)*, remains the same.

2.2.3.4. Remedial actions

Examine possible causes

A. Causes that do not warrant corrective action (but which do require that the current measurement be discarded) are:

1. Chance failure where the precision is actually in control
2. Glitch in setting up or operating the measurement process
3. Error in recording of data

B. Changes in instrument performance can be due to:

1. Degradation in electronics or mechanical components
2. Changes in environmental conditions
3. Effect of a new or inexperienced operator

Repeat measurements

Repeat the measurement sequence to establish whether or not the out-of-control signal was simply a chance occurrence, glitch, or whether it flagged a permanent change or trend in the process.

Assign new value to test item

With high precision processes, for which the uncertainty must be guaranteed, new values should be assigned to the test items based on new measurement data.

Check for degradation

Examine the patterns of recent standard deviations. If the process is gradually drifting out of control because of degradation in instrumentation or artifacts, instruments may need to be repaired or replaced.

2.3. Calibration

The purpose of this section is to outline the procedures for calibrating artifacts and instruments while guaranteeing the 'goodness' of the calibration results. Calibration is a measurement process that assigns values to the property of an artifact or to the response of an instrument relative to reference standards or to a designated measurement process. The purpose of calibration is to eliminate or reduce bias in the user's measurement system relative to the reference base. The calibration procedure compares an "unknown" or test item(s) or instrument with reference standards according to a specific algorithm.

What are the issues for calibration?

1. Artifact or instrument calibration
2. Reference base
3. Reference standard(s)

What is artifact (single-point) calibration?

1. Purpose
2. Assumptions
3. Bias
4. Calibration model

What are calibration designs?

1. Purpose
2. Assumptions
3. Properties of designs
4. Restraint
5. Check standard in a design
6. Special types of bias (left-right effect & linear drift)
7. Solutions to calibration designs
8. Uncertainty of calibrated values

Catalog of calibration designs

1. Mass weights
2. Gage blocks
3. Electrical standards - saturated standard cells, zeners, resistors
4. Roundness standards
5. Angle blocks

6. Indexing tables
7. Humidity cylinders

Control of artifact calibration

1. Control of the precision of the calibrating instrument
2. Control of bias and long-term variability

What is instrument calibration over a regime?

1. Models for instrument calibration
2. Data collection
3. Assumptions
4. What can go wrong with the calibration procedure?
5. Data analysis and model validation
6. Calibration of future measurements
7. Uncertainties of calibrated values
    1. From propagation of error for a quadratic calibration
    2. From check standard measurements for a linear calibration
    3. Comparison of check standard technique and propagation of error

Control of instrument calibration

1. Control chart for linear calibration
2. Critical values of t* statistic

2.3.1. Issues in calibration

Calibration reduces bias

Calibration is a measurement process that assigns values to the property of an artifact or to the response of an instrument relative to reference standards or to a designated measurement process. The purpose of calibration is to eliminate or reduce bias in the user's measurement system relative to the reference base.

Artifact & instrument calibration

The calibration procedure compares an "unknown" or test item(s) or instrument with reference standards according to a specific algorithm. Two general types of calibration are considered in this Handbook:

- artifact calibration at a single point
- instrument calibration over a regime

Types of calibration not discussed

The procedures in this Handbook are appropriate for calibrations at secondary or lower levels of the traceability chain where reference standards for the unit already exist. Calibration from first principles of physics and reciprocity calibration are not discussed.

2.3.1.1. Reference base

Ultimate authority

The most critical element of any measurement process is the relationship between a single measurement and the reference base for the unit of measurement. The reference base is the ultimate source of authority for the measurement unit.

Base and derived units of measurement

The base units of measurement in the Le Systeme International d'Unites (SI) are (Taylor):

- kilogram - mass
- meter - length
- second - time
- ampere - electric current
- kelvin - thermodynamic temperature
- mole - amount of substance
- candela - luminous intensity

These units are maintained by the Bureau International des Poids et Mesures in Paris. Local reference bases for these units and SI derived units such as:

- pascal - pressure
- newton - force
- hertz - frequency
- ohm - resistance
- degrees Celsius - Celsius temperature, etc.

are maintained by national and regional standards laboratories.

Other sources

Consensus values from interlaboratory tests or instrumentation/standards as maintained in specific environments may serve as reference bases for other units of measurement.

2.3.1.2. Reference standards

Primary reference standards

A reference standard for a unit of measurement is an artifact that embodies the quantity of interest in a way that ties its value to the reference base.

At the highest level, a primary reference standard is assigned a value by direct comparison with the reference base. Mass is the only unit of measurement that is defined by an artifact. The kilogram is defined as the mass of a platinum-iridium kilogram that is maintained by the Bureau International des Poids et Mesures in Sevres, France.

Primary reference standards for other units come from realizations of the units embodied in artifact standards. For example, the reference base for length is the meter which is defined as the length of the path by light in vacuum during a time interval of 1/299,792,458 of a second.

Secondary reference standards

Secondary reference standards are calibrated by comparing with primary standards using a high precision comparator and making appropriate corrections for non-ideal conditions of measurement.

Secondary reference standards for mass are stainless steel kilograms, which are calibrated by comparing with a primary standard on a high precision balance and correcting for the buoyancy of air. In turn these weights become the reference standards for assigning values to test weights.

Secondary reference standards for length are gage blocks, which are calibrated by comparing with primary gage block standards on a mechanical comparator and correcting for temperature. In turn, these gage blocks become the reference standards for assigning values to test sets of gage blocks.

2.3.2. What is artifact (single-point) calibration?

Purpose

Artifact calibration is a measurement process that assigns values to the property of an artifact relative to a reference standard(s). The purpose of calibration is to eliminate or reduce bias in the user's measurement system relative to the reference base.

The calibration procedure compares an "unknown" or test item(s) with a reference standard(s) of the same nominal value (hence, the term single-point calibration) according to a specific algorithm called a calibration design.

Assumptions

The calibration procedure is based on the assumption that individual readings on test items and reference standards are subject to:

- Bias that is a function of the measuring system or instrument
- Random error that may be uncontrollable

What is bias?

The operational definition of bias is that it is the difference between values that would be assigned to an artifact by the client laboratory and the laboratory maintaining the reference standards. Values, in this sense, are understood to be the long-term averages that would be achieved in both laboratories.

Calibration model for eliminating bias requires a reference standard that is very close in value to the test item

One approach to eliminating bias is to select a reference standard that is almost identical to the test item; measure the two artifacts with a comparator type of instrument; and take the difference of the two measurements to cancel the bias. The only requirement on the instrument is that it be linear over the small range needed for the two artifacts.

The test item has value X^*, as yet to be assigned, and the reference standard has an assigned value R^*. Given a measurement, X, on the test item and a measurement, R, on the reference standard,

$$X = Bias + X^* + error_1$$
$$R = Bias + R^* + error_2,$$

the difference between the test item and the reference is estimated by

$$D = X - R,$$

and the value of the test item is reported as

$$\hat{Test} = X^* = D + R^*.$$

Need for redundancy leads to calibration designs

A deficiency in relying on a single difference to estimate D is that there is no way of assessing the effect of random errors. The obvious solution is to:

- Repeat the calibration measurements J times
- Average the results
- Compute a standard deviation from the J results

Schedules of redundant intercomparisons involving measurements on several reference standards and test items in a connected sequence are called calibration designs and are discussed in later sections.

2.3.3. What are calibration designs?

Calibration designs are redundant schemes for intercomparing reference standards and test items

Calibration designs are redundant schemes for intercomparing reference standards and test items in such a way that the values can be assigned to the test items based on known values of reference standards. Artifacts that traditionally have been calibrated using calibration designs are:

- mass weights
- resistors
- voltage standards
- length standards
- angle blocks
- indexing tables
- liquid-in-glass thermometers, etc.

Outline of section

The topics covered in this section are:

- Designs for elimination of left-right bias and linear drift
- Solutions to calibration designs
- Uncertainties of calibrated values

A catalog of calibration designs is provided in the next section.

Assumptions for calibration designs include demands on the quality of the artifacts

The assumptions that are necessary for working with calibration designs are that:

- Random errors associated with the measurements are independent.
- All measurements come from a distribution with the same standard deviation.
- Reference standards and test items respond to the measuring environment in the same manner.
- Handling procedures are consistent from item to item.
- Reference standards and test items are stable during the time of measurement.
- Bias is canceled by taking the difference between measurements on the test item and the reference standard.

Important concept - Restraint

The restraint is the known value of the reference standard or, for designs with two or more reference standards, the restraint is the summation of the values of the reference standards.

Requirements & properties of designs

Basic requirements are:

- The differences must be nominally zero.
- The design must be solvable for individual items given the restraint.

It is possible to construct designs which do not have these properties. This will happen, for example, if reference standards are only compared among themselves and test items are only compared among themselves without any intercomparisons.

Practical considerations determine a 'good' design

We do not apply 'optimality' criteria in constructing calibration designs because the construction of a 'good' design depends on many factors, such as convenience in manipulating the test items, time, expense, and the maximum load of the instrument.

- The number of measurements should be small.
- The degrees of freedom should be greater than three.
- The standard deviations of the estimates for the test items should be small enough for their intended purpose.

Check standard in a design

Designs listed in this Handbook have provision for a check standard in each series of measurements. The check standard is usually an artifact, of the same nominal size, type, and quality as the items to be calibrated. Check standards are used for:

- Controlling the calibration process
- Quantifying the uncertainty of calibrated results

Estimates that can be computed from a design

Calibration designs are solved by a restrained least-squares technique (Zelen) which gives the following estimates:

- Values for individual reference standards
- Values for individual test items
- Value for the check standard
- Repeatability standard deviation and degrees of freedom
- Standard deviations associated with values for reference standards and test items

2.3.3.1. Elimination of special types of bias

Assumptions which may be violated

Two of the usual assumptions relating to calibration measurements are not always valid and result in biases. These assumptions are:

- Bias is canceled by taking the difference between the measurement on the test item and the measurement on the reference standard
- Reference standards and test items remain stable throughout the measurement sequence

Ideal situation

In the ideal situation, bias is eliminated by taking the difference between a measurement X on the test item and a measurement R on the reference standard. However, there are situations where the ideal is not satisfied:

- Left-right (or constant instrument) bias
- Bias caused by instrument drift

2. Measurement Process Characterization
2.3. Calibration
2.3.3. What are calibration designs?
2.3.3.1. Elimination of special types of bias

2.3.3.1.1. Left-right (constant instrument) bias

Left-right bias *which is not eliminated by differencing*

A situation can exist in which a bias, P, which is constant and independent of the direction of measurement, is introduced by the measurement instrument itself. This type of bias, which has been observed in measurements of standard voltage cells (Eicke & Cameron) and is not eliminated by reversing the direction of the current, is shown in the following equations.

$$Y_1 = X - R + P + error_1$$
$$Y_2 = R - X + P + error_2$$

Elimination of left-right bias requires two measurements in reverse direction

The difference between the test and the reference can be estimated without bias only by taking the difference between the two measurements shown above where P cancels in the differencing so that

$$D = Y_1 - Y_2 = 2X - 2R.$$

*The value of the test item depends on the known value of the reference standard, R**

The test item, X, can then be estimated without bias by

$$\hat{Test} = X^* = \frac{1}{2}(Y_1 - Y_2) + R^*$$

and P can be estimated by

$$\hat{P} = \frac{1}{2}(Y_1 + Y_2).$$

Calibration designs that are left-right balanced

This type of scheme is called left-right balanced and the principle is extended to create a catalog of left-right balanced designs for intercomparing reference standards among themselves. These designs are appropriate *ONLY* for comparing reference standards in the same environment, or enclosure, and are not appropriate for comparing, say, across standard voltage cells in two boxes.

1. Left-right balanced design for a group of 3 artifacts
2. Left-right balanced design for a group of 4 artifacts
3. Left-right balanced design for a group of 5 artifacts
4. Left-right balanced design for a group of 6 artifacts

2. Measurement Process Characterization
2.3. Calibration
2.3.3. What are calibration designs?
2.3.3.1. Elimination of special types of bias

2.3.3.1.2. Bias caused by instrument drift

Bias caused by linear drift *over the time of measurement*

The requirement that reference standards and test items be stable during the time of measurement cannot always be met because of changes in temperature caused by body heat, handling, etc.

Representation of linear drift

Linear drift for an even number of measurements is represented by

..., -5d, -3d, -1d, +1d, +3d, +5d, ...

and for an odd number of measurements by

..., -3d, -2d, -1d, 0d, +1d, +2d, +3d, ...

Assumptions for drift elimination

The effect can be mitigated by a drift-elimination scheme (Cameron/Hailes) which assumes:

- Linear drift over time
- Equally spaced measurements in time

Example of drift-elimination scheme

An example is given by substitution weighing where scale deflections on a balance are observed for X, a test weight, and R, a reference weight.

$$Y_1 = X - 3d_1 + error_1$$
$$Y_2 = R - 1d_2 + error_2$$
$$Y_3 = R + 1d_3 + error_3$$
$$Y_4 = X + 3d_4 + error_4$$

Estimates of drift-free difference and size of drift

The drift-free difference between the test and the reference is estimated by

$$D = \frac{1}{2}\{(Y_1 - Y_2) - (Y_3 - Y_4)\}$$

and the size of the drift is estimated by

$$\hat{d} = \frac{1}{4}\{-Y_1 + Y_2 - Y_3 + Y_4\}$$

Calibration designs for eliminating linear drift

This principle is extended to create a catalog of drift-elimination designs for multiple reference standards and test items. These designs are listed under calibration designs for gauge blocks because they have traditionally been used to counteract the effect of temperature build-up in the comparator during calibration.

2.3.3.2. Solutions to calibration designs

Solutions for designs listed in the catalog

Solutions for all designs that are cataloged in this Handbook are included with the designs. Solutions for other designs can be computed from the instructions on the following page given some familiarity with matrices.

Measurements for the 1,1,1 design

The use of the tables shown in the catalog are illustrated for three artifacts; namely, a reference standard with known value R^* and a check standard and a test item with unknown values. All artifacts are of the same nominal size. The design is referred to as a 1,1,1 design for

- n = 3 difference measurements
- m = 3 artifacts

Convention for showing the measurement sequence and identifying the reference and check standards

The convention for showing the measurement sequence is shown below. Nominal values are underlined in the first line showing that this design is appropriate for comparing three items of the same nominal size such as three one-kilogram weights. The reference standard is the first artifact, the check standard is the second, and the test item is the third.

	1	1	1
Y(1) =	+	-	
Y(2) =	+		-
Y(3) =		+	-
Restraint	+		
Check standard		+	

Limitation of this design

This design has degrees of freedom

$$v = n - m + 1 = 1$$

Convention for showing least-squares estimates for individual items

The table shown below lists the coefficients for finding the estimates for the individual items. The estimates are computed by taking the cross-product of the appropriate column for the item of interest with the column of measurement data and dividing by the divisor shown at the top of the table.

```
                    SOLUTION MATRIX
                      DIVISOR = 3

OBSERVATIONS      1      1      1

   Y(1)           0     -2     -1
   Y(2)           0     -1     -2
   Y(3)           0      1     -1
   R*             3      3      3
```

Solutions for individual items from the table above

For example, the solution for the reference standard is shown under the first column; for the check standard under the second column; and for the test item under the third column. Notice that the estimate for the reference standard is guaranteed to be R*, regardless of the measurement results, because of the restraint that is imposed on the design. The estimates are as follows:

$$\hat{R^*} = \frac{1}{3}(0 \cdot Y_1 + 0 \cdot Y_2 + 0 \cdot Y_3) + R^*$$

$$\hat{Chk} = \frac{1}{3}(-2 \cdot Y_1 - 1 \cdot Y_2 + 1 \cdot Y_3) + R^*$$

$$\hat{Test} = \frac{1}{3}(-1 \cdot Y_1 - 2 \cdot Y_2 - 1 \cdot Y_3) + R^*$$

Convention for showing standard deviations for individual items and combinations of items

The standard deviations are computed from two tables of factors as shown below. The standard deviations for combinations of items include appropriate covariance terms.

```
FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT  FACTOR
      K1       1   1   1
 1  0.0000    +
 1  0.8165        +
 1  0.8165            +
 2  1.4142        +   +
 1  0.8165        +

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT  FACTOR

      K2       1   1   1
 1  0.0000    +
 1  1.4142        +
 1  1.4142            +
 2  2.4495        +   +
 1  1.4142        +
```

Unifying equation

The standard deviation for each item is computed using the unifying equation:

$$s_{test} = \sqrt{K_1^2 s_1^2 + K_2^2 s_{days}^2}$$

Standard deviations for 1,1,1 design from the

For the 1,1,1 design, the standard deviations are:

$$s_{R^*} = 0$$

tables of factors

$$s_{chk} = \sqrt{(0.8165\ s_1)^2 + (1.4142\ s_{days})^2} = \sqrt{\frac{2}{3}s_1^2 + 2s_{days}^2}$$

$$s_{test} = \sqrt{(0.8165\ s_1)^2 + (1.4142\ s_{days})^2} = \sqrt{\frac{2}{3}s_1^2 + 2s_{days}^2}$$

$$s_{chk+test} = \sqrt{(1.4142\ s_1)^2 + (2.4495\ s_{days})^2} = \sqrt{2s_1^2 + 6s_{days}^2}$$

Process standard deviations must be known from historical data

In order to apply these equations, we need an estimate of the standard deviation, s_{days}, that describes day-to-day changes in the measurement process. This standard deviation is in turn derived from the level-2 standard deviation, s_2, for the check standard. This standard deviation is estimated from historical data on the check standard; it can be negligible, in which case the calculations are simplified.

The repeatability standard deviation s_1, is estimated from historical data, usually from data of several designs.

Steps in computing standard deviations

The steps in computing the standard deviation for a test item are:

- Compute the repeatability standard deviation from the design or historical data.
- Compute the standard deviation of the check standard from historical data.
- Locate the factors, K_1 and K_2 for the **check standard**; for the 1,1,1 design the factors are 0.8165 and 1.4142, respectively, where the check standard entries are last in the tables.
- Apply the unifying equation to the check standard to estimate the standard deviation for days. Notice that the standard deviation of the check standard is the same as the level-2 standard deviation, s_2, that is referred to on some pages. The equation for the between-days standard deviation from the unifying equation is

  $$s_{days} = \frac{1}{K_2}\sqrt{s_2^2 - K_1^2 s_1^2}$$

  .

  Thus, for the example above

  $$s_{days} = \frac{1}{\sqrt{2}}\sqrt{s_2^2 - \frac{2}{3}s_1^2}$$

  .
- This is the number that is entered into the NIST mass calibration software as the between-time standard deviation. If you are using this software, this is the only computation that you need to make because the standard deviations for the test items are computed automatically by the software.

- If the computation under the radical sign gives a negative number, set $s_{days}=0$. (This is possible and indicates that there is no contribution to uncertainty from day-to-day effects.)

- For completeness, the computations of the standard deviations for the test item and for the sum of the test and the check standard using the appropriate factors are shown below.

$$s_{test} = \sqrt{\frac{2}{3}s_1^2 + 2s_{days}^2} = \sqrt{\frac{2}{3}s_1^2 + 2\frac{1}{2}\left(s_2^2 - \frac{2}{3}s_1^2\right)} = s_2$$

$$s_{test+chk} = \sqrt{2s_1^2 + 6s_{days}^2} = \sqrt{2s_1^2 + 6\frac{1}{2}\left(s_2^2 - \frac{2}{3}s_1^2\right)} = \sqrt{3}s_2$$

2.3.3.2.1. General matrix solutions to calibration designs

Requirements

Solutions for all designs that are cataloged in this Handbook are included with the designs. Solutions for other designs can be computed from the instructions below given some familiarity with matrices. The matrix manipulations that are required for the calculations are:

- transposition (indicated by ')
- multiplication
- inversion

Notation

- n = number of difference measurements
- m = number of artifacts
- (n - m + 1) = degrees of freedom
- X= (nxm) design matrix
- r'= (mx1) vector identifying the restraint
- v_i' = (mx1) vector identifying ith item of interest consisting of a 1 in the ith position and zeros elsewhere
- R*= value of the reference standard
- Y= (mx1) vector of observed difference measurements

Convention for showing the measurement sequence

The convention for showing the measurement sequence is illustrated with the three measurements that make up a 1,1,1 design for 1 reference standard, 1 check standard, and 1 test item. Nominal values are underlined in the first line .

	1	1	1
Y(1) =	+	-	
Y(2) =	+		-
Y(3) =		+	-

Matrix algebra for solving a design

The (mxn) design matrix X is constructed by replacing the pluses (+), minues (-) and blanks with the entries 1, -1, and 0 respectively.

The (mxm) matrix of normal equations, $X'X$, is formed and augmented by the restraint vector to form an (m+1)x(m+1) matrix, A:

$$A = \begin{bmatrix} X'X & r' \\ r & 0 \end{bmatrix}$$

Inverse of design matrix

The A matrix is inverted and shown in the form:

$$A^{-1} = \begin{bmatrix} Q & h' \\ h & 0 \end{bmatrix}$$

where Q is an mxm matrix that, when multiplied by s^2, yields the usual variance-covariance matrix.

Estimates of values of individual artifacts

The least-squares estimates for the values of the individual artifacts are contained in the (mx1) matrix, B, where

$$B = QX'Y + h'R^*$$

where Q is the upper left element of the A^{-1} matrix shown above. The structure of the individual estimates is contained in the QX' matrix; i.e. the estimate for the ith item can be computed from XQ and Y by

- Cross multiplying the ith column of XQ with Y
- And adding R^**(nominal test)/(nominal restraint)*

Clarify with an example

We will clarify the above discussion with an example from the mass calibration process at NIST. In this example, two NIST kilograms are compared with a customer's unknown kilogram.

The design matrix, X, is

$$X = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

The first two columns represent the two NIST kilograms while the third column represents the customers kilogram (i.e., the kilogram being calibrated).

The measurements obtained, i.e., the Y matrix, are

$$Y = \begin{bmatrix} -0.3800 \\ -1.5900 \\ -1.2150 \end{bmatrix}$$

The measurements are the differences between two measurements, as specified by the design matrix, measured in grams. That is, $Y(1)$ is the difference in measurement between NIST kilogram one and NIST kilogram two, $Y(2)$ is the difference in measurement between NIST kilogram one and the customer kilogram, and $Y(3)$ is the difference in measurement between NIST kilogram two and the customer kilogram.

The value of the reference standard, R^*, is 0.82329.

Then

$$X'X = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

If there are three weights with known values for weights one and two, then

r = [1 1 0]

Thus

$$A = \begin{bmatrix} 2 & -1 & -1 & 1 \\ -1 & 2 & -1 & 1 \\ -1 & -1 & 2 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

and so

$$A^{-1} = \frac{1}{6}\begin{bmatrix} 1 & -1 & 0 & 3 \\ -1 & 1 & 0 & 3 \\ 0 & 0 & 3 & 3 \\ 3 & 3 & 3 & 0 \end{bmatrix}$$

From A^{-1}, we have

$$Q = \frac{1}{6}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

We then compute QX'

$$QX' = \frac{1}{6}\begin{bmatrix} 2 & 1 & -1 \\ -2 & -1 & 1 \\ 0 & -3 & -3 \end{bmatrix}$$

We then compute $B = QX'Y + h'R^*$

$$B = \frac{1}{6}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}\begin{bmatrix} -0.3800 \\ -1.5900 \\ -1.2150 \end{bmatrix} + 0.82329\begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$$

This yields the following least-squares coefficient estimates:

$$B = \begin{bmatrix} 0.2225 \\ 0.6008 \\ 1.8141 \end{bmatrix}$$

Standard deviations of estimates

The standard deviation for the ith item is:

$$s_{item_i} = \sqrt{v_i Q v_i' s_1^2 + v_i D v_i' s_{days}^2}$$

where

$$D = (QX'X)(QX'X)'$$

The process standard deviation, which is a measure of the overall precision of the (NIST) mass calibrarion process,

$$s_1 = \sqrt{\frac{1}{n-m+1} Y'(I - XQX')Y}$$

is the residual standard deviation from the design, and s_{days} is the standard deviation for days, which can only be estimated from check standard measurements.

Example

We continue the example started above. Since n = 3 and m = 3, the formula reduces to:

$$s_1 = \sqrt{Y'(I - XQX')Y}$$

Substituting the values shown above for X, Y, and Q results in

$$(I - XQX') = \begin{bmatrix} 0.3333 & -0.3333 & 0.3333 \\ -0.3333 & 0.3333 & -0.3333 \\ 0.3333 & -0.3333 & 0.3333 \end{bmatrix}$$

and

$$Y'(I - XQX')Y = 0.0000083333$$

Finally, taking the square root gives

$$s_1 = 0.002887$$

The next step is to compute the standard deviation of item 3 (the customers kilogram), that is s_{item_3}. We start by substitituting the values for X and Q and computing D

$$D = (QX'X)(QX'X)' = \begin{bmatrix} 0.5000 & -0.5000 & 0.0000 \\ -0.5000 & 0.5000 & 0.0000 \\ 0.0000 & 0.0000 & 1.5000 \end{bmatrix}$$

Next, we substitute v_i' = [0 0 1] and $s_{days}^2 = 0.02111^2$ (this value is taken from a check standard and not computed from the values given in this example).

We obtain the following computations

$$v_i Q v_i' = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \frac{1}{6} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} [0 \ 0 \ 1] = 0.5$$

and

$$v_i D v_i' = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0.5000 & -0.5000 & 0.0000 \\ -0.5000 & 0.5000 & 0.0000 \\ 0.0000 & 0.0000 & 1.5000 \end{bmatrix} [0 \ 0 \ 1] = 1.5$$

and

$$s_{item_i} = \sqrt{v_i Q v_i' s_1^2 + v_i D v_i' s_{days}^2}$$

$$s_{item_3} = \sqrt{0.5(0.002887)^2 + 1.5(0.0211)^2} = 0.02593$$

2.3.3.3. Uncertainties of calibrated values

Uncertainty analysis follows the ISO principles

This section discusses the calculation of uncertainties of calibrated values from calibration designs. The discussion follows the guidelines in the section on classifying and combining components of uncertainty. Two types of evaluations are covered.

1. type A evaluations of time-dependent sources of random error
2. type B evaluations of other sources of error

The latter includes, but is not limited to, uncertainties from sources that are not replicated in the calibration design such as uncertainties of values assigned to reference standards.

Uncertainties for test items

Uncertainties associated with calibrated values for test items from designs require calculations that are specific to the individual designs. The steps involved are outlined below.

Outline for the section on uncertainty analysis

- Historical perspective
- Assumptions
- Example of more realistic model
- Computation of repeatability standard deviations
- Computation of level-2 standard deviations
- Combination of repeatability and level-2 standard deviations
- Example of computations for 1,1,1,1 design
- Type B uncertainty associated with the restraint
- Expanded uncertainty of calibrated values

2. Measurement Process Characterization
2.3. Calibration
2.3.3. What are calibration designs?
2.3.3.3. Uncertainties of calibrated values

2.3.3.3.1. Type A evaluations for calibration designs

Change over time

Type A evaluations for calibration processes must take into account changes in the measurement process that occur over time.

Historically, uncertainties considered only instrument imprecision

Historically, computations of uncertainties for calibrated values have treated the precision of the comparator instrument as the primary source of random uncertainty in the result. However, as the precision of instrumentation has improved, effects of other sources of variability have begun to show themselves in measurement processes. This is not universally true, but for many processes, instrument imprecision (short-term variability) cannot explain all the variation in the process.

Effects of environmental changes

Effects of humidity, temperature, and other environmental conditions which cannot be closely controlled or corrected must be considered. These tend to exhibit themselves over time, say, as between-day effects. The discussion of between-day (level-2) effects relating to gauge studies carries over to the calibration setting, but the computations are not as straightforward.

Assumptions which are specific to this section

The computations in this section depend on specific assumptions:

1. Short-term effects associated with instrument response
   - come from a single distribution
   - vary randomly from measurement to measurement within a design.
2. Day-to-day effects
   - come from a single distribution
   - vary from artifact to artifact but remain constant for a single calibration
   - vary from calibration to calibration

These

These assumptions have proved useful for characterizing

assumptions have proved useful but may need to be expanded in the future

high precision measurement processes, but more complicated models may eventually be needed which take the relative magnitudes of the test items into account. For example, in mass calibration, a 100 g weight can be compared with a summation of 50g, 30g and 20 g weights in a single measurement. A sophisticated model might consider the size of the effect as relative to the nominal masses or volumes.

Example of the two models for a design for calibrating test item using 1 reference standard

To contrast the simple model with the more complicated model, a measurement of the difference between X, the test item, with unknown and yet to be determined value, X^*, and a reference standard, R, with known value, R^*, and the reverse measurement are shown below.

Model (1) takes into account only instrument imprecision so that:

(1)

$$Y_1 = X - R + error_1$$
$$Y_2 = R - X + error_2$$

with the error terms random errors that come from the imprecision of the measuring instrument.

Model (2) allows for both instrument imprecision and level-2 effects such that:

(2)

$$Y_1 = (X + \Delta_X) - (R + \Delta_R) + error_1$$
$$Y_2 = (R + \Delta_R) - (X + \Delta_X) + error_2$$

where the delta terms explain small changes in the values of the artifacts that occur over time. For both models, the value of the test item is estimated as

$$\hat{Test} = X^* = \frac{1}{2}(Y_1 - Y_2) + R^*$$

Standard deviations from both models

For model (l), the standard deviation of the test item is

$$s_{test} = \frac{s_1}{\sqrt{2}}$$

For model (2), the standard deviation of the test item is

$$s_{test} = \sqrt{\frac{s_1^2}{2} + \frac{s_2^2}{2}}$$

.

Note on relative contributions of both components to uncertainty

In both cases, s_1 is the repeatability standard deviation that describes the precision of the instrument and s_2 is the level-2 standard deviation that describes day-to-day changes. One thing to notice in the standard deviation for the test item is the contribution of s_2 relative to the total uncertainty. If s_2 is large relative to s_1, or dominates, the uncertainty will not be appreciably reduced by adding measurements to the calibration design.

2. Measurement Process Characterization
2.3. Calibration
2.3.3. What are calibration designs?
2.3.3.3. Uncertainties of calibrated values

2.3.3.3.2. Repeatability and level-2 standard deviations

Repeatability standard deviation comes from the data of a single design

The repeatability standard deviation of the instrument can be computed in two ways.

1. It can be computed as the residual standard deviation from the design and should be available as output from any software package that reduces data from calibration designs. The matrix equations for this computation are shown in the section on solutions to calibration designs. The standard deviation has degrees of freedom

$$v = n - m + 1$$

for n difference measurements and m items. Typically the degrees of freedom are very small. For two differences measurements on a reference standard and test item, the degrees of freedom is $v=1$.

A more reliable estimate comes from pooling over historical data

2. A more reliable estimate of the standard deviation can be computed by pooling variances from K calibrations (and then taking its square root) using the same instrument (assuming the instrument is in statistical control). The formula for the pooled estimate is

$$s_1 = \sqrt{\frac{1}{\sum_k \nu_k} \sum_k \nu_k s_k^2}$$

Level-2 standard deviation is estimated from check standard measurements

The level-2 standard deviation cannot be estimated from the data of the calibration design. It cannot generally be estimated from repeated designs involving the test items. The best mechanism for capturing the day-to-day effects is a check standard, which is treated as a test item and included in each calibration design. Values of the check standard, estimated over time from the calibration design,

are used to estimate the standard deviation.

Assumptions

The check standard value must be stable over time, and the measurements must be in statistical control for this procedure to be valid. For this purpose, it is necessary to keep a historical record of values for a given check standard, and these values should be kept by instrument and by design.

Computation of level-2 standard deviation

Given K historical check standard values,

$$C_1, C_2, \cdots, C_K$$

the standard deviation of the check standard values is computed as

$$s_c = s_2 = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(C_k - \overline{C_.}\right)^2}$$

where

$$\overline{C_.} = \frac{1}{K}\sum_{k=1}^{K} C_k$$

with degrees of freedom $v = K - 1$.

2. Measurement Process Characterization
2.3. Calibration
2.3.3. What are calibration designs?
2.3.3.3. Uncertainties of calibrated values

2.3.3.3.3. Combination of repeatability and level-2 standard deviations

Standard deviation of test item depends on several factors

The final question is how to combine the repeatability standard deviation and the standard deviation of the check standard to estimate the standard deviation of the test item. This computation depends on:

- structure of the design
- position of the check standard in the design
- position of the reference standards in the design
- position of the test item in the design

Derivations require matrix algebra

Tables for estimating standard deviations for all test items are reported along with the solutions for all designs in the catalog. The use of the tables for estimating the standard deviations for test items is illustrated for the 1,1,1,1 design. Matrix equations can be used for deriving estimates for designs that are not in the catalog.

The check standard for each design is either an additional test item in the design, other than the test items that are submitted for calibration, or it is a construction, such as the difference between two reference standards as estimated by the design.

2.3.3.3.4. Calculation of standard deviations for 1,1,1,1 design

Design with 2 reference standards and 2 test items

An example is shown below for a 1,1,1,1 design for two reference standards, R_1 and R_2, and two test items, X_1 and X_2, and six difference measurements. The restraint, R^*, is the sum of values of the two reference standards, and the check standard, which is independent of the restraint, is the difference between the values of the reference standards. The design and its solution are reproduced below.

Check standard is the difference between the 2 reference standards

OBSERVATIONS	1	1	1	1
Y(1)	+	-		
Y(2)	+		-	
Y(3)	+			-
Y(4)		+	-	
Y(5)		+		-
Y(6)			+	-
RESTRAINT	+	+		
CHECK STANDARD	+	-		

DEGREES OF FREEDOM = 3

SOLUTION MATRIX
DIVISOR = 8

OBSERVATIONS	1	1	1	1
Y(1)	2	-2	0	0
Y(2)	1	-1	-3	-1
Y(3)	1	-1	-1	-3
Y(4)	-1	1	-3	-1
Y(5)	-1	1	-1	-3
Y(6)	0	0	2	-2
R*	4	4	4	4

Explanation of solution matrix

The solution matrix gives values for the test items of

$$X_1^* = \frac{1}{8}(-3Y_2 - Y_3 - 3Y_4 - Y_5 + 2Y_6) + \frac{1}{2}R^*$$

$$X_2^* = \frac{1}{8}(-Y_2 - 3Y_3 - Y_4 - 3Y_5 - 2Y_6) + \frac{1}{2}R^*$$

Factors for computing contributions of repeatability and level-2 standard deviations to uncertainty

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR K_1	1	1	1	1
1	0.3536	+			
1	0.3536		+		
1	0.6124			+	
1	0.6124				+
0	0.7071	+	-		

FACTORS FOR LEVEL-2 STANDARD DEVIATIONS

WT	FACTOR K_2	1	1	1	1
1	0.7071	+			
1	0.7071		+		
1	1.2247			+	
1	1.2247				+
0	1.4141	+	-		

The first table shows factors for computing the contribution of the repeatability standard deviation to the total uncertainty. The second table shows factors for computing the contribution of the between-day standard deviation to the uncertainty. Notice that the check standard is the last entry in each table.

Unifying equation

The unifying equation is:

$$s_{test} = \sqrt{K_1^2 s_1^2 + K_2^2 s_{days}^2}$$

Standard deviations are computed using the factors from the tables with the unifying equation

The steps in computing the standard deviation for a test item are:

- Compute the repeatability standard deviation from historical data.
- Compute the standard deviation of the check standard from historical data.
- Locate the factors, K_1 and K_2, for the check standard.
- Compute the between-day variance (using the unifying equation for the check standard). For this example,

$$s_2^2 = \left\{ \left(\sqrt{.5}s_1\right)^2 + \left(\sqrt{2}s_{days}\right)^2 \right\} = \left\{ \frac{1}{2}s_1^2 + 2s_{days}^2 \right\}$$

implies

$$\Rightarrow$$

$$s_{days}^2 = \frac{1}{2}\left\{ s_2^2 - \frac{1}{2}s_1^2 \right\}$$

.

- If this variance estimate is negative, set s_{days} = 0. (This is possible and indicates that there is no contribution to uncertainty from day-to-day effects.)
- Locate the factors, K_1 and K_2, for the test items, and compute the standard deviations using the unifying equation. For this example,

$$s_{x_1} = \sqrt{\frac{3}{8}s_1^2 + \frac{3}{2}s_{days}^2} = \sqrt{\frac{3}{8}s_1^2 + \frac{3}{2}\frac{1}{2}\{s_2^2 - \frac{1}{2}s_1^2\}} = \sqrt{\frac{3}{4}s_2^2}$$

and

$$s_{x_2} = s_{x_1}$$

2. Measurement Process Characterization
2.3. Calibration
2.3.3. What are calibration designs?
2.3.3.3. Uncertainties of calibrated values

2.3.3.3.5. Type B uncertainty

Type B uncertainty associated with the restraint

The reference standard is assumed to have known value, R^*, for the purpose of solving the calibration design. For the purpose of computing a standard uncertainty, it has a type B uncertainty that contributes to the uncertainty of the test item.

The value of R^* comes from a higher-level calibration laboratory or process, and its value is usually reported along with its uncertainty, U. If the laboratory also reports the k factor for computing U, then the standard deviation of the restraint is

$$s_{R^*} = \frac{U}{k}$$

If k is not reported, then a conservative way of proceeding is to assume $k = 2$.

Situation where the test is different in size from the reference

Usually, a reference standard and test item are of the same nominal size and the calibration relies on measuring the small difference between the two; for example, the intercomparison of a reference kilogram compared with a test kilogram. The calibration may also consist of an intercomparison of the reference with a summation of artifacts where the summation is of the same nominal size as the reference; for example, a reference kilogram compared with 500 g + 300 g + 200 g test weights.

Type B uncertainty for the test artifact

The type B uncertainty that accrues to the test artifact from the uncertainty of the reference standard is proportional to their nominal sizes; i.e.,

$$s_B = \frac{Nominal\ Test}{Nominal\ Restraint} s_{R^*}$$

2.3.3.3.6. Expanded uncertainties

Standard uncertainty

The standard uncertainty for the test item is

$$u = \sqrt{s_{test}^2 + \left(\frac{Nominal\ Test}{Nominal\ Restraint}\right)^2 s_{R^*}^2}$$

Expanded uncertainty

The expanded uncertainty is computed as

$$U = ku$$

where k is either the critical value from the t table for degrees of freedom v or k is set equal to 2.

Problem of the degrees of freedom

The calculation of degrees of freedom, v, can be a problem. Sometimes it can be computed using the Welch-Satterthwaite approximation and the structure of the uncertainty of the test item. Degrees of freedom for the standard deviation of the restraint is assumed to be infinite. The coefficients in the Welch-Satterthwaite formula must all be positive for the approximation to be reliable.

Standard deviation for test item from the 1,1,1,1 design

For the 1,1,1,1 design, the standard deviation of the test items can be rewritten by substituting in the equation

$$s_{X_1} = s_{X_2} = \sqrt{\frac{3}{8}s_1^2 + \frac{3}{2}s_{days}^2} = \sqrt{\frac{3}{8}s_1^2 + \frac{3}{2}\left\{\frac{1}{2}s_2^2 - \frac{1}{4}s_1^2\right\}} = \frac{\sqrt{3}}{2}s_2$$

so that the degrees of freedom depends only on the degrees of freedom in the standard deviation of the check standard. This device may not work satisfactorily for all designs.

Standard uncertainty from the 1,1,1,1 design

To complete the calculation shown in the equation at the top of the page, the nominal value of the test item (which is equal to 1) is divided by the nominal value of the restraint (which is also equal to 1), and the result is squared. Thus, the standard uncertainty is

$$u = \sqrt{\frac{3}{4}s_2^2 + s_{R^*}^2}$$

Degrees of

Therefore, the degrees of freedom is approximated as

freedom using the Welch-Satterthwaite approximation

$$\nu = \frac{u^4}{\frac{\frac{9}{16}s_2^4}{n-1}}$$

where *n - 1* is the degrees of freedom associated with the check standard uncertainty. Notice that the standard deviation of the restraint drops out of the calculation because of an infinite degrees of freedom.

2.3.4. Catalog of calibration designs

Important concept - Restraint

The designs are constructed for measuring differences among reference standards and test items, singly or in combinations. Values for individual standards and test items can be computed from the design only if the value (called the *restraint* = R^*) of one or more reference standards is known. The methodology for constructing and solving calibration designs is described briefly in matrix solutions and in more detail in a NIST publication. (Cameron et al.).

Designs listed in this catalog

Designs are listed by traditional subject area although many of the designs are appropriate generally for intercomparisons of artifact standards.

- Designs for mass weights
- Drift-eliminating designs for gage blocks
- Left-right balanced designs for electrical standards
- Designs for roundness standards
- Designs for angle blocks
- Drift-eliminating design for thermometers in a bath
- Drift-eliminating designs for humidity cylinders

Properties of designs in this catalog

Basic requirements are:

1. The differences must be nominally zero.
2. The design must be solvable for individual items given the restraint.

Other desirable properties are:

1. The number of measurements should be small.
2. The degrees of freedom should be greater than zero.
3. The standard deviations of the estimates for the test items should be small enough for their intended purpose.

Information:

Given

Design

- n = number of difference measurements
- m = number of artifacts (reference standards + test items) to be calibrated

Solution

Factors for computing standard deviations

the following information is shown for each design:

- Design matrix -- (n x m)
- Vector that identifies standards in the restraint -- (1 x m)
- Degrees of freedom = (n - m + 1)
- Solution matrix for given restraint -- (n x m)
- Table of factors for computing standard deviations

Convention for showing the measurement sequence

Nominal sizes of standards and test items are shown at the top of the design. Pluses (+) indicate items that are measured together; and minuses (-) indicate items are not measured together. The difference measurements are constructed from the design of pluses and minuses. For example, a 1,1,1 design for one reference standard and two test items of the same nominal size with three measurements is shown below:

	1	1	1
Y(1) =	+	-	
Y(2) =	+		-
Y(3) =	+	-	

Solution matrix

Example and interpretation

The cross-product of the column of difference measurements and R^* with a column from the solution matrix, divided by the named divisor, gives the value for an individual item. For example,

Solution matrix
Divisor = 3

	1	1	1
Y(1)	0	-2	-1
Y(2)	0	-1	-2
Y(3)	0	+1	-1
R*	+3	+3	+3

implies that estimates for the restraint and the two test items are:

$$\hat{R^*} = \frac{1}{3}\{0 \cdot Y_1 + 0 \cdot Y_2 + 0 \cdot Y_3 + 3R^*\} = R^*$$

$$\hat{Test_1} = \frac{1}{3}\{-2Y_1 - Y_2 + Y_3 + 3R^*\}$$

$$\hat{Test_2} = \frac{1}{3}\{-Y_1 - 2Y_2 - Y_3 + 3R^*\}$$

Interpretation of table of factors

The factors in this table provide information on precision. The repeatability standard deviation, s_1, is multiplied by the appropriate factor to obtain the standard deviation for an

individual item or combination of items. For example,

```
          Sum     Factor      1       1
1
          1       0.0000      +
          1       0.8166              +
          1       0.8166
+
          2       1.4142               +
+
```

implies that the standard deviations for the estimates are:

$$s_{R^*} = 0$$

$$s_{test_1} = 0.8661 \cdot s_1$$

$$s_{test_2} = 0.8661 \cdot s_1$$

$$s_{test_{1+2}} = 1.4142 \cdot s_1$$

2.3.4.1. Mass weights

Tie to kilogram reference standards

Near-accurate mass measurements require a sequence of designs that relate the masses of individual weights to a reference kilogram(s) standard (Jaeger & Davis). Weights generally come in sets, and an entire set may require several series to calibrate all the weights in the set.

Example of weight set

A 5,3,2,1 weight set would have the following weights:

1000 g

500g, 300g, 200g, 100g

50g, 30g 20g, 10g

5g, 3g, 2g, 1g

0.5g, 0.3g, 0.2g, 0.1g

Depiction of a design with three series for calibrating a 5,3,2,1 weight set with weights between 1 kg and 10 g

First series using

The calibrations start with a comparison of the one kilogram test weight with the reference kilograms (see the graphic above). The 1,1,1,1 design requires two kilogram reference

1,1,1,1 design

standards with known values, *R1** and *R2**. The fourth kilogram in this design is actually a summation of the 500, 300, 200 g weights which becomes the restraint in the next series.

The restraint for the first series is the known average mass of the reference kilograms,

$$R^{\bullet} = \frac{R_1^{\bullet} + R_2^{\bullet}}{2}$$

The design assigns values to all weights including the individual reference standards. For this design, the check standard is not an artifact standard but is defined as the difference between the values assigned to the reference kilograms by the design; namely,

$$C = \left(\hat{R}_1^{*} - \hat{R}_2^{*} \right)$$

2nd series using 5,3,2,1,1,1 design

The second series is a 5,3,2,1,1,1 design where the restraint over the 500g, 300g and 200g weights comes from the value assigned to the summation in the first series; i.e.,

$$R^{\bullet} = \sum 1 = \hat{X}_{500} + \hat{X}_{300} + \hat{X}_{200}$$

The weights assigned values by this series are:

- 500g, 300g, 200 g and 100g test weights
- 100 g check standard (2nd 100g weight in the design)
- Summation of the 50g, 30g, 20g weights.

Other starting points

The calibration sequence can also start with a 1,1,1 design. This design has the disadvantage that it does **not** have provision for a check standard.

Better choice of design

A better choice is a 1,1,1,1,1 design which allows for two reference kilograms and a kilogram check standard which occupies the 4th position among the weights. This is preferable to the 1,1,1,1 design but has the disadvantage of requiring the laboratory to maintain three kilogram standards.

Important detail

The solutions are only applicable for the restraints as shown.

Designs for decreasing weight sets

1. 1,1,1 design
2. 1,1,1,1 design
3. 1,1,1,1,1 design
4. 1,1,1,1,1,1 design
5. 2,1,1,1 design
6. 2,2,1,1,1 design

7. 2,2,2,1,1 design
8. 5,2,2,1,1,1 design
9. 5,2,2,1,1,1,1 design
10. 5,3,2,1,1,1 design
11. 5,3,2,1,1,1,1 design
12. 5,3,2,2,1,1,1 design
13. 5,4,4,3,2,2,1,1 design
14. 5,5,2,2,1,1,1,1 design
15. 5,5,3,2,1,1,1 design
16. 1,1,1,1,1,1,1,1 design
17. 3,2,1,1,1 design

Design for pound weights

1. 1,2,2,1,1 design

Designs for increasing weight sets

1. 1,1,1 design
2. 1,1,1,1 design
3. 5,3,2,1,1 design
4. 5,3,2,1,1,1 design
5. 5,2,2,1,1,1 design
6. 3,2,1,1,1 design

2.3.4.1.1. Design for 1,1,1

Design 1,1,1

```
OBSERVATIONS    1    1    1

     Y(1)       +    -
     Y(2)       +         -
     Y(3)            +    -


RESTRAINT       +

CHECK STANDARD      +

DEGREES OF FREEDOM   =   1

                    SOLUTION MATRIX
                      DIVISOR   =   3

OBSERVATIONS       1       1       1

   Y(1)            0      -2      -1
   Y(2)            0      -1      -2
   Y(3)            0       1      -1
   R*              3       3       3

   R* = value of reference weight


FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
WT   FACTOR
                   1    1    1
 1   0.0000        +
 1   0.8165             +
 1   0.8165                  +
 2   1.4142             +    +
 1   0.8165             +

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS
WT   FACTOR
                   1    1    1
 1   0.0000        +
 1   1.4142             +
 1   1.4142                  +
 2   2.4495             +    +
 1   1.4142             +
```

Explanation of notation and interpretation of tables

2.3.4.1.2. Design for 1,1,1,1

```
Design 1,1,1,1

 OBSERVATIONS   1
1   1   1


     Y(1)       +
-
     Y(2)       +
-
     Y(3)       +
-
     Y(4)
+   -
     Y(5)
+       -
     Y(6)
+   -

 RESTRAINT      +
+

 CHECK STANDARD +
-

 DEGREES OF
FREEDOM  =  3


SOLUTION MATRIX

DIVISOR  =  8

 OBSERVATIONS
1      1      1
1


    Y(1)
2     -2      0
0
    Y(2)
1     -1     -3
-1
    Y(3)
1     -1     -1
-3
    Y(4)
-1     1     -3
-1
    Y(5)
-1     1     -1
-3
    Y(6)
0      0      2
-2
    R*
4      4      4
```

```
4

    R* = sum of
two reference
standards


 FACTORS FOR
REPEATABILITY
STANDARD
DEVIATIONS
 WT  FACTOR
        K1      1
1   1   1
  1  0.3536    +
  1  0.3536
+
  1  0.6124
+
  1  0.6124
+
  0  0.7071    +
-


 FACTORS FOR
BETWEEN-DAY
STANDARD
DEVIATIONS
 WT  FACTOR
        K2      1
1   1   1
  1  0.7071    +
  1  0.7071
+
  1  1.2247
+
  1  1.2247
+
  0  1.4141    +
-
```

Explanation of notation and interpretation of tables

2.3.4.1.3. Design for 1,1,1,1,1

CASE 1: CHECK STANDARD = DIFFERENCE BETWEEN FIRST TWO WEIGHTS

```
 OBSERVATIONS    1    1    1    1
1

      Y(1)        +    -
      Y(2)        +         -
      Y(3)        +              -
      Y(4)        +
-
      Y(5)             +    -
      Y(6)             +         -
      Y(7)             +
-
      Y(8)                  +    -
      Y(9)                  +
-
      Y(10)                       +
-

 RESTRAINT       +    +
 CHECK STANDARD  +    -
 DEGREES OF FREEDOM   =    6
```

```
SOLUTION MATRIX
DIVISOR =  10

 OBSERVATIONS       1       1
1       1       1

    Y(1)             2      -2
0       0       0
    Y(2)             1      -1
-3      -1      -1
    Y(3)             1      -1
-1      -3      -1
    Y(4)             1      -1
-1      -1      -3
    Y(5)            -1       1
-3      -1      -1
    Y(6)            -1       1
-1      -3      -1
    Y(7)            -1       1
-1      -1      -3
    Y(8)             0       0
2      -2       0
    Y(9)             0       0
```

CASE 2: CHECK STANDARD = FOURTH WEIGHT

```
 OBSERVATIONS    1    1    1    1
1

      Y(1)        +    -
      Y(2)        +         -
      Y(3)        +              -
      Y(4)        +
-
      Y(5)             +    -
      Y(6)             +         -
      Y(7)             +
-
      Y(8)                  +    -
      Y(9)                  +
-
      Y(10)                       +
-

 RESTRAINT       +    +
 CHECK STANDARD                  +
 DEGREES OF FREEDOM   =    6
```

```
SOLUTION MATRIX
DIVISOR =  10

 OBSERVATIONS       1       1
1       1       1

    Y(1)             2      -2
0       0       0
    Y(2)             1      -1
-3      -1      -1
    Y(3)             1      -1
-1      -3      -1
    Y(4)             1      -1
-1      -1      -3
    Y(5)            -1       1
-3      -1      -1
    Y(6)            -1       1
-1      -3      -1
    Y(7)            -1       1
-1      -1      -3
    Y(8)             0       0
2      -2       0
    Y(9)             0       0
2       0      -2
```

```
2       0      -2
    Y(10)          0       0
0       2      -2
    R*             5       5
5       5       5

    R* = sum of two reference
standards
```

```
 FACTORS FOR REPEATABILITY
STANDARD DEVIATIONS
 WT  FACTOR
        K1      1   1   1   1
1
  1  0.3162    +
  1  0.3162        +
  1  0.5477            +
  1  0.5477                +
  1  0.5477
+
  2  0.8944                +
+
  3  1.2247            +   +
+
  0  0.6325    +   -
```

```
 FACTORS FOR BETWEEN-DAY
STANDARD DEVIATIONS
 WT  FACTOR
        K2      1   1   1   1
1
  1  0.7071    +
  1  0.7071        +
  1  1.2247            +
  1  1.2247                +
  1  1.2247
+
  2  2.0000                +
+
  3  2.7386            +   +
+
  0  1.4142     +   -
```

```
    Y(10)          0       0
0       2      -2
    R*             5       5
5       5       5

    R* = sum of two reference
standards
```

```
 FACTORS FOR REPEATABILITY
STANDARD DEVIATIONS
 WT  FACTOR
        K1      1   1   1   1
1
  1  0.3162    +
  1  0.3162        +
  1  0.5477            +
  1  0.5477                +
  1  0.5477
+
  2  0.8944                +
+
  3  1.2247            +   +
+
  1  0.5477                +
```

```
 FACTORS FOR BETWEEN-DAY
STANDARD DEVIATIONS
 WT  FACTOR
        K2      1   1   1   1
1
  1  0.7071    +
  1  0.7071        +
  1  1.2247            +
  1  1.2247                +
  1  1.2247
+
  2  2.0000                +
+
  3  2.7386            +   +
+
  1  1.2247                +
```

Explanation of notation and interpretation of tables

2.3.4.1.4. Design for 1,1,1,1,1,1

Design 1,1,1,1,1,1

OBSERVATIONS	1	1	1	1	1	1
X(1)	+	-				
X(2)	+		-			
X(3)	+			-		
X(4)	+				-	
X(5)	+					-
X(6)		+	-			
X(7)		+		-		
X(8)		+			-	
X(9)		+				-
X(10)			+	-		
X(11)			+		-	
X(12)			+			-
X(13)				+	-	
X(14)				+		-
X(15)					+	-
RESTRAINT	+	+				
CHECK STANDARD					+	

DEGREES OF FREEDOM = 10

SOLUTION MATRIX
DIVISOR = 8

OBSERVATIONS	1	1	1	1	1	1
Y(1)	1	-1	0	0	0	0
Y(2)	1	0	-1	0	0	0
Y(3)	1	0	0	-1	0	0
Y(4)	1	0	0	0	-1	0
Y(5)	2	1	1	1	1	0
Y(6)	0	1	-1	0	0	0
Y(7)	0	1	0	-1	0	0
Y(8)	0	1	0	0	-1	0
Y(9)	1	2	1	1	1	0
Y(10)	0	0	1	-1	0	0
Y(11)	0	0	1	0	-1	0
Y(12)	1	1	2	1	1	0
Y(13)	0	0	0	1	-1	0
Y(14)	1	1	1	2	1	0
Y(15)	1	1	1	1	2	0
R*	6	6	6	6	6	6

R* = sum of two reference standards

FACTORS FOR COMPUTING REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	1	1	1	1	1	1
1	0.2887	+					

```
 1  0.2887        +
 1  0.5000            +
 1  0.5000                +
 1  0.5000                    +
 1  0.5000                        +
 2  0.8165            +   +
 3  1.1180            +   +   +
 4  1.4142            +   +   +   +
 1  0.5000                    +


FACTORS FOR COMPUTING BETWEEN-DAY STANDARD DEVIATIONS
WT  FACTOR
                1   1   1   1   1   1
 1  0.7071     +
 1  0.7071         +
 1  1.2247             +
 1  1.2247                 +
 1  1.2247                     +
 1  1.2247                         +
 2  2.0000             +   +
 3  2.7386             +   +   +
 4  3.4641             +   +   +   +
 1  1.2247                     +
```

Explanation of notation and interpretation of tables

2.3.4.1.5. Design for 2,1,1,1

```
Design 2,1,1,1

 OBSERVATIONS   2   1   1   1

     Y(1)       +   -       -
     Y(2)       +       -   -
     Y(3)       +   -   -
     Y(4)           +       -
     Y(5)           +   -
     Y(6)               +   -

 RESTRAINT      +

 CHECK STANDARD         +

DEGREES OF FREEDOM  =      3


                     SOLUTION MATRIX
                      DIVISOR  =  4

 OBSERVATIONS      2      1      1      1

    Y(1)           0     -1      0     -1
    Y(2)           0      0     -1     -1
    Y(3)           0     -1     -1      0
    Y(4)           0      1      0     -1
    Y(5)           0      1     -1      0
    Y(6)           0      0      1     -1
    R*             4      2      2      2

    R* = value of the reference standard


 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 WT  FACTOR
                2   1   1   1
  2  0.0000     +
  1  0.5000         +
  1  0.5000             +
  1  0.5000                 +
  2  0.7071         +   +
  3  0.8660         +   +   +
  1  0.5000             +


 FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS
 WT  FACTOR
                2   1   1   1
  2  0.0000     +
  1  1.1180         +
  1  1.1180             +
  1  1.1180                 +
  2  1.7321         +   +
```

```
  3  2.2913        +    +    +
  1  1.1180             +
```

[Explanation of notation and interpretation of tables]

2. Measurement Process Characterization
2.3. Calibration
2.3.4. Catalog of calibration designs
2.3.4.1. Mass weights

2.3.4.1.6. Design for 2,2,1,1,1

Design 2,2,1,1,1

OBSERVATIONS	2	2	1	1	1
Y(1)	+	-	-	+	
Y(2)	+	-		-	+
Y(3)	+	-	+		-
Y(4)	+	-			
Y(5)	+		-	-	
Y(6)	+		-		-
Y(7)	+			-	-
Y(8)		+	-	-	
Y(9)		+	-		-
Y(10)		+		-	-
RESTRAINT	+	+	+		
CHECK STANDARD				+	

DEGREES OF FREEDOM = 6

SOLUTION MATRIX
DIVISOR = 275

OBSERVATIONS	2	2	1	1	1
Y(1)	47	-3	-44	66	11
Y(2)	25	-25	0	-55	55
Y(3)	3	-47	44	-11	-66
Y(4)	25	-25	0	0	0
Y(5)	29	4	-33	-33	22
Y(6)	29	4	-33	22	-33
Y(7)	7	-18	11	-44	-44
Y(8)	4	29	-33	-33	22
Y(9)	4	29	-33	22	-33
Y(10)	-18	7	11	-44	-44
R*	110	110	55	55	55

R* = sum of three reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	2	2	1	1	1
2	0.2710	+				
2	0.2710		+			
1	0.3347			+		
1	0.4382				+	
1	0.4382					+
2	0.6066				+	+
3	0.5367			+	+	+
1	0.4382				+	

```
FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS
WT   FACTOR
                2   2   1   1   1
 2   0.8246     +
 2   0.8246         +
 1   0.8485             +
 1   1.0583                 +
 1   1.0583                     +
 2   1.5748                 +   +
 3   1.6971             +   +   +
 1   1.0583                 +
```

Explanation of notation and interpretation of tables

2.3.4.1.7. Design for 2,2,2,1,1

Design 2,2,2,1,1

OBSERVATIONS	2	2	2	1	1
Y(1)	+	-			
Y(2)	+		-		
Y(3)		+	-		
Y(4)	+			-	-
Y(5)		+		-	-
Y(6)			+	-	-
Y(7)				+	-
RESTRAINT	+	+			
CHECK STANDARD			+		

DEGREES OF FREEDOM = 3

SOLUTION MATRIX
DIVISOR = 16

OBSERVATIONS	2	2	2	1	1
Y(1)	4	-4	0	0	0
Y(2)	2	-2	-6	-1	-1
Y(3)	-2	2	-6	-1	-1
Y(4)	2	-2	-2	-3	-3
Y(5)	-2	2	-2	-3	-3
Y(6)	0	0	4	-2	-2
Y(7)	0	0	0	8	-8
R*	8	8	8	4	4

R* = sum of the two reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	2	2	2	1	1
2	0.3536	+				
2	0.3536		+			
2	0.6124			+		
1	0.5863				+	
1	0.5863					+
2	0.6124				+	+
4	1.0000			+	+	+
1	0.5863				+	

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT	FACTOR	2	2	2	1	1
2	0.7071	+				
2	0.7071		+			

```
2  1.2247            +
1  1.0607               +
1  1.0607                    +
2  1.5811                 +  +
4  2.2361            +    +  +
1  1.0607                 +
```

Explanation of notation and interpretation of tables

2.3.4.1.8. Design for 5,2,2,1,1,1

Design 5,2,2,1,1,1

OBSERVATIONS	5	2	2	1	1	1
Y(1)	+	-	-	-	-	+
Y(2)	+	-	-	-	+	-
Y(3)	+	-	-	+	-	-
Y(4)	+	-		-	-	-
Y(5)	+		-	-	-	-
Y(6)		+	-	+	-	
Y(7)		+	-	-		+
Y(8)		+	-		+	-
RESTRAINT	+	+	+	+		
CHECK STANDARD					+	

DEGREES OF FREEDOM = 3

SOLUTION MATRIX
DIVISOR = 70

OBSERVATIONS	5	2	2	1	1	1
Y(1)	15	-8	-8	1	1	21
Y(2)	15	-8	-8	1	21	1
Y(3)	5	-12	-12	19	-1	-1
Y(4)	0	2	12	-14	-14	-14
Y(5)	0	12	2	-14	-14	-14
Y(6)	-5	8	-12	9	-11	-1
Y(7)	5	12	-8	-9	1	11
Y(8)	0	10	-10	0	10	-10
R*	35	14	14	7	7	7

R* = sum of the four reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	5	2	2	1	1	1
5	0.3273	+					
2	0.3854		+				
2	0.3854			+			
1	0.4326				+		
1	0.4645					+	
1	0.4645						+
1	0.4645					+	

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT	FACTOR	5	2	2	1	1	1
5	1.0000	+					

```
2  0.8718        +
2  0.8718            +
1  0.9165                +
1  1.0198                    +
1  1.0198                        +
1  1.0198                    +
```

Explanation of notation and interpretation of tables

2.3.4.1.9. Design for 5,2,2,1,1,1,1

Design 5,2,2,1,1,1,1

OBSERVATIONS	5	2	2	1	1	1	1
Y(1)	+	-	-	-			
Y(2)	+	-	-		-		
Y(3)	+	-	-			-	
Y(4)	+	-	-				-
Y(5)		+		+	-	-	-
Y(6)			+	+	-	-	-
Y(7)		+	+	-	-	-	-
Y(8)					+	-	
Y(9)					+		-
Y(10)						+	-
RESTRAINT	+	+	+	+			
CHECK STANDARD						+	

DEGREES OF FREEDOM = 4

SOLUTION MATRIX
DIVISOR = 60

OBSERVATIONS	5	2	2	1	1	1	1
Y(1)	12	0	0	-12	0	0	0
Y(2)	6	-4	-4	2	-12	3	3
Y(3)	6	-4	-4	2	3	-12	3
Y(4)	6	-4	-4	2	3	3	-12
Y(5)	-6	28	-32	10	-6	-6	-6
Y(6)	-6	-32	28	10	-6	-6	-6
Y(7)	6	8	8	-22	-6	-6	-6
Y(8)	0	0	0	0	15	-15	0
Y(9)	0	0	0	0	15	0	-15
Y(10)	0	0	0	0	0	15	-15
R*	30	12	12	6	6	6	6

R* = sum of the four reference standards

```
FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
WT  FACTOR
                5   2   2   1   1   1   1
 5  0.3162      +
 2  0.7303          +
 2  0.7303              +
 1  0.4830                  +
 1  0.4472                      +
 1  0.4472                          +
 1  0.4472                              +
 2  0.5477                      +   +
 3  0.5477                      +   +   +
 1  0.4472                          +
```

```
FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS
WT  FACTOR
                5   2   2   1   1   1   1
 5  1.0000      +
 2  0.8718          +
 2  0.8718              +
 1  0.9165                  +
 1  1.0198                      +
 1  1.0198                          +
 1  1.0198                              +
 2  1.4697                      +   +
 3  1.8330                      +   +   +
 1  1.0198                          +
```

Explanation of notation and interpretation of tables

2.3.4.1.10. Design for 5,3,2,1,1,1

OBSERVATIONS	5	3	2	1	1	1
Y(1)	+	-	-	+	-	
Y(2)	+	-	-		+	-
Y(3)	+	-	-	-		+
Y(4)	+	-	-			
Y(5)	+		-	-	-	-
Y(6)		+	-	+	-	-
Y(7)		+	-	-	+	-
Y(8)		+	-	-	-	+
Y(9)			+	-	-	
Y(10)			+	-		-
Y(11)			+		-	-
RESTRAINT	+	+	+			
CHECK STANDARD				+		

DEGREES OF FREEDOM = 6

SOLUTION MATRIX
DIVISOR = 920

OBSERVATIONS	5	3	2	1	1	1
Y(1)	100	-68	-32	119	-111	4
Y(2)	100	-68	-32	4	119	-111
Y(3)	100	-68	-32	-111	4	119
Y(4)	100	-68	-32	4	4	4
Y(5)	60	-4	-56	-108	-108	-108
Y(6)	-20	124	-104	128	-102	-102
Y(7)	-20	124	-104	-102	128	-102
Y(8)	-20	124	-104	-102	-102	128
Y(9)	-20	-60	80	-125	-125	-10
Y(10)	-20	-60	80	-125	-10	-125
Y(11)	-20	-60	80	-10	-125	-125
R*	460	276	184	92	92	92

R* = sum of the three reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	5	3	2	1	1	1
5	0.2331	+					
3	0.2985		+				
2	0.2638			+			
1	0.3551				+		
1	0.3551					+	
1	0.3551						+
2	0.5043				+	+	
3	0.6203				+	+	+
1	0.3551					+	

```
FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS
WT   FACTOR
                5   3   2   1   1   1
 5   0.8660     +
 3   0.8185         +
 2   0.8485             +
 1   1.0149                 +
 1   1.0149                     +
 1   1.0149                         +
 2   1.4560                 +   +
 3   1.8083                 +   +   +
 1   1.0149                     +
```

Explanation of notation and interpretation of tables

2.3.4.1.11. Design for 5,3,2,1,1,1,1

Design 5,3,2,1,1,1,1

OBSERVATIONS	5	3	2	1	1	1	1
Y(1)	+	-	-				
Y(2)	+	-		-	-		
Y(3)	+	-				-	-
Y(4)	+		-	-	-	-	
Y(5)	+		-	-	-		-
Y(6)	+		-	-		-	-
Y(7)	+		-		-	-	-
Y(8)		+	-	-			
Y(9)		+	-		-		
Y(10)		+	-			-	
Y(11)		+	-				-
RESTRAINT	+	+	+				
CHECK STANDARD						+	

DEGREES OF FREEDOM = 5

SOLUTION MATRIX
DIVISOR = 40

OBSERVATIONS	5	3	2	1	1	1	1
Y(1)	20	-4	-16	12	12	12	12
Y(2)	0	-4	4	-8	-8	2	2
Y(3)	0	-4	4	2	2	-8	-8
Y(4)	0	0	0	-5	-5	-10	10
Y(5)	0	0	0	-5	-5	10	-10
Y(6)	0	0	0	-10	10	-5	-5
Y(7)	0	0	0	10	-10	-5	-5
Y(8)	0	4	-4	-12	8	3	3
Y(9)	0	4	-4	8	-12	3	3
Y(10)	0	4	-4	3	3	-12	8
Y(11)	0	4	-4	3	3	8	-12
R*	20	12	8	4	4	4	4

R* = sum of the three reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	5	3	2	1	1	1	1
5	0.5000	+						
3	0.2646		+					
2	0.4690			+				
1	0.6557				+			
1	0.6557					+		
1	0.6557						+	
1	0.6557							+
2	0.8485				+	+		
3	1.1705				+	+	+	
4	1.3711				+	+	+	+
1	0.6557						+	

FACTORS FOR LEVEL-2 STANDARD DEVIATIONS

WT	FACTOR	5	3	2	1	1	1	1
5	0.8660	+						
3	0.8185		+					
2	0.8485			+				
1	1.0149				+			
1	1.0149					+		
1	1.0149						+	
1	1.0149							+
2	1.4560				+	+		
3	1.8083				+	+	+	
4	2.1166				+	+	+	+
1	1.0149						+	

Explanation of notation and interpretation of tables

HOME TOOLS & AIDS SEARCH BACK NEXT

2.3.4.1.12. Design for 5,3,2,2,1,1,1

OBSERVATIONS	5	3	2	2	1	1	1
Y(1)	+	-	-				
Y(2)	+	-		-			
Y(3)	+		-	-	-		
Y(4)	+		-	-		-	
Y(5)	+		-	-			-
Y(6)		+		-	-		
Y(7)		+		-		-	
Y(8)		+		-			-
Y(9)		+			-	-	-
Y(10)					+	-	
Y(11)						+	-
Y(12)					-		+
RESTRAINT	+	+	+				
CHECK STANDARDS						+	

DEGREES OF FREEDOM = 6

SOLUTION MATRIX
DIVISOR = 10

OBSERVATIONS	5	3	2	2	1	1
Y(1)	2	0	-2	2	0	0
Y(2)	0	-6	6	-4	-2	-2
Y(3)	1	1	-2	0	-1	1
Y(4)	1	1	-2	0	1	-1
Y(5)	1	1	-2	0	1	1
Y(6)	-1	1	0	-2	-1	1
Y(7)	-1	1	0	-2	1	-1
Y(8)	-1	1	0	-2	1	1
Y(9)	0	-2	2	2	-4	-4
Y(10)	0	0	0	0	2	-2
Y(11)	0	0	0	0	0	2
Y(12)	0	0	0	0	-2	0
R*	5	3	2	2	1	1

R* = sum of the three reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	5	3	2	2	1	1	1
5	0.3162	+						
3	0.6782		+					
2	0.7483			+				
2	0.6000				+			
1	0.5831					+		
1	0.5831						+	
1	0.5831							+
3	0.8124				+	+		
4	1.1136				+	+	+	
1	0.5831						+	

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT	FACTOR	5	3	2	2	1	1	1
5	0.8660	+						
3	0.8185		+					
2	0.8485			+				
2	1.0583				+			
1	1.0149					+		
1	1.0149						+	
1	1.0149							+
3	1.5067				+	+		
4	1.8655				+	+	+	
1	1.0149						+	

Explanation of notation and interpretation of tables

2. Measurement Process Characterization
2.3. Calibration
2.3.4. Catalog of calibration designs
2.3.4.1. Mass weights

2.3.4.1.13. Design for 5,4,4,3,2,2,1,1

OBSERVATIONS	5	4	4	3	2	2	1	1
Y(1)	+	+		-	-	-	-	-
Y(2)	+		+	-	-	-	-	-
Y(3)	+	-					-	
Y(4)	+		-					-
Y(5)	+		-				-	
Y(6)	+	-			-			
Y(7)	+			-		-	-	
Y(8)	+			-	-	-		
Y(9)	+	-		-				
Y(10)	+		-		-			
Y(11)	+		-	-				
Y(12)	+	-	-					
RESTRAINT							+	+
CHECK STANDARD							+	-

DEGREES OF FREEDOM = 5

SOLUTION MATRIX
DIVISOR = 916

OBSERVATIONS	5	4	4	3	2	2	1	1
Y(1)	232	325	123	8	-37	135	-1	1
Y(2)	384	151	401	108	73	105	101	-101
Y(3)	432	84	308	236	168	204	-144	144
Y(4)	608	220	196	400	440	-120	408	-408
Y(5)	280	258	30	136	58	234	-246	246
Y(6)	24	-148	68	64	-296	164	-8	8
Y(7)	-104	-122	-142	28	214	-558	-118	118
Y(8)	-512	-354	-382	-144	-250	-598	18	-18
Y(9)	76	-87	139	-408	55	443	51	-51
Y(10)	-128	26	-210	-36	-406	194	-110	110
Y(11)	-76	87	-139	-508	-55	473	-51	51
Y(12)	-300	-440	-392	116	36	-676	100	-100
R*	1224	696	720	516	476	120	508	408

R* = sum of the two reference standards (for going-up calibrations)

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	5	4	4	3	2	2	1	1
5	1.2095	+							
4	0.8610		+						
4	0.9246			+					
3	0.9204				+				
2	0.8456					+			
2	1.4444						+		
1	0.5975							+	
1	0.5975								+
4	1.5818					+	+		
7	1.7620				+	+	+		
11	2.5981			+	+	+	+		
15	3.3153		+	+	+	+	+		
20	4.4809	+	+	+	+	+	+		
0	1.1950							+	-

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT	FACTOR	5	4	4	3	2	2	1	1
5	2.1380	+							
4	1.4679		+						
4	1.4952			+					
3	1.2785				+				
2	1.2410					+			
2	1.0170						+		
1	0.7113							+	
1	0.7113								+
4	1.6872					+	+		
7	2.4387				+	+	+		
11	3.4641			+	+	+	+		
15	4.4981		+	+	+	+	+		
20	6.2893	+	+	+	+	+	+		
0	1.4226							+	-

Explanation of notation and interpretation of tables

2.3.4.1.14. Design for 5,5,2,2,1,1,1,1

Design 5,5,2,2,1,1,1,1

OBSERVATIONS	5	5	2	2	1	1	1	1
Y(1)	+		-	-	-			
Y(2)		+	-	-		-		
Y(3)	+		-	-			-	
Y(4)		+	-	-				-
Y(5)			+	+	-	-	-	-
Y(6)			+		-		-	
Y(7)			+			-		-
Y(8)				+	-			-
Y(9)				+		-	-	
Y(10)					+		-	
Y(11)						+		-
RESTRAINT	+	+						
CHECK STANDARD							+	

DEGREES OF FREEDOM = 4

SOLUTION MATRIX
DIVISOR = 120

OBSERVATIONS	5	5	2	2	1	1
Y(1)	30	-30	-12	-12	-22	-10
Y(2)	-30	30	-12	-12	-10	-22
Y(3)	30	-30	-12	-12	10	-2
Y(4)	-30	30	-12	-12	-2	10
Y(5)	0	0	6	6	-12	-12
Y(6)	-30	30	33	-27	-36	24
Y(7)	30	-30	33	-27	24	-36
Y(8)	0	0	-27	33	-18	6
Y(9)	0	0	-27	33	6	-18
Y(10)	0	0	0	0	32	8
Y(11)	0	0	0	0	8	32
R*	60	60	24	24	12	12

1	1
10	-2
-2	10
22	-10
10	-22
12	-12
36	-24
24	-36
6	-18
18	-6
32	-8
-8	-32
12	12

R* = sum of the two reference standards

FACTORS FOR COMPUTING REPEATABILITY STANDARD DEVIATIONS

WT	FACTOR	5	5	2	2	1	1	1	1
5	0.6124	+							
5	0.6124		+						
2	0.5431			+					
2	0.5431				+				
1	0.5370					+			
1	0.5370						+		
1	0.5370							+	
1	0.5370								+
2	0.6733					+	+		
4	0.8879				+	+	+		
6	0.8446			+	+	+	+		
11	1.0432		+	+	+	+	+		
16	0.8446	+	+	+	+	+	+		
1	0.5370							+	

FACTORS FOR COMPUTING LEVEL-2 STANDARD DEVIATIONS

WT	FACTOR	5	5	2	2	1	1	1	1
5	0.7071	+							
5	0.7071		+						
2	1.0392			+					
2	1.0392				+				
1	1.0100					+			
1	1.0100						+		
1	1.0100							+	
1	1.0100								+
2	1.4422					+	+		
4	1.8221				+	+	+		
6	2.1726			+	+	+	+		
11	2.2847		+	+	+	+	+		
16	2.1726	+	+	+	+	+	+		
1	1.0100							+	

Explanation of notation and interpretation of tables

2.3.4.1.15. Design for 5,5,3,2,1,1,1

```
 OBSERVATIONS    5    5    3
2    1    1    1

     Y(1)        +         -
-
     Y(2)             +    -
-
     Y(3)        +
-    -    -    -
     Y(4)             +
-    -    -    -
     Y(5)        +         -
-    -
     Y(6)        +         -
-         -
     Y(7)        +         -
-    -
     Y(8)             +    -
-    -
     Y(9)             +    -
-         -
     Y(10)            +    -
-    -

 RESTRAINT       +    +

 CHECK STANDARD
+

 DEGREES OF FREEDOM  =
4


SOLUTION MATRIX

DIVISOR =  10

 OBSERVATIONS       5
5       3       2       1
1       1


    Y(1)             1      -
1      -2      -3       1
1       1
    Y(2)            -1
1      -2      -3       1
1       1
    Y(3)             1      -
1       2      -2      -1
-1      -1
    Y(4)            -1
1       2      -2      -1
-1      -1
    Y(5)             1      -
1      -1       1      -2
```

```
-2      3
    Y(6)            1     -
1     -1      1     -2
3     -2
    Y(7)            1     -
1     -1      1       3
-2     -2
    Y(8)           -1
1     -1      1     -2
-2      3
    Y(9)           -1
1     -1      1     -2
3     -2
    Y(10)          -1
1     -1      1       3
-2     -2
    R*              5
5      3      2       1
1      1

    R* = sum of the two
reference standards


 FACTORS FOR REPEATABILITY
STANDARD DEVIATIONS
 WT  FACTOR
                 5   5   3
2   1   1   1
  5  0.3162    +
  5  0.3162        +
  3  0.4690            +
  2  0.5657
+
  1  0.6164
+
  1  0.6164
+
  1  0.6164
+
  3  0.7874
+   +
  6  0.8246            +
+   +
 11  0.8832        +   +
+   +
 16  0.8246    +   +   +
+   +
  1  0.6164
+


 FACTORS FOR BETWEEN-DAY
STANDARD DEVIATIONS
 WT  FACTOR
                 5   5   3
2   1   1   1
  5  0.7071    +
  5  0.7071        +
  3  1.0863            +
  2  1.0392
+
  1  1.0100
+
  1  1.0100
+
  1  1.0100
+
  3  1.4765
+   +
  6  1.9287            +
+   +
 11  2.0543        +   +
+   +
 16  1.9287    +   +   +
+   +
  1  1.0100
+
```

[Explanation of notation and interpretation of tables]

2.3.4.1.16. Design for 1,1,1,1,1,1,1,1 weights

```
 OBSERVATIONS    1   1   1   1   1   1   1   1

     Y(1)        +       -
     Y(2)        +           -
     Y(3)        +               -
     Y(4)        +                   -
     Y(5)        +                       -
     Y(6)        +                           -
     Y(7)            +   -
     Y(8)            +       -
     Y(9)            +           -
     Y(10)           +               -
     Y(11)           +                   -
     Y(12)           +                       -

 RESTRAINT       +   +

 CHECK STANDARD                          +

DEGREES OF FREEDOM  =  5
```

SOLUTION MATRIX
DIVISOR = 12

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	1	-1	-6	0	0	0	0	0
Y(2)	1	-1	0	-6	0	0	0	0
Y(3)	1	-1	0	0	-6	0	0	0
Y(4)	1	-1	0	0	0	-6	0	0
Y(5)	1	-1	0	0	0	0	-6	0
Y(6)	1	-1	0	0	0	0	0	-6
Y(7)	-1	1	-6	0	0	0	0	0
Y(8)	-1	1	0	-6	0	0	0	0
Y(9)	-1	1	0	0	-6	0	0	0
Y(10)	-1	1	0	0	0	-6	0	0
Y(11)	-1	1	0	0	0	0	-6	0
Y(12)	-1	1	0	0	0	0	0	-6
R*	6	6	6	6	6	6	6	6

R* = sum of the two reference standards

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT	K1	1	1	1	1	1	1	1	1
1	0.2887	+							
1	0.2887		+						
1	0.7071			+					
1	0.7071				+				
1	0.7071					+			
1	0.7071						+		
1	0.7071							+	
1	0.7071								+
2	1.0000			+	+				
3	1.2247			+	+	+			
4	1.4142			+	+	+	+		
5	1.5811			+	+	+	+	+	
6	1.7321			+	+	+	+	+	+
1	0.7071							+	

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT	K2	1	1	1	1	1	1	1	1
1	0.7071	+							
1	0.7071		+						
1	1.2247			+					
1	1.2247				+				
1	1.2247					+			
1	1.2247						+		
1	1.2247							+	
1	1.2247								+
2	2.0000			+	+				
3	2.7386			+	+	+			
4	3.4641			+	+	+	+		
5	4.1833			+	+	+	+	+	
6	4.8990			+	+	+	+	+	+
1	1.2247							+	

Explanation of notation and interpretation of tables

2.3.4.1.17. Design for 3,2,1,1,1 weights

```
OBSERVATIONS    3   2   1   1   1

    Y(1)        +   -   -
    Y(2)        +   -       -
    Y(3)        +   -           -
    Y(4)        +       -   -   -
    Y(5)            +   -   -
    Y(6)            +   -       -
    Y(7)            +       -   -
    Y(8)                +   -
    Y(9)                +       -
    Y(10)                   +   -

RESTRAINT       +   +

CHECK STANDARD              +

DEGREES OF FREEDOM = 6
```

SOLUTION MATRIX
DIVISOR = 25

OBSERVATIONS	3	2	1	1	1
Y(1)	3	-3	-4	1	1
Y(2)	3	-3	1	-4	1
Y(3)	3	-3	1	1	-4
Y(4)	1	-1	-3	-3	-3
Y(5)	-2	2	-4	-4	1
Y(6)	-2	2	-4	1	-4
Y(7)	-2	2	1	-4	-4
Y(8)	0	0	5	-5	0
Y(9)	0	0	5	0	-5
Y(10)	0	0	0	5	-5
R*	15	10	5	5	5

R* = sum of the two reference standards

```
FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

 WT     K1      3   2   1   1   1
 3   0.2530     +
 2   0.2530         +
 1   0.4195             +
 1   0.4195                 +
 1   0.4195                     +
 2   0.5514             +   +
 3   0.6197             +   +   +
 1   0.4195                 +

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS
```

WT	K2	3	2	1	1	1
3	0.7211	+				
2	0.7211		+			
1	1.0392			+		
1	1.0392				+	
1	1.0392					+
2	1.5232			+	+	
3	1.9287			+	+	+
1	1.0392				+	

[Explanation of notation and interpretation of tables]

2.3.4.1.18. Design for 10-and 20-pound weights

```
OBSERVATIONS    1   2   2   1   1

    Y(1)        +           -
    Y(2)        +               -
    Y(3)        +   -           +
    Y(4)        +       -   +
    Y(5)        +   -       +
    Y(6)        +       -       +
    Y(7)            +   -

RESTRAINT       +

CHECK STANDARD              +

DEGREES OF FREEDOM = 3


                        SOLUTION MATRIX
                          DIVISOR  =   24

OBSERVATIONS       1      2      2      1      1

   Y(1)            0    -12    -12    -16     -8
   Y(2)            0    -12    -12     -8    -16
   Y(3)            0     -9     -3     -4      4
   Y(4)            0     -3     -9      4     -4
   Y(5)            0     -9     -3      4     -4
   Y(6)            0     -3     -9     -4      4
   Y(7)            0      6     -6      0      0
   R*             24     48     48     24     24

   R* = Value of the reference standard

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

WT    K1       1   2   2   1   1
 2  0.9354        +
 2  0.9354            +
 1  0.8165                +
 1  0.8165                    +
 4  1.7321        +   +
 5  2.3805        +   +   +
 6  3.0000        +   +   +   +
 1  0.8165                +

FACTORS FOR BETWEEN-DAY STANDARD DEVIATIONS

WT    K2       1   2   2   1   1
 2  2.2361        +
 2  2.2361            +
 1  1.4142                +
 1  1.4142                    +
 4  4.2426        +   +
 5  5.2915        +   +   +
```

```
6  6.3246        +    +    +    +
1  1.4142                  +
```

[Explanation of notation and interpretation of tables]

2.3.4.2. Drift-elimination designs for gauge blocks

Tie to the defined unit of length

The unit of length in many industries is maintained and disseminated by gauge blocks. The highest accuracy calibrations of gauge blocks are done by laser intererometry which allows the transfer of the unit of length to a gauge piece. Primary standards laboratories maintain master sets of English gauge blocks and metric gauge blocks which are calibrated in this manner. Gauge blocks ranging in sizes from 0.1 to 20 inches are required to support industrial processes in the United States.

Mechanical comparison of gauge blocks

However, the majority of gauge blocks are calibrated by comparison with master gauges using a mechanical comparator specifically designed for measuring the small difference between two blocks of the same nominal length. The measurements are temperature corrected from readings taken directly on the surfaces of the blocks. Measurements on 2 to 20 inch blocks require special handling techniques to minimize thermal effects. A typical calibration involves a set of 81 gauge blocks which are compared one-by-one with master gauges of the same nominal size.

Calibration designs for gauge blocks

Calibration designs allow comparison of several gauge blocks of the same nominal size to one master gauge in a manner that promotes economy of operation and minimizes wear on the master gauge. The calibration design is repeated for each size until measurements on all the blocks in the test sets are completed.

Problem of thermal drift

Measurements on gauge blocks are subject to drift from heat build-up in the comparator. This drift must be accounted for in the calibration experiment or the lengths assigned to the blocks will be contaminated by the drift term.

Elimination of linear drift

The designs in this catalog are constructed so that the solutions are immune to linear drift if the measurements are equally spaced over time. The size of the drift is the average of the n difference measurements. Keeping track of drift from design to design is useful because a marked change from its usual range of values may indicate a problem with the measurement system.

Assumption for Doiron designs

Mechanical measurements on gauge blocks take place successively with one block being inserted into the comparator followed by a second block and so on. This scenario leads to the assumption that the individual measurements are subject to drift (Doiron). Doiron lists designs meeting this criterion which also allow for:

- two master blocks, *R1* and *R2*
- one check standard = difference between *R1* and *R2*
- one - nine test blocks

Properties of drift-elimination designs that use 1 master block

The designs are constructed to:

- Be immune to linear drift
- Minimize the standard deviations for test blocks (as much as possible)
- Spread the measurements on each block throughout the design
- Be completed in 5-10 minutes to keep the drift at the 5 nm level

Caution

Because of the large number of gauge blocks that are being intercompared and the need to eliminate drift, the Doiron designs are not completely balanced with respect to the test blocks. Therefore, the standard deviations are not equal for all blocks. If all the blocks are being calibrated for use in one facility, it is easiest to quote the largest of the standard deviations for all blocks rather than try to maintain a separate record on each block.

Definition of master block and check standard

At the National Institute of Standards and Technology (NIST), the first two blocks in the design are NIST masters which are designated *R1* and *R2*, respectively. The *R1* block is a steel block, and the *R2* block is a chrome-carbide block. If the test blocks are steel, the reference is *R1*; if the test blocks are chrome-carbide, the reference is *R2*. The check standard is always the difference between *R1* and *R2* as estimated from the design and is independent of *R1* and *R2*. The designs are listed in this section of the catalog as:

1. Doiron design for 3 gauge blocks - 6 measurements
2. Doiron design for 3 gauge blocks - 9 measurements
3. Doiron design for 4 gauge blocks - 8 measurements
4. Doiron design for 4 gauge blocks - 12 measurements
5. Doiron design for 5 gauge blocks - 10 measurements
6. Doiron design for 6 gauge blocks - 12 measurements
7. Doiron design for 7 gauge blocks - 14 measurements
8. Doiron design for 8 gauge blocks - 16 measurements
9. Doiron design for 9 gauge blocks - 18 measurements
10. Doiron design for 10 gauge blocks - 20 measurements
11. Doiron design for 11 gauge blocks - 22 measurements

Properties of designs that use 2 master blocks

Historical designs for gauge blocks (Cameron and Hailes) work on the assumption that the difference measurements are contaminated by linear drift. This assumption is more restrictive and covers the case of drift in successive measurements but produces fewer designs. The Cameron/Hailes designs meeting this criterion allow for:

- two reference (master) blocks, *R1* and *R2*
- check standard = difference between the two master blocks

and assign equal uncertainties to values of all test blocks.

The designs are listed in this section of the catalog as:

1. Cameron-Hailes design for 2 masters + 2 test blocks
2. Cameron-Hailes design for 2 masters + 3 test blocks
3. Cameron-Hailes design for 2 masters + 4 test blocks
4. Cameron-Hailes design for 2 masters + 5 test blocks

Important concept - check standard

The check standards for the designs in this section are not artifact standards but constructions from the design. The value of one master block or the average of two master blocks is the restraint for the design, and values for the masters, *R1* and *R2*, are estimated from a set of measurements taken according to the design. The check standard value is the difference between the estimates, *R1* and *R2*. Measurement control is exercised by comparing the current value of the check standard with its historical average.

2.3.4.2.1. Doiron 3-6 Design

```
Doiron 3-6 design

OBSERVATIONS   1   1   1

    Y(1)       +   -
    Y(2)       -       +
    Y(3)           +   -
    Y(4)       -   +
    Y(5)           -   +
    Y(6)       +       -

RESTRAINT      +

CHECK STANDARD     +

DEGREES OF FREEDOM  =      4

                  SOLUTION MATRIX
                   DIVISOR  =  6

OBSERVATIONS     1      1      1

   Y(1)          0      -2     -1
   Y(2)          0       1      2
   Y(3)          0       1     -1
   Y(4)          0       2      1
   Y(5)          0      -1      1
   Y(6)          0      -1     -2
   R*            6       6      6

   R* = Value of the reference standard

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 NOM FACTOR
                1   1   1
  1  0.0000     +
  1  0.5774         +
  1  0.5774             +
  1  0.5774         +
```

Explanation of notation and interpretation of tables

2.3.4.2.2. Doiron 3-9 Design

```
 Doiron 3-9 Design

 OBSERVATIONS   1   1   1

     Y(1)       +   -
     Y(2)           -   +
     Y(3)       +       -
     Y(4)       -   +
     Y(5)       -       +
     Y(6)           +   -
     Y(7)       -   +
     Y(8)           -   +
     Y(9)       +       -

 RESTRAINT      +

 CHECK STANDARD     +

 DEGREES OF FREEDOM  =      7


               SOLUTION MATRIX
                DIVISOR  =  9

 OBSERVATIONS     1       1      1

    Y(1)          0      -2     -1
    Y(2)          0      -1      1
    Y(3)          0      -1     -2
    Y(4)          0       2      1
    Y(5)          0       1      2
    Y(6)          0       1     -1
    Y(7)          0       2      1
    Y(8)          0      -1      1
    Y(9)          0      -1     -2
    R(1)          9       9      9


 FACTORS FOR COMPUTING REPEATABILITY STANDARD
DEVIATIONS
 NOM FACTOR
                1   1   1
  1  0.0000     +
  1  0.4714         +
  1  0.4714             +
  1  0.4714         +
```

Explanation of notation and interpretation of tables

2.3.4.2.3. Doiron 4-8 Design

```
Doiron 4-8 Design

OBSERVATIONS    1    1    1    1

    Y(1)        +    -
    Y(2)                  +    -
    Y(3)        -                +
    Y(4)             +    -
    Y(5)        -    +
    Y(6)                  -    +
    Y(7)        +                -
    Y(8)             -    +

RESTRAINT       +


CHECK STANDARD      +

DEGREES OF FREEDOM  =    5

                SOLUTION MATRIX
                 DIVISOR  =  8

OBSERVATIONS      1        1       1       1

   Y(1)           0       -3      -2      -1
   Y(2)           0        1       2      -1
   Y(3)           0        1       2       3
   Y(4)           0        1      -2      -1
   Y(5)           0        3       2       1
   Y(6)           0       -1      -2       1
   Y(7)           0       -1      -2      -3
   Y(8)           0       -1       2       1
   R*             8        8       8       8

   R* = Value of the reference standard


FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
NOM FACTOR
                1    1    1    1
 1  0.0000     +
 1  0.6124          +
 1  0.7071               +
 1  0.6124                    +
 1  0.6124          +
```

Explanation of notation and interpretation of tables

2.3.4.2.4. Doiron 4-12 Design

Doiron 4-12 Design

OBSERVATIONS	1	1	1	1
Y(1)	+	-		
Y(2)	+			+
Y(3)			+	-
Y(4)	-	+		
Y(5)		+	-	
Y(6)		-		+
Y(7)	+		-	
Y(8)		+		-
Y(9)	+			-
Y(10)		-	+	
Y(11)	-		+	
Y(12)			-	+
RESTRAINT	+			
CHECK STANDARD		+		

DEGREES OF FREEDOM = 9

SOLUTION MATRIX
DIVISOR = 8

OBSERVATIONS	1	1	1	1
Y(1)	0	-2	-1	-1
Y(2)	0	1	1	2
Y(3)	0	0	1	-1
Y(4)	0	2	1	1
Y(5)	0	1	-1	0
Y(6)	0	-1	0	1
Y(7)	0	-1	-2	-1
Y(8)	0	1	0	-1
Y(9)	0	-1	-1	-2
Y(10)	0	-1	1	0
Y(11)	0	1	2	1
Y(12)	0	0	-1	1
R*	6	6	6	4

R* = Value of the reference standard

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

NOM	FACTOR	1	1	1	1
1	0.0000	+			
1	0.5000		+		
1	0.5000			+	
1	0.5000				+
1	0.5000		+		

Explanation of notation and interpretation of tables

2.3.4.2.5. Doiron 5-10 Design

Doiron 5-10 Design

OBSERVATIONS	1	1	1	1	1
Y(1)	+	-			
Y(2)			-	+	
Y(3)	+		-		
Y(4)		-			+
Y(5)			-	+	
Y(6)	+			-	
Y(7)	-		+		
Y(8)			+		-
Y(9)	-			+	
Y(10)		+			-
RESTRAINT	+				
CHECK STANDARD		+			

DEGREES OF FREEDOM = 6

SOLUTION MATRIX
DIVISOR = 90

OBSERVATIONS	1	1	1	1	1
Y(1)	0	-50	-10	-10	-30
Y(2)	0	20	4	-14	30
Y(3)	0	-10	-29	-11	-15
Y(4)	0	-20	5	5	15
Y(5)	0	0	-18	18	0
Y(6)	0	-10	-11	-29	-15
Y(7)	0	10	29	11	15
Y(8)	0	-20	14	-4	-30
Y(9)	0	10	11	29	15
Y(10)	0	20	-5	-5	-15
R*	90	90	90	90	90

R* = Value of the reference standard

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

NOM	FACTOR	1	1	1	1	1
1	0.0000	+				
1	0.7454		+			
1	0.5676			+		
1	0.5676				+	
1	0.7071					+
1	0.7454		+			

Explanation of notation and interpretation of tables

2.3.4.2.6. Doiron 6-12 Design

Doiron 6-12 Design

OBSERVATIONS	1	1	1	1	1	1
Y(1)	+	-				
Y(2)			-	+		
Y(3)		-			+	
Y(4)				-	+	
Y(5)		-		+		
Y(6)			+	-		
Y(7)	+				-	
Y(8)		+		-		
Y(9)		+	-			
Y(10)	-			+		
Y(11)		+			-	
Y(12)	-		+			
RESTRAINT	+					
CHECK STANDARD		+				

DEGREES OF FREEDOM = 7

SOLUTION MATRIX
DIVISOR = 360

OBSERVATIONS	1	1	1	1	1	1
Y(1)	0	-136	-96	-76	-72	-76
Y(2)	0	-4	-24	-79	72	11
Y(3)	0	-20	-120	-35	0	55
Y(4)	0	4	24	-11	-72	79
Y(5)	0	-60	0	75	0	-15
Y(6)	0	20	120	-55	0	35
Y(7)	0	-76	-96	-61	-72	-151
Y(8)	0	64	24	4	-72	4
Y(9)	0	40	-120	-20	0	-20
Y(10)	0	72	72	72	144	72
Y(11)	0	60	0	15	0	-75
Y(12)	0	76	96	151	72	61
R*	360	360	360	360	360	360

R* = Value of the reference standard

```
FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
NOM FACTOR
                1   1   1   1   1   1
 1  0.0000      +
 1  0.6146          +
 1  0.7746              +
 1  0.6476                  +
 1  0.6325                      +
 1  0.6476                          +
 1  0.6146          +
```

Explanation of notation and interpretation of tables

2.3.4.2.7. Doiron 7-14 Design

Doiron 7-14 Design

OBSERVATIONS	1	1	1	1	1	1	1
Y(1)	+	-					
Y(2)				-		+	
Y(3)			+		-		
Y(4)		+					-
Y(5)	+					-	
Y(6)			-		+		
Y(7)		+		-			
Y(8)			+				-
Y(9)					+	-	
Y(10)	-						+
Y(11)		-				+	
Y(12)			-	+			
Y(13)	-			+			
Y(14)					-		+

RESTRAINT +

CHECK STANDARD +

DEGREES OF FREEDOM = 8

PARAMETER VALUES
DIVISOR = 1015

OBSERVATIONS	1	1	1	1	1	1	1
Y(1)	0	-406	-203	-203	-203	-203	-203
Y(2)	0	0	-35	-210	35	210	0
Y(3)	0	0	175	35	-175	-35	0
Y(4)	0	203	-116	29	-116	29	-261
Y(5)	0	-203	-229	-214	-264	-424	-174
Y(6)	0	0	-175	-35	175	35	0
Y(7)	0	203	-61	-221	-26	-11	29
Y(8)	0	0	305	90	130	55	-145
Y(9)	0	0	220	15	360	-160	145
Y(10)	0	203	319	174	319	174	464
Y(11)	0	-203	26	11	61	221	-29
Y(12)	0	0	-360	160	-220	-15	-145
Y(13)	0	203	264	424	229	214	174
Y(14)	0	0	-130	-55	-305		

```
-90    145
    R*        1015      1015   1015   1015   1015
1015   1015

    R* = Value of the reference standard


 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 NOM FACTOR
                1   1   1   1   1   1   1
  1  0.0000     +
  1  0.6325         +
  1  0.7841             +
  1  0.6463                 +
  1  0.7841                     +
  1  0.6463                         +
  1  0.6761                             +
  1  0.6325         +
```

Explanation of notation and interpretation of tables

2.3.4.2.8. Doiron 8-16 Design

```
 Doiron 8-16 Design

 OBSERVATIONS   1   1   1   1   1   1   1   1

     Y(1)       +   -
     Y(2)                       +       -
     Y(3)               -               +
     Y(4)       -               +
     Y(5)           +           -
     Y(6)                               -   +
     Y(7)           -   +
     Y(8)       -                   +
     Y(9)                       -       +
     Y(10)      -           +
     Y(11)          +                       -
     Y(12)              -       +
     Y(13)      -           +
     Y(14)                  -               +
     Y(15)          +       -
     Y(16)          +                   -

 RESTRAINT      +

 CHECK STANDARD     +

 DEGREES OF FREEDOM  =      9

                          SOLUTION MATRIX
                           DIVISOR = 2852

 OBSERVATIONS  1          1      1      1      1
1      1      1

    Y(1)       0      -1392   -620   -472   -516
-976   -824   -916
    Y(2)       0         60    248    -78     96
878   -112   -526
    Y(3)       0        352    124   -315    278
255    864    289
    Y(4)       0        516    992    470   1396
706    748    610
    Y(5)       0       -356    620     35    286
-979    -96   -349
    Y(6)       0         92      0     23   -138
253   -552    667
    Y(7)       0       -148   -992    335   -522
-407   -104    -81
    Y(8)       0       -416    372    113    190
995     16    177
    Y(9)       0        308   -248    170   -648
134    756    342
    Y(10)      0        472    620    955    470
585    640    663
    Y(11)      0        476   -124   -191    -94
-117   -128   -703
    Y(12)      0       -104   -620   -150    404
-286      4   -134
```

```
    Y(13)      0         472    620    955    470
585    640    663
    Y(14)      0         444    124   -292    140
508    312    956
    Y(15)      0         104    620    150   -404
286     -4    134
    Y(16)      0         568   -124   -168   -232
136   -680    -36
    R*      2852        2852   2852   2852   2852
2852   2852   2852

    R* = value of reference block


 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 WT  FACTOR
                1   1   1   1   1   1   1   1
  1  0.0000     +
  1  0.6986         +
  1  0.7518             +
  1  0.5787                 +
  1  0.6996                     +
  1  0.8313                         +
  1  0.7262                             +
  1  0.7534                                 +
  1  0.6986         +
```

[Explanation of notation and interpretation of tables]

2.3.4.2.9. Doiron 9-18 Design

```
Doiron 9-18 Design

OBSERVATIONS   1   1   1   1   1   1   1   1   1

    Y(1)       +   -
    Y(2)                                   -       +
    Y(3)           +   -
    Y(4)                   -   +
    Y(5)                           +       -
    Y(6)           -       +
    Y(7)               +                   -
    Y(8)                                   +   -
    Y(9)       -               +
    Y(10)                  +       -
    Y(11)      -                           +
    Y(12)                          -       +
    Y(13)      -           +
    Y(14)          +       -
    Y(15)                      -       +
    Y(16)      +                           -
    Y(17)              -           +
    Y(18)                              +   -

RESTRAINT      +

CHECK STANDARD     +

DEGREES OF FREEDOM  =     10

                          SOLUTION MATRIX
                           DIVISOR = 8247

 OBSERVATIONS   1         1      1      1      1
1      1      1      1

    Y(1)         0     -3680  -2305  -2084  -1175   -
1885  -1350  -1266   -654
    Y(2)         0      -696  -1422   -681  -1029
-984  -2586   -849   1203
    Y(3)         0      1375  -3139    196   -491   -
1279  -1266   -894   -540
    Y(4)         0      -909   -222  -1707   1962
-432    675    633    327
    Y(5)         0       619   1004    736   -329
2771   -378  -1674   -513
    Y(6)         0     -1596   -417   1140    342
303     42    186     57
    Y(7)         0       955   2828    496   -401
971  -1689   -411   -525
    Y(8)         0       612    966    741   1047
1434    852   2595  -1200
    Y(9)         0      1175   1666   1517   3479
1756   2067   2085   1038
    Y(10)        0       199  -1276   1036   -239   -
3226   -801  -1191   -498
    Y(11)        0       654   1194    711   1038
1209   1719   1722   2922
```

```
    Y(12)       0        91    494    -65  -1394
887    504   2232    684
    Y(13)       0      2084   1888   3224   1517
2188   1392   1452    711
    Y(14)       0      1596    417  -1140   -342
-303    -42   -186    -57
    Y(15)       0       175    950   -125  -1412
437   2238    486    681
    Y(16)       0      -654  -1194   -711  -1038  -
1209  -1719  -1722  -2922
    Y(17)       0      -420  -2280    300     90
2250   -423    483     15
    Y(18)       0        84    456    -60    -18
-450   1734  -1746     -3
    R*       8247      8247   8247   8247   8247
8247   8247   8247   8247

    R* = Value of the reference standard


 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 NOM FACTOR
                1   1   1   1   1   1   1   1   1
  1  0.0000    +
  1  0.6680        +
  1  0.8125            +
  1  0.6252                +
  1  0.6495                    +
  1  0.8102                        +
  1  0.7225                            +
  1  0.7235                                +
  1  0.5952                                    +
  1  0.6680        +
```

[Explanation of notation and interpretation of tables]

2.3.4.2.10. Doiron 10-20 Design

```
 Doiron 10-20 Design

 OBSERVATIONS   1   1   1   1   1   1   1   1   1
1

     Y(1)       +   -
     Y(2)                           +   -
     Y(3)                   -
+
     Y(4)                       +   -
     Y(5)           +                       -
     Y(6)               +   -
     Y(7)                               +   -
     Y(8)       -
+
     Y(9)                           +   -
     Y(10)                                  +
-
     Y(11)                      +       -
     Y(12)          +   -
     Y(13)              +
-
     Y(14)                          -           +
     Y(15)              +                   -
     Y(16)      +               -
     Y(17)              -                   +
     Y(18)                  +   -
     Y(19)      -                           +
     Y(20)          -               +

 RESTRAINT      +

 CHECK STANDARD     +

 DEGREES OF FREEDOM  =     11

                            SOLUTION MATRIX
                          DIVISOR  =  33360

 OBSERVATIONS 1          1      1      1      1
1      1      1      1

    Y(1)       0     -15300  -9030  -6540  -5970  -
9570  -7770  -6510  -9240
    Y(2)       0       1260   1594   1716   3566
3470   9078  -5678    -24
    Y(3)       0       -960  -2856  -7344  -2664  -
1320  -1992  -1128    336
    Y(4)       0      -3600  -1536    816   5856  -
9120  -1632  -1728  -3744
    Y(5)       0       6060    306  -1596   -906  -
1050   -978  -2262  -8376
    Y(6)       0       2490   8207  -8682  -1187
1165   2769   2891    588
    Y(7)       0      -2730    809  -1494   -869  -
2885    903   6557  -8844
    Y(8)       0       5580   7218  11412   6102
```

```
6630   6366   5514   8472
    Y(9)      0      1800  -2012   -408   -148
7340  -7524  -1916   1872
    Y(10)     0      3660   1506  -3276    774
3990   2382   3258   9144
    Y(11)     0     -1800  -3548    408   5708  -
1780  -9156  -3644  -1872
    Y(12)     0      6270  -9251  -3534  -1609
455  -3357  -3023    516
    Y(13)     0       960   2856   7344   2664
1320   1992   1128   -336
    Y(14)     0      -330   -391    186  -2549  -
7925  -2457   1037   6996
    Y(15)     0      2520   8748   3432   1572
1380   1476  -5796    -48
    Y(16)     0     -5970  -7579  -8766 -15281  -
9425  -9573  -6007  -6876
    Y(17)     0     -1260  -7154  -1716   1994
2090   7602    118     24
    Y(18)     0       570   2495   9990  -6515  -
1475  -1215    635   1260
    Y(19)     0      6510   9533   6642   6007
7735   9651  15329   8772
    Y(20)     0     -5730     85   1410   3455
8975   3435   1225   1380
    R*    33360      33360  33360  33360  33360
33360  33360  33360  33360

    R* = Value of the reference standard


 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 NOM FACTOR
                1   1   1   1   1   1   1   1   1
1
  1  0.0000    +
  1  0.6772        +
  1  0.7403            +
  1  0.7498                +
  1  0.6768                    +
  1  0.7456                        +
  1  0.7493                            +
  1  0.6779                                +
  1  0.7267                                    +
  1  0.6961
+
  1  0.6772        +
```

[Explanation of notation and interpretation of tables]

2.3.4.2.11. Doiron 11-22 Design

```
 Doiron 11-22 Design

 OBSERVATIONS   1   1   1   1   1   1   1   1   1
1   1

     Y(1)       +   -
     Y(2)                       +   -
     Y(3)                               +
-
     Y(4)                   +                   -
     Y(5)           +   -
     Y(6)
+   -
     Y(7)                           -       +
     Y(8)       -               +
     Y(9)               +   -
     Y(10)      +                   -
     Y(11)          +                       -
     Y(12)              -                   +
     Y(13)                          +   -
     Y(14)          -
+
     Y(15)                  +   -
     Y(16)                                  +
-
     Y(17)                              +   -
     Y(18)      -
+
     Y(19)              +                       -
     Y(20)                      -           +
     Y(21)                  -
+
     Y(22)                          +
-

 RESTRAINT      +


 CHECK STANDARD     +


 DEGREES OF FREEDOM  =     12


                            SOLUTION MATRIX
                            DIVISOR = 55858

 OBSERVATIONS      1      1      1      1      1
1      1      1      1      1

    Y(1)      0    -26752 -18392 -15532  -9944  -
8778 -14784 -15466 -16500 -10384 -17292
    Y(2)      0      1166   1119   3976  12644 -
11757  -1761   2499   1095  -2053   1046
    Y(3)      0      5082   4446   3293   4712
160   5882  15395   3527  -9954    487
    Y(4)      0      -968  -1935  10496   2246   -
635  -4143   -877 -13125   -643  -1060
    Y(5)      0      8360 -18373  -8476  -3240  -
3287  -8075  -1197  -9443  -1833  -2848
    Y(6)      0     -6908  -7923  -9807  -2668
```

```
431  -4753  -1296 -10224   9145 -18413
    Y(7)      0      1716   3084   6091    404  -
2452 -10544  -2023  15073    332   5803
    Y(8)      0      9944  13184  15896  24476
11832  13246  14318  13650   9606  12274
    Y(9)      0      2860  12757 -11853  -2712
145   3585    860    578   -293  -2177
    Y(10)     0     -8778 -12065 -11920 -11832 -
23589 -15007 -11819 -12555 -11659 -11228
    Y(11)     0     11286   1729   -271  -4374  -
3041  -3919 -14184   -180  -3871   1741
    Y(12)     0     -3608 -13906  -4734     62
2942  11102   2040  -2526    604  -2566
    Y(13)     0     -6006 -10794  -7354  -1414
8582 -18954  -6884 -10862  -1162  -6346
    Y(14)     0     -9460   1748   6785   2330
2450   2790     85   6877   4680  16185
    Y(15)     0      5588  10824  19965  -8580
88   6028   1485  11715   2904  10043
    Y(16)     0      -792   5803   3048   1376
1327   5843   1129  15113  -1911 -10100
    Y(17)     0      -682   6196   3471  -1072
3188  15258 -10947   6737  -1434   2023
    Y(18)     0     10384  12217  12510   9606
11659  12821  14255  13153  24209  15064
    Y(19)     0      1892  10822  -1357   -466   -
490   -558    -17 -12547   -936  -3237
    Y(20)     0      5522   3479    -93 -10158    -
13   5457  15332   3030   4649   3277
    Y(21)     0      1760  -3868 -13544  -3622   -
692  -1700   -252  -1988   2554  11160
    Y(22)     0     -1606   -152   -590   2226
11930   2186  -2436   -598 -12550  -3836
    R*    55858     55858  55858  55858  55858
55858  55858  55858  55858  55858  55858

    R* = Value of the reference standard

 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 NOM FACTOR
                1   1   1   1   1   1   1   1   1
1   1
  1  0.0000    +
  1  0.6920        +
  1  0.8113            +
  1  0.8013                +
  1  0.6620                    +
  1  0.6498                        +
  1  0.7797                            +
  1  0.7286                                +
  1  0.8301                                    +
  1  0.6583
+
  1  0.6920        +
```

Explanation of notation and interpretation of tables

2.3.4.3. Designs for electrical quantities

Standard cells

Banks of saturated standard cells that are nominally one volt are the basis for maintaining the unit of voltage in many laboratories.

Bias problem

It has been observed that potentiometer measurements of the difference between two saturated standard cells, connected in series opposition, are effected by a thermal emf which remains constant even when the direction of the circuit is reversed.

Designs for eliminating bias

A calibration design for comparing standard cells can be constructed to be left-right balanced so that:

- A constant bias, P, does not contaminate the estimates for the individual cells.
- P is estimated as the average of difference measurements.

Designs for electrical quantities

Designs are given for the following classes of electrical artifacts. These designs are left-right balanced and may be appropriate for artifacts other than electrical standards.

- Saturated standard reference cells
- Saturated standard test cells
- Zeners
- Resistors

Standard cells in a single box

Left-right balanced designs for comparing standard cells among themselves where the restraint is over all reference cells are listed below. These designs are **not** appropriate for assigning values to test cells.

Estimates for individual standard cells and the bias term, P, are shown under the heading, 'SOLUTION MATRIX'. These designs also have the advantage of requiring a change of connections to only one cell at a time.

1. Design for 3 standard cells
2. Design for 4 standard cells
3. Design for 5 standard cells
4. Design for 6 standard cells

Test cells Calibration designs for assigning values to test cells in a common environment on the basis of comparisons with reference cells with known values are shown below. The designs in this catalog are left-right balanced.

1. Design for 4 test cells and 4 reference cells
2. Design for 8 test cells and 8 reference cells

Zeners Increasingly, zeners are replacing saturated standard cells as artifacts for maintaining and disseminating the volt. Values are assigned to test zeners, based on a group of reference zeners, using calibration designs.

1. Design for 4 reference zeners and 2 test zeners
2. Design for 4 reference zeners and 3 test zeners

Standard resistors Designs for comparing standard resistors that are used for maintaining and disseminating the ohm are listed in this section.

1. Design for 3 reference resistors and 1 test resistor
2. Design for 4 reference resistors and 1 test resistor

2.3.4.3.1. Left-right balanced design for 3 standard cells

Design 1,1,1

```
                  CELLS
OBSERVATIONS    1   1   1
    Y(1)        +   -
    Y(2)        +       -
    Y(3)            +   -
    Y(4)        -   +
    Y(5)        -       +
    Y(6)            -   +

RESTRAINT       +   +   +

DEGREES OF FREEDOM  =  3
```

```
                      SOLUTION MATRIX
                        DIVISOR = 6
OBSERVATIONS       1      1      1      P
   Y(1)            1     -1      0      1
   Y(2)            1      0     -1      1
   Y(3)            0      1     -1      1
   Y(4)           -1      1      0      1
   Y(5)           -1      0      1      1
   Y(6)            0     -1      1      1
   R*              2      2      2      0

   R* = AVERAGE VALUE OF 3 REFERENCE CELLS

   P = LEFT-RIGHT BIAS
```

```
FACTORS FOR COMPUTING STANDARD DEVIATIONS
  V  FACTOR      CELLS
                1   1   1
  1  0.3333     +
  1  0.3333         +
  1  0.3333             +
```

Explanation of notation and interpretation of tables

2.3.4.3.2. Left-right balanced design for 4 standard cells

Design 1,1,1,1

```
OBSERVATIONS    1   1   1   1
    Y(1)        +   -
    Y(2)        +       -
    Y(3)            +   -
    Y(4)            +       -
    Y(5)                +   -
    Y(6)        -       +
    Y(7)            -   +
    Y(8)            -       +
    Y(9)        -           +
    Y(10)               -   +
    Y(11)       -   +
    Y(12)       +           -

RESTRAINT       +   +   +   +


DEGREES OF FREEDOM  =  8
```

SOLUTION MATRIX
DIVISOR = 8

OBSERVATIONS	1	1	1	1	P
Y(1)	1	-1	0	0	1
Y(2)	1	0	-1	0	1
Y(3)	0	1	-1	0	1
Y(4)	0	1	0	-1	1
Y(5)	0	0	1	-1	1
Y(6)	-1	0	1	0	1
Y(7)	0	-1	1	0	1
Y(8)	0	-1	0	1	1
Y(9)	-1	0	0	1	1
Y(10)	0	0	-1	1	1
Y(11)	-1	1	0	0	1
Y(12)	1	0	0	-1	1
R*	2	2	2	2	0

R* = AVERAGE VALUE OF 4 REFERENCE CELLS

P = LEFT-RIGHT BIAS

FACTORS FOR COMPUTING STANDARD DEVIATIONS

V	FACTOR	CELLS			
		1	1	1	1
1	0.3062	+			
1	0.3062		+		
1	0.3062			+	
1	0.3062				+

Explanation of notation and interpretation of tables

2.3.4.3.3. Left-right balanced design for 5 standard cells

Design 1,1,1,1,1

```
OBSERVATIONS     1   1   1   1   1

      Y(1)       +   -
      Y(2)       +       -
      Y(3)           +   -
      Y(4)           +       -
      Y(5)               +   -
      Y(6)               +       -
      Y(7)                   +   -
      Y(8)       -           +
      Y(9)       -                   +
      Y(10)          -               +

  RESTRAINT      +   +   +   +   +

  DEGREES OF FREEDOM  =  5
```

SOLUTION MATRIX
DIVISOR = 5

OBSERVATIONS	1	1	1	1	1	P
Y(1)	1	-1	0	0	0	1
Y(2)	1	0	-1	0	0	1
Y(3)	0	1	-1	0	0	1
Y(4)	0	1	0	-1	0	1
Y(5)	0	0	1	-1	0	1
Y(6)	0	0	1	0	-1	1
Y(7)	0	0	0	1	-1	1
Y(8)	-1	0	0	1	0	1
Y(9)	-1	0	0	0	1	1
Y(10)	0	-1	0	0	1	1
R*	1	1	1	1	1	0

R* = AVERAGE VALUE OF 5 REFERENCE CELLS

P = LEFT-RIGHT BIAS

FACTORS FOR COMPUTING REPEATABILITY STANDARD DEVIATIONS

V	FACTOR	CELLS 1	1	1	1	1
1	0.4000	+				
1	0.4000		+			
1	0.4000			+		
1	0.4000				+	
1	0.4000					+

Explanation of notation and interpretation of tables

2.3.4.3.4. Left-right balanced design for 6 standard cells

Design 1,1,1,1,1,1

```
                        CELLS
OBSERVATIONS     1   1   1   1   1   1
     Y(1)        +   -
     Y(2)        +       -
     Y(3)            +   -
     Y(4)            +       -
     Y(5)                +   -
     Y(6)                +       -
     Y(7)                    +   -
     Y(8)                    +       -
     Y(9)                        +   -
     Y(10)       -               +
     Y(11)       -                   +
     Y(12)           -               +
     Y(13)       +               -
     Y(14)           +               -
     Y(15)               +               -

 RESTRAINT       +   +   +   +   +   +

 DEGREES OF FREEDOM = 9
```

SOLUTION MATRIX
DIVISOR = 6

OBSERVATIONS	1	1	1	1	1	1
Y(1)	1	-1	0	0	0	0
Y(2)	1	0	-1	0	0	0
Y(3)	0	1	-1	0	0	0
Y(4)	0	1	0	-1	0	0
Y(5)	0	0	1	-1	0	0
Y(6)	0	0	1	0	-1	0
Y(7)	0	0	0	1	-1	0
Y(8)	0	0	0	1	0	-1
Y(9)	0	0	0	0	1	-1
Y(10)	-1	0	0	0	1	0
Y(11)	-1	0	0	0	0	1
Y(12)	0	-1	0	0	0	1

P
1
1
1
1
1
1
1
1
1
1

```
    Y(13)          1      0      0     -1      0      0
1
    Y(14)          0      1      0      0     -1      0
1
    Y(15)          0      0      1      0      0     -1
1
    R*             1      1      1      1      1      1
0

    R* = AVERAGE VALUE OF 6 REFERENCE CELLS

    P = LEFT-RIGHT BIAS


 FACTORS FOR COMPUTING STANDARD DEVIATIONS
  V  FACTOR            CELLS
                  1   1   1   1   1   1
  1  0.3727      +
  1  0.3727          +
  1  0.3727              +
  1  0.3727                  +
  1  0.3727                      +
  1  0.3727                          +
```

Explanation of notation and interpretation of tables

2.3.4.3.5. Left-right balanced design for 4 references and 4 test items

Design for 4 references and 4 test items.

```
 OBSERVATIONS    1   1   1   1   1   1   1   1

     Y(1)        +                   -
     Y(2)        +                           -
     Y(3)                +                   -
     Y(4)                +           -
     Y(5)            +                   -
     Y(6)            +                           -
     Y(7)                        +               -
     Y(8)                        +           -
     Y(9)        -                           +
     Y(10)       -                                   +
     Y(11)               -                           +
     Y(12)               -                   +
     Y(13)           -               +
     Y(14)           -                           +
     Y(15)                       -               +
     Y(16)                       -   +

 RESTRAINT       +   +   +   +

 DEGREES OF FREEDOM = 8
```

SOLUTION MATRIX
DIVISOR = 16

OBSERVATIONS	1	1	1	1	1	1	1	P
Y(1)	3	-1	-1	-1	-4	0	0	1
Y(2)	3	-1	-1	-1	0	0	-4	1
Y(3)	-1	-1	3	-1	0	0	-4	1
Y(4)	-1	-1	3	-1	-4	0	0	1
Y(5)	-1	3	-1	-1	0	-4	0	1
Y(6)	-1	3	-1	-1	0	0	0	1
Y(7)	-1	-1	-1	3	0	0	0	1
Y(8)	-1	-1	-1	3	0	-4	0	1
Y(9)	-3	1	1	1	0	4	0	1
Y(10)	-3	1	1	1	0	0	0	1
Y(11)	1	1	-3	1	0	0	0	1
Y(12)	1	1	-3	1	0	4	0	1

```
0      0     1
    Y(13)          1     -3      1      1      4      0
0      0     1
    Y(14)          1     -3      1      1      0      0
4      0     1
    Y(15)          1      1      1     -3      0      0
4      0     1
    Y(16)          1      1      1     -3      4      0
0      0     1
    R*             4      4      4      4      4      4
4      4     0

    R* = AVERAGE VALUE OF REFERENCE CELLS

    P = ESTIMATE OF LEFT-RIGHT BIAS


 FACTORS FOR COMPUTING STANDARD DEVIATIONS
  V  FACTORS              CELLS
                    1   1   1   1   1   1   1   1
  1  0.4330         +
  1  0.4330             +
  1  0.4330                 +
  1  0.4330                     +
  1  0.5000                         +
  1  0.5000                             +
  1  0.5000                                 +
  1  0.5000                                     +
```

[Explanation of notation and interpretation of tables]

2.3.4.3.6. Design for 8 references and 8 test items

Design for 8 references and 8 test items.

	TEST CELLS								REFERENCE		
CELLS											
OBSERVATIONS	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1							
Y(1)	+								-		
Y(2)	-									+	
Y(3)		-									+
Y(4)		+									
-											
Y(5)			+								
-											
Y(6)			-								
+											
Y(7)				-							
+											
Y(8)				+							
-											
Y(9)					+				-		
Y(10)						+				-	
Y(11)						-					+
Y(12)							-				
+											
Y(13)							+				
-											
Y(14)								+			
-											
Y(15)								-			
+											
Y(16)					-						
+											
RESTRAINT									+	+	+
+	+	+	+	+							

DEGREES OF FREEDOM = 0

SOLUTION MATRIX FOR TEST CELLS
DIVISOR = 16

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	8	4	0	-4	-6	6	2	-2
Y(2)	-8	4	0	-4	-6	6	2	-2
Y(3)	4	-8	-4	0	2	6	-6	-2
Y(4)	4	8	-4	0	2	6	-6	-2
Y(5)	0	-4	8	4	2	-2	-6	6
Y(6)	0	-4	-8	4	2	-2	-6	6
Y(7)	-4	0	4	-8	-6	-2	2	6
Y(8)	-4	0	4	8	-6	-2	2	6

```
2       6
    Y(9)           -6      -2       2       6       8      -4
0       4
    Y(10)          -6       6       2      -2      -4       8
4       0
    Y(11)          -6       6       2      -2      -4      -8
4       0
    Y(12)           2       6      -6      -2       0       4
-8      -4
    Y(13)           2       6      -6      -2       0       4
8      -4
    Y(14)           2      -2      -6       6       4       0
-4      8
    Y(15)           2      -2      -6       6       4       0
-4     -8
    Y(16)          -6      -2       2       6      -8      -4
0       4
    R               2       2       2       2       2       2
2       2

                          SOLUTION MATRIX FOR REFERENCE
CELLS
                                      DIVISOR = 16
 OBSERVATIONS       1       1       1       1       1       1
1       1      P

    Y(1)           -7       7       5       3       1      -1
-3      -5      1
    Y(2)           -7       7       5       3       1      -1
-3      -5      1
    Y(3)            3       5       7      -7      -5      -3
-1       1      1
    Y(4)            3       5       7      -7      -5      -3
-1       1      1
    Y(5)            1      -1      -3      -5      -7       7
5       3      1
    Y(6)            1      -1      -3      -5      -7       7
5       3      1
    Y(7)           -5      -3      -1       1       3       5
7      -7      1
    Y(8)           -5      -3      -1       1       3       5
7      -7      1
    Y(9)           -7      -5      -3      -1       1       3
5       7      1
    Y(10)          -5      -7       7       5       3       1
-1      -3      1
    Y(11)          -5      -7       7       5       3       1
-1      -3      1
    Y(12)           1       3       5       7      -7      -5
-3      -1      1
    Y(13)           1       3       5       7      -7      -5
-3      -1      1
    Y(14)           3       1      -1      -3      -5      -7
7       5      1
    Y(15)           3       1      -1      -3      -5      -7
7       5      1
    Y(16)          -7      -5      -3      -1       1       3
5       7      1
    R*              2       2       2       2       2       2
2       2      0

    R* = AVERAGE VALUE OF 8 REFERENCE CELLS

    P = ESTIMATE OF LEFT-RIGHT BIAS


FACTORS FOR COMPUTING STANDARD DEVIATIONS FOR TEST CELLS
  V  FACTORS           TEST CELLS
                  1   1   1   1   1   1   1   1
  1  1.1726       +
  1  1.1726           +
  1  1.1726               +
  1  1.1726                   +
  1  1.1726                       +
  1  1.1726                           +
  1  1.1726                               +
  1  1.1726                                   +
```

[Explanation of notation and interpretation of tables]

2.3.4.3.7. Design for 4 reference zeners and 2 test zeners

Design for 4 references zeners and 2 test zeners.

```
                        ZENERS
 OBSERVATIONS    1   1   1   1   1   1

      Y(1)       +                   -
      Y(2)       +                       -
      Y(3)           +               -
      Y(4)           +                   -
      Y(5)               +           -
      Y(6)               +               -
      Y(7)                   +       -
      Y(8)                   +           -
      Y(9)                   -       +
      Y(10)                  -           +
      Y(11)              -           +
      Y(12)              -               +
      Y(13)          -               +
      Y(14)          -                   +
      Y(15)      -                   +
      Y(16)      -                       +

 RESTRAINT       +   +   +   +

 CHECK STANDARD  +   -

 DEGREES OF FREEDOM  =  10
```

SOLUTION MATRIX
DIVISOR = 16

OBSERVATIONS	1	1	1	1	1	1	P
Y(1)	3	-1	-1	-1	-2	0	1
Y(2)	3	-1	-1	-1	0	-2	1
Y(3)	-1	3	-1	-1	-2	0	1
Y(4)	-1	3	-1	-1	0	-2	1
Y(5)	-1	-1	3	-1	-2	0	1
Y(6)	-1	-1	3	-1	0	-2	1
Y(7)	-1	-1	-1	3	-2	0	1
Y(8)	-1	-1	-1	3	0	-2	1
Y(9)	1	1	1	-3	2	0	1
Y(10)	1	1	1	-3	0	2	

1						
	Y(11)	1	1	-3	1	2	0
1							
	Y(12)	1	1	-3	1	0	2
1							
	Y(13)	1	-3	1	1	2	0
1							
	Y(14)	1	-3	1	1	0	2
1							
	Y(15)	-3	1	1	1	2	0
1							
	Y(16)	-3	1	1	1	0	2
1							
	R*	4	4	4	4	4	4
0							

R* = AVERAGE VALUE OF 4 REFERENCE STANDARDS

P = LEFT-RIGHT EFFECT

FACTORS FOR COMPUTING STANDARD DEVIATIONS

V	FACTORS	ZENERS					
		1	1	1	1	1	P
1	0.4330	+					
1	0.4330		+				
1	0.4330			+			
1	0.4330				+		
1	0.3536					+	
1	0.3536						+
1	0.2500						+

Explanation of notation and interpretation of tables

NIST SEMATECH HOME TOOLS & AIDS SEARCH BACK NEXT

2.3.4.3.8. Design for 4 reference zeners and 3 test zeners

Design for 4 references and 3 test zeners.

```
                        ZENERS

 OBSERVATIONS    1    1    1    1    1    1    1

      Y(1)       -    +
      Y(2)       -         +
      Y(3)            +         -
      Y(4)       +                              -
      Y(5)       +                         -
      Y(6)       +                    -
      Y(7)            -               +
      Y(8)            -                    +
      Y(9)            -                         +
      Y(10)                -                     +
      Y(11)                -                +
      Y(12)                -          +
      Y(13)                     +     -
      Y(14)                     +          -
      Y(15)                     +               -
      Y(16)                +    -
      Y(17)           +         -
      Y(18)      -         +


 RESTRAINT       +    +    +    +

 CHECK STANDARD  +    -


 DEGREES OF FREEDOM = 11


                              SOLUTION MATRIX
                              DIVISOR  = 1260

 OBSERVATIONS      1       1       1       1       1       1
1     P


     Y(1)        -196     196     -56      56       0       0
0    70
     Y(2)        -160     -20     160      20       0       0
0    70
     Y(3)          20     160     -20    -160       0       0
0    70
     Y(4)         143     -53     -17     -73       0       0    -
315    70
     Y(5)         143     -53     -17     -73       0    -315
0    70
     Y(6)         143     -53     -17     -73    -315       0
0    70
     Y(7)          53    -143      73      17     315       0
0    70
     Y(8)          53    -143      73      17       0     315
0    70
```

```
    Y(9)          53   -143     73     17      0      0
315   70
    Y(10)         17     73   -143     53      0      0
315   70
    Y(11)         17     73   -143     53      0    315
0   70
    Y(12)         17     73   -143     53    315      0
0   70
    Y(13)        -73    -17    -53    143   -315      0
0   70
    Y(14)        -73    -17    -53    143      0   -315
0   70
    Y(15)        -73    -17    -53    143      0      0   -
315   70
    Y(16)         56    -56    196   -196      0      0
0   70
    Y(17)         20    160    -20   -160      0      0
0   70
    Y(18)       -160    -20    160     20      0      0
0   70
    R*           315    315    315    315    315    315
315   0
```

R* = Average value of the 4 reference zeners

P = left-right effect

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

V	K1	1	1	1	1	1	1	1
1	0.5000					+		
1	0.5000						+	
1	0.5000							+
2	0.7071					+	+	
3	0.8660					+	+	+
0	0.5578	+	-					

Explanation of notation and interpretation of tables

2.3.4.3.9. Design for 3 references and 1 test resistor

Design 1,1,1,1

```
 OBSERVATIONS    1    1    1    1

     Y(1)        +    -
     Y(2)        +         -
     Y(3)        +              -
     Y(4)        -              +
     Y(5)        -         +
     Y(6)        -    +

 RESTRAINT       +    +    +

 DEGREES OF FREEDOM = 3
```

SOLUTION MATRIX
DIVISOR = 6

OBSERVATIONS	1	1	1	1
Y(1)	1	-2	1	1
Y(2)	1	1	-2	1
Y(3)	0	0	0	-3
Y(4)	0	0	0	3
Y(5)	-1	-1	2	-1
Y(6)	-1	2	-1	-1
R	2	2	2	2

R = AVERAGE VALUE OF 3 REFERENCE RESISTORS

FACTORS FOR COMPUTING STANDARD DEVIATIONS

OHM FACTORS		RESISTORS			
		1	1	1	1
1	0.3333	+			
1	0.5270		+		
1	0.5270			+	
1	0.7817				+

Explanation of notation and interpretation of tables

2.3.4.3.10. Design for 4 references and 1 test resistor

Design 1,1,1,1,1

```
OBSERVATIONS    1    1    1    1    1

     Y(1)       +                   -
     Y(2)            +              -
     Y(3)                 +         -
     Y(4)                      +    -
     Y(5)                      -    +
     Y(6)                 -         +
     Y(7)            -              +
     Y(8)       -                   +

RESTRAINT       +    +    +    +

DEGREES OF FREEDOM   =       4


                          SOLUTION MATRIX
                            DIVISOR = 8

OBSERVATIONS       1       1       1       1       1

   Y(1)            3      -1      -1      -1      -1
   Y(2)           -1       3      -1      -1      -1
   Y(3)           -1      -1       3      -1      -1
   Y(4)           -1      -1      -1       3      -1
   Y(5)            1       1       1      -3       1
   Y(6)            1       1      -3       1       1
   Y(7)            1      -3       1       1       1
   Y(8)           -3       1       1       1       1
   R               2       2       2       2       2

   R = AVERAGE VALUE OF REFERENCE RESISTORS


FACTORS FOR COMPUTING STANDARD DEVIATIONS
OHM FACTORS
                1    1    1    1    1
 1   0.6124     +
 1   0.6124          +
 1   0.6124               +
 1   0.6124                    +
 1   0.3536                         +
```

Explanation of notation and interpretation of tables

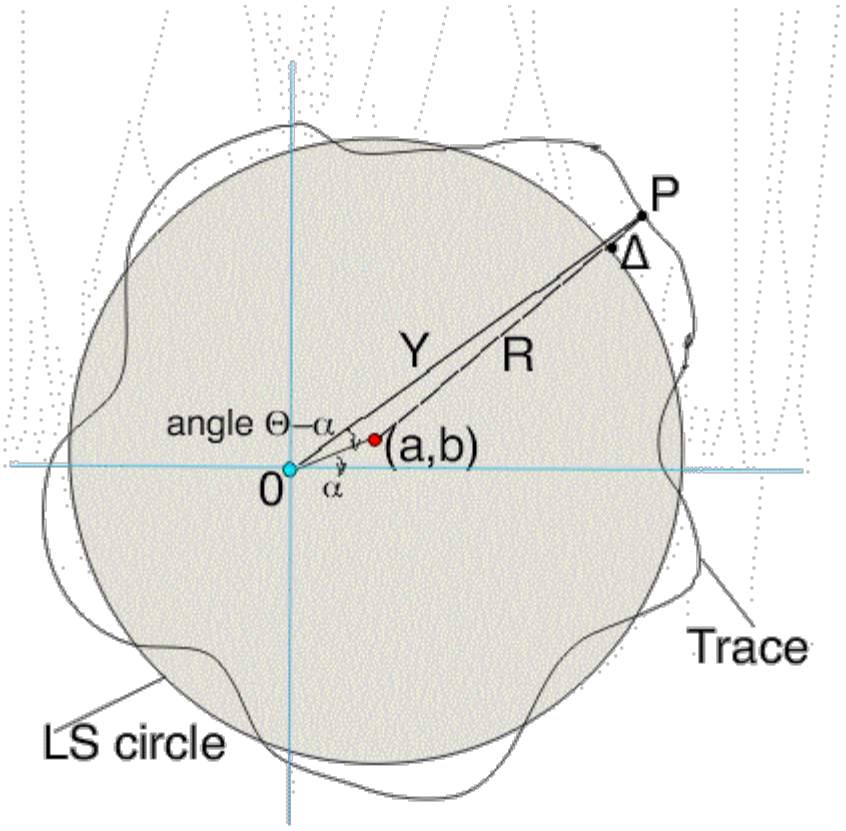
2.3.4.4. Roundness measurements

Roundness measurements

Measurements of roundness require 360° traces of the workpiece made with a turntable-type instrument or a stylus-type instrument. A least squares fit of points on the trace to a circle define the parameters of noncircularity of the workpiece. A diagram of the measurement method is shown below.

The diagram shows the trace and Y, the distance from the spindle center to the trace at the angle.

A least squares circle fit to data at equally spaced angles gives estimates of P - R, the noncircularity, where R = radius of the circle and P = distance from the center of the circle to the trace.



Low precision measurements

Some measurements of roundness do not require a high level of precision, such as measurements on cylinders, spheres, and ring gages where roundness is not of primary importance. For this purpose, a single trace is made of the workpiece.

Weakness of single trace method

The weakness of this method is that the deviations contain both the spindle error and the workpiece error, and these two errors cannot be separated with the single trace. Because the spindle error is usually small and within known limits, its effect can be ignored except when

the most precise measurements are needed.

High precision measurements

High precision measurements of roundness are appropriate where an object, such as a hemisphere, is intended to be used primarily as a roundness standard.

Measurement method

The measurement sequence involves making multiple traces of the roundness standard where the standard is rotated between traces. Least-squares analysis of the resulting measurements enables the noncircularity of the spindle to be separated from the profile of the standard.

Choice of measurement method

A synopsis of the measurement method and the estimation technique are given in this chapter for:

- Single-trace method
- Multiple-trace method

The reader is encouraged to obtain a copy of the publication on roundness (Reeve) for a more complete description of the measurement method and analysis.

2. Measurement Process Characterization
2.3. Calibration
2.3.4. Catalog of calibration designs
2.3.4.4. Roundness measurements

2.3.4.4.1. Single-trace roundness design

Low precision measurements

Some measurements of roundness do not require a high level of precision, such as measurements on cylinders, spheres, and ring gages where roundness is not of primary importance. The diagram of the measurement method shows the trace and *Y*, the distance from the spindle center to the trace at the angle. A least-squares circle fit to data at equally spaced angles gives estimates of *P* - *R*, the noncircularity, where *R* = radius of the circle and *P* = distance from the center of the circle to the trace.

Single trace method

For this purpose, a single trace covering exactly 360° is made of the workpiece and measurements Y_i at angles θ_i of the distance between the center of the spindle and the trace, are made at

$$\theta_i \{i = 1, \cdots, N\}$$

equally spaced angles. A least-squares circle fit to the data gives the following estimators of the parameters of the circle.

$$\hat{R} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

$$\hat{a} = \frac{2}{N}\sum_{i=1}^{N} Y_i \cos(\theta_i)$$

$$\hat{b} = \frac{2}{N}\sum_{i=1}^{N} Y_i \sin(\theta_i)$$

.

Noncircularity of workpiece

The deviation of the trace from the circle at angle θ_i, which defines the noncircularity of the workpiece, is estimated by:

$$\hat{\Delta} = Y_i - \hat{R} - \hat{a}\cos(\theta_i) - \hat{b}\sin(\theta_i)$$

Weakness of single trace method

The weakness of this method is that the deviations contain both the spindle error and the workpiece error, and these two errors cannot be separated with the single trace. Because the spindle error is usually small and within known limits, its effect can be ignored except when the most precise measurements are needed.

2.3.4.4.2. Multiple-trace roundness designs

High precision measurements

High precision roundness measurements are required when an object, such as a hemisphere, is intended to be used primarily as a roundness standard. The method outlined on this page is appropriate for either a turntable-type instrument or a spindle-type instrument.

Measurement method

The measurement sequence involves making multiple traces of the roundness standard where the standard is rotated between traces. Least-squares analysis of the resulting measurements enables the noncircularity of the spindle to be separated from the profile of the standard. The reader is referred to the publication on the subject (Reeve) for details covering measurement techniques and analysis.

Method of n traces

The number of traces that are made on the workpiece is arbitrary but should not be less than four. The workpiece is centered as well as possible under the spindle. The mark on the workpiece which denotes the zero angular position is aligned with the zero position of the spindle as shown in the graph. A trace is made with the workpiece in this position. The workpiece is then rotated clockwise by *360/n* degrees and another trace is made. This process is continued until *n* traces have been recorded.

Mathematical model for estimation

For $i = 1,...,n$, the ith angular position is denoted by

$$\theta_i = \frac{360(i-1)}{n}\text{deg}$$

Definition of terms relating to distances to the least squares circle

The deviation from the least squares circle (LSC) of the workpiece at the θ_i position is α_i.

The deviation of the spindle from its LSC at the α_i position is β_i.

Terms relating to

For the jth graph, let the three parameters that define the LSC be given by

parameters of least squares circle

$$R_j, a_j, b_j$$

defining the radius *R*, *a*, and *b* as shown in the graph. In an idealized measurement system these parameters would be constant for all *j*. In reality, each rotation of the workpiece causes it to shift a small amount vertically and horizontally. To account for this shift, separate parameters are needed for each trace.

Correction for obstruction to stylus

Let Y_{ij} be the observed distance (in polar graph units) from the center of the *j*th graph to the point on the curve that corresponds to the θ_i position of the spindle. If *K* is the magnification factor of the instrument in microinches/polar graph unit and δ is the angle between the lever arm of the stylus and the tangent to the workpiece at the point of contact (which normally can be set to zero if there is no obstruction), the transformed observations to be used in the estimation equations are:

$$Z_{ij} = K\cos(\delta)Y_{ij}$$

.

Estimates for parameters

The estimation of the individual parameters is obtained as a least-squares solution that requires six restraints which essentially guarantee that the sum of the vertical and horizontal deviations of the spindle from the center of the LSC are zero. The expressions for the estimators are as follows:

$$\hat{\alpha}_i = \sum_{j=1}^{n}\sum_{k=1}^{n} t_{k-i+j+n} Z_{kj}$$

$$\hat{\beta}_i = \sum_{j=1}^{n}\sum_{k=1}^{n} t_{k-i+1+n} Z_{kj}$$

$$\hat{R}_j = \sum_{k=1}^{n} Z_{kj}$$

$$\hat{a}_j = \frac{2}{n}\sum_{k=1}^{n} Z_{kj}\cos(\theta_k)$$

$$\hat{b}_j = \frac{2}{n}\sum_{k=1}^{n} Z_{kj}\sin(\theta_k)$$

where

$$t_m = \frac{n-3}{n^2}; m = 1$$

$$t_m = -\frac{1}{n^2}(1 + 2\cos(\theta_m)); 2 \le m \le n$$

Finally, the standard deviations of the profile estimators are given by:

$$s_{\alpha_j} = s_{\beta_j} = \frac{\sqrt{n-3}}{n} s; (i = 1, \cdots, n.)$$

Computation of standard deviation

The computation of the residual standard deviation of the fit requires, first, the computation of the predicted values,

$$\hat{Z}_{ij} = \hat{\alpha}_{i+j-1} - \hat{\beta}_i + \hat{R}_j + \hat{a}_j \cos(\theta_i) + \hat{b}_j \sin(\theta_i)$$

The residual standard deviation with $v = n*n - 5n + 6$ degrees of freedom is

$$s = \sqrt{\frac{1}{n^2 - 5n + 6} \sum_{j=1}^{n} \sum_{i=1}^{n} \left(Z_{ij} - \hat{Z}_{ij} \right)^2}$$

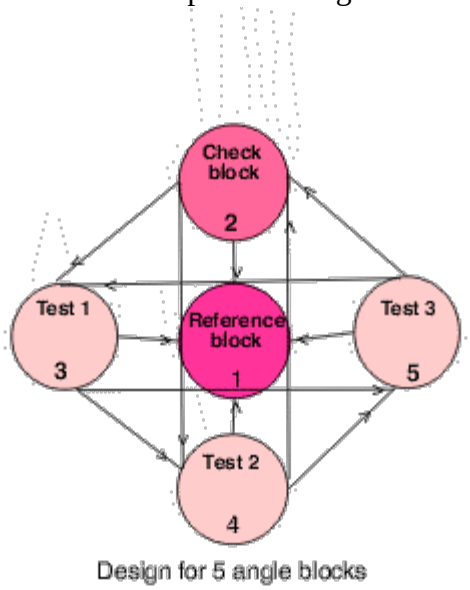
2.3.4.5. Designs for angle blocks

Purpose

The purpose of this section is to explain why calibration of angle blocks of the same size in groups is more efficient than calibration of angle blocks individually.

Calibration schematic for five angle blocks showing the reference as block 1 in the center of the diagram, the check standard as block 2 at the top; and the test blocks as blocks 3, 4, and 5.

A schematic of a calibration scheme for 1 reference block, 1 check standard, and three test blocks is shown below. The reference block, *R*, is shown in the center of the diagram and the check standard, *C*, is shown at the top of the diagram.



Design for 5 angle blocks

Block sizes

Angle blocks normally come in sets of

1, 3, 5, 20, and 30 seconds

1, 3, 5, 20, 30 minutes

1, 3, 5, 15, 30, 45 degrees

and blocks of the same nominal size from 4, 5 or 6 different sets can be calibrated simultaneously using one of the designs shown in this catalog.

- Design for 4 angle blocks

- Design for 5 angle blocks
- Design for 6 angle blocks

Restraint

The solution to the calibration design depends on the known value of a reference block, which is compared with the test blocks. The reference block is designated as block 1 for the purpose of this discussion.

Check standard

It is suggested that block 2 be reserved for a check standard that is maintained in the laboratory for quality control purposes.

Calibration scheme

A calibration scheme developed by Charles Reeve (Reeve) at the National Institute of Standards and Technology for calibrating customer angle blocks is explained on this page. The reader is encouraged to obtain a copy of the publication for details on the calibration setup and quality control checks for angle block calibrations.

Series of measurements for calibrating 4, 5, and 6 angle blocks simultaneously

For all of the designs, the measurements are made in groups of seven starting with the measurements of blocks in the following order: 2-3-2-1-2-4-2. Schematically, the calibration design is completed by counter-clockwise rotation of the test blocks about the reference block, one-at-a-time, with 7 readings for each series reduced to 3 difference measurements. For *n* angle blocks (including the reference block), this amounts to *n - 1* series of 7 readings. The series for 4, 5, and 6 angle blocks are shown below.

Measurements for 4 angle blocks

```
Series 1: 2-3-2-1-2-4-2
Series 2: 4-2-4-1-4-3-4
Series 3: 3-4-3-1-3-2-3
```

Measurements for 5 angle blocks (see diagram)

```
Series 1: 2-3-2-1-2-4-2
Series 2: 5-2-5-1-5-3-5
Series 3: 4-5-4-1-4-2-4
Series 4: 3-4-3-1-3-5-3
```

Measurements for 6 angle blocks

```
Series 1: 2-3-2-1-2-4-2
Series 2: 6-2-6-1-6-3-6
Series 3: 5-6-5-1-5-2-5
Series 4: 4-5-4-1-4-6-4
Series 5: 3-4-3-1-3-5-3
```

Equations for the measurements in the first series showing error sources

The equations explaining the seven measurements for the first series in terms of the errors in the measurement system are:

$$
\begin{aligned}
Z_{11} &= B + X_1 + \quad\; error_{11} \\
Z_{12} &= B + X_2 + \quad d + error_{12} \\
Z_{13} &= B + X_3 + 2d + error_{13} \\
Z_{14} &= B + X_4 + 3d + error_{14} \\
Z \quad &= B + X \quad + 4d + error
\end{aligned}
$$

$$Z_{16} = B + X_6^{15} + 5d + error_{16}^{15}$$
$$Z_{17} = B + X_7^{5} + 6d + error_{17}$$

with B a bias associated with the instrument, d is a linear drift factor, X is the value of the angle block to be determined; and the error terms relate to random errors of measurement.

Calibration procedure depends on difference measurements

The check block, C, is measured before and after each test block, and the difference measurements (which are not the same as the difference measurements for calibrations of mass weights, gage blocks, etc.) are constructed to take advantage of this situation. Thus, the 7 readings are reduced to 3 difference measurements for the first series as follows:

$$Y_{11} = (Z_{11} - 2Z_{12} + Z_{13})/2$$
$$Y_{12} = (Z_{13} - 2Z_{14} + Z_{15})/2$$
$$Y_{13} = (Z_{15} - 2Z_{16} + X_{17})/2$$

For all series, there are $3(n - 1)$ difference measurements, with the first subscript in the equations above referring to the series number. The difference measurements are free of drift and instrument bias.

Design matrix

As an example, the design matrix for $n = 4$ angle blocks is shown below.

1	1	1	1
0	1	-1	0
-1	1	0	0
0	1	0	-1
0	-1	0	1
-1	0	0	1
0	0	-1	1
0	0	1	-1
-1	0	1	0
0	-1	1	0

The design matrix is shown with the solution matrix for identification purposes only because the least-squares solution is weighted (Reeve) to account for the fact that test blocks are measured twice as many times as the reference block. The weight matrix is not shown.

Solutions to the calibration designs measurements

Solutions to the angle block designs are shown on the following pages. The solution matrix and factors for the repeatability standard deviation are to be interpreted as explained in solutions to calibration designs . As an example, the solution for the design for *n=4* angle blocks is as follows:

The solution for the reference standard is shown under the first column of the solution matrix; for the check standard under the second column; for the first test block under the third column; and for the second test block under the fourth column. Notice that the estimate for the reference block is guaranteed to be R^*, regardless

of the measurement results, because of the restraint that is imposed on the design. Specifically,

$$\hat{Reference} = 0 + R^*$$

$$\hat{Check} = \frac{1}{24}\begin{pmatrix} 2.272Y_{11} + 9.352Y_{12} + 2.272Y_{13} \\ -5.052Y_{21} + 7.324Y_{22} - 1.221Y_{23} \\ -1.221Y_{31} + 7.324Y_{32} - 5.052Y_{33} \end{pmatrix} + R^*$$

$$\hat{Test}_1 = \frac{1}{24}\begin{pmatrix} -5.052Y_{11} + 7.324Y_{12} - 1.221Y_{13} \\ -1.221Y_{21} + 7.324Y_{22} - 5.052Y_{23} \\ +2.272Y_{31} + 9.352Y_{32} + 2.272Y_{33} \end{pmatrix} + R^*$$

$$\hat{Test}_2 = \frac{1}{24}\begin{pmatrix} -1.221Y_{11} + 7.324Y_{12} - 5.052Y_{13} \\ +2.272Y_{21} + 9.352Y_{22} + 2.272Y_{23} \\ -5.052Y_{31} + 7.324Y_{32} - 1.221Y_{33} \end{pmatrix} + R^*$$

Solutions are correct only for the restraint as shown.

Calibrations can be run for top and bottom faces of blocks

The calibration series is run with the blocks all face "up" and is then repeated with the blocks all face "down", and the results averaged. The difference between the two series can be large compared to the repeatability standard deviation, in which case a between-series component of variability must be included in the calculation of the standard deviation of the reported average.

Calculation of standard deviations when the blocks are measured in two orientations

For *n* blocks, the differences between the values for the blocks measured in the top (denoted by "t") and bottom (denoted by "b") positions are denoted by:

$$\delta_i = \hat{X}_i^t - \hat{X}_i^t \qquad \text{for } i = I, ..., n$$

The standard deviation of the average (for each block) is calculated from these differences to be:

$$s_{avg\ test} = \sqrt{\frac{1}{4(n-1)}\sum_{i=1}^{n}\delta_i^2}$$

Standard deviations when the blocks are measured in only one

If the blocks are measured in only one orientation, there is no way to estimate the between-series component of variability and the standard deviation for the value of each block is computed as

$$s_{test} = K_1 s_1$$

orientation	where K_1 is shown under "Factors for computing repeatability standard deviations" for each design and s_1 is the repeatability standard deviation as estimated from the design. Because this standard deviation may seriously underestimate the uncertainty, a better approach is to estimate the standard deviation from the data on the check standard over time. An expanded uncertainty is computed according to the ISO guidelines.

2. Measurement Process Characterization
2.3. Calibration
2.3.4. Catalog of calibration designs
2.3.4.5. Designs for angle blocks

2.3.4.5.1. Design for 4 angle blocks

DESIGN MATRIX

	1	1	1	1
Y(1)	0	1	-1	0
Y(2)	-1	1	0	0
Y(3)	0	1	0	-1
Y(4)	0	-1	0	1
Y(5)	-1	0	0	1
Y(6)	0	0	-1	1
Y(7)	0	0	1	-1
Y(8)	-1	0	1	0
Y(9)	0	-1	1	0

REFERENCE +

CHECK STANDARD +

DEGREES OF FREEDOM = 6

SOLUTION MATRIX
DIVISOR = 24

OBSERVATIONS	1	1	1	1
Y(11)	0	2.2723000	-5.0516438	-1.2206578
Y(12)	0	9.3521166	7.3239479	7.3239479
Y(13)	0	2.2723000	-1.2206578	-5.0516438
Y(21)	0	-5.0516438	-1.2206578	2.2723000
Y(22)	0	7.3239479	7.3239479	9.3521166
Y(23)	0	-1.2206578	-5.0516438	2.2723000
Y(31)	0	-1.2206578	2.2723000	-5.0516438
Y(32)	0	7.3239479	9.3521166	7.3239479
Y(33)	0	-5.0516438	2.2723000	-1.2206578
R*	1	1.	1.	1.

R* = VALUE OF REFERENCE ANGLE BLOCK

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

SIZE	K1	1	1	1	1
1	0.0000	+			
1	0.9749		+		
1	0.9749			+	
1	0.9749				+
1	0.9749		+		

Explanation of notation and interpretation of tables

2. Measurement Process Characterization
2.3. Calibration
2.3.4. Catalog of calibration designs
2.3.4.5. Designs for angle blocks

2.3.4.5.2. Design for 5 angle blocks

DESIGN MATRIX

1	1	1	1	1
0	1	-1	0	0
-1	1	0	0	0
0	1	0	-1	0
0	-1	0	0	1
-1	0	0	0	1
0	0	-1	0	1
0	0	0	1	-1
-1	0	0	1	0
0	-1	0	1	0
0	0	1	-1	0
-1	0	1	0	0
0	0	1	0	-1

REFERENCE +

CHECK STANDARD +

DEGREES OF FREEDOM = 8

SOLUTION MATRIX
DIVISOR = 24

OBSERVATIONS	1	1	1	1	1
Y(11)	0.00000	3.26463	-5.48893	-0.21200	-1.56370
Y(12)	0.00000	7.95672	5.38908	5.93802	4.71618
Y(13)	0.00000	2.48697	-0.89818	-4.80276	-0.78603
Y(21)	0.00000	-5.48893	-0.21200	-1.56370	3.26463
Y(22)	0.00000	5.38908	5.93802	4.71618	7.95672
Y(23)	0.00000	-0.89818	-4.80276	-0.78603	2.48697
Y(31)	0.00000	-0.21200	-1.56370	3.26463	-5.48893
Y(32)	0.00000	5.93802	4.71618	7.95672	5.38908
Y(33)	0.00000	-4.80276	-0.78603	2.48697	-0.89818
Y(41)	0.00000	-1.56370	3.26463	-5.48893	-0.21200
Y(42)	0.00000	4.71618	7.95672	5.38908	5.93802
Y(43)	0.00000	-0.78603	2.48697	-0.89818	-4.80276
R*	1.	1.	1.	1.	1.

R* = VALUE OF REFERENCE ANGLE BLOCK

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

SIZE	K1	1	1	1	1	1
1	0.0000	+				
1	0.7465		+			
1	0.7465			+		
1	0.7456				+	
1	0.7456					+
1	0.7465		+			

Explanation of notation and interpretation of tables

2. Measurement Process Characterization
2.3. Calibration
2.3.4. Catalog of calibration designs
2.3.4.5. Designs for angle blocks

2.3.4.5.3. Design for 6 angle blocks

DESIGN MATRIX

1	1	1	1	1	1
0	1	-1	0	0	0
-1	1	0	0	0	0
0	1	0	-1	0	0
0	-1	0	0	0	1
-1	0	0	0	0	1
0	0	-1	0	0	1
0	0	0	0	1	-1
-1	0	0	0	1	0
0	-1	0	0	1	0
0	0	0	1	-1	0
-1	0	0	1	0	0
0	0	0	1	0	-1
0	0	1	-1	0	0
-1	0	1	0	0	0
0	0	1	0	-1	0

REFERENCE +

CHECK STANDARD +

DEGREES OF FREEDOM = 10

SOLUTION MATRIX
DIVISOR = 24

OBSERVATIONS	1	1	1	1	1
Y(11)	0.0000	3.2929	-5.2312	-0.7507	-0.6445
-0.6666					
Y(12)	0.0000	6.9974	4.6324	4.6495	3.8668
3.8540					
Y(13)	0.0000	3.2687	-0.7721	-5.2098	-0.6202
-0.6666					
Y(21)	0.0000	-5.2312	-0.7507	-0.6445	-0.6666
3.2929					
Y(22)	0.0000	4.6324	4.6495	3.8668	3.8540
6.9974					
Y(23)	0.0000	-0.7721	-5.2098	-0.6202	-0.6666
3.2687					
Y(31)	0.0000	-0.7507	-0.6445	-0.6666	3.2929
-5.2312					
Y(32)	0.0000	4.6495	3.8668	3.8540	6.9974
4.6324					
Y(33)	0.0000	-5.2098	-0.6202	-0.6666	3.2687
-0.7721					
Y(41)	0.0000	-0.6445	-0.6666	3.2929	-5.2312
-0.7507					
Y(42)	0.0000	3.8668	3.8540	6.9974	4.6324
4.6495					
Y(43)	0.0000	-0.6202	-0.6666	3.2687	-0.7721
-5.2098					
Y(51)	0.0000	-0.6666	3.2929	-5.2312	-0.7507
-0.6445					

```
  Y(52)      0.0000     3.8540     6.9974     4.6324     4.6495
3.8668
  Y(53)      0.0000    -0.6666     3.2687    -0.7721    -5.2098
-0.6202
  R*         1.         1.         1.         1.         1.
1.

  R* = VALUE OF REFERENCE ANGLE BLOCK


 FACTORS FOR REPEATABILITY STANDARD DEVIATIONS
 SIZE  K1
                1   1   1   1   1   1
  1  0.0000     +
  1  0.7111         +
  1  0.7111             +
  1  0.7111                 +
  1  0.7111                     +
  1  0.7111                         +
  1  0.7111         +
```

Explanation of notation and interpretation of tables

2.3.4.6. Thermometers in a bath

Measurement sequence

Calibration of liquid in glass thermometers is usually carried out in a controlled bath where the temperature in the bath is increased steadily over time to calibrate the thermometers over their entire range. One way of accounting for the temperature drift is to measure the temperature of the bath with a standard resistance thermometer at the beginning, middle and end of each run of K test thermometers. The test thermometers themselves are measured twice during the run in the following time sequence:

$$R_1, T_1, T_2, \cdots, T_K, R_2, T'_K, \cdots, T'_2, T'_1, R_3$$

where R_1, R_2, R_3 represent the measurements on the standard resistance thermometer and T_1, T_2, ... , T_K and T'_1, T'_2, ... , T'_K represent the pair of measurements on the K test thermometers.

Assumptions regarding temperature

The assumptions for the analysis are that:

- Equal time intervals are maintained between measurements on the test items.
- Temperature increases by ΔT with each interval.
- A temperature change of ϕ is allowed for the reading of the resistance thermometer in the middle of the run.

Indications for test thermometers

It can be shown (Cameron and Hailes) that the average reading for a test thermometer is its indication at the temperature implied by the average of the three resistance readings. The standard deviation associated with this indication is calculated from difference readings where

$$d_i = T_i - T'_i$$

is the difference for the ith thermometer. This difference is an estimate of $\phi + 2(K - i)\Delta$.

Estimates of drift

The estimates of the shift due to the resistance thermometer and temperature drift are given by:

$$\hat{\phi} = \frac{2}{K(K+1)}\left(3\sum_{i=1}^{K} id_i - (K+1)\sum_{i=1}^{K} d_i\right)$$

$$\hat{\Delta} = \frac{3}{K(K^2-1)}\left((K+1)\sum_{i=1}^{K} d_i - 2\sum_{i=1}^{K} id_i\right)$$

Standard deviations

The residual variance is given by

$$s^2 = \frac{1}{(K-2)}\sum_{i=1}^{K}\left(d_i - \hat{\phi} - 2(K-i)\hat{\Delta}\right)^2 .$$

The standard deviation of the indication assigned to the ith test thermometer is

$$s_{test} = \frac{s}{\sqrt{2}}$$

and the standard deviation for the estimates of shift and drift are

$$s_{\phi} = \frac{\sqrt{2}(2K-1)}{K(K+1)}s$$

$$s_{\Delta} = \frac{\sqrt{3}}{K(K^2-1)}s$$

respectively.

2.3.4.7. Humidity standards

Humidity standards

The calibration of humidity standards usually involves the comparison of reference weights with cylinders containing moisture. The designs shown in this catalog are drift-eliminating and may be suitable for artifacts other than humidity cylinders.

List of designs

- 2 reference weights and 3 cylinders

2.3.4.7.1. Drift-elimination design for 2 reference weights and 3 cylinders

```
OBSERVATIONS    1   1   1   1   1

    Y(1)        +   -
    Y(2)                    +   -
    Y(3)                +   -
    Y(4)            +   -
    Y(5)        -                   +
    Y(6)        -               +
    Y(7)                +           -
    Y(8)            +           -
    Y(9)            -               +
    Y(10)       +       -

RESTRAINT       +   +

CHECK STANDARD  +   -

DEGREES OF FREEDOM = 6
```

SOLUTION MATRIX
DIVISOR = 10

OBSERVATIONS	1	1	1	1	1
Y(1)	2	-2	0	0	0
Y(2)	0	0	0	2	-2
Y(3)	0	0	2	-2	0
Y(4)	-1	1	-3	-1	-1
Y(5)	-1	1	1	1	3
Y(6)	-1	1	1	3	1
Y(7)	0	0	2	0	-2
Y(8)	-1	1	-1	-3	-1
Y(9)	1	-1	1	1	3
Y(10)	1	-1	-3	-1	-1
R*	5	5	5	5	5

R* = average value of the two reference weights

FACTORS FOR REPEATABILITY STANDARD DEVIATIONS

```
WT    K1        1   1   1   1   1
 1  0.5477                    +
 1  0.5477                +
 1  0.5477            +
 2  0.8944            +   +
 3  1.2247            +   +   +
 0  0.6325    +   -
```

Explanation of notation and interpretation of tables

2.3.5. Control of artifact calibration

Purpose

The purpose of statistical control in the calibration process is to guarantee the 'goodness' of calibration results within predictable limits and to validate the statement of uncertainty of the result. Two types of control can be imposed on a calibration process that makes use of statistical designs:

1. Control of instrument precision or short-term variability
2. Control of bias and long-term variability
    - Example of a Shewhart control chart
    - Example of an EWMA control chart

Short-term standard deviation

The short-term standard deviation from each design is the basis for controlling instrument precision. Because the measurements for a single design are completed in a short time span, this standard deviation estimates the basic precision of the instrument. Designs should be chosen to have enough measurements so that the standard deviation from the design has at least 3 degrees of freedom where the degrees of freedom are $(n - m + 1)$ with

- n = number of difference measurements
- m = number of artifacts.

Check standard

Measurements on a check standard provide the mechanism for controlling the bias and long-term variability of the calibration process. The check standard is treated as one of the test items in the calibration design, and its value as computed from each calibration run is the basis for accepting or rejecting the calibration. All designs cataloged in this Handbook have provision for a check standard.

The check standard should be of the same type and geometry as items that are measured in the designs. These artifacts must be stable and available to the calibration process on a continuing basis. There should be a check standard at each critical level of measurement. For example, for mass calibrations there should be check standards at the 1 kg; 100 g, 10 g, 1 g, 0.1 g levels, etc. For gage blocks, there should be check standards at all nominal lengths.

A check standard can also be a mathematical construction,

such as the computed difference between the calibrated values of two reference standards in a design.

Database of check standard values

The creation and maintenance of the database of check standard values is an important aspect of the control process. The results from each calibration run are recorded in the database. The best way to record this information is in one file with one line (row in a spreadsheet) of information in fixed fields for each calibration run. A list of typical entries follows:

1. Date
2. Identification for check standard
3. Identification for the calibration design
4. Identification for the instrument
5. Check standard value
6. Repeatability standard deviation from design
7. Degrees of freedom
8. Operator identification
9. Flag for out-of-control signal
10. Environmental readings (if pertinent)

2.3.5.1. Control of precision

Control parameters from historical data

A modified control chart procedure is used for controlling instrument precision. The procedure is designed to be implemented in real time after a baseline and control limit for the instrument of interest have been established from the database of short-term standard deviations. A separate control chart is required for each instrument -- except where instruments are of the same type with the same basic precision, in which case they can be treated as one.

The baseline is the process standard deviation that is pooled from $k = 1, ..., K$ individual repeatability standard deviations, s_k , in the database, each having ν_k degrees of freedom. The pooled repeatability standard deviation is

$$s_1 = \sqrt{\frac{1}{\nu}\sum_{k=1}^{K}\nu_k s_k^2}$$

with degrees of freedom

$$\nu = \sum_{k=1}^{K}\nu_k \quad .$$

Control procedure is invoked in real-time for each calibration run

The control procedure compares each new repeatability standard deviation that is recorded for the instrument with an upper control limit, *UCL*. Usually, only the upper control limit is of interest because we are primarily interested in detecting degradation in the instrument's precision. A possible complication is that the control limit is dependent on the degrees of freedom in the new standard deviation and is computed as follows:

$$UCL = s_1\sqrt{F_{\alpha,\nu_{new},\nu}} \quad .$$

The quantity under the radical is the upper α percentage point from the *F* table where α is chosen small to be, say, 0.05. The other two terms refer to the degrees of freedom in the new standard deviation and the degrees of freedom in the process standard deviation.

Limitation of graphical method

The graphical method of plotting every new estimate of repeatability on a control chart does not work well when the *UCL* can change with each calibration design, depending on the degrees of freedom. The algebraic equivalent is to test if the new standard deviation exceeds its control limit, in which case the short-term precision is judged to be out of control and the current calibration run is rejected. For more guidance, see Remedies and strategies for dealing with out-of-control signals.

As long as the repeatability standard deviations are in control, there is reason for confidence that the precision of the instrument has not degraded.

Case study: Mass balance precision

It is recommended that the repeatability standard deviations be plotted against time on a regular basis to check for gradual degradation in the instrument. Individual failures may not trigger a suspicion that the instrument is in need of adjustment or tuning.

2.3.5.1.1. Example of control chart for precision

Example of a control chart for precision of a mass balance

Mass calibrations usually start with the comparison of kilograms standards using a high precision balance as a comparator. Many of the measurements at the kilogram level that were made at NIST between 1975 and 1989 were made on balance #12 using a 1,1,1,1 calibration design. The redundancy in the calibration design produces estimates for the individual kilograms and a repeatability standard deviation with three degrees of freedom for each calibration run. These standard deviations estimate the precision of the balance.

Need for monitoring precision

The precision of the balance is monitored to check for:

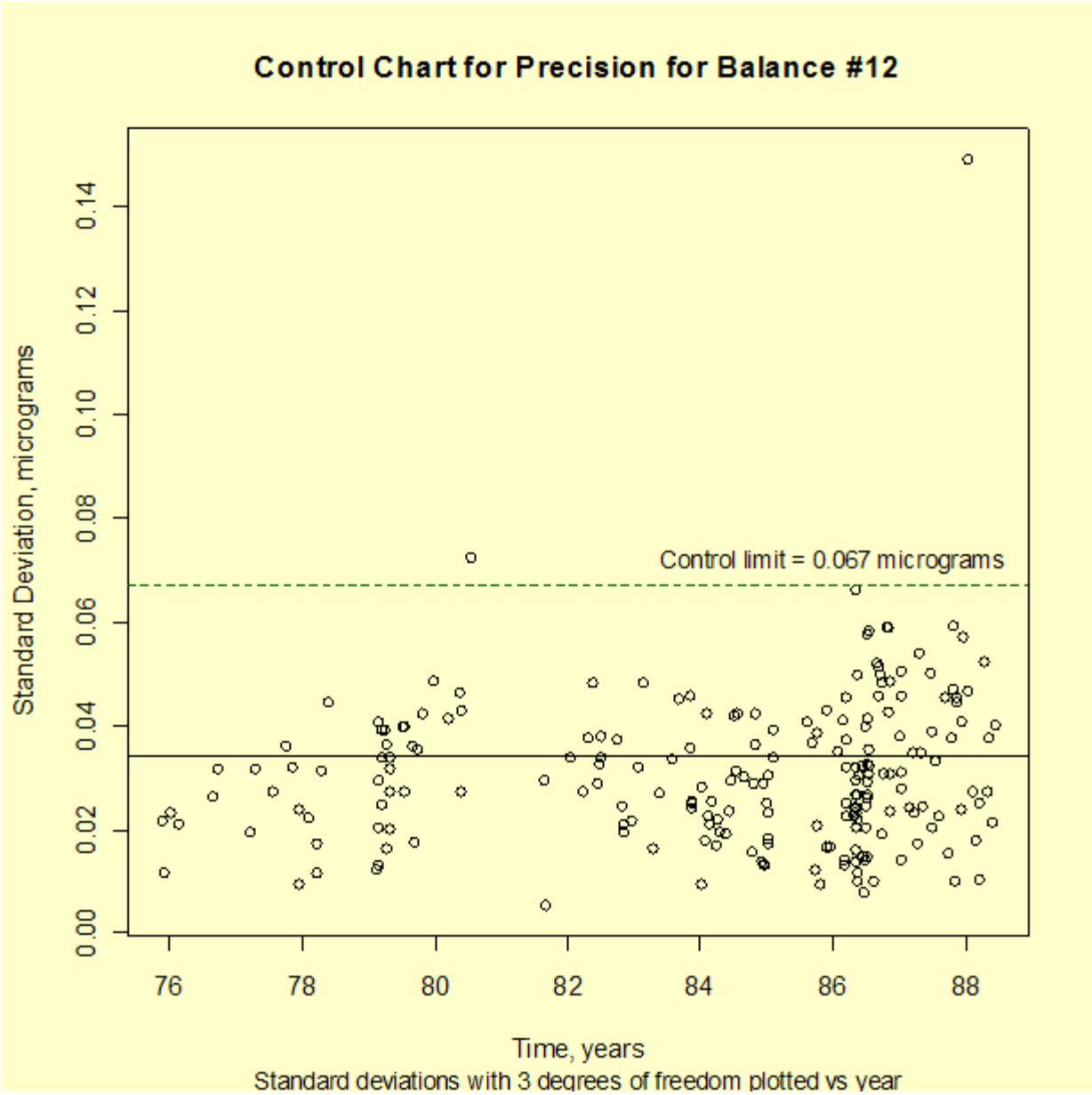
1. Slow degradation in the balance
2. Anomalous behavior at specific times

Monitoring technique for standard deviations

The standard deviations over time and many calibrations are tracked and monitored using a control chart for standard deviations. The database and control limits are updated on a yearly or bi-yearly basis and standard deviations for each calibration run in the next cycle are compared with the control limits. In this case, the standard deviations from 117 calibrations between 1975 and 1985 were pooled to obtain a repeatability standard deviation with v = 3*117 = 351 degrees of freedom, and the control limits were computed at the 1 % significance level.

Control chart for precision

The following control chart for precision for balance #12 can be generated using both Dataplot code and R code.



Interpretation of the control chart

The control chart shows that the precision of the balance remained in control through the first five months of 1988 with only two violations of the control limits. For those occasions, the calibrations were discarded and repeated. Clearly, for the second violation, something significant occurred that invalidated the calibration results.

Further interpretation of the control chart

However, it is also clear from the pattern of standard deviations over time that the precision of the balance was gradually degrading and more and more points were approaching the control limits. This finding led to a decision to replace this balance for high accuracy calibrations.

2. Measurement Process Characterization
2.3. Calibration
2.3.5. Control of artifact calibration

2.3.5.2. Control of bias and long-term variability

Control parameters are estimated using historical data

A control chart procedure is used for controlling bias and long-term variability. The procedure is designed to be implemented in real time after a baseline and control limits for the check standard of interest have been established from the database of check standard values. A separate control chart is required for each check standard. The control procedure outlined here is based on a Shewhart control chart with upper and lower control limits that are symmetric about the average. The EWMA control procedure that is sensitive to small changes in the process is discussed on another page.

For a Shewhart control procedure, the average and standard deviation of historical check standard values are the parameters of interest

The check standard values are denoted by

$$C_k \, (k=1, \cdots, K)$$

The baseline is the process average which is computed from the check standard values as

$$\bar{C} = \frac{1}{K} \sum_{k=1}^{K} C_k$$

The process standard deviation is

$$s_2 = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} (C_k - \bar{C})^2}$$

with K - 1 degrees of freedom.

The control limits depend on the t distribution and the degrees of freedom in the process standard deviation

If $\bar{C}$ has been computed from historical data, the upper and lower control limits are:

$$UCL = \overline{C} + t_{1-\alpha/2,\, K-1} \cdot s_2$$

$$LCL = \overline{C} - t_{1-\alpha/2,\, K-1} \cdot s_2$$

where $t_{1\text{-}\alpha/2,\, K\text{-}1}$ denotes the 1-α/2 critical value from the *t*

table with $v = K - 1$ degrees of freedom.

Sample code

Sample code for computing the t value for a conservative case where $\alpha = 0.05$, $J = 6$, and $K = 6$, is available for both Dataplot and R.

Simplification for large degrees of freedom

It is standard practice to use a value of 3 instead of a critical value from the t table, given the process standard deviation has large degrees of freedom, say, $v > 15$.

The control procedure is invoked in real-time and a failure implies that the current calibration should be rejected

The control procedure compares the check standard value, C, from each calibration run with the upper and lower control limits. This procedure should be implemented in real time and does not necessarily require a graphical presentation. The check standard value can be compared algebraically with the control limits. The calibration run is judged to be out-of-control if either:

$$C > UCL$$

or

$$C < LCL$$

Actions to be taken

If the check standard value exceeds one of the control limits, the process is judged to be out of control and the current calibration run is rejected. The best strategy in this situation is to repeat the calibration to see if the failure was a chance occurrence. Check standard values that remain in control, especially over a period of time, provide confidence that no new biases have been introduced into the measurement process and that the long-term variability of the process has not changed.

Out-of-control signals that recur require investigation

Out-of-control signals, particularly if they recur, can be symptomatic of one of the following conditions:

- Change or damage to the reference standard(s)
- Change or damage to the check standard
- Change in the long-term variability of the calibration process

For more guidance, see Remedies and strategies for dealing with out-of-control signals.

Caution - be sure to plot the data

If the tests for control are carried out algebraically, it is recommended that, at regular intervals, the check standard values be plotted against time to check for drift or anomalies in the measurement process.

2.3.5.2.1. Example of Shewhart control chart for mass calibrations

Example of a control chart for mass calibrations at the kilogram level

Mass calibrations usually start with the comparison of four kilogram standards using a high precision balance as a comparator. Many of the measurements at the kilogram level that were made at NIST between 1975 and 1989 were made on balance #12 using a 1,1,1,1 calibration design. The restraint for this design is the known average of two kilogram reference standards. The redundancy in the calibration design produces individual estimates for the two test kilograms and the two reference standards.

Check standard

There is no slot in the 1,1,1,1 design for an artifact check standard when the first two kilograms are reference standards; the third kilogram is a test weight; and the fourth is a summation of smaller weights that act as the restraint in the next series. Therefore, the check standard is a computed difference between the values of the two reference standards as estimated from the design. The convention with mass calibrations is to report the correction to nominal, in this case the correction to 1000 g, as shown in the control charts below.

Need for monitoring

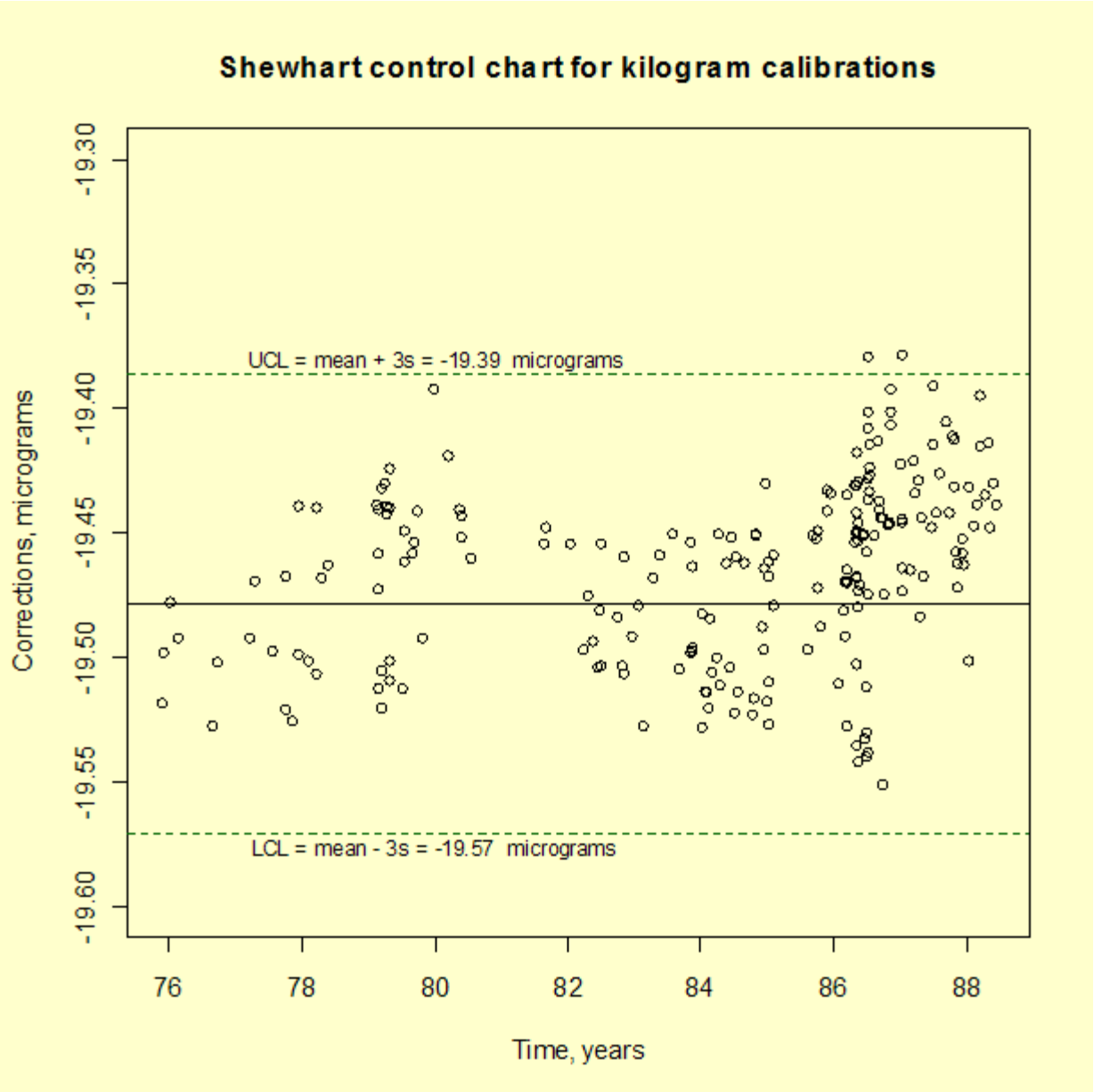
The kilogram check standard is monitored to check for:

1. Long-term degradation in the calibration process
2. Anomalous behavior at specific times

Monitoring technique for check standard values

Check standard values over time and many calibrations are tracked and monitored using a Shewhart control chart. The database and control limits are updated when needed and check standard values for each calibration run in the next cycle are compared with the control limits. In this case, the values from 117 calibrations between 1975 and 1985 were averaged to obtain a baseline and process standard deviation with $\nu = 116$ degrees of freedom. Control limits are computed with a factor of $k = 3$ to identify truly anomalous data points.

Control chart of kilogram check standard measurements showing a change in the process after 1985



Interpretation of the control chart

The control chart shows only two violations of the control limits. For those occasions, the calibrations were discarded and repeated. The configuration of points is unacceptable if many points are close to a control limit and there is an unequal distribution of data points on the two sides of the control chart -- indicating a change in either:

- process average which may be related to a change in the reference standards

or

- variability which may be caused by a change in the instrument precision or may be the result of other factors on the measurement process.

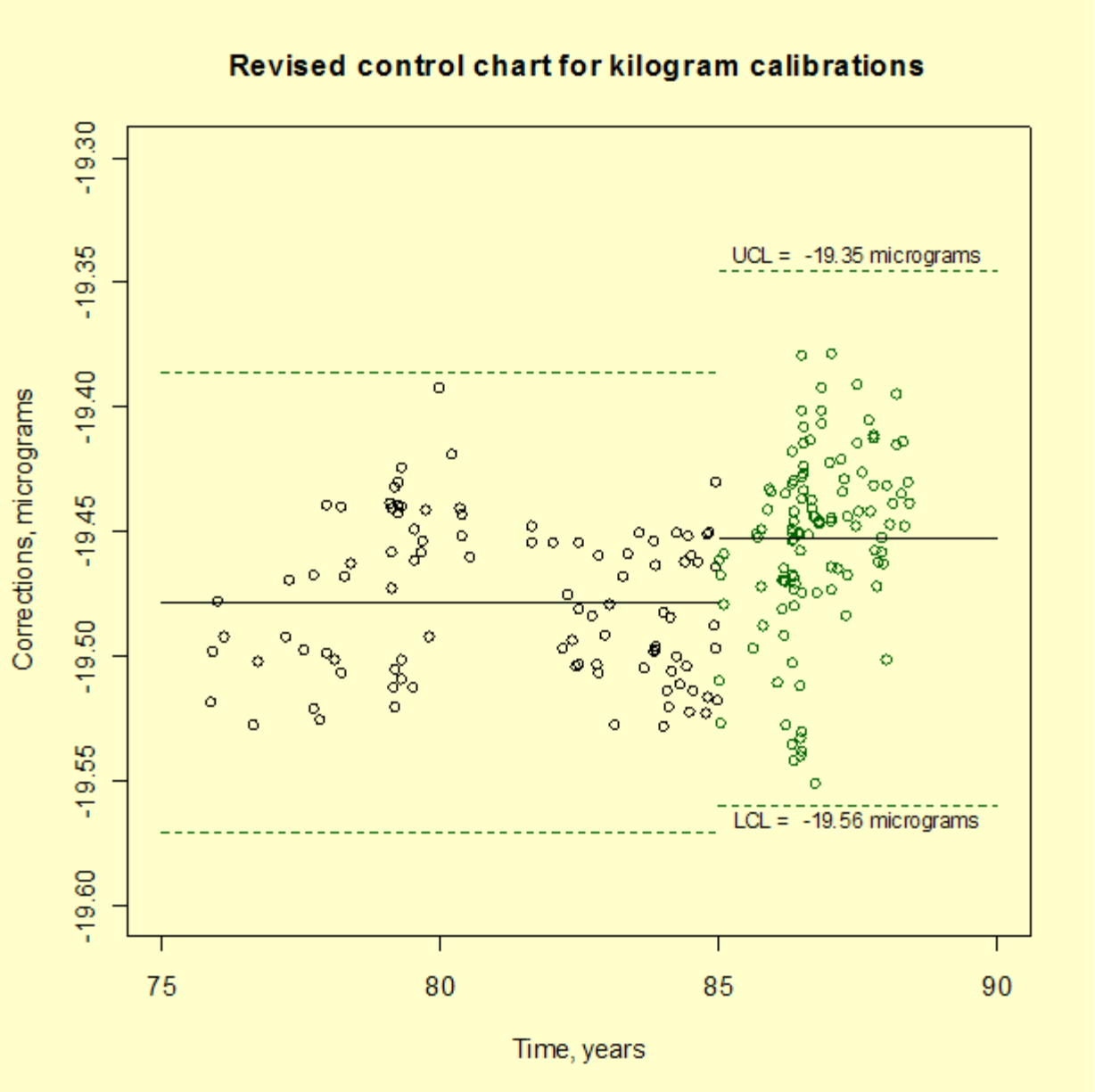
Small changes only become obvious over time

Unfortunately, it takes time for the patterns in the data to emerge because individual violations of the control limits do not necessarily point to a permanent shift in the process. The Shewhart control chart is not powerful for detecting small changes, say of the order of at most one standard deviation, which appears to be approximately the case in this application. This level of change might seem insignificant, but the calculation of uncertainties for the calibration process depends on the control limits.

Re-establishing the limits based on recent data and EWMA option

If the limits for the control chart are re-calculated based on the data after 1985, the extent of the change is obvious. Because the exponentially weighted moving average (EWMA) control chart is capable of detecting small changes, it may be a better choice for a high precision process that is producing many control values.

Revised control chart based on check standard measurements after 1985



Sample code

The original and revised Shewhart control charts can be generated using both Dataplot code and R code.

2.3.5.2.2. Example of EWMA control chart for mass calibrations

Small changes only become obvious over time

Unfortunately, it takes time for the patterns in the data to emerge because individual violations of the control limits do not necessarily point to a permanent shift in the process. The Shewhart control chart is not powerful for detecting small changes, say of the order of at most one standard deviation, which appears to be the case for the calibration data shown on the previous page. The EWMA (exponentially weighted moving average) control chart is better suited for this purpose.

Explanation of EWMA statistic at the kilogram level

The exponentially weighted moving average (EWMA) is a statistic for monitoring the process that averages the data in a way that gives less and less weight to data as they are further removed in time from the current measurement. The EWMA statistic at time *t* is computed recursively from individual data points which are ordered in time to be

$$Y_1, Y_2, \cdots, Y_t$$

where the first EWMA statistic is the average of historical data.

$$EWMA_{t+1} = \lambda Y_t + (1-\lambda) EWMA_t$$

Control mechanism for EWMA

The EWMA control chart can be made sensitive to small changes or a gradual drift in the process by the choice of the weighting factor, λ. A weighting factor between 0.2 - 0.3 has been suggested for this purpose (Hunter), and 0.15 is another popular choice.

Limits for the control chart

The target or center line for the control chart is the average of historical data. The upper (UCL) and lower (LCL) limits are

$$UCL = EWMA_1 + sk\sqrt{\frac{\lambda}{2-\lambda}}$$

$$LCL = EWMA_1 - sk\sqrt{\frac{\lambda}{2-\lambda}}$$

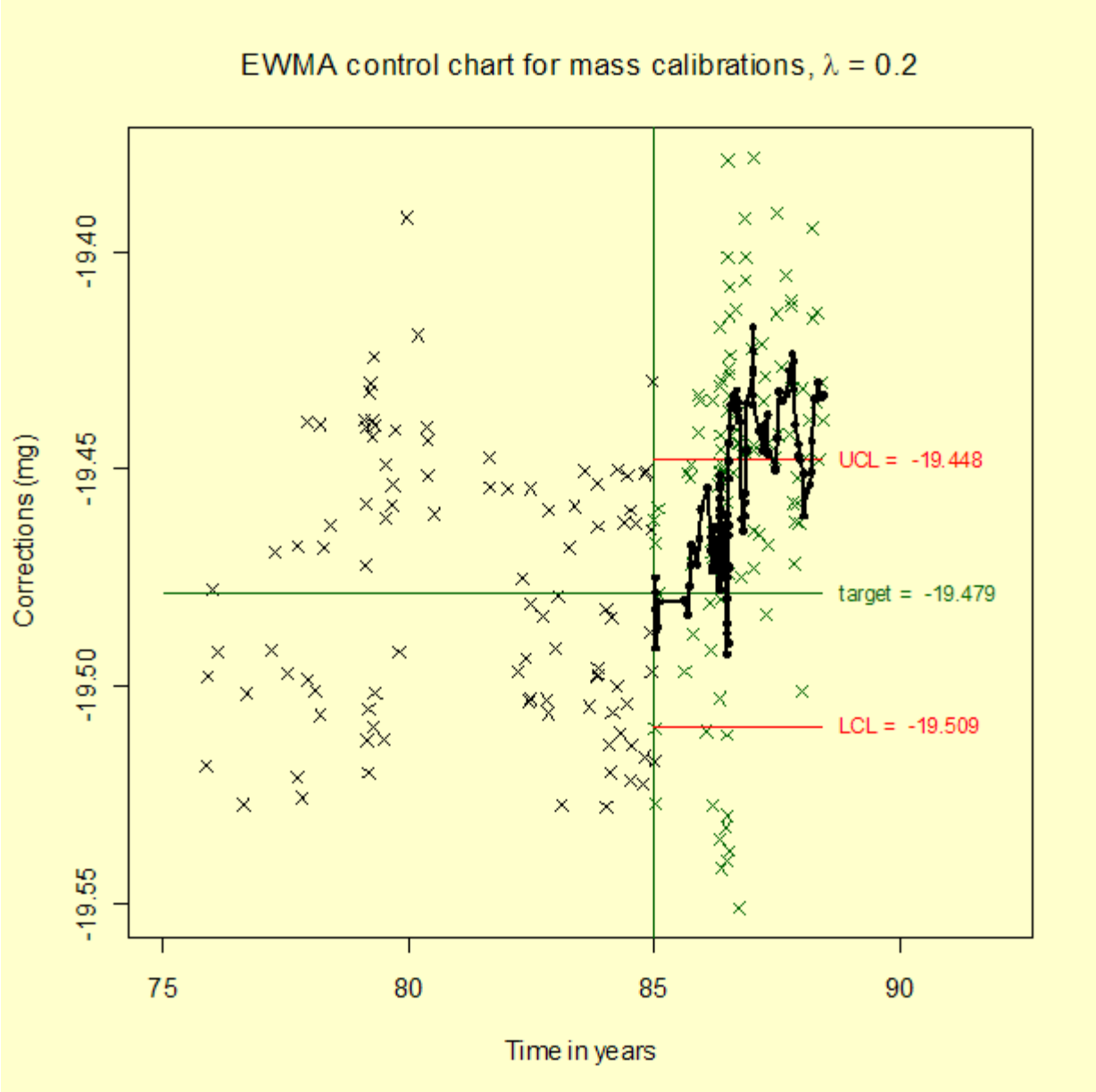
where *s* is the standard deviation of the historical data; the function under the radical is a good approximation to the component of the standard deviation of the EWMA statistic that is a function of time; and *k* is the multiplicative factor, defined in the same manner as for the Shewhart control chart, which is usually taken to be 3.

Example of EWMA chart for check

The target (average) and process standard deviation are computed from the check standard data taken prior to 1985. The computation of the EWMA statistic begins with the data taken at the start of 1985. In the control chart below, the control data after 1985 are shown in green,

standard data for kilogram calibrations showing multiple violations of the control limits for the EWMA statistics

and the EWMA statistics are shown as black dots superimposed on the raw data. The control limits are calculated according to the equation above where the process standard deviation, s = 0.03065 mg and k = 3. The EWMA statistics, and not the raw data, are of interest in looking for out-of-control signals. Because the EWMA statistic is a weighted average, it has a smaller standard deviation than a single control measurement, and, therefore, the EWMA control limits are narrower than the limits for a Shewhart control chart.



The EWMA control chart for mass calibrations can be generated using both Dataplot code and

R code.

Interpretation of the control chart

The EWMA control chart shows many violations of the control limits starting at approximately the mid-point of 1986. This pattern emerges because the process average has actually shifted about one standard deviation, and the EWMA control chart is sensitive to small changes.

2.3.6. Instrument calibration over a regime

Topics

This section discusses the creation of a calibration curve for calibrating instruments (gauges) whose responses cover a large range. Topics are:

- Models for instrument calibration
- Data collection
- Assumptions
- Conditions that can invalidate the calibration procedure
- Data analysis and model validation
- Calibration of future measurements
- Uncertainties of calibrated values

Purpose of instrument calibration

Instrument calibration is intended to eliminate or reduce bias in an instrument's readings over a range for all continuous values. For this purpose, reference standards with known values for selected points covering the range of interest are measured with the instrument in question. Then a functional relationship is established between the values of the standards and the corresponding measurements. There are two basic situations.

Instruments which require correction for bias

- The instrument reads in the same units as the reference standards. The purpose of the calibration is to identify and eliminate any bias in the instrument relative to the defined unit of measurement. For example, optical imaging systems that measure the width of lines on semiconductors read in micrometers, the unit of interest. Nonetheless, these instruments must be calibrated to values of reference standards if line width measurements across the industry are to agree with each other.

Instruments whose measurements act as surrogates for other measurements

- The instrument reads in different units than the reference standards. The purpose of the calibration is to convert the instrument readings to the units of interest. An example is densitometer measurements that act as surrogates for measurements of radiation dosage. For this purpose, reference standards are irradiated at several dosage levels and then measured by radiometry. The same reference standards are measured by densitometer. The calibrated results of future densitometer readings on medical devices are the basis for deciding if the devices have been sterilized at the proper radiation level.

Basic steps for correcting the instrument for bias

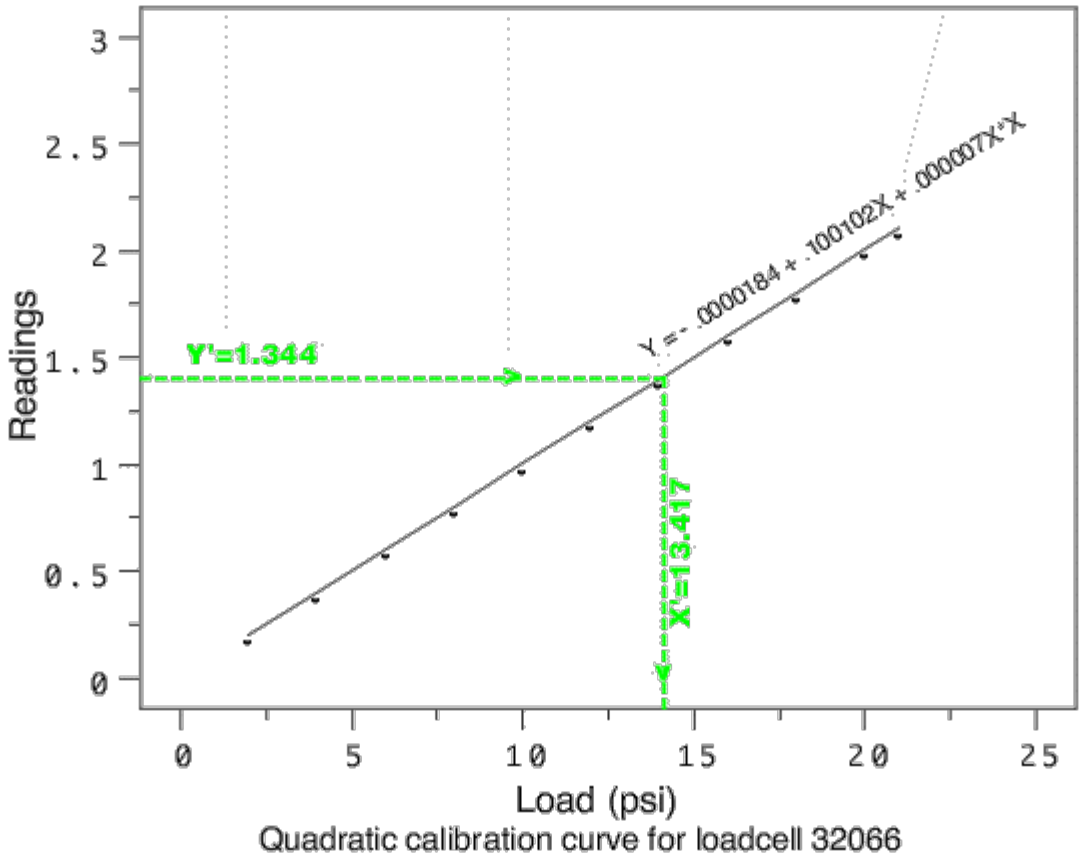
The calibration method is the same for both situations and requires the following basic steps:

- Selection of reference standards with known values to cover the range of interest.
- Measurements on the reference standards with the instrument to be calibrated.
- Functional relationship between the measured and known values of the reference standards (usually a least-squares fit to the data) called a **calibration curve**.
- Correction of all measurements by the inverse of the calibration curve.

Schematic example of a calibration curve and resulting value

A schematic explanation is provided by the figure below for load cell calibration. The loadcell measurements (shown as *) are plotted on the *y*-axis against the corresponding values of known load shown on the *x*-axis.

A quadratic fit to the loadcell data produces the calibration curve that is shown as the solid line. For a future measurement with the load cell, *Y' = 1.344* on the *y*-axis, a dotted line is drawn through *Y'* parallel to the *x*-axis. At the point where it intersects the calibration curve, another dotted line is drawn parallel to the y-axis. Its point of intersection with the *x*-axis at *X' = 13.417* is the calibrated value.



2.3.6.1. Models for instrument calibration

Notation

The following notation is used in this chapter in discussing models for calibration curves.

- Y denotes a measurement on a reference standard
- X denotes the known value of a reference standard
- ε denotes measurement error.
- a, b and c denote coefficients to be determined

Possible forms for calibration curves

There are several models for calibration curves that can be considered for instrument calibration. They fall into the following classes:

- Linear:

$$Y = a + bX + \varepsilon$$

- Quadratic:

$$Y = a + bX + cX^2 + \varepsilon$$

- Power:

$$Y = aX^b\varepsilon$$

- Non-linear:

$$Y = g(X) + \varepsilon$$

Special case of linear model - no calibration required

An instrument requires no calibration if

$$a=0 \text{ and } b=1$$

i.e., if measurements on the reference standards agree with their known values given an allowance for measurement error, the instrument is already calibrated. Guidance on collecting data, estimating and testing the coefficients is given on other pages.

Advantages of the linear

The linear model ISO 11095 is widely applied to instrument calibration because it has several advantages

model

over more complicated models.

- Computation of coefficients and standard deviations is easy.
- Correction for bias is easy.
- There is often a theoretical basis for the model.
- The analysis of uncertainty is tractable.

***Warning** on excluding the intercept term from the model*

It is often tempting to exclude the intercept, a, from the model because a zero stimulus on the x-axis should lead to a zero response on the y-axis. However, the correct procedure is to fit the full model and test for the significance of the intercept term.

Quadratic model and higher order polynomials

Responses of instruments or measurement systems which cannot be linearized, and for which no theoretical model exists, can sometimes be described by a quadratic model (or higher-order polynomial). An example is a load cell where force exerted on the cell is a non-linear function of load.

Disadvantages of quadratic models

Disadvantages of quadratic and higher-order polynomials are:

- They may require more reference standards to capture the region of curvature.
- There is rarely a theoretical justification; however, the adequacy of the model can be tested statistically.
- The correction for bias is more complicated than for the linear model.
- The uncertainty analysis is difficult.

Warning

A plot of the data, although always recommended, is not sufficient for identifying the correct model for the calibration curve. Instrument responses may not appear non-linear over a large interval. If the response and the known values are in the same units, differences from the known values should be plotted versus the known values.

Power model treated as a linear model

The power model is appropriate when the measurement error is proportional to the response rather than being additive. It is frequently used for calibrating instruments that measure dosage levels of irradiated materials.

The power model is a special case of a non-linear model that can be linearized by a natural logarithm transformation to

$$Y = \log_e(a) + b \cdot \log_e(X) + \log_e(\varepsilon)$$

so that the model to be fit to the data is of the familiar

linear form

$$W = a' + bZ + e$$

where W, Z and e are the transforms of the variables, Y, X and the measurement error, respectively, and a' is the natural logarithm of a.

Non-linear models and their limitations

Instruments whose responses are not linear in the coefficients can sometimes be described by non-linear models. In some cases, there are theoretical foundations for the models; in other cases, the models are developed by trial and error. Two classes of non-linear functions that have been shown to have practical value as calibration functions are:

1. Exponential
2. Rational

Non-linear models are an important class of calibration models, but they have several significant limitations.

- The model itself may be difficult to ascertain and verify.
- There can be severe computational difficulties in estimating the coefficients.
- Correction for bias cannot be applied algebraically and can only be approximated by interpolation.
- Uncertainty analysis is very difficult.

Example of an exponential function

An exponential function is shown in the equation below. Instruments for measuring the ultrasonic response of reference standards with various levels of defects (holes) that are submerged in a fluid are described by this function.

$$Y = \frac{e^{-aX}}{b + cX} + \varepsilon$$

Example of a rational function

A rational function is shown in the equation below. Scanning electron microscope measurements of line widths on semiconductors are described by this function (Kirby).

$$Y = \frac{a + bX + cX^2}{a_1 + b_1 X + c_1 X^2} + \varepsilon$$

2.3.6.2. Data collection

Data collection

The process of collecting data for creating the calibration curve is critical to the success of the calibration program. General rules for designing calibration experiments apply, and guidelines that are adequate for the calibration models in this chapter are given below.

Selection of reference standards

A minimum of five reference standards is required for a linear calibration curve, and ten reference standards should be adequate for more complicated calibration models.

The optimal strategy in selecting the reference standards is to space the reference standards at points corresponding to equal increments on the *y*-axis, covering the range of the instrument. Frequently, this strategy is not realistic because the person producing the reference materials is often not the same as the person who is creating the calibration curve. Spacing the reference standards at equal intervals on the *x*-axis is a good alternative.

Exception to the rule above - bracketing

If the instrument is not to be calibrated over its entire range, but only over a very short range for a specific application, then it may not be necessary to develop a complete calibration curve, and a bracketing technique (ISO 11095) will provide satisfactory results. The bracketing technique assumes that the instrument is linear over the interval of interest, and, in this case, only two reference standards are required -- one at each end of the interval.

Number of repetitions on each reference standard

A minimum of two measurements on each reference standard is required and four is recommended. The repetitions should be separated in time by days or weeks. These repetitions provide the data for determining whether a candidate model is adequate for calibrating the instrument.

2.3.6.3. Assumptions for instrument calibration

Assumption regarding reference values	The basic assumption regarding the reference values of artifacts that are measured in the calibration experiment is that they are known without error. In reality, this condition is rarely met because these values themselves usually come from a measurement process. Systematic errors in the reference values will always bias the results, and random errors in the reference values can bias the results.
Rule of thumb	It has been shown by Bruce Hoadly, in an internal NIST publication, that the best way to mitigate the effect of random fluctuations in the reference values is to plan for a large spread of values on the x-axis relative to the precision of the instrument.
Assumptions regarding measurement errors	The basic assumptions regarding measurement errors associated with the instrument are that they are: • free from outliers • independent • of equal precision • from a normal distribution.

2.3.6.4. What can go wrong with the calibration procedure

Calibration procedure may fail to eliminate bias

There are several circumstances where the calibration curve will not reduce or eliminate bias as intended. Some are discussed on this page. A critical exploratory analysis of the calibration data should expose such problems.

Lack of precision

Poor instrument precision or unsuspected day-to-day effects may result in standard deviations that are large enough to jeopardize the calibration. There is nothing intrinsic to the calibration procedure that will improve precision, and the best strategy, before committing to a particular instrument, is to estimate the instrument's precision in the environment of interest to decide if it is good enough for the precision required.

Outliers in the calibration data

Outliers in the calibration data can seriously distort the calibration curve, particularly if they lie near one of the endpoints of the calibration interval.

- Isolated outliers (single points) should be deleted from the calibration data.
- An entire day's results which are inconsistent with the other data should be examined and rectified before proceeding with the analysis.

Systematic differences among operators

It is possible for different operators to produce measurements with biases that differ in sign and magnitude. This is not usually a problem for automated instrumentation, but for instruments that depend on line of sight, results may differ significantly by operator. To diagnose this problem, measurements by different operators on the same artifacts are plotted and compared. Small differences among operators can be accepted as part of the imprecision of the measurement process, but large systematic differences among operators require resolution. Possible solutions are to retrain the operators or maintain separate calibration curves by operator.

Lack of system

The calibration procedure, once established, relies on the instrument continuing to respond in the same way over time.

control	If the system drifts or takes unpredictable excursions, the calibrated values may not be properly corrected for bias, and depending on the direction of change, the calibration may further degrade the accuracy of the measurements. To assure that future measurements are properly corrected for bias, the calibration procedure should be coupled with a statistical control procedure for the instrument.
Example of differences among repetitions in the calibration data	An important point, but one that is rarely considered, is that there can be differences in responses from repetition to repetition that will invalidate the analysis. A plot of the aggregate of the calibration data may not identify changes in the instrument response from day-to-day. What is needed is a plot of the fine structure of the data that exposes any day to day differences in the calibration data.
***Warning** - calibration can fail because of day-to-day changes*	A straight-line fit to the aggregate data will produce a 'calibration curve'. However, if straight lines fit separately to each day's measurements show very disparate responses, the instrument, at best, will require calibration on a daily basis and, at worst, may be sufficiently lacking in control to be usable.

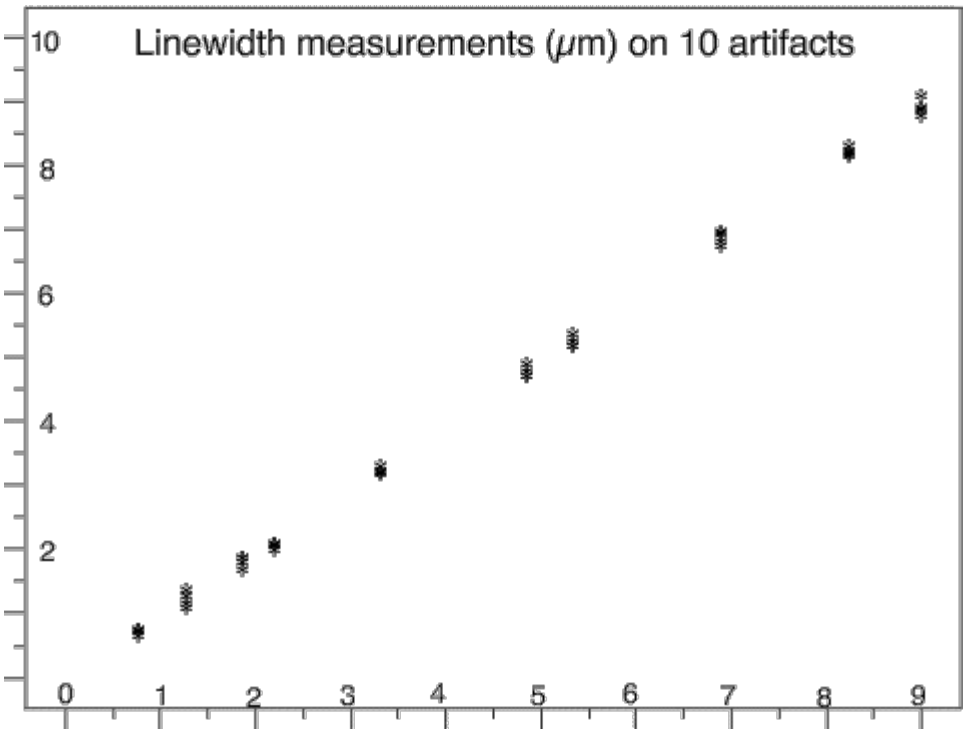
2. Measurement Process Characterization
2.3. Calibration
2.3.6. Instrument calibration over a regime
2.3.6.4. What can go wrong with the calibration procedure

2.3.6.4.1. Example of day-to-day changes in calibration

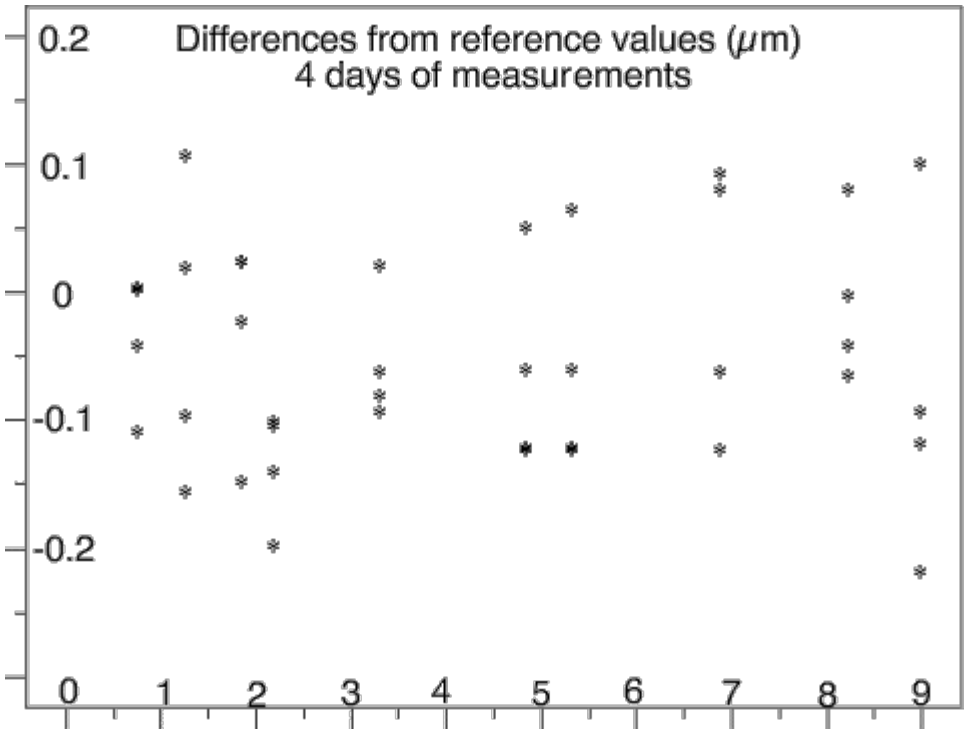
Calibration data over 4 days

Line width measurements on 10 NIST reference standards were made with an optical imaging system on each of four days. The four data points for each reference value appear to overlap in the plot because of the wide spread in reference values relative to the precision. The plot suggests that a linear calibration line is appropriate for calibrating the imaging system.

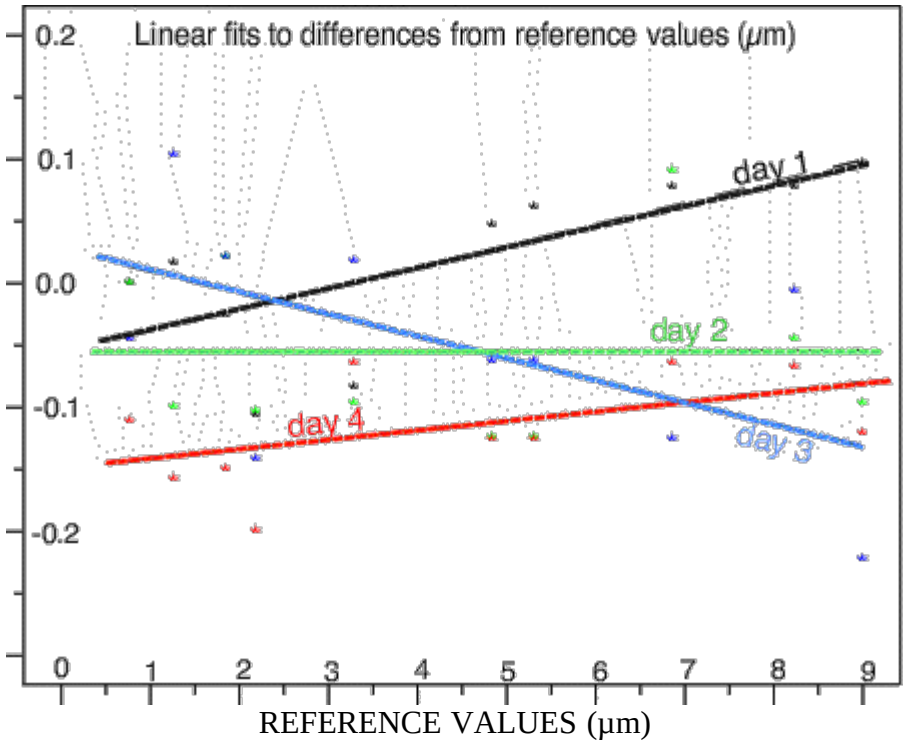
This plot shows measurements made on 10 reference materials repeated on four days with the 4 points for each day overlapping



This plot shows the differences between each measurement and the corresponding reference value. Because days are not identified, the



plot gives no indication of problems in the control of the imaging system from from day to day.



This plot, with linear calibration lines fit to each day's measurements individually, shows how the response of the imaging system changes dramatically from day to day. Notice that the slope of the calibration line goes from positive on day 1 to negative on day 3.

Interpretation of calibration findings

Given the lack of control for this measurement process, any calibration procedure built on the average of the calibration data will fail to properly correct the system on some days and invalidate resulting measurements. There is no good solution to this problem except daily calibration.

2.3.6.5. Data analysis and model validation

First step - plot the calibration data

If the model for the calibration curve is not known from theoretical considerations or experience, it is necessary to identify and validate a model for the calibration curve. To begin this process, the calibration data are plotted as a function of known values of the reference standards; this plot should suggest a candidate model for describing the data. A linear model should always be a consideration. If the responses and their known values are in the same units, a plot of differences between responses and known values is more informative than a plot of the data for exposing structure in the data.

Warning - regarding statistical software

Once an initial model has been chosen, the coefficients in the model are estimated from the data using a statistical software package. It is impossible to over-emphasize the importance of using reliable and documented software for this analysis.

Output required from a software package

The software package will use the method of least squares for estimating the coefficients. The software package should also be capable of performing a 'weighted' fit for situations where errors of measurement are non-constant over the calibration interval. The choice of weights is usually the responsibility of the user. The software package should, at the minimum, provide the following information:

- Coefficients of the calibration curve
- Standard deviations of the coefficients
- Residual standard deviation of the fit
- F-ratio for goodness of fit (if there are repetitions on the y-axis at each reference value)

Typical analysis of a quadratic fit

Load cell measurements are modeled as a quadratic function of known loads as shown below. There are three repetitions at each load level for a total of 33 measurements.

```
Parameter estimates for model y = a + b*x +
c*x*x + e:
```

```
Parameter  Estimate  Std. Error     t-value
Pr(>|t|)
a        -1.840e-05   2.451e-05      -0.751
0.459
b         1.001e-01   4.839e-06   20687.891
<2e-16
c         7.032e-06   2.014e-07      34.922
<2e-16

Residual standard error = 3.764e-05 (30 degrees
of freedom)
Multiple R-squared = 1
Adjusted R-squared = 1


Analysis of variance table:

Source of     Degrees of   Sum of         Mean
Variation      Freedom     Squares       Square
F-Ratio      Pr(>F)

Model              2         12.695       6.3475
4.48e+09    <2.2e-16
Residual          30       4.2504e-08    1.4170e-
09

(Lack of fit)      8       4.7700e-09    5.9625e-
10    0.3477       0.9368
(Pure error)      22       3.7733e-08    1.7151e-
09

Total             32         12.695
```

The analyses shown above can be reproduced using Dataplot code and R code.

Note: Dataplot reports a probability associated with the F-ratio (for example, 6.334 % for the lack-of-fit test), where a probability greater than 95 % indicates an F-ratio that is significant at the 5 % level. R reports a *p*-value that corresponds to the probability greater than the F-ratio, so a value less than 0.05 would indicate significance at the 5 % level. Other software may report in other ways; therefore, it is necessary to check the interpretation for each package.

The F-ratio is used to test the goodness of the fit to the data

The F-ratio provides information on the model as a good descriptor of the data. The F-ratio is compared with a critical value from the F-table. An F-ratio smaller than the critical value indicates that all significant structure has been captured by the model.

F-ratio < 1 always indicates a good fit

For the load cell analysis, a plot of the data suggests a linear fit. However, the linear fit gives a very large F-ratio. For the quadratic fit, the F-ratio is 0.3477 with $v_1 = 8$ and $v_2 = 22$ degrees of freedom. The critical value of F(0.05, 8, 20) = 2.45 indicates that the quadratic function is sufficient for describing the data. A fact to keep in mind is that an F-ratio < 1 does not need to be checked against a critical value; it always indicates a good fit to the data.

The t-values are used to test the significance of

The *t*-values can be compared with critical values from a *t*-table. However, for a test at the 5 % significance level, a *t*-value < 2 is a good indicator of non-significance. The *t*-value for the intercept term, a, is < 2 indicating that the

individual coefficients	intercept term is not significantly different from zero. The *t*-values for the linear and quadratic terms are significant indicating that these coefficients are needed in the model. If the intercept is dropped from the model, the analysis is repeated to obtain new estimates for the coefficients, b and c.
Residual standard deviation	The residual standard deviation estimates the standard deviation of a single measurement with the load cell.
Further considerations and tests of assumptions	The residuals (differences between the measurements and their fitted values) from the fit should also be examined for outliers and structure that might invalidate the calibration curve. They are also a good indicator of whether basic assumptions of normality and equal precision for all measurements are valid. If the initial model proves inappropriate for the data, a strategy for improving the model is followed.

2.3.6.5.1. Data on load cell #32066

Three repetitions on a load cell at eleven known loads

X	Y
2.	0.20024
2.	0.20016
2.	0.20024
4.	0.40056
4.	0.40045
4.	0.40054
6.	0.60087
6.	0.60075
6.	0.60086
8.	0.80130
8.	0.80122
8.	0.80127
10.	1.00173
10.	1.00164
10.	1.00173
12.	1.20227
12.	1.20218
12.	1.20227
14.	1.40282
14.	1.40278
14.	1.40279
16.	1.60344
16.	1.60339
16.	1.60341
18.	1.80412
18.	1.80409
18.	1.80411
20.	2.00485
20.	2.00481
20.	2.00483
21.	2.10526
21.	2.10524
21.	2.10524

2.3.6.6. Calibration of future measurements

Purpose

The purpose of creating the calibration curve is to correct future measurements made with the same instrument to the correct units of measurement. The calibration curve can be applied many, many times before it is discarded or reworked as long as the instrument remains in statistical control. Chemical measurements are an exception where frequently the calibration curve is used only for a single batch of measurements, and a new calibration curve is created for the next batch.

Notation

The notation for this section is as follows:

- Y' denotes a future measurement.
- X' denotes the associated calibrated value.
- $\hat{a}, \hat{b}, \hat{c}$ are the estimates of the coefficients, a, b, c.
- s_a, s_b, s_c are standard deviations of the coefficients, a, b, c.

Procedure

To apply a correction to a future measurement, Y^*, to obtain the calibration value X^* requires the inverse of the calibration curve.

Linear calibration line

The inverse of the calibration line for the linear model

$$Y = a + bX + \varepsilon$$

gives the calibrated value

$$X' = \frac{Y' - \hat{a}}{\hat{b}}$$

Tests for the intercept and slope of calibration

Before correcting for the calibration line by the equation above, the intercept and slope should be tested for $a=0$, and $b=1$. If both

$$\left|\frac{\hat{a}}{s_a}\right| < t_{1-\alpha/2,\nu} \quad \text{and} \quad \left|\frac{\hat{b}-1}{s_b}\right| < t_{1-\alpha/2,\nu}$$

curve -- If both conditions hold, no calibration is needed.

there is no need for calibration. If, on the other hand only the test for $a=0$ *fails*, the error is constant; if only the test for $b=1$ *fails*, the errors are related to the size of the reference standards.

Table look-up for t-factor

The factor, $t_{1-\alpha/2,\,\nu}$, is found in the *t*-table where ν is the degrees of freedom for the residual standard deviation from the calibration curve, and α is chosen to be small, say, 0.05.

Quadratic calibration curve

The inverse of the calibration curve for the quadratic model

$$Y = a + bX + cX^2 + \varepsilon$$

requires a root

$$X' = \frac{-\hat{b} \pm \sqrt{\hat{b}^2 - 4\hat{c}\left(\hat{a} - Y'\right)}}{2\hat{c}}$$

The correct root (+ or -) can usually be identified from practical considerations.

Power curve

The inverse of the calibration curve for the power model

$$Y = aX^b\varepsilon$$

gives the calibrated value

$$X' = \exp\left(\frac{\log_e(Y') - \log_e\left(\hat{a}\right)}{\hat{b}}\right)$$

where b and the natural logarithm of a are estimated from the power model transformed to a linear function.

Non-linear and other calibration curves

For more complicated models, the inverse for the calibration curve is obtained by interpolation from a graph of the function or from predicted values of the function.

2.3.6.7. Uncertainties of calibrated values

Purpose

The purpose is to quantify the uncertainty of a 'future' result that has been corrected by the calibration curve. In principle, the uncertainty quantifies any possible difference between the calibrated value and its reference base (which normally depends on reference standards).

Explanation in terms of reference artifacts

Measurements of interest are future measurements on unknown artifacts, but one way to look at the problem is to ask: If a measurement is made on one of the reference standards and the calibration curve is applied to obtain the calibrated value, how well will this value agree with the 'known' value of the reference standard?

Difficulties

The answer is not easy because of the intersection of two uncertainties associated with

1. the calibration curve itself because of limited data
2. the 'future' measurement

If the calibration experiment were to be repeated, a slightly different calibration curve would result even for a system in statistical control. An exposition of the intersection of the two uncertainties is given for the calibration of proving rings (Hockersmith and Ku).

ISO approach to uncertainty can be based on check standards or propagation of error

General procedures for computing an uncertainty based on ISO principles of uncertainty analysis are given in the chapter on modeling.

Type A uncertainties for calibrated values from calibration curves can be derived from

- check standard values
- propagation of error

An example of type A uncertainties of calibrated values from a linear calibration curve are analyzed from measurements on linewidth check standards. Comparison of the uncertainties from check standards and propagation of error for the linewidth calibration data are also illustrated.

An example of the derivation of propagation of error type A

uncertainties for calibrated values from a quadratic calibration curve for loadcells is discussed on the next page.

2. Measurement Process Characterization
2.3. Calibration
2.3.6. Instrument calibration over a regime
2.3.6.7. Uncertainties of calibrated values

2.3.6.7.1. Uncertainty for quadratic calibration using propagation of error

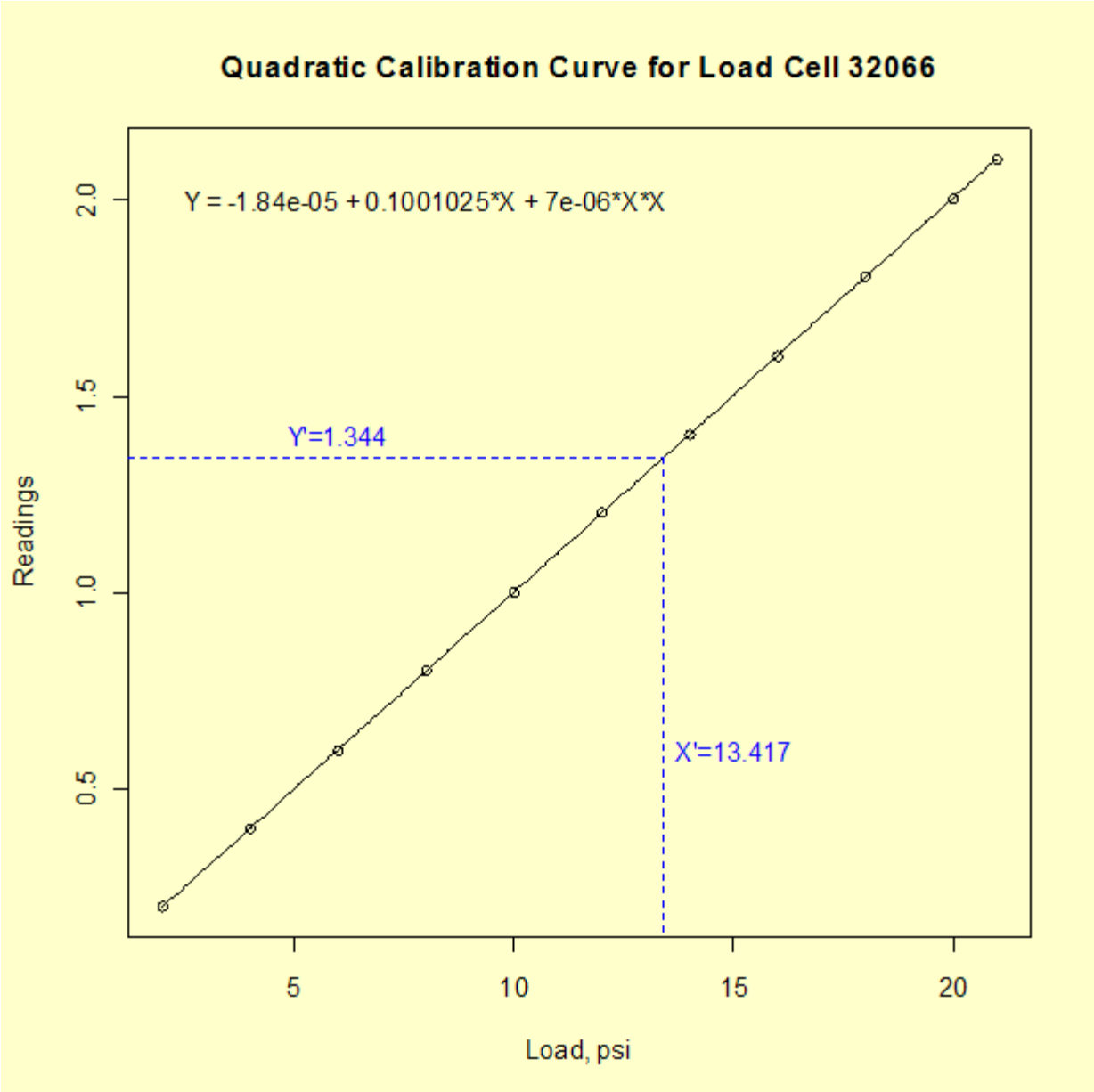
Propagation of error for uncertainty of calibrated values of loadcells

The purpose of this page is to show the propagation of error for calibrated values of a loadcell based on a quadratic calibration curve where the model for instrument response is

$$Y = a + bX + cX^2 + \varepsilon$$

The calibration data are instrument responses at known loads (psi), and estimates of the quadratic coefficients, a, b, c, and their associated standard deviations are shown with the analysis.

A graph of the calibration curve showing a measurement Y' corrected to X', the proper load (psi), is shown below.



Uncertainty of the calibrated value X'

The uncertainty to be evaluated is the uncertainty of the calibrated value, *X'*, computed for any future measurement, *Y'*, made with the calibrated instrument where

$$X' = \frac{-\hat{b} \pm \sqrt{\hat{b}^2 - 4\hat{c}(\hat{a} - Y')}}{2\hat{c}}$$

Partial derivatives

The partial derivatives are needed to compute uncertainty.

$$\frac{\partial X'}{\partial Y'} = \frac{1}{\sqrt{\hat{b}^2 - 4\hat{c}(\hat{a} - Y')}}$$

$$\frac{\partial X'}{\partial \hat{a}} = \frac{-1}{\sqrt{\hat{b}^2 - 4\hat{c}(\hat{a} - Y')}}$$

$$\frac{\partial X'}{\partial \hat{b}} = \frac{-1 + \frac{\hat{b}}{\sqrt{\hat{b}^2 - 4\hat{c}(\hat{a} - Y')}}}{2\hat{c}}$$

$$\frac{\partial X'}{\partial \hat{c}} = \frac{-\hat{a} + Y'}{\hat{c}\sqrt{\hat{b}^2 - 4\hat{c}(\hat{a} - Y')}} - \frac{-\hat{b} + \sqrt{\hat{b}^2 - 4\hat{c}(\hat{a} - Y')}}{2\hat{c}^2}$$

The variance of the calibrated value from propagation of error

The variance of X' is defined from propagation of error as follows:

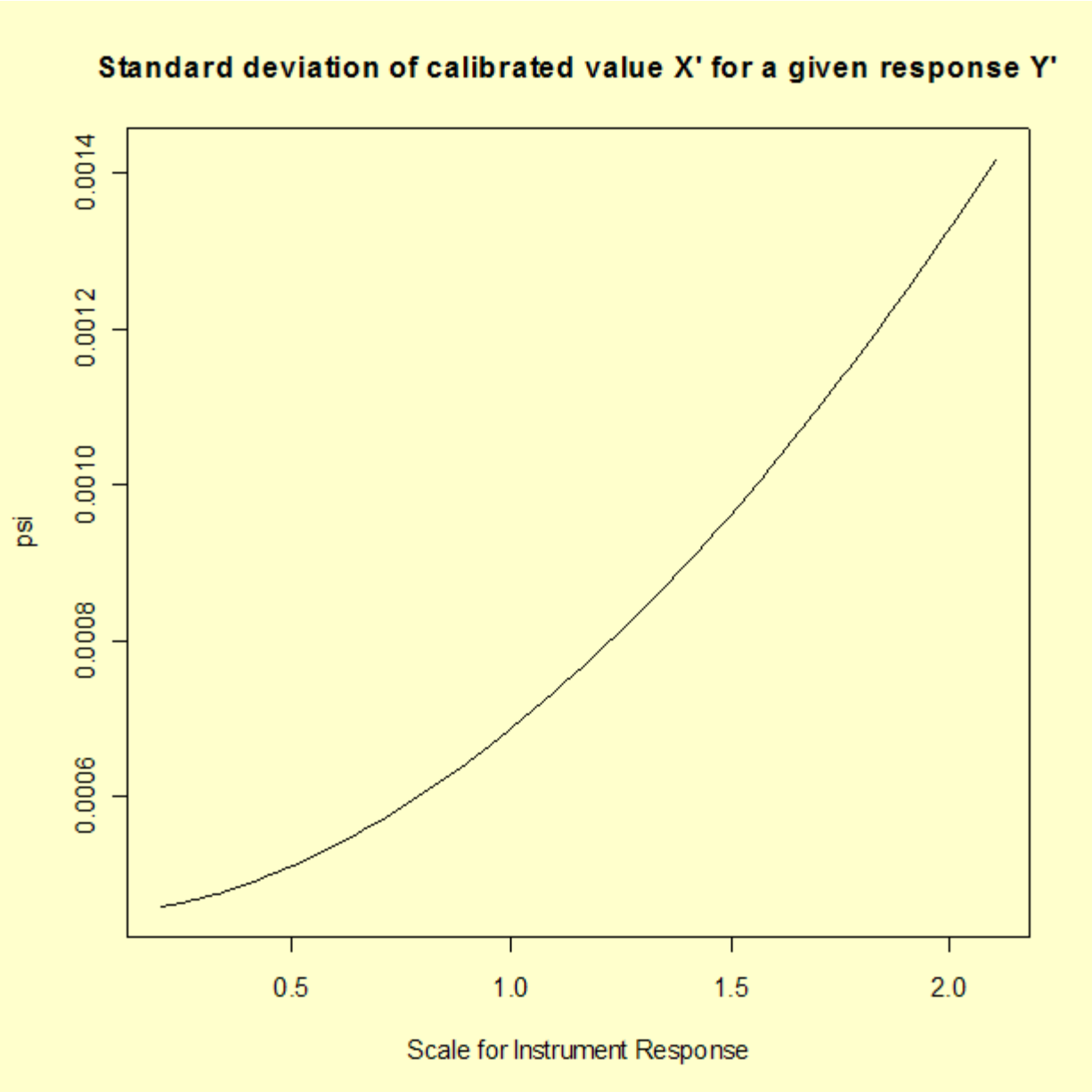
$$u^2 = \left(\frac{\partial X'}{\partial Y'}\right)^2 (s_{Y'})^2 + \left(\frac{\partial X'}{\partial \hat{a}}\right)^2 (s_{\hat{a}})^2 + \left(\frac{\partial X'}{\partial \hat{b}}\right)^2 (s_{\hat{b}})^2 + \left(\frac{\partial X'}{\partial \hat{c}}\right)^2 (s_{\hat{c}})^2$$

The values of the coefficients and their respective standard deviations from the quadratic fit to the calibration curve are substituted in the equation. The standard deviation of the measurement, *Y*, may not be the same as the standard deviation from the fit to the calibration data if the measurements to be corrected are taken with a different system; here we assume that the instrument to be calibrated has a standard deviation that is essentially the same as the instrument used for collecting the calibration data and the residual standard deviation from the quadratic fit is the appropriate estimate.

```
a  = -0.183980e-04
sa =  0.2450e-04
b  =  0.100102
sb =  0.4838e-05
c  =  0.703186e-05
sc =  0.2013e-06
sy =  0.0000376353
```

Graph showing the standard deviations of calibrated values X' for given instrument responses Y' ignoring covariance terms in the propagation of error

The standard deviation expressed above is not easily interpreted but it is easily graphed. A graph showing standard deviations of calibrated values, *X'*, as a function of instrument response, *Y'*, is shown below.



Problem with propagation of error

The propagation of errors shown above is not complete because it ignores the covariances among the coefficients, a, b, c. Unfortunately, some statistical software packages do not display these covariance terms with the other output from the analysis.

Covariance terms for loadcell data

The variance-covariance terms for the loadcell data set are shown below.

	a	b	c
a	6.0049021-10		
b	-1.0759599-10	2.3408589-11	
c	4.0191106-12	-9.5051441-13	4.0538705-14

The diagonal elements are the variances of the coefficients, a, b, c, respectively, and the off-diagonal elements are the covariance terms.

Recomputation of the standard deviation of X'

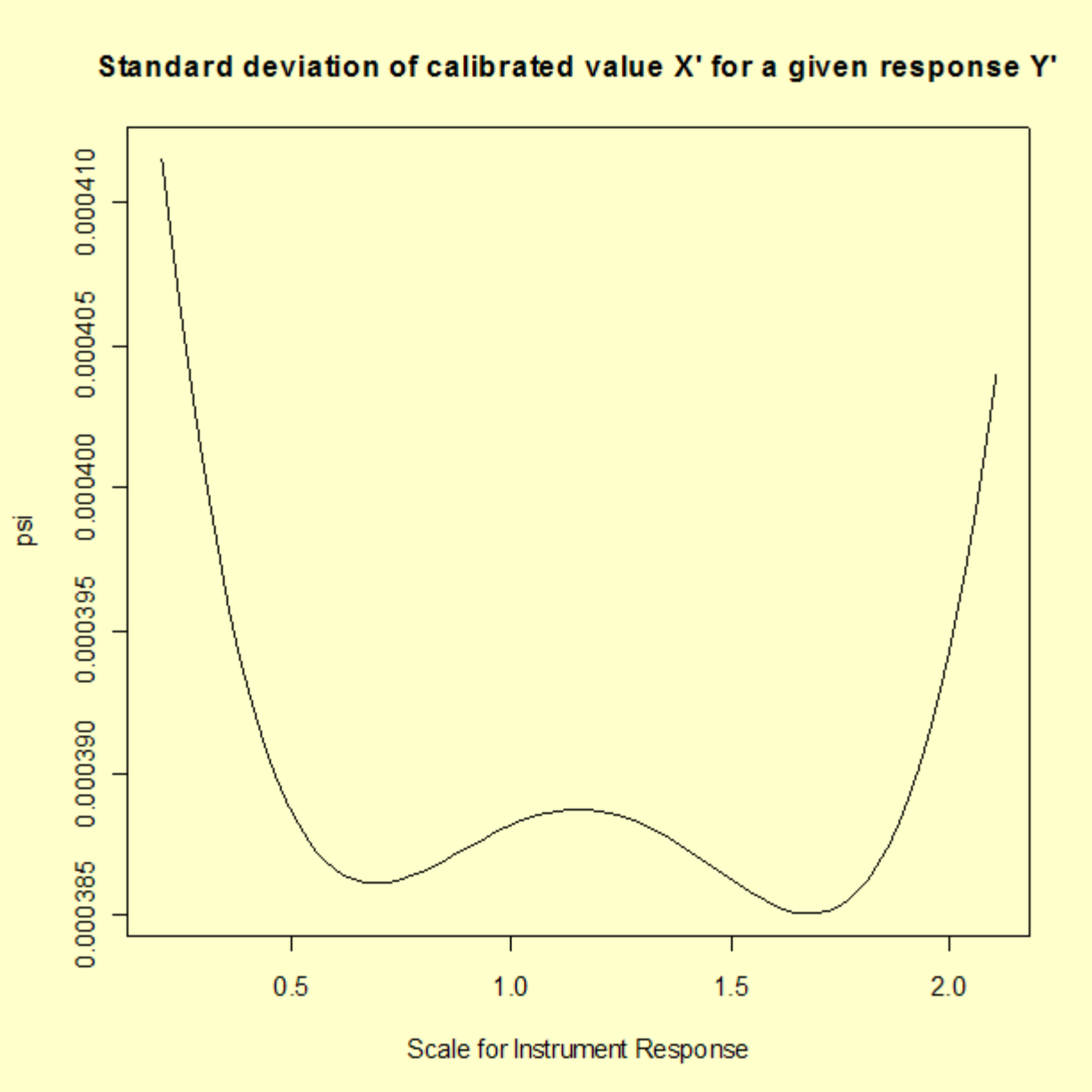
To account for the covariance terms, the variance of X' is redefined by adding the covariance terms. Appropriate substitutions are made; the standard deviations are recomputed and graphed as a function of instrument response.

$$u^2 = u^2 + 2\left(\frac{\partial X'}{\partial \hat{a}}\right)\left(\frac{\partial X'}{\partial \hat{b}}\right)s_{\hat{a}\hat{b}} + 2\left(\frac{\partial X'}{\partial \hat{a}}\right)\left(\frac{\partial X'}{\partial \hat{c}}\right)s_{\hat{a}\hat{c}} + 2\left(\frac{\partial X'}{\partial \hat{b}}\right)\left(\frac{\partial X'}{\partial \hat{c}}\right)s_{\hat{b}\hat{c}}$$

```
sab = -1.0759599e-10
sac =  4.0191106e-12
sbc = -9.5051441e-13
```

The graph below shows the correct estimates for the standard deviation of X' and gives a means for assessing the loss of accuracy that can be incurred by ignoring covariance terms. In this case, the uncertainty is reduced by including covariance terms, some of which are negative.

Graph showing the standard deviations of calibrated values, X', for given instrument responses, Y', with covariance terms included in the propagation of error



Sample code

The results in this section can be generated using R code.

2.3.6.7.2. Uncertainty for linear calibration using check standards

Check standards provide a mechanism for calculating uncertainties

The easiest method for calculating type A uncertainties for calibrated values from a calibration curve requires periodic measurements on check standards. The check standards, in this case, are artifacts at the lower, mid-point and upper ends of the calibration curve. The measurements on the check standard are made in a way that randomly samples the output of the calibration procedure.

Calculation of check standard values

The check standard values are the raw measurements on the artifacts corrected by the calibration curve. The standard deviation of these values should estimate the uncertainty associated with calibrated values. The success of this method of estimating the uncertainties depends on adequate sampling of the measurement process.

Measurements corrected by a linear calibration curve

As an example, consider measurements of linewidths on photomask standards, made with an optical imaging system and corrected by a linear calibration curve. The three control measurements were made on reference standards with values at the lower, mid-point, and upper end of the calibration interval.

Compute the calibration standard deviation

For the linewidth data, the regression equation from the calibration experiment is

$$Y = a + bX + \epsilon$$

and the estimated regression coefficients are the following.

$$\hat{a} = 0.2357$$
$$\hat{b} = 0.9870$$

Next, we calculate the difference between the "predicted" X from the regression fit and the observed X.

$$W_i = \frac{(Y_i - \hat{a})}{\hat{b}} - X_i$$

Finally, we find the calibration standard deviation by calculating the standard deviation of the computed differences.

$$S = \sqrt{\frac{\Sigma (W_i - \overline{W})^2}{n-1}}$$

The calibration standard deviation for the linewidth data is 0.119 µm.

The calculations in this section can be completed using Dataplot code and R code.

Comparison with propagation of error

The standard deviation, 0.119 µm, can be compared with a propagation of error analysis.

Other sources of uncertainty

In addition to the type A uncertainty, there may be other contributors to the uncertainty such as the uncertainties of the values of the reference materials from which the calibration curve was derived.

2. Measurement Process Characterization
2.3. Calibration
2.3.6. Instrument calibration over a regime
2.3.6.7. Uncertainties of calibrated values

2.3.6.7.3. Comparison of check standard analysis and propagation of error

Propagation of error for the linear calibration

The analysis of uncertainty for calibrated values from a linear calibration line can be addressed using propagation of error. On the previous page, the uncertainty was estimated from check standard values.

Estimates from calibration data

The calibration data consist of 40 measurements with an optical imaging system on 10 linewidth artifacts. A linear fit to the data gives a calibration curve with the following estimates for the intercept, *a*, and the slope, *b*:

Parameter	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
a	0.2357623	0.02430034	9.702014	7.860745e-12
b	0.9870377	0.00344058	286.881171	5.354121e-65

with the following covariance matrix.

	a	b
a	5.905067e-04	-7.649453e-05
b	-7.649453e-05	1.183759e-05

The results shown above can be generated with R code.

Propagation of error

The propagation of error is performed for the equation

$$X' = \frac{Y' - \hat{a}}{\hat{b}}$$

so that the squared uncertainty of a calibrated value, X', is

$$u^2 = \left[\frac{\partial X'}{\partial Y'}\right]^2 s^2_{Y'} + \left[\frac{\partial X'}{\partial \hat{a}}\right]^2 s^2_{\hat{a}} + \left[\frac{\partial X'}{\partial \hat{b}}\right]^2 s^2_{\hat{b}} + 2\left[\frac{\partial X'}{\partial \hat{a}}\right]\left[\frac{\partial X'}{\partial \hat{b}}\right] s_{\hat{a}\hat{b}}$$

where

$$\frac{\partial X'}{\partial Y'} = \frac{1}{\hat{b}}$$

$$\frac{\partial X'}{\partial \hat{a}} = \frac{-1}{\hat{b}}$$

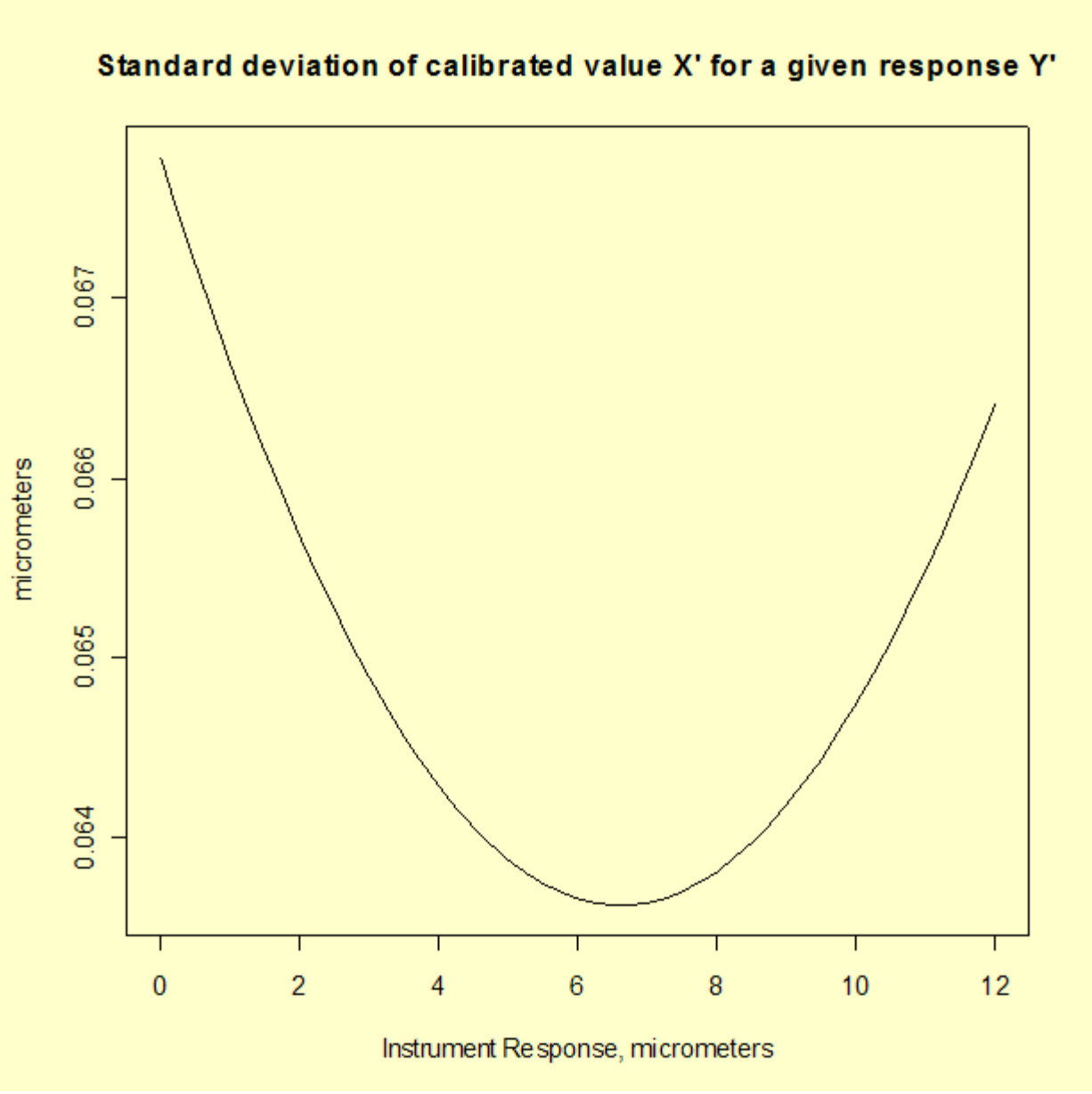
$$\frac{\partial X'}{\partial \hat{b}} = \frac{-(Y' - \hat{a})}{\hat{b}^2}$$

The uncertainty of the calibrated value, X',

$$u^2 = \left(\frac{1}{\hat{b}}\right)^2 s^2_{Y'} + \left(\frac{-1}{\hat{b}}\right)^2 s^2_{\hat{a}} + \left(\frac{-(Y' - \hat{a})}{\hat{b}^2}\right)^2 s^2_{\hat{b}} + 2\left(\frac{-1}{\hat{b}}\right)\left(\frac{-(Y' - \hat{a})}{\hat{b}^2}\right) s_{\hat{a}\hat{b}}$$

is dependent on the value of the instrument reponse Y'.

Graph showing standard deviation of calibrated value X' plotted as a function of instrument response Y' for a linear calibration



Comparison of check standard analysis and propagation of error

Comparison of the analysis of check standard data, which gives a standard deviation of 0.119 µm, and propagation of error, which gives a maximum standard deviation of 0.068 µm, suggests that the propagation of error may underestimate the type A uncertainty. The check standard measurements are undoubtedly sampling some sources of variability that do not appear in the formal propagation of error formula.

2.3.7. Instrument control for linear calibration

Purpose

The purpose of the control program is to guarantee that the calibration of an instrument does not degrade over time.

Approach

This is accomplished by exercising quality control on the instrument's output in much the same way that quality control is exercised on components in a process using a modification of the Shewhart control chart.

Check standards needed for the control program

For linear calibration, it is sufficient to control the end-points and the middle of the calibration interval to ensure that the instrument does not drift out of calibration. Therefore, check standards are required at three points; namely,

- at the lower-end of the regime
- at the mid-range of the regime
- at the upper-end of the regime

Data collection

One measurement is needed on each check standard for each checking period. It is advisable to start by making control measurements at the start of each day or as often as experience dictates. The time between checks can be lengthened if the instrument continues to stay in control.

Definition of control value

To conform to the notation in the section on instrument corrections, X^* denotes the known value of a standard, and X denotes the measurement on the standard.

A control value is defined as the difference

$$W = X^* - X$$

If the calibration is perfect, control values will be randomly distributed about zero and fall within appropriate upper and lower limits on a control chart.

Calculation of control limits

The upper and lower control limits (Croarkin and Varner)) are, respectively,

$$l_{upper} = +\frac{s}{\hat{b}} t^{\bullet}_{\alpha/2}(\nu)$$

$$l_{lower} = -\frac{s}{\hat{b}} t^{\bullet}_{\alpha/2}(\nu)$$

where s is the residual standard deviation of the fit from the calibration experiment, and $\hat{b}$ is the estimated slope of the linear calibration curve.

*Values t**

The critical value, $t^{\bullet}_{\alpha/2}$, can be found in the t* table; ν is the degrees of freedom for the residual standard deviation; and α is equal to 0.05.

*Determining t**

For the case where α = 0.05 and ν = 38, the critical value of the t* statistic is 2.497575.

R code and Dataplot code can be used to determine t* critical values using a standard t-table for the ζ quantile and ν degrees of freedom where ζ is computed as

$$\zeta = \frac{1}{2}\left[1 - e^{\left(\frac{\ln(1-\alpha)}{m}\right)}\right]$$

where m is the number of check standards.

Sensitivity to departure from linearity

If

$$l_{lower} \le W \le l_{upper}$$

the instrument is in statistical control. Statistical control in this context implies not only that measurements are repeatable within certain limits but also that instrument response remains linear. The test is sensitive to departures from linearity.

Control chart for a system corrected by a linear calibration curve

An example of measurements of line widths on photomask standards, made with an optical imaging system and corrected by a linear calibration curve, are shown as an example. The three control measurements were made on reference standards with values at the lower, mid-point, and upper end of the calibration interval.

2.3.7.1. Control chart for a linear calibration line

Purpose

Line widths of three photomask reference standards (at the low, middle and high end of the calibration line) were measured on six days with an optical imaging system that had been calibrated from similar measurements on 10 reference artifacts. The control values and limits for the control chart , which depend on the intercept and slope of the linear calibration line, monitor the calibration and linearity of the optical imaging system.

Initial calibration experiment

The initial calibration experiment consisted of 40 measurements (not shown here) on 10 artifacts and produced a linear calibration line with:

- Intercept = 0.2357
- Slope = 0.9870
- Residual standard deviation = 0.06203 micrometers
- Degrees of freedom = 38

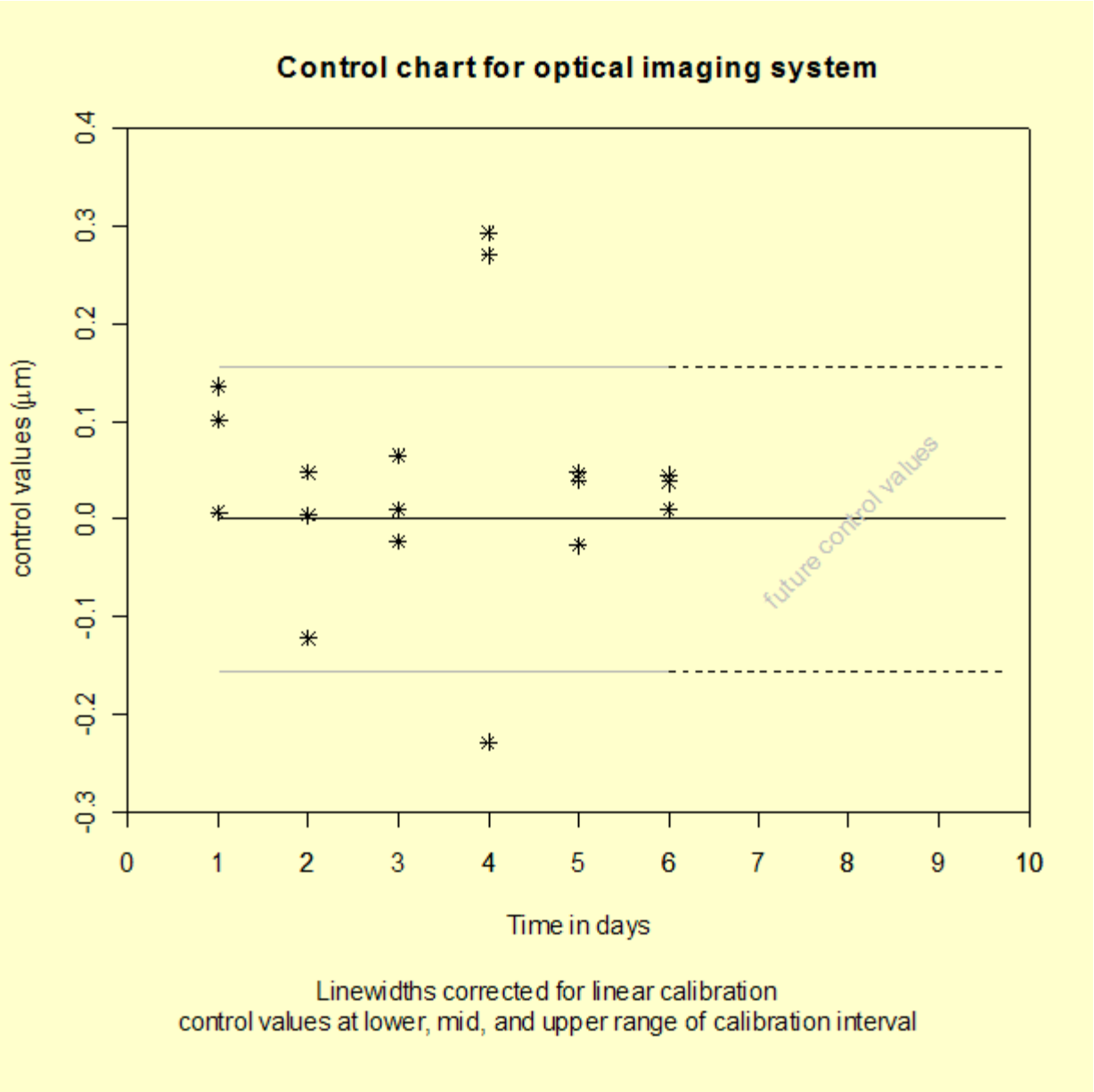
Line width measurements made with an optical imaging system

The control measurements, *Y*, and known values, *X*, for the three artifacts at the upper, mid-range, and lower end (U, M, L) of the calibration line are shown in the following table:

DAY	POSITION	X	Y
1	L	0.76	1.12
1	M	3.29	3.49
1	U	8.89	9.11
2	L	0.76	0.99
2	M	3.29	3.53
2	U	8.89	8.89
3	L	0.76	1.05
3	M	3.29	3.46
3	U	8.89	9.02
4	L	0.76	0.76
4	M	3.29	3.75
4	U	8.89	9.30
5	L	0.76	0.96
5	M	3.29	3.53
5	U	8.89	9.05
6	L	0.76	1.03
6	M	3.29	3.52
6	U	8.89	9.02

Control chart

The control chart shown below can be generated using both Dataplot code and R code.



Interpretation of control chart

The control measurements show no evidence of drift and are within the control limits except on the fourth day when all three control values are outside the limits. The cause of the problem on that day cannot be diagnosed from the data at hand, but all measurements made on that day, including workload items, should be rejected and remeasured.

2.4. Gauge R & R studies

The purpose of this section is to outline the steps that can be taken to characterize the performance of gauges and instruments used in a production setting in terms of errors that affect the measurements.

What are the issues for a gauge R & R study?

What are the design considerations for the study?

1. Artifacts
2. Operators
3. Gauges, parameter levels, configurations

How do we collect data for the study?

How do we quantify variability of measurements?

1. Repeatability
2. Reproducibility
3. Stability

How do we identify and analyze bias?

1. Resolution
2. Linearity
3. Hysteresis
4. Drift
5. Differences among gauges
6. Differences among geometries, configurations

Remedies and strategies

How do we quantify uncertainties of measurements made with the gauges?

2.4.1. What are the important issues?

Basic issues

The basic issue for the study is the behavior of gauges in a particular environment with respect to:

- Repeatability
- Reproducibility
- Stability
- Bias

Strategy

The strategy is to conduct and analyze a study that examines the behavior of similar gauges to see if:

- They exhibit different levels of precision;
- Instruments in the same environment produce equivalent results;
- Operators in the same environment produce equivalent results;
- Responses of individual gauges are affected by configuration or geometry changes or changes in setup procedures.

Other goals

Other goals are to:

- Test the resolution of instruments
- Test the gauges for linearity
- Estimate differences among gauges (bias)
- Estimate differences caused by geometries, configurations
- Estimate operator biases
- Incorporate the findings in an uncertainty budget

2.4.2. Design considerations

Design considerations

Design considerations for a gauge study are choices of:

- Artifacts (check standards)
- Operators
- Gauges
- Parameter levels
- Configurations, etc.

Selection of artifacts or check standards

The artifacts for the study are check standards or test items of a type that are typically measured with the gauges under study. It may be necessary to include check standards for different parameter levels if the gauge is a multi-response instrument. The discussion of check standards should be reviewed to determine the suitability of available artifacts.

Number of artifacts

The number of artifacts for the study should be Q $(Q > 2)$. Check standards for a gauge study are needed only for the limited time period (two or three months) of the study.

Selection of operators

Only those operators who are trained and experienced with the gauges should be enlisted in the study, with the following constraints:

- If there is a small number of operators who are familiar with the gauges, they should all be included in the study.
- If the study is intended to be representative of a large pool of operators, then a random sample of L $(L > 2)$ operators should be chosen from the pool.
- If there is only one operator for the gauge type, that operator should make measurements on K $(K > 2)$ days.

Selection of gauges

If there is only a small number of gauges in the facility, then all gauges should be included in the study.

If the study is intended to represent a larger pool of gauges, then a random sample of I $(I > 3)$ gauges should be chosen for the study.

Limit the

If the gauges operate at several parameter levels (for

initial study

example; frequencies), an initial study should be carried out at 1 or 2 levels before a larger study is undertaken.

If there are differences in the way that the gauge can be operated, an initial study should be carried out for one or two configurations before a larger study is undertaken.

NIST SEMATECH HOME TOOLS & AIDS SEARCH BACK NEXT

2.4.3. Data collection for time-related sources of variability

Time-related analysis

The purpose of this page is to present several options for collecting data for estimating time-dependent effects in a measurement process.

Time intervals

The following levels of time-dependent errors are considered in this section based on the characteristics of many measurement systems and should be adapted to a specific measurement situation as needed.

1. Level-1 Measurements taken over a short time to capture the precision of the gauge
2. Level-2 Measurements taken over days (of other appropriate time increment)
3. Level-3 Measurements taken over runs separated by months

Time intervals

- Simple design for 2 levels of random error
- Nested design for 2 levels of random error
- Nested design for 3 levels of random error

In all cases, data collection and analysis are straightforward, and there is no reason to estimate interaction terms when dealing with time-dependent errors. Two levels should be sufficient for characterizing most measurement systems. Three levels are recommended for measurement systems where sources of error are not well understood and have not previously been studied.

2.4.3.1. Simple design

Constraints on time and resources

In planning a gauge study, particularly for the first time, it is advisable to start with a simple design and progress to more complicated and/or labor intensive designs after acquiring some experience with data collection and analysis. The design recommended here is appropriate as a preliminary study of variability in the measurement process that occurs over time. It requires about two days of measurements separated by about a month with two repetitions per day.

Relationship to 2-level and 3-level nested designs

The disadvantage of this design is that there is minimal data for estimating variability over time. A 2-level nested design and a 3-level nested design, both of which require measurments over time, are discussed on other pages.

Plan of action

Choose at least $Q = 10$ work pieces or check standards, which are essentially identical insofar as their expected responses to the measurement method. Measure each of the check standards twice with the same gauge, being careful to randomize the order of the check standards.

After about a month, repeat the measurement sequence, randomizing anew the order in which the check standards are measured.

Notation

Measurements on the check standards are designated:

$$Y_{11}, Y_{12}$$
$$Y_{21}, Y_{22}$$

with the first index identifying the month of measurement and the second index identifying the repetition number.

Analysis of data

The level-1 standard deviation, which describes the basic precision of the gauge, is

$$s_1 = \sqrt{\frac{1}{4Q}\sum_{i=1}^{Q}\{(Y_{11} - Y_{12})^2 + (Y_{21} - Y_{22})^2\}}$$

with $v_1 = 2Q$ degrees of freedom.

The level-2 standard deviation, which describes the variability of the measurement process over time, is

$$s_2 = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q}\left\{\frac{(Y_{11}+Y_{12})-(Y_{21}+Y_{22})}{2}\right\}^2}$$

with $v_2 = Q$ degrees of freedom.

Relationship to uncertainty for a test item

The standard deviation that defines the uncertainty for a single measurement on a test item, often referred to as the reproducibility standard deviation (ASTM), is given by

$$s_R = \sqrt{s_{days}^2 + s_1^2} = \frac{s_2}{\sqrt{2}}$$

The time-dependent component is

$$s_{days} = \sqrt{\frac{1}{2}s_2^2 - s_1^2}$$

There may be other sources of uncertainty in the measurement process that must be accounted for in a formal analysis of uncertainty.

2.4.3.2. 2-level nested design

Check standard measurements for estimating time-dependent sources of variability

Measurements on a check standard are recommended for studying the effect of sources of variability that manifest themselves over time. Data collection and analysis are straightforward, and there is no reason to estimate interaction terms when dealing with time-dependent errors. The measurements can be made at one of two levels. Two levels should be sufficient for characterizing most measurement systems. Three levels are recommended for measurement systems for which sources of error are not well understood and have not previously been studied.

Time intervals in a nested design

The following levels are based on the characteristics of many measurement systems and should be adapted to a specific measurement situation as needed.

- Level-1 Measurements taken over a short term to estimate gauge precision
- Level-2 Measurements taken over days (of other appropriate time increment)

Definition of number of measurements at each level

The following symbols are defined for this chapter:

- Level-1 J $(J > 1)$ repetitions
- Level-2 K $(K > 2)$ days

Schedule for making measurements

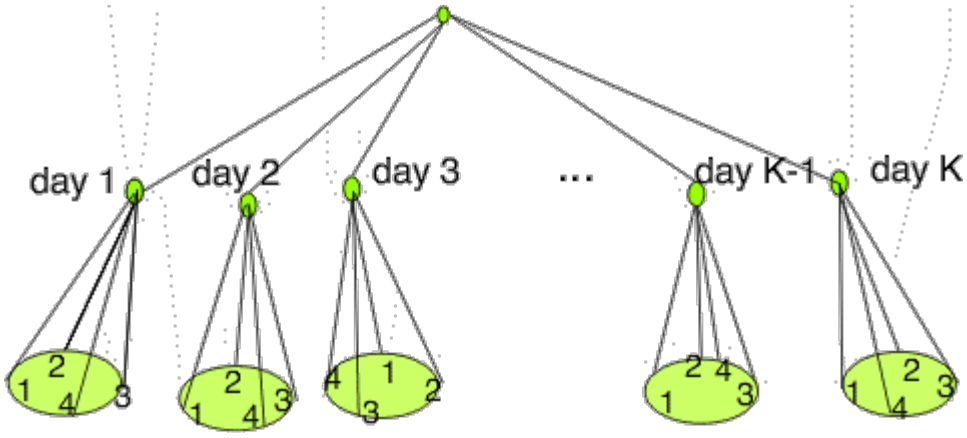
A schedule for making check standard measurements over time (once a day, twice a week, or whatever is appropriate for sampling all conditions of measurement) should be set up and adhered to. The check standard measurements should be structured in the same way as values reported on the test items. For example, if the reported values are averages of two repetitions made within 5 minutes of each other, the check standard values should be averages of the two measurements made in the same manner.

Exception

One exception to this rule is that there should be at least $J = 2$ repetitions per day, etc. Without this redundancy, there is no way to check on the short-term precision of the measurement system.

Depiction of schedule for making check standard

measurements with 4 repetitions per day over K days on the surface of a silicon wafer



K days - 4 repetitions

2-level design for check standard measurements

Operator considerations

The measurements should be taken with *ONE* operator. Operator is not usually a consideration with automated systems. However, systems that require decisions regarding line edge or other feature delineations may be operator dependent.

Case Study: Resistivity check standard

Results should be recorded along with pertinent environmental readings and identifications for significant factors. The best way to record this information is in one file with one line or row (on a spreadsheet) of information in fixed fields for each check standard measurement.

Data analysis of gauge precision

The check standard measurements are represented by

$$Y_{kj} \left(k = 1, \cdots K; j = 1, \cdots J\right)$$

for the *j*th repetition on the *k*th day. The mean for the *k*th day is

$$\bar{Y}_{k\bullet} = \frac{1}{J} \sum_{j=1}^{J} Y_{kj}$$

and the (level-1) standard deviation for gauge precision with $v = J - 1$ degrees of freedom is

$$s_k = \sqrt{\frac{1}{J-1} \sum_{j=1}^{J} \left(Y_{kj} - \bar{Y}_{k\bullet} \right)^2} .$$

Pooling increases the reliability of the estimate of the standard deviation

The pooled level-1 standard deviation with $v = K(J - 1)$ degrees of freedom is

$$s_1 = \sqrt{\frac{1}{K} \sum_{k=1}^{K} s_k^2} .$$

Data analysis of process

The level-2 standard deviation of the check standard represents the process variability. It is computed with $v = K - 1$ degrees of freedom as:

(level-2) standard deviation

$$s_{chkstd} = s_2 = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{Y}_{k\bullet} - \bar{Y}_{\bullet\bullet}\right)^2}$$

where

$$\bar{Y}_{\bullet\bullet} = \frac{1}{K}\sum_{k=1}^{K}\bar{Y}_{k\bullet}$$

Relationship to uncertainty for a test item

The standard deviation that defines the uncertainty for a single measurement on a test item, often referred to as the reproducibility standard deviation (ASTM), is given by

$$s_R = \sqrt{s_{day}^2 + s_1^2} = \sqrt{s_2^2 + \frac{J-1}{J}s_1^2}$$

The time-dependent component is

$$s_{day} = \sqrt{s_2^2 - \frac{1}{J}s_1^2}$$

There may be other sources of uncertainty in the measurement process that must be accounted for in a formal analysis of uncertainty.

2.4.3.3. 3-level nested design

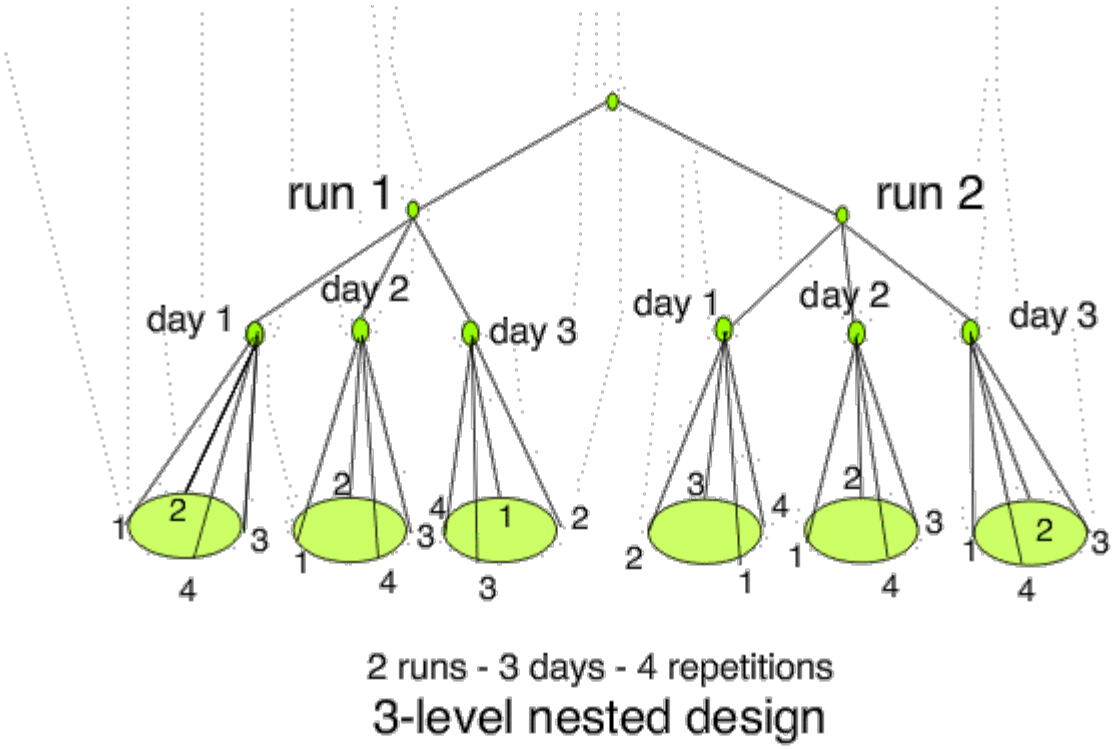
Advantages of nested designs

A nested design is recommended for studying the effect of sources of variability that manifest themselves over time. Data collection and analysis are straightforward, and there is no reason to estimate interaction terms when dealing with time-dependent errors. Nested designs can be run at several levels. Three levels are recommended for measurement systems where sources of error are not well understood and have not previously been studied.

Time intervals in a nested design

The following levels are based on the characteristics of many measurement systems and should be adapted to a specific measurement situation as need be. A typical design is shown below.

- Level-1 Measurements taken over a short-time to capture the precision of the gauge
- Level-2 Measurements taken over days (or other appropriate time increment)
- Level-3 Measurements taken over runs separated by months



Definition of number of measurements at each level

The following symbols are defined for this chapter:

- Level-1 *J (J > 1)* repetitions
- Level-2 *K (K > 2)* days
- Level-3 *L (L > 2)* runs

For the design shown above, *J = 4; K = 3 and L = 2*. The design can be repeated for:

- *Q (Q > 2)* check standards
- *I (I > 3)* gauges if the intent is to characterize several similar gauges

2-level nested design

The design can be truncated at two levels to estimate repeatability and day-to-day variability if there is no reason to estimate longer-term effects. The analysis remains the same through the first two levels.

Advantages

This design has advantages in ease of use and computation. The number of repetitions at each level need not be large because information is being gathered on several check standards.

Operator considerations

The measurements should be made with *ONE* operator. Operator is not usually a consideration with automated systems. However, systems that require decisions regarding

line edge or other feature delineations may be operator dependent. If there is reason to believe that results might differ significantly by operator, 'operators' can be substituted for 'runs' in the design. Choose *L (L > 2)* operators at random from the pool of operators who are capable of making measurements at the same level of precision. (Conduct a small experiment with operators making repeatability measurements, if necessary, to verify comparability of precision among operators.) Then complete the data collection and analysis as outlined. In this case, the level-3 standard deviation estimates operator effect.

Caution

Be sure that the design is truly nested; i.e., that each operator reports results for the same set of circumstances, particularly with regard to day of measurement so that each operator measures every day, or every other day, and so forth.

Randomize on gauges

Randomize with respect to gauges for each check standard; i.e., choose the first check standard and randomize the gauges; choose the second check standard and randomize gauges; and so forth.

Record results in a file

Record the average and standard deviation from each group of *J* repetitions by:

- check standard
- gauge

Case Study: Resistivity Gauges

Results should be recorded along with pertinent environmental readings and identifications for significant factors. The best way to record this information is in one file with one line or row (on a spreadsheet) of information in fixed fields for each check standard measurement. A list of typical entries follows.

1. Month
2. Day
3. Year
4. Operator identification
5. Check standard identification
6. Gauge identification
7. Average of J repetitions
8. Short-term standard deviation from *J* repetitions
9. Degrees of freedom
10. Environmental readings (if pertinent)

2.4.4. Analysis of variability

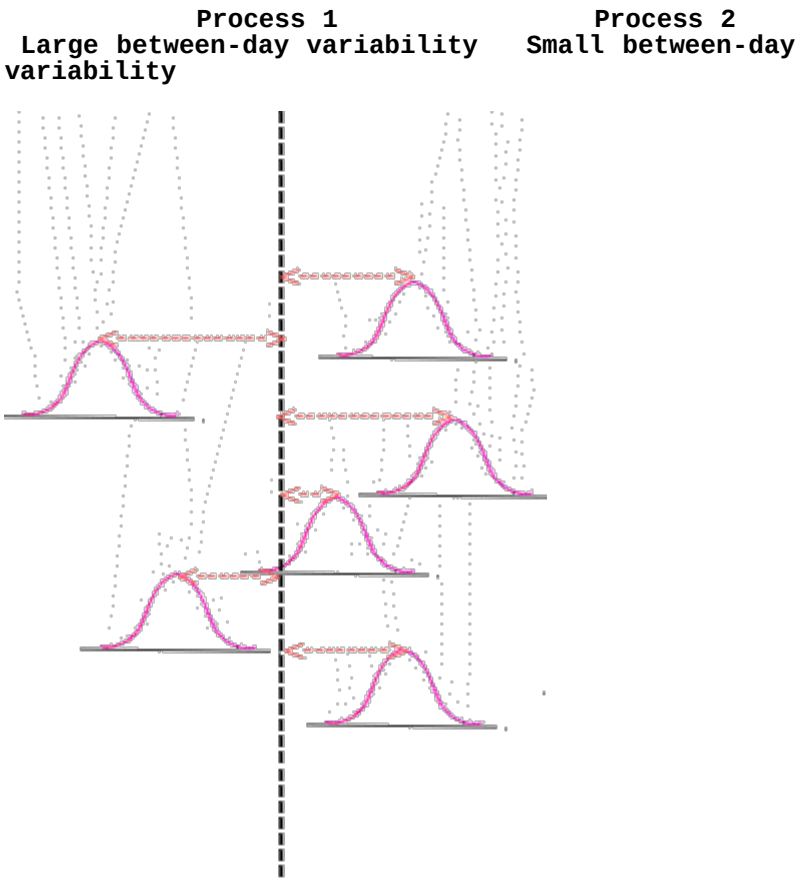
Analysis of variability from a nested design

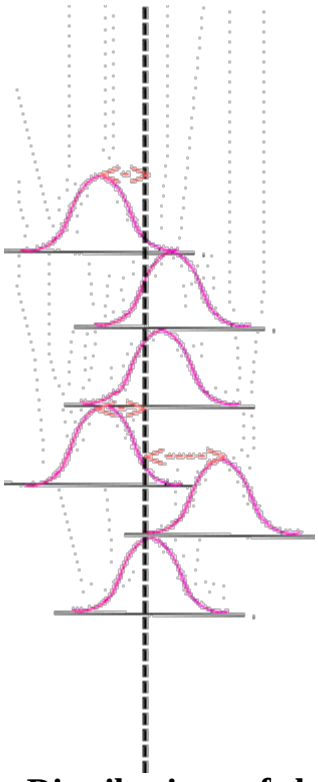
The purpose of this section is to show the effect of various levels of time-dependent effects on the variability of the measurement process with standard deviations for each level of a 3-level nested design.

- Level 1 - repeatability/short-term precision
- Level 2 - reproducibility/day-to-day
- Level 3 - stability/run-to-run

The graph below depicts possible scenarios for a 2-level design (short-term repetitions and days) to illustrate the concepts.

Depiction of 2 measurement processes with the same short-term variability over 6 days where process 1 has large between-day variability and process 2 has negligible between-day variability





Distributions of short-term measurements over 6 days where distances from centerlines illustrate between-day variability

Hint on using tabular method of analysis

An easy way to begin is with a 2-level table with *J* columns and *K* rows for the repeatability/reproducibility measurements and proceed as follows:

1. Compute an average for each row and put it in the *J+1* column.
2. Compute the level-1 (repeatability) standard deviation for each row and put it in the *J+2* column.
3. Compute the grand average and the level-2 standard deviation from data in the *J+1* column.
4. Repeat the table for each of the *L* runs.
5. Compute the level-3 standard deviation from the *L* grand averages.

Level-1: LK repeatability standard deviations can be computed from the data

The measurements from the nested design are denoted by

$$Y_{lkj}\,(l=1,\cdots,L;k=1,\cdots,K;j=1,\cdots,J)$$

Equations corresponding to the tabular analysis are shown below. Level-1 repeatability standard deviations, s_{1lk}, are pooled over the *K* days and *L* runs. Individual standard deviations with *(J - 1)* degrees of freedom each are computed from *J* repetitions as

$$s_{1lk} = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}(Y_{lkj} - \bar{Y}_{lk.})^2}$$

where

$$\bar{Y}_{lk\bullet} = \frac{1}{J}\sum_{j=1}^{J} Y_{lkj}$$

Level-2: L reproducibility standard deviations can be computed from the data

The level-2 standard deviation, s_{2l}, is pooled over the L runs. Individual standard deviations with *(K - 1)* degrees of freedom each are computed from K daily averages as

$$s_{2l} = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(Y_{lk.} - \bar{Y}_{l..})^2}$$

where

$$\bar{Y}_{l\bullet\bullet} = \frac{1}{K}\sum_{k=1}^{K}\bar{Y}_{lk\bullet}$$

Level-3: A single global standard deviation can be computed from the L-run averages

A level-3 standard deviation with *(L - 1)* degrees of freedom is computed from the L-run averages as

$$s_3 = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}\left(\bar{Y}_{l\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}\right)^2}$$

where

$$\bar{Y}_{..} = \frac{1}{L}\sum_{l=1}^{L}\bar{Y}_{l..}$$

Relationship to uncertainty for a test item

The standard deviation that defines the uncertainty for a single measurement on a test item is given by

$$s_R = \sqrt{s_{runs}^2 + s_{days}^2 + s_1^2} = \sqrt{s_3^2 + \frac{K-1}{K}s_2^2 + \frac{J-1}{J}s_1^2}$$

where the pooled values, s_1 and s_2, are the usual

$$s_1 = \sqrt{\frac{\sum_{l=1}^{L}\sum_{k=1}^{K} s_{1lk}^2}{LK}}$$

and

$$s_2 = \sqrt{\frac{1}{L}\sum_{l=1}^{L} s_{2l}^2}$$

There may be other sources of uncertainty in the measurement process that must be accounted for in a formal analysis of uncertainty.

2.4.4.1. Analysis of repeatability

Case study: Resistivity probes

The repeatability quantifies the basic precision for the gauge. A level-1 repeatability standard deviation is computed for each group of J repetitions, and a graphical analysis is recommended for deciding if repeatability is dependent on the check standard, the operator, or the gauge. Two graphs are recommended. These should show:

- Plot of repeatability standard deviations versus check standard with day coded
- Plot of repeatability standard deviations versus check standard with gauge coded

Typically, we expect the standard deviation to be gauge dependent -- in which case there should be a separate standard deviation for each gauge. If the gauges are all at the same level of precision, the values can be combined over all gauges.

Repeatability standard deviations can be pooled over operators, runs, and check standards

A repeatability standard deviation from J repetitions is not a reliable estimate of the precision of the gauge. Fortunately, these standard deviations can be pooled over days; runs; and check standards, if appropriate, to produce a more reliable precision measure. The table below shows a mechanism for pooling. The pooled repeatability standard deviation, s_1, has $LK(J - 1)$ degrees of freedom for measurements taken over:

- J repetitions
- K days
- L runs

Basic pooling rules

The table below gives the mechanism for pooling repeatability standard deviations over days and runs. The pooled value is an average of weighted variances and is shown as the last entry in the right-hand column of the table. The pooling can also cover check standards, if appropriate.

View of entire dataset from the nested design

To illustrate the calculations, a subset of data collected in a nested design for one check standard (#140) and one probe (#2362) are shown below. The measurements are resistivity (ohm.cm) readings with six repetitions per day. The individual level-1 standard deviations from the six repetitions and degrees of freedom are recorded in the last two columns of the database.

Run	Wafer	Probe	Month	Day	Op	Temp	Average	Stddev	df
1	140	2362	3	15	1	23.08	96.0771	0.1024	5
1	140	2362	3	17	1	23.00	95.9976	0.0943	5
1	140	2362	3	18	1	23.01	96.0148	0.0622	5
1	140	2362	3	22	1	23.27	96.0397	0.0702	5
1	140	2362	3	23	2	23.24	96.0407	0.0627	5
1	140	2362	3	24	2	23.13	96.0445	0.0622	5
2	140	2362	4	12	1	22.88	96.0793	0.0996	5
2	140	2362	4	18	2	22.76	96.1115	0.0533	5
2	140	2362	4	19	2	22.79	96.0803	0.0364	5
2	140	2362	4	19	1	22.71	96.0411	0.0768	5
2	140	2362	4	20	2	22.84	96.0988	0.1042	5
2	140	2362	4	21	1	22.94	96.0482	0.0868	5

Pooled repeatability standard deviations over days, runs

Source of Variability	Degrees of Freedom	Standard Deviations	Sum of Squares (SS)
Probe 2362	ν_i	s_{1i}	$SS_i = \nu_i \cdot s_{1i}^2$
run 1 - day 1	5	0.1024	0.05243
run 1 - day 2	5	0.0943	0.04446
run 1 - day 3	5	0.0622	0.01934
run 1 - day 4	5	0.0702	0.02464
run 1 - day 5	5	0.0627	0.01966
run 1 - day 6	5	0.0622	0.01934
run 2 - day 1	5	0.0996	0.04960
run 2 - day 2	5	0.0533	0.01420
run 2 - day 3	5	0.0364	0.00662
run 2 - day 4	5	0.0768	0.02949
run 2 - day 5	5	0.1042	0.05429
run 2 - day 6	5	0.0868	0.03767
$\nu = \sum \nu_i \rightarrow$ gives the total degrees of freedom for s_1	**60**	$SS = \sum SS_i \rightarrow$ gives the total sum of squares for s_1	**0.37176**
The pooled value of s_1 is given by			**0.07871**

$s_1 = \sqrt{SS/\nu} \rightarrow$	

The calculations displayed in the table above can be generated using both Dataplot code and R code.

2.4.4.2. Analysis of reproducibility

Case study: Resistivity gauges

Day-to-day variability can be assessed by a graph of check standard values (averaged over *J* repetitions) versus day with a separate graph for each check standard. Graphs for all check standards should be plotted on the same page to obtain an overall view of the measurement situation.

Pooling results in more reliable estimates

The level-2 standard deviations with (*K* - 1) degrees of freedom are computed from the check standard values for days and pooled over runs as shown in the table below. The pooled level-2 standard deviation has degrees of freedom
L(*K* - 1) for measurements made over:

- *K* days
- *L* runs

Mechanism for pooling

The table below gives the mechanism for pooling level-2 standard deviations over runs. The pooled value is an average of weighted variances and is the last entry in the right-hand column of the table. The pooling can be extended in the same manner to cover check standards, if appropriate.

The table was generated using a subset of data (shown on previous page) collected in a nested design on one check standard (#140) with probe (#2362) over six days. The data are analyzed for between-day effects. The level-2 standard deviations and pooled level-2 standard deviations over runs 1 and 2 are:

Level-2 standard deviations for a single gauge pooled over runs

Source of variability	Standard deviations	Degrees of freedom	Sum of squares
Days	s_{2i}	v_i	$SS_i = v_i \cdot {s_{2i}}^2$
Run 1 Run 2	0.027280 0.027560	5 5 ------ -	0.003721 0.003798 ----------

Sum		10	0.007519
Pooled value $s_2 = \sqrt{\sum SS_i / \sum v_i}$			0.02742

Relationship to day effect

The level-2 standard deviation is related to the standard deviation for between-day precision and gauge precision by

$$s_{days} = \sqrt{s_2^2 - \frac{1}{j} s_1^2}$$

The size of the day effect can be calculated by subtraction using the formula above once the other two standard deviations have been estimated reliably.

Computation of variance component for days

For our example, the variance component for between days is -0.00028072. The negative number for the variance is interpreted as meaning that the variance component for days is zero. However, with only 10 degrees of freedom for the level-2 standard deviation, this estimate is not necessarily reliable. The standard deviation for days over the entire database shows a significant component for days.

Sample code

The calculations included in this section can be implemented using both
Dataplot code and R code.

2.4.4.3. Analysis of stability

Case study: Resistivity probes

Run-to-run variability can be assessed graphically by a plot of check standard values (averaged over J repetitions) versus time with a separate graph for each check standard. Data on all check standards should be plotted on one page to obtain an overall view of the measurement situation.

Advantage of pooling

A level-3 standard deviation with (L - 1) degrees of freedom is computed from the run averages. Because there will rarely be more than two runs per check standard, resulting in one degree of freedom per check standard, it is prudent to have three or more check standards in the design to take advantage of pooling. The mechanism for pooling over check standards is shown in the table below. The pooled standard deviation has $Q(L - 1)$ degrees and is shown as the last entry in the right-hand column of the table.

Example of pooling

The following table shows how the level-3 standard deviations for a single gauge (probe #2362) are pooled over check standards. The table can be reproduced using R code.

Level-3 standard deviations for a single gauge pooled over check standards

Source of variability	Standard deviation	Degrees of freedom	Sum of squares
Level-3	s_{3i}	v_i	$SS_i = v_i \cdot {s_{3i}}^2$
Chk std 138	0.0223	1	0.0004973
Chk std 139	0.0027	1	0.0000073
Chk std 140	0.0289	1	0.0008352
Chk std 141	0.0133	1	0.0001769
Chk std 142	0.0205	1	0.0004203
Sum		5	0.0019370
Pooled value $s_3 = \sqrt{\sum SS_i / \sum v_i}$			**0.0197**

Level-3 standard deviations

A subset of data collected in a nested design on one check standard (#140) with probe (#2362) for six days and two runs is analyzed for between-run effects. The level-3 standard deviation, computed from the averages of two runs, is 0.02885 with one degree of freedom. Dataplot code and R code can be used to perform the calculations for this data.

Relationship to long-term changes, days and gauge precision

The size of the between-run effect can be calculated by subtraction using the standard deviations for days and gauge precision as

$$s_{runs} = \sqrt{s_3^2 - \frac{1}{K}s_2^2} = \sqrt{s_3^2 - \frac{1}{K}s_{days}^2 - \frac{1}{KJ}s_1^2}$$

2.4.4.4.4. Example of calculations

Example of repeatability calculations

Short-term standard deviations based on

- $J = 6$ repetitions with 5 degrees of freedom
- $K = 6$ days
- $L = 2$ runs

were recorded with a probing instrument on $Q = 5$ wafers. The standard deviations were pooled over $K = 6$ days and $L = 2$ runs to give 60 degrees of freedom for each wafer. The pooling of repeatability standard deviations over the 5 wafers is demonstrated in the table below.

Pooled repeatability standard deviation for a single gauge

Source of variability	Sum of Squares (SS)	Degrees of freedom (DF)	Std Devs
Repeatability	$SS_i = \nu_i \cdot s_{Ii}^2$	ν_i	$s_I = \sqrt{SS/\nu} \rightarrow$
Wafer #138	0.48115	60	
Wafer #139	0.69209	60	
Wafer #140	0.48483	60	
Wafer #141	1.21752	60	
Wafer #142	0.30076	60	
SUM	**3.17635**	**300**	**0.10290**

2.4.5. Analysis of bias

Definition of bias

The terms 'bias' and 'systematic error' have the same meaning in this handbook. Bias is defined (VIM) as the difference between the measurement result and its unknown 'true value'. It can often be estimated and/or eliminated by calibration to a reference standard.

Potential problem

Calibration relates output to 'true value' in an ideal environment. However, it may not assure that the gauge reacts properly in its working environment. Temperature, humidity, operator, wear, and other factors can introduce bias into the measurements. There is no single method for dealing with this problem, but the gauge study is intended to uncover biases in the measurement process.

Sources of bias

Sources of bias that are discussed in this Handbook include:

- Lack of gauge resolution
- Lack of linearity
- Drift
- Hysteresis
- Differences among gauges
- Differences among geometries
- Differences among operators
- Remedial actions and strategies

2.4.5.1. Resolution

Resolution	Resolution (MSA) is the ability of the measurement system to detect and faithfully indicate small changes in the characteristic of the measurement result.
Definition from (MSA) manual	The resolution of the instrument is δ if there is an equal probability that the indicated value of any artifact, which differs from a reference standard by less than δ, will be the same as the indicated value of the reference.
Good versus poor	A small δ implies good resolution -- the measurement system can discriminate between artifacts that are close together in value. A large δ implies poor resolution -- the measurement system can only discriminate between artifacts that are far apart in value.
Warning	The number of digits displayed does not indicate the resolution of the instrument.
Manufacturer's statement of resolution	Resolution as stated in the manufacturer's specifications is usually a function of the least-significant digit (LSD) of the instrument and other factors such as timing mechanisms. This value should be checked in the laboratory under actual conditions of measurement.
Experimental determination of resolution	To make a determination in the laboratory, select several artifacts with known values over a range from close in value to far apart. Start with the two artifacts that are farthest apart and make measurements on each artifact. Then, measure the two artifacts with the second largest difference, and so forth, until two artifacts are found which repeatedly give the same result. The difference between the values of these two artifacts estimates the resolution.
Consequence of poor resolution	No useful information can be gained from a study on a gauge with poor resolution relative to measurement needs.

2.4.5.2. Linearity of the gauge

Definition of linearity for gauge studies

Linearity is given a narrow interpretation in this Handbook to indicate that gauge response increases in equal increments to equal increments of stimulus, or, if the gauge is biased, that the bias remains constant throughout the course of the measurement process.

Data collection and repetitions

A determination of linearity requires Q $(Q > 4)$ reference standards that cover the range of interest in fairly equal increments and J $(J > 1)$ measurements on each reference standard. One measurement is made on each of the reference standards, and the process is repeated J times.

Plot of the data

A test of linearity starts with a plot of the measured values versus corresponding values of the reference standards to obtain an indication of whether or not the points fall on a straight line with slope equal to 1 -- indicating linearity.

Least-squares estimates of bias and slope

A least-squares fit of the data to the model

$$Y = a + bX + \text{measurement error}$$

where Y is the measurement result and X is the value of the reference standard, produces an estimate of the intercept, a, and the slope, b.

Output from software package

The intercept and bias are estimated using a statistical software package that should provide the following information:

- Estimates of the intercept and slope, $\hat{a}, \hat{b}$
- Standard deviations of the intercept and slope
- Residual standard deviation of the fit
- F-test for goodness of fit

Test for linearity

Tests for the slope and bias are described in the section on instrument calibration. If the slope is different from one, the gauge is non-linear and requires calibration or repair. If the intercept is different from zero, the gauge has a bias.

Causes of non-linearity

The reference manual on Measurement Systems Analysis (MSA) lists possible causes of gauge non-linearity that should be investigated if the gauge shows symptoms of non-linearity.

1. Gauge not properly calibrated at the lower and upper ends of the operating range
2. Error in the value of X at the maximum or minimum range
3. Worn gauge
4. Internal design problems (electronics)

Note - on artifact calibration

The requirement of linearity for artifact calibration is not so stringent. Where the gauge is used as a comparator for measuring small differences among test items and reference standards of the same nominal size, as with calibration designs, the only requirement is that the gauge be linear over the small on-scale range needed to measure both the reference standard and the test item.

Situation where the calibration of the gauge is neglected

Sometimes it is not economically feasible to correct for the calibration of the gauge (Turgel and Vecchia). In this case, the bias that is incurred by neglecting the calibration is estimated as a component of uncertainty.

2.4.5.3. Drift

Definition	Drift can be defined (VIM) as a slow change in the response of a gauge.
Instruments used as comparators for calibration	Short-term drift can be a problem for comparator measurements. The cause is frequently heat build-up in the instrument during the time of measurement. It would be difficult, and probably unproductive, to try to pinpoint the extent of such drift with a gauge study. The simplest solution is to use drift-free designs for collecting calibration data. These designs mitigate the effect of linear drift on the results. Long-term drift should not be a problem for comparator measurements because such drift would be constant during a calibration design and would cancel in the difference measurements.
Instruments corrected by linear calibration	For instruments whose readings are corrected by a linear calibration line, drift can be detected using a control chart technique and measurements on three or more check standards.
Drift in direct reading instruments and uncertainty analysis	For other instruments, measurements can be made on a daily basis on two or more check standards over a preset time period, say, one month. These measurements are plotted on a time scale to determine the extent and nature of any drift. Drift rarely continues unabated at the same rate and in the same direction for a long time period. Thus, the expectation from such an experiment is to document the maximum change that is likely to occur during a set time period and plan adjustments to the instrument accordingly. A further impact of the findings is that uncorrected drift is treated as a type A component in the uncertainty analysis.

2.4.5.4. Differences among gauges

Purpose

A gauge study should address whether gauges agree with one another and whether the agreement (or disagreement) is consistent over artifacts and time.

Data collection

For each gauge in the study, the analysis requires measurements on

- *Q (Q > 2)* check standards
- *K (K > 2)* days

The measurements should be made by a single operator.

Data reduction

The steps in the analysis are:

1. Measurements are averaged over days by artifact/gauge configuration.
2. For each artifact, an average is computed over gauges.
3. Differences from this average are then computed for each gauge.
4. If the design is run as a 3-level design, the statistics are computed separately for each run.

Data from a gauge study

The data in the table below come from resistivity (ohm.cm) measurements on *Q = 5* artifacts on *K = 6* days. Two runs were made which were separated by about a month's time. The artifacts are silicon wafers and the gauges are four-point probes specifically designed for measuring resistivity of silicon wafers. Differences from the wafer means are shown in the table.

Biases for 5 probes from a gauge study with 5 artifacts on 6 days

Table of biases for probes and silicon wafers (ohm.cm)

	Wafers				
Probe	138	139	140	141	142
1	0.02476	-0.00356	0.04002	0.03938	0.00620
181	0.01076	0.03944	0.01871	-0.01072	0.03761
182	0.01926	0.00574	-0.02008	0.02458	

```
-0.00439

  2062   -0.01754   -0.03226   -0.01258   -0.02802
-0.00110

  2362   -0.03725   -0.00936   -0.02608   -0.02522
-0.03830
```

Plot of differences among probes

A graphical analysis can be more effective for detecting differences among gauges than a table of differences. The differences are plotted versus artifact identification with each gauge identified by a separate plotting symbol. For ease of interpretation, the symbols for any one gauge can be connected by dotted lines.

Interpretation

Because the plots show differences from the average by artifact, the center line is the zero-line, and the differences are estimates of bias. Gauges that are *consistently* above or below the other gauges are biased high or low, respectively, relative to the average. The best estimate of bias for a particular gauge is its average bias over the *Q* artifacts. For this data set, notice that probe #2362 is consistently biased low relative to the other probes.

Strategies for dealing with differences among gauges

Given that the gauges are a random sample of like-kind gauges, the best estimate in any situation is an average over all gauges. In the usual production or metrology setting, however, it may only be feasible to make the measurements on a particular piece with one gauge. Then, there are two methods of dealing with the differences among gauges.

1. Correct each measurement made with a particular gauge for the bias of that gauge and report the standard deviation of the correction as a type A uncertainty.
2. Report each measurement as it occurs and assess a type A uncertainty for the differences among the gauges.

2.4.5.5. Geometry/configuration differences

How to deal with configuration differences

The mechanism for identifying and/or dealing with differences among geometries or configurations in an instrument is basically the same as dealing with differences among the gauges themselves.

Example of differences among wiring configurations

An example is given of a study of configuration differences for a single gauge. The gauge, a 4-point probe for measuring resistivity of silicon wafers, can be wired in several ways. Because it was not possible to test all wiring configurations during the gauge study, measurements were made in only two configurations as a way of identifying possible problems.

Data on wiring configurations and a plot of differences between the 2 wiring configurations

Measurements were made on six wafers over six days (except for 5 measurements on wafer 39) with probe #2062 wired in two configurations. This sequence of measurements was repeated after about a month resulting in two runs. Differences between measurements in the two configurations on the same day are shown in the following table.

Differences between wiring configurations

Wafer	Day	Probe	Run 1	Run 2
17.	1	2062.	-0.0108	0.0088
17.	2	2062.	-0.0111	0.0062
17.	3	2062.	-0.0062	0.0074
17.	4	2062.	0.0020	0.0047
17.	5	2062.	0.0018	0.0049
17.	6	2062.	0.0002	0.0000
39.	1	2062.	-0.0089	0.0075
39.	3	2062.	-0.0040	-0.0016
39.	4	2062.	-0.0022	0.0052

```
   39.    5       2062.    -0.0012
0.0085
   39.    6       2062.    -0.0034    -
0.0018

   63.    1       2062.    -0.0016
0.0092
   63.    2       2062.    -0.0111
0.0040
   63.    3       2062.    -0.0059
0.0067
   63.    4       2062.    -0.0078
0.0016
   63.    5       2062.    -0.0007
0.0020
   63.    6       2062.     0.0006
0.0017

  103.    1       2062.    -0.0050
0.0076
  103.    2       2062.    -0.0140
0.0002
  103.    3       2062.    -0.0048
0.0025
  103.    4       2062.     0.0018
0.0045
  103.    5       2062.     0.0016    -
0.0025
  103.    6       2062.     0.0044
0.0035

  125.    1       2062.    -0.0056
0.0099
  125.    2       2062.    -0.0155
0.0123
  125.    3       2062.    -0.0010
0.0042
  125.    4       2062.    -0.0014
0.0098
  125.    5       2062.     0.0003
0.0032
  125.    6       2062.    -0.0017
0.0115
```

Test of difference between configurations

Because there are only two configurations, a *t*-test is used to decide if there is a difference. If

$$t = \left| \frac{\sqrt{N}}{s_{diff}} Avg_{diff} \right| > 2$$

the difference between the two configurations is statistically significant.

The average and standard deviation computed from the 29 differences in each run are shown in the table below along with the *t*-values which confirm that the differences are significant for both runs.

Average differences between wiring configurations

Run	Probe		Average	Std dev	N
	t				
1	2062	-	0.00383	0.00514	29
	-4.0				
2	2062	+	0.00489	0.00400	29
	+6.6				

Unexpected result

The data reveal a wiring bias for both runs that changes direction between runs. This is a somewhat disturbing finding, and further study of the gauges is needed. Because neither wiring configuration is preferred or known to give the 'correct' result, the differences are treated as a component of the measurement uncertainty.

2.4.5.6. Remedial actions and strategies

Variability	The variability of the gauge in its normal operating mode needs to be examined in light of measurement requirements. If the standard deviation is too large, relative to requirements, the uncertainty can be reduced by making repeated measurements and taking advantage of the standard deviation of the average (which is reduced by a factor of $1/\sqrt{n}$ when *n* measurements are averaged).
Causes of excess variability	If multiple measurements are not economically feasible in the workload, then the performance of the gauge must be improved. Causes of variability which should be examined are: • Wear • Environmental effects such as humidity • Temperature excursions • Operator technique
Resolution	There is no remedy for a gauge with insufficient resolution. The gauge will need to be replaced with a better gauge.
Lack of linearity	Lack of linearity can be dealt with by correcting the output of the gauge to account for bias that is dependent on the level of the stimulus. Lack of linearity can be tolerated (left uncorrected) if it does not increase the uncertainty of the measurement result beyond its requirement.
Drift	It would be very difficult to correct a gauge for drift unless there is sufficient history to document the direction and size of the drift. Drift can be tolerated if it does not increase the uncertainty of the measurement result beyond its requirement.
Differences among gauges or configurations	Significant differences among gauges/configurations can be treated in one of two ways: 1. By correcting each measurement for the bias of the specific gauge/configuration.

2. By accepting the difference as part of the uncertainty of the measurement process.

Differences among operators

Differences among operators can be viewed in the same way as differences among gauges. However, an operator who is incapable of making measurements to the required precision because of an untreatable condition, such as a vision problem, should be re-assigned to other tasks.

2.4.6. Quantifying uncertainties from a gauge study

Gauge studies can be used as the basis for uncertainty assessment

One reason for conducting a gauge study is to quantify uncertainties in the measurement process that would be difficult to quantify under conditions of actual measurement.

This is a reasonable approach to take if the results are truly representative of the measurement process in its working environment. Consideration should be given to all sources of error, particularly those sources of error which do not exhibit themselves in the short-term run.

Potential problem with this approach

The potential problem with this approach is that the calculation of uncertainty depends totally on the gauge study. If the measurement process changes its characteristics over time, the standard deviation from the gauge study will not be the correct standard deviation for the uncertainty analysis. One way to try to avoid such a problem is to carry out a gauge study both before and after the measurements that are being characterized for uncertainty. The 'before' and 'after' results should indicate whether or not the measurement process changed in the interim.

Uncertainty analysis requires information about the specific measurement

The computation of uncertainty depends on the particular measurement that is of interest. The gauge study gathers the data and estimates standard deviations for sources that contribute to the uncertainty of the measurement result. However, specific formulas are needed to relate these standard deviations to the standard deviation of a measurement result.

General guidance

The following sections outline the general approach to uncertainty analysis and give methods for combining the standard deviations into a final uncertainty:

1. Approach
2. Methods for type A evaluations
3. Methods for type B evaluations
4. Propagation of error
5. Error budgets and sensitivity coefficients
6. Standard and expanded uncertainties
7. Treatment of uncorrected biases

Type A evaluations of random error

Data collection methods and analyses of random sources of uncertainty are given for the following:

1. Repeatability of the gauge
2. Reproducibility of the measurement process
3. Stability (very long-term) of the measurement process

Biases - Rule of thumb

The approach for biases is to estimate the maximum bias from a gauge study and compute a standard uncertainty from the maximum bias assuming a suitable distribution. The formulas shown below assume a uniform distribution for each bias.

Determining resolution

If the resolution of the gauge is δ, the standard uncertainty for resolution is

$$s_{resolution} = \delta / \sqrt{3}$$

Determining non-linearity

If the maximum departure from linearity for the gauge has been determined from a gauge study, and it is reasonable to assume that the gauge is equally likely to be engaged at any point within the range tested, the standard uncertainty for linearity is

$$s_{linearity} = Max\left|Y_{observed} - Y_{fitted}\right| / \sqrt{3}$$

Hysteresis

Hysteresis, as a performance specification, is defined (NCSL RP-12) as the maximum difference between the upscale and downscale readings on the same artifact during a full range traverse in each direction. The standard uncertainty for hysteresis is

$$s_{hysteresis} = Max\left|Y_{upscale} - Y_{downscale}\right| / \sqrt{3}$$

Determining drift

Drift in direct reading instruments is defined for a specific time interval of interest. The standard uncertainty for drift is

$$s_{drift} = Max\left|Y_0 - Y_t\right| / \sqrt{3}$$

where Y_0 and Y_t are measurements at time zero and t, respectively.

Other biases

Other sources of bias are discussed as follows:

1. Differences among gauges
2. Differences among configurations

Case study:

A case study on type A uncertainty analysis from a gauge

Type A uncertainties from a gauge study	study is recommended as a guide for bringing together the principles and elements discussed in this section. The study in question characterizes the uncertainty of resistivity measurements made on silicon wafers.

NIST SEMATECH HOME TOOLS & AIDS SEARCH BACK NEXT

2.5. Uncertainty analysis

Uncertainty measures 'goodness' of a test result

This section discusses the uncertainty of measurement results. Uncertainty is a measure of the 'goodness' of a result. Without such a measure, it is impossible to judge the fitness of the value as a basis for making decisions relating to health, safety, commerce or scientific excellence.

Contents

1. What are the issues for uncertainty analysis?
2. Approach to uncertainty analysis
    1. Steps
3. Type A evaluations
    1. Type A evaluations of random error
        1. Time-dependent components
        2. Measurement configurations
    2. Type A evaluations of material inhomogeneities
        1. Data collection and analysis
    3. Type A evaluations of bias
        1. Treatment of inconsistent bias
        2. Treatment of consistent bias
        3. Treatment of bias with sparse data
4. Type B evaluations
    1. Assumed distributions
5. Propagation of error considerations
    1. Functions of a single variable
    2. Functions of two variables
    3. Functions of several variables
6. Error budgets and sensitivity coefficients
    1. Sensitivity coefficients for measurements on the test item
    2. Sensitivity coefficients for measurements on a check standard
    3. Sensitivity coefficients for measurements with a 2-level design
    4. Sensitivity coefficients for measurements with a 3-level design
    5. Example of error budget
7. Standard and expanded uncertainties
    1. Degrees of freedom
8. Treatment of uncorrected bias

1. Computation of revised uncertainty

2.5.1. Issues

Issues for uncertainty analysis

Evaluation of uncertainty is an ongoing process that can consume time and resources. It can also require the services of someone who is familiar with data analysis techniques, particularly statistical analysis. Therefore, it is important for laboratory personnel who are approaching uncertainty analysis for the first time to be aware of the resources required and to carefully lay out a plan for data collection and analysis.

Problem areas

Some laboratories, such as test laboratories, may not have the resources to undertake detailed uncertainty analyses even though, increasingly, quality management standards such as the ISO 9000 series are requiring that all measurement results be accompanied by statements of uncertainty.

Other situations where uncertainty analyses are problematical are:

- One-of-a-kind measurements
- Dynamic measurements that depend strongly on the application for the measurement

Directions being pursued

What can be done in these situations? There is no definitive answer at this time. Several organizations, such as the National Conference of Standards Laboratories (NCSL) and the International Standards Organization (ISO) are investigating methods for dealing with this problem, and there is a document in draft that will recommend a simplified approach to uncertainty analysis based on results of interlaboratory tests.

Relationship to interlaboratory test results

Many laboratories or industries participate in interlaboratory studies where the test method itself is evaluated for:

- repeatability within laboratories
- reproducibility across laboratories

These evaluations do not lead to uncertainty statements because the purpose of the interlaboratory test is to evaluate, and then improve, the test method as it is

applied across the industry. The purpose of uncertainty analysis is to evaluate the result of a particular measurement, in a particular laboratory, at a particular time. However, the two purposes are related.

Default recommendation for test laboratories

If a test laboratory has been party to an interlaboratory test that follows the recommendations and analyses of an American Society for Testing Materials standard (ASTM E691) or an ISO standard (ISO 5725), the laboratory can, as a default, represent its standard uncertainty for a single measurement as the reproducibility standard deviation as defined in ASTM E691 and ISO 5725. This standard deviation includes components for within-laboratory repeatability common to all laboratories and between-laboratory variation.

Drawbacks of this procedure

The standard deviation computed in this manner describes a future single measurement made at a laboratory randomly drawn from the group and leads to a prediction interval (Hahn & Meeker) rather than a confidence interval. It is not an ideal solution and may produce either an unrealistically small or unacceptably large uncertainty for a particular laboratory. The procedure can reward laboratories with poor performance or those that do not follow the test procedures to the letter and punish laboratories with good performance. Further, the procedure does not take into account sources of uncertainty other than those captured in the interlaboratory test. Because the interlaboratory test is a snapshot at one point in time, characteristics of the measurement process over time cannot be accurately evaluated. Therefore, it is a strategy to be used only where there is no possibility of conducting a realistic uncertainty investigation.

2.5.2. Approach

Procedures in this chapter

The procedures in this chapter are intended for test laboratories, calibration laboratories, and scientific laboratories that report results of measurements from ongoing or well-documented processes.

Pertinent sections

The following pages outline methods for estimating the individual uncertainty components, which are consistent with materials presented in other sections of this Handbook, and rules and equations for combining them into a final expanded uncertainty. The general framework is:

1. ISO Approach
2. Outline of steps to uncertainty analysis
3. Methods for type A evaluations
4. Methods for type B evaluations
5. Propagation of error considerations
6. Uncertainty budgets and sensitivity coefficients
7. Standard and expanded uncertainties
8. Treatment of uncorrected bias

Specific situations are outlined in other places in this chapter

Methods for calculating uncertainties for specific results are explained in the following sections:

- Calibrated values of artifacts
- Calibrated values from calibration curves
  - From propagation of error
  - From check standard measurements
  - Comparison of check standards and propagation of error
- Gauge R & R studies
- Type A components for resistivity measurements
- Type B components for resistivity measurements

ISO definition of uncertainty

Uncertainty, as defined in the ISO Guide to the Expression of Uncertainty in Measurement (GUM) and the International Vocabulary of Basic and General Terms in Metrology (VIM), is a

> "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be

attributed to the measurand."

Consistent with historical view of uncertainty

This definition is consistent with the well-established concept that an uncertainty statement assigns credible limits to the accuracy of a reported value, stating to what extent that value may differ from its reference value (Eisenhart). In some cases, reference values will be traceable to a national standard, and in certain other cases, reference values will be consensus values based on measurements made according to a specific protocol by a group of laboratories.

Accounts for both random error and bias

The estimation of a possible discrepancy takes into account both random error and bias in the measurement process. The distinction to keep in mind with regard to random error and bias is that random errors cannot be corrected, and biases can, theoretically at least, be corrected or eliminated from the measurement result.

Relationship to precision and bias statements

Precision and bias are properties of a measurement method. Uncertainty is a property of a *specific result* for a single test item that depends on a specific measurement configuration (laboratory/instrument/operator, etc.). It depends on the repeatability of the instrument; the reproducibility of the result over time; the number of measurements in the test result; and all sources of random and systematic error that could contribute to disagreement between the result and its reference value.

Handbook follows the ISO approach

This Handbook follows the ISO approach (GUM) to stating and combining components of uncertainty. To this basic structure, it adds a statistical framework for estimating individual components, particularly those that are classified as type A uncertainties.

Basic ISO tenets

The ISO approach is based on the following rules:

- Each uncertainty component is quantified by a standard deviation.
- All biases are assumed to be corrected and any uncertainty is the uncertainty of the correction.
- Zero corrections are allowed if the bias cannot be corrected and an uncertainty is assessed.
- All uncertainty intervals are symmetric.

ISO approach to classifying sources of error

Components are grouped into two major categories, depending on the source of the data and not on the type of error, and each component is quantified by a standard deviation. The categories are:

- Type A - components evaluated by statistical

methods
- Type B - components evaluated by other means (or in other laboratories)

Interpretation of this classification

One way of interpreting this classification is that it distinguishes between information that comes from sources local to the measurement process and information from other sources -- although this interpretation does not always hold. In the computation of the final uncertainty it makes no difference how the components are classified because the ISO guidelines treat type A and type B evaluations in the same manner.

Rule of quadrature

All uncertainty components (standard deviations) are combined by root-sum-squares (quadrature) to arrive at a 'standard uncertainty', u, which is the standard deviation of the reported value, taking into account all sources of error, both random and systematic, that affect the measurement result.

Expanded uncertainty for a high degree of confidence

If the purpose of the uncertainty statement is to provide coverage with a high level of confidence, an expanded uncertainty is computed as

$$U = k\,u$$

where k is chosen to be the $t_{1-\alpha/2,\nu}$ critical value from the *t*-table with ν degrees of freedom.

For large degrees of freedom, it is suggested to use $k = 2$ to approximate 95% coverage. Details for these calculations are found under degrees of freedom.

Type B evaluations

Type B evaluations apply to random errors and biases for which there is little or no data from the local process, and to random errors and biases from other measurement processes.

2.5.2.1. Steps

Steps in uncertainty analysis - define the result to be reported

The first step in the uncertainty evaluation is the definition of the result to be reported for the test item for which an uncertainty is required. The computation of the standard deviation depends on the number of repetitions on the test item and the range of environmental and operational conditions over which the repetitions were made, in addition to other sources of error, such as calibration uncertainties for reference standards, which influence the final result. If the value for the test item cannot be measured directly, but must be calculated from measurements on secondary quantities, the equation for combining the various quantities must be defined. The steps to be followed in an uncertainty analysis are outlined for two situations:

Outline of steps to be followed in the evaluation of uncertainty for a single quantity

A. **Reported value involves measurements on one quantity.**

1. Compute a type A standard deviation for random sources of error from:
   - Replicated results for the test item.
   - Measurements on a check standard.
   - Measurements made according to a 2-level designed experiment
   - Measurements made according to a 3-level designed experiment
2. Make sure that the collected data and analysis cover all sources of random error such as:
   - instrument imprecision
   - day-to-day variation
   - long-term variation

   and bias such as:
   - differences among instruments
   - operator differences.
3. Compute a standard deviation for each type B component of uncertainty.
4. Combine type A and type B standard deviations into a

standard uncertainty for the reported result using sensitivity factors.

5. Compute an expanded uncertainty.

Outline of steps to be followed in the evaluation of uncertainty involving several secondary quantities

B. - **Reported value involves more than one quantity.**

1. Write down the equation showing the relationship between the quantities.
    - Write-out the propagation of error equation and do a preliminary evaluation, if possible, based on propagation of error.
2. If the measurement result **can be replicated directly**, regardless of the number of secondary quantities in the individual repetitions, treat the uncertainty evaluation as in (A.1) to (A.5) above, being sure to evaluate all sources of random error in the process.
3. If the measurement result **cannot be replicated directly**, treat *each measurement quantity* as in (A.1) and (A.2) and:
    - Compute a standard deviation for each measurement quantity.
    - Combine the standard deviations for the individual quantities into a standard deviation for the reported result via propagation of error.
4. Compute a standard deviation for each type B component of uncertainty.
5. Combine type A and type B standard deviations into a standard uncertainty for the reported result.
6. Compute an expanded uncertainty.
7. Compare the uncerainty derived by propagation of error with the uncertainty derived by data analysis techniques.

2.5.3. Type A evaluations

Type A evaluations apply to both error and bias

Type A evaluations can apply to both random error and bias. The only requirement is that the calculation of the uncertainty component be based on a statistical analysis of data. The distinction to keep in mind with regard to random error and bias is that:

- random errors cannot be corrected
- biases can, theoretically at least, be corrected or eliminated from the result.

Caveat for biases

The ISO guidelines are based on the assumption that all biases are corrected and that the only uncertainty from this source is the uncertainty of the correction. The section on type A evaluations of bias gives guidance on how to assess, correct and calculate uncertainties related to bias.

Random error and bias require different types of analyses

How the source of error affects the reported value and the context for the uncertainty determines whether an analysis of random error or bias is appropriate.

Consider a laboratory with several instruments that can reasonably be assumed to be representative of all similar instruments. Then the differences among these instruments can be considered to be a random effect if the uncertainty statement is intended to apply to the result of any instrument, selected at random, from this batch.

If, on the other hand, the uncertainty statement is intended to apply to one specific instrument, then the bias of this instrument relative to the group is the component of interest.

The following pages outline methods for type A evaluations of:

1. Random errors
2. Bias

2.5.3.1. Type A evaluations of random components

Type A evaluations of random components

Type A sources of uncertainty fall into three main categories:

1. Uncertainties that reveal themselves over time
2. Uncertainties caused by specific conditions of measurement
3. Uncertainties caused by material inhomogeneities

Time-dependent changes are a primary source of random errors

One of the most important indicators of random error is time, with the root cause perhaps being environmental changes over time. Three levels of time-dependent effects are discussed in this section.

Many possible configurations may exist in a laboratory for making measurements

Other sources of uncertainty are related to measurement configurations within the laboratory. Measurements on test items are usually made on a single day, with a single operator, on a single instrument, etc. If the intent of the uncertainty is to characterize all measurements made in the laboratory, the uncertainty should account for any differences due to:

1. instruments
2. operators
3. geometries
4. other

Examples of causes of differences within a laboratory

Examples of causes of differences within a well-maintained laboratory are:

1. Differences among instruments for measurements of derived units, such as sheet resistance of silicon, where the instruments cannot be directly calibrated to a reference base
2. Differences among operators for optical measurements that are not automated and depend strongly on operator sightings
3. Differences among geometrical or electrical

configurations of the instrumentation

Calibrated instruments do not fall in this class	Calibrated instruments do not normally fall in this class because uncertainties associated with the instrument's calibration are reported as type B evaluations, and the instruments in the laboratory should agree within the calibration uncertainties. Instruments whose responses are not directly calibrated to the defined unit are candidates for type A evaluations. This covers situations in which the measurement is defined by a test procedure or standard practice using a specific instrument type.
Evaluation depends on the context for the uncertainty	How these differences are treated depends primarily on the context for the uncertainty statement. The differences, depending on the context, will be treated either as random differences, or as bias differences.
Uncertainties due to inhomogeneities	Artifacts, electrical devices, and chemical substances, etc. can be inhomogeneous relative to the quantity that is being characterized by the measurement process. If this fact is known beforehand, it may be possible to measure the artifact very carefully at a specific site and then direct the user to also measure at this site. In this case, there is no contribution to measurement uncertainty from inhomogeneity. However, this is not always possible, and measurements may be destructive. As an example, compositions of chemical compounds may vary from bottle to bottle. If the reported value for the lot is established from measurements on a few bottles drawn at random from the lot, this variability must be taken into account in the uncertainty statement. Methods for testing for inhomogeneity and assessing the appropriate uncertainty are discussed on another page.

2.5.3.1.1. Type A evaluations of time-dependent effects

Time-dependent changes are a primary source of random errors

One of the most important indicators of random error is time. Effects not specifically studied, such as environmental changes, exhibit themselves over time. Three levels of time-dependent errors are discussed in this section. These can be usefully characterized as:

1. Level-1 or short-term errors (repeatability, imprecision)
2. Level-2 or day-to-day errors (reproducibility)
3. Level-3 or long-term errors (stability - which may not be a concern for all processes)

Day-to-day errors can be the dominant source of uncertainty

With instrumentation that is exceedingly precise in the short run, changes over time, often caused by small environmental effects, are frequently the dominant source of uncertainty in the measurement process. The uncertainty statement is not 'true' to its purpose if it describes a situation that cannot be reproduced over time. The customer for the uncertainty is entitled to know the range of possible results for the measurement result, independent of the day or time of year when the measurement was made.

Two levels may be sufficient

Two levels of time-dependent errors are probably sufficient for describing the majority of measurement processes. Three levels may be needed for new measurement processes or processes whose characteristics are not well understood.

Measurements on test item are used to assess uncertainty only when no other data are available

Repeated measurements on the test item generally do not cover a sufficient time period to capture day-to-day changes in the measurement process. The standard deviation of these measurements is quoted as the estimate of uncertainty only if no other data are available for the assessment. For J short-term measurements, this standard deviation has $\nu = J - 1$ degrees of freedom.

A check standard is the best device for capturing all sources of random error

The best approach for capturing information on time-dependent sources of uncertainties is to intersperse the workload with measurements on a check standard taken at set intervals over the life of the process. The standard deviation of the check standard measurements estimates the overall temporal component of uncertainty directly -- thereby obviating the estimation of individual components.

Nested design for estimating type A uncertainties

Case study: Temporal uncertainty from a 3-level nested design

A less-efficient method for estimating time-dependent sources of uncertainty is a designed experiment. Measurements can be made specifically for estimating two or three levels of errors. There are many ways to do this, but the easiest method is a nested design where J short-term measurements are replicated on K days and the entire operation is then replicated over L runs (months, etc.). The analysis of these data leads to:

- s_1 = standard deviation with $(J-1)$ degrees of freedom for short-term errors
- s_2 = standard deviation with $(K-1)$ degrees of freedom for day-to-day errors
- s_3 = standard deviation with $(L-1)$ degrees of freedom for very long-term errors

Approaches given in this chapter

The computation of the uncertainty of the reported value for a test item is outlined for situations where temporal sources of uncertainty are estimated from:

1. measurements on the test item itself
2. measurements on a check standard
3. measurements from a 2-level nested design (gauge study)
4. measurements from a 3-level nested design (gauge study)

2. Measurement Process Characterization
2.5. Uncertainty analysis
2.5.3. Type A evaluations
2.5.3.1. Type A evaluations of random components

2.5.3.1.2. Measurement configuration within the laboratory

Purpose of this page

The purpose of this page is to outline options for estimating uncertainties related to the specific measurement configuration under which the test item is measured, given other possible measurement configurations. Some of these may be controllable and some of them may not, such as:

- instrument
- operator
- temperature
- humidity

The effect of uncontrollable environmental conditions in the laboratory can often be estimated from check standard data taken over a period of time, and methods for calculating components of uncertainty are discussed on other pages. Uncertainties resulting from controllable factors, such as operators or instruments chosen for a specific measurement, are discussed on this page.

First, decide on context for uncertainty

The approach depends primarily on the context for the uncertainty statement. For example, if instrument effect is the question, one approach is to regard, say, the instruments in the laboratory as a random sample of instruments of the same type and to compute an uncertainty that applies to all results regardless of the particular instrument on which the measurements are made. The other approach is to compute an uncertainty that applies to results using a specific instrument.

Next, evaluate whether or not there are differences

To treat instruments as a random source of uncertainty requires that we first determine if differences due to instruments are significant. The same can be said for operators, etc.

Plan for collecting data

To evaluate the measurement process for instruments, select a random sample of I $(I > 4)$ instruments from those available. Make measurements on Q $(Q > 2)$ artifacts with each instrument.

Graph showing differences among instruments

For a graphical analysis, differences from the average for each artifact can be plotted versus artifact, with instruments individually identified by a special plotting symbol. The plot is examined to determine if some instruments always read high or low relative to the other instruments and if this behavior is consistent across artifacts. If there are systematic and significant differences among instruments, a type A uncertainty for instruments is computed. Notice that in the graph for resistivity probes, there are differences among the probes with probes #4 and #5, for example, consistently reading low relative to the other probes. A standard deviation that describes the differences among the probes is included as a component of the uncertainty.

Standard deviation for instruments

Given the measurements,

$$Y_{11}, Y_{12}, \cdots Y_{1I}; \cdots; Y_{Q1}, Y_{Q2}, \cdots Y_{QI}$$

for each of Q artifacts and I instruments, the pooled standard deviation that describes the differences among instruments is:

$$s_{inst} = \sqrt{\frac{1}{Q}\frac{1}{I-1}\sum_{q=1}^{Q}\sum_{i=1}^{I}(Y_{qi} - \bar{Y}_{q\cdot})^2}$$

where

$$\bar{Y}_{q\cdot} = \frac{1}{I}\sum_{i=1}^{I} Y_{qi}$$

Example of resistivity measurements on silicon wafers

A two-way table of resistivity measurements (ohm.cm) with 5 probes on 5 wafers (identified as: 138, 139, 140, 141, 142) is shown below. Standard deviations for probes with 4 degrees of freedom each are shown for each wafer. The pooled standard deviation over all wafers, with 20 degrees of freedom, is the type A standard deviation for instruments.

	Wafers				
Probe	138	139	140	141	142
1	95.1548	99.3118	96.1018	101.1248	94.2593
281	95.1408	99.3548	96.0805	101.0747	94.2907
. 283	95.1493	99.3211	96.0417	101.1100	94.2487
2062	95.1125	99.2831	96.0492	101.0574	94.2520
2362	95.0928	99.3060	96.0357	101.0602	94.2148

```
Std dev     0.02643   0.02612   0.02826    0.03038
0.02711
DF                4         4         4          4
4


Pooled standard deviation  =  0.02770      DF =
20
```

2.5.3.2. Material inhomogeneity

Purpose of this page

The purpose of this page is to outline methods for assessing uncertainties related to material inhomogeneities. Artifacts, electrical devices, and chemical substances, etc. can be inhomogeneous relative to the quantity that is being characterized by the measurement process.

Effect of inhomogeneity on the uncertainty

Inhomogeneity can be a factor in the uncertainty analysis where

1. an artifact is characterized by a single value and the artifact is inhomogeneous over its surface, etc.
2. a lot of items is assigned a single value from a few samples from the lot and the lot is inhomogeneous from sample to sample.

An unfortunate aspect of this situation is that the uncertainty from inhomogeneity may dominate the uncertainty. If the measurement process itself is very precise and in statistical control, the total uncertainty may still be unacceptable for practical purposes because of material inhomogeneities.

Targeted measurements can eliminate the effect of inhomogeneity

It may be possible to measure an artifact very carefully at a specific site and direct the user to also measure at this site. In this case there is no contribution to measurement uncertainty from inhomogeneity.

Example

Silicon wafers are doped with boron to produce desired levels of resistivity (ohm.cm). Manufacturing processes for semiconductors are not yet capable (at least at the time this was originally written) of producing 2" diameter wafers with constant resistivity over the surfaces. However, because measurements made at the center of a wafer by a certification laboratory can be reproduced in the industrial setting, the inhomogeneity is not a factor in the uncertainty analysis -- as long as only the center-point of the wafer is used for future measurements.

Random

Random inhomogeneities are assessed using statistical

inhomogeneities

methods for quantifying random errors. An example of inhomogeneity is a chemical compound which cannot be sufficiently homogenized with respect to isotopes of interest. Isotopic ratio determinations, which are destructive, must be determined from measurements on a few bottles drawn at random from the lot.

Best strategy

The best strategy is to draw a sample of bottles from the lot for the purpose of identifying and quantifying between-bottle variability. These measurements can be made with a method that lacks the accuracy required to certify isotopic ratios, but is precise enough to allow between-bottle comparisons. A second sample is drawn from the lot and measured with an accurate method for determining isotopic ratios, and the reported value for the lot is taken to be the average of these determinations. There are therefore two components of uncertainty assessed:

1. component that quantifies the imprecision of the average
2. component that quantifies how much an individual bottle can deviate from the average.

Systematic inhomogeneities

Systematic inhomogeneities require a somewhat different approach. Roughness can vary systematically over the surface of a 2" square metal piece lathed to have a specific roughness profile. The certification laboratory can measure the piece at several sites, but unless it is possible to characterize roughness as a mathematical function of position on the piece, inhomogeneity must be assessed as a source of uncertainty.

Best strategy

In this situation, the best strategy is to compute the reported value as the average of measurements made over the surface of the piece and assess an uncertainty for departures from the average. The component of uncertainty can be assessed by one of several methods for evaluating bias -- depending on the type of inhomogeneity.

Standard method

The simplest approach to the computation of uncertainty for systematic inhomogeneity is to compute the maximum deviation from the reported value and, assuming a uniform, normal or triangular distribution for the distribution of inhomogeneity, compute the appropriate standard deviation. Sometimes the approximate shape of the distribution can be inferred from the inhomogeneity measurements. The standard deviation for inhomogeneity assuming a uniform distribution is:

$$s_{inh} = \frac{1}{\sqrt{3}} MaxDeviation$$

2.5.3.2.1. Data collection and analysis

Purpose of this page

The purpose of this page is to outline methods for:

- collecting data
- testing for inhomogeneity
- quantifying the component of uncertainty

Balanced measurements at 2-levels

The simplest scheme for identifying and quantifying the effect of inhomogeneity of a measurement result is a balanced (equal number of measurements per cell) 2-level nested design. For example, *K* bottles of a chemical compound are drawn at random from a lot and *J (J > 1)* measurements are made per bottle. The measurements are denoted by

$$Y_{11}, Y_{12}, \cdots Y_{1j}; \cdots; Y_{k1}, Y_{k2}, \cdots Y_{kj}$$

where the *k* index runs over bottles and the *j* index runs over repetitions within a bottle.

Analysis of measurements

The between (bottle) variance is calculated using an analysis of variance technique that is repeated here for convenience.

$$s^2_{inh} = \frac{1}{K-1}\sum_{k=1}^{K}(\bar{Y}_{k\cdot} - \bar{Y}_{\cdot\cdot})^2 - \frac{1}{KJ(J-1)}\sum_{k=1}^{K}\sum_{j=1}^{J}(Y_{kj} - \bar{Y}_{k\cdot})^2$$

where

$$\bar{Y}_{k\cdot} = \frac{1}{J}\sum_{j=1}^{J} Y_{kj}$$

and

$$\bar{Y}_{\cdot\cdot} = \frac{1}{K}\sum_{k=1}^{K} \bar{Y}_{k\cdot}$$

Between bottle variance may be negative

If this variance is negative, there is no contribution to uncertainty, and the bottles are equivalent with regard to their chemical compositions. Even if the variance is positive, inhomogeneity still may not be statistically significant, in which case it is not required to be included

as a component of the uncertainty.

If the between-bottle variance is statistically significantly (i.e., judged to be greater than zero), then inhomogeneity contributes to the uncertainty of the reported value.

Certification, reported value and associated uncertainty

The purpose of assessing inhomogeneity is to be able to assign a value to the entire batch based on the average of a few bottles, and the determination of inhomogeneity is usually made by a less accurate method than the certification method. The reported value for the batch would be the **average of *N* repetitions on *Q* bottles** using the certification method.

The uncertainty calculation is summarized below for the case where the only contribution to uncertainty from the measurement method itself is the repeatability standard deviation, s_1 associated with the certification method. For more complicated scenarios, see the pages on uncertainty budgets.

If $s_{inh}^2 \leq 0 \Rightarrow s_{reported\ value} = \frac{1}{\sqrt{QN}} s_1$

If $s_{inh}^2 > 0$, we need to distinguish two cases and their interpretations:

1. The standard deviation

$$s_{reported\ value} = \sqrt{\frac{Q+1}{Q} s_{inh}^2 + \frac{1}{QN} s_1^2}$$

   leads to an interval that covers the difference between the reported value and the average for a bottle selected at random from the batch.

2. The standard deviation

$$s_{reported\ value} = \sqrt{\frac{Q+1}{Q} s_{inh}^2 + \frac{QN+1}{QN} s_1^2}$$

   allows one to test the instrument using a single measurement. The prediction interval for the difference between the reported value and a single measurement, made with the same precision as the certification measurements, on a bottle selected at random from the batch. This is appropriate when the instrument under test is similar to the certification instrument. If the difference is not within the interval, the user's instrument is in need of calibration.

Relationship to prediction intervals

When the standard deviation for inhomogeneity is included in the calculation, as in the last two cases above, the uncertainty interval becomes a *prediction interval* (Hahn & Meeker) and is interpreted as characterizing a future measurement on a bottle drawn at random from the lot.

2.5.3.3. Type A evaluations of bias

Sources of bias relate to the specific measurement environment

The sources of bias discussed on this page cover specific measurement configurations. Measurements on test items are usually made on a single day, with a single operator, with a single instrument, etc. Even if the intent of the uncertainty is to characterize only those measurements made in one specific configuration, the uncertainty must account for any significant differences due to:

1. instruments
2. operators
3. geometries
4. other

Calibrated instruments do not fall in this class

Calibrated instruments do not normally fall in this class because uncertainties associated with the instrument's calibration are reported as type B evaluations, and the instruments in the laboratory should agree within the calibration uncertainties. Instruments whose responses are not directly calibrated to the defined unit are candidates for type A evaluations. This covers situations where the measurement is defined by a test procedure or standard practice using a specific instrument type.

The best strategy is to correct for bias and compute the uncertainty of the correction

This problem was treated on the foregoing page as an analysis of random error for the case where the uncertainty was intended to apply to all measurements for all configurations. If measurements for only *one configuration* are of interest, such as measurements made with a specific instrument, or if a smaller uncertainty is required, the differences among, say, instruments are treated as biases. The *best strategy* in this situation is to correct all measurements made with a specific instrument to the average for the instruments in the laboratory and compute a type A uncertainty for the correction. This strategy, of course, relies on the assumption that the instruments in the laboratory represent a random sample of all instruments of a specific type.

Only limited comparisons can be made

However, suppose that it is possible to make comparisons among, say, only two instruments and neither is known to be 'unbiased'. This scenario requires a different strategy

among sources of possible bias

because the average will not necessarily be an unbiased result. The best strategy if there is a significant difference between the instruments, and this should be tested, is to apply a 'zero' correction and assess a type A uncertainty of the correction.

Guidelines for treatment of biases

The discussion above is intended to point out that there are many possible scenarios for biases and that they should be treated on a case-by-case basis. A plan is needed for:

- gathering data
- testing for bias (graphically and/or statistically)
- estimating biases
- assessing uncertainties associated with significant biases.

caused by:

- instruments
- operators
- configurations, geometries, etc.
- inhomogeneities

Plan for testing for assessing bias

Measurements needed for assessing biases among instruments, say, requires a random sample of I $(I > 1)$ instruments from those available and measurements on Q $(Q > 2)$ artifacts with each instrument. The same can be said for the other sources of possible bias. General strategies for dealing with significant biases are given in the table below.

Data collection and analysis for assessing biases related to:

- lack of resolution of instrument
- non-linearity of instrument
- drift

are addressed in the section on gauge studies.

Sources of data for evaluating this type of bias

Databases for evaluating bias may be available from:

- check standards
- gauge R and R studies
- control measurements

Strategies for assessing corrections and uncertainties associated with significant biases

Type of bias	Examples	Type of correction	Uncertainty
1. Inconsistent	Sign change (+ to -) Varying magnitude	Zero	Based on maximum bias

2. Consistent	Instrument bias ~ same magnitude over many artifacts	Bias (for a single instrument) = difference from average over several instruments	Standard deviation of correction
3. Not correctable because of sparse data - consistent or inconsistent	Limited testing; e.g., only 2 instruments, operators, configurations, etc.	Zero	Standard deviation of correction
4. Not correctable - consistent	Lack of resolution, non-linearity, drift, material inhomogeneity	Zero	Based on maximum bias

Strategy for no significant bias If there is no significant bias over time, there is no correction and no contribution to uncertainty.

2.5.3.3.1. Inconsistent bias

Strategy for inconsistent bias -- apply a zero correction

If there is significant bias but it changes direction over time, a zero correction is assumed and the standard deviation of the correction is reported as a type A uncertainty; namely,

$$s_{correction} = \frac{1}{\sqrt{3}} MaxBias$$

Computations based on uniform or normal distribution

The equation for estimating the standard deviation of the correction assumes that biases are uniformly distributed between {-max |bias|, + max |bias|}. This assumption is quite conservative. It gives a larger uncertainty than the assumption that the biases are normally distributed. If normality is a more reasonable assumption, substitute the number '3' for the 'square root of 3' in the equation above.

Example of change in bias over time

The results of resistivity measurements with five probes on five silicon wafers are shown below for probe #283, which is the probe of interest at this level with the artifacts being 1 ohm.cm wafers. The bias for probe #283 is negative for run 1 and positive for run 2 with the runs separated by a two-month time period. The correction is taken to be zero.

Table of biases (ohm.cm) for probe 283

Wafer	Probe	Run 1	Run 2
11	283	0.0000340	-0.0001841
26	283	-0.0001000	0.0000861
42	283	0.0000181	0.0000781
131	283	-0.0000701	0.0001580
208	283	-0.0000240	0.0001879
Average	283	-0.0000284	0.0000652

A conservative assumption is that the bias could fall somewhere within the limits ± a, with a = maximum bias or 0.0000652 ohm.cm. The standard deviation of the correction is included as a type A systematic component of the uncertainty.

$$s_{correction} = \frac{1}{\sqrt{3}} MaxBias = 0.000038\ ohm.cm$$

2.5.3.3.2. Consistent bias

Consistent bias

Bias that is significant and persists consistently over time for a specific instrument, operator, or configuration should be corrected if it can be reliably estimated from repeated measurements. Results with the instrument of interest are then corrected to:

$$\textit{Corrected result} = \textit{Measurement} - \textit{Estimate of bias}$$

The example below shows how bias can be identified graphically from measurements on five artifacts with five instruments and estimated from the differences among the instruments.

Graph showing consistent bias for probe #5

An analysis of bias for five instruments based on measurements on five artifacts shows differences from the average for each artifact plotted versus artifact with instruments individually identified by a special plotting symbol. The plot is examined to determine if some instruments always read high or low relative to the other instruments, and if this behavior is consistent across artifacts. Notice that on the graph for resistivity probes, probe #2362, (#5 on the graph), which is the instrument of interest for this measurement process, consistently reads low relative to the other probes. This behavior is consistent over 2 runs that are separated by a two-month time period.

Strategy - correct for bias

Because there is significant and consistent bias for the instrument of interest, the measurements made with that instrument should be corrected for its average bias relative to the other instruments.

Computation of bias

Given the measurements,

$$Y_{qi}\,(q=1,\cdots,Q;i=1,\cdots,I)$$

on Q artifacts with I instruments, the average bias for instrument, I' say, is

$$\bar{B}_{I'} = \frac{1}{Q}\sum_{q=1}^{Q}\left(Y_{qI'} - \bar{Y}_{q\cdot}\right)$$

where

$$\bar{Y}_{q\cdot} = \frac{1}{I}\sum_{i=1}^{I} Y_{qi}$$

Computation of correction

The correction that should be made to measurements made with instrument I' is

$$Y_{corrected} = Y_{measured} - \bar{B}_{I'}$$

Type A uncertainty of the correction

The type A uncertainty of the correction is the standard deviation of the average bias or

$$s_{correction} = \frac{1}{\sqrt{Q}} s_{bias} = \frac{1}{\sqrt{Q}} \sqrt{\frac{1}{(Q-1)} \sum_{q=1}^{Q} \left(Y_{qI'} - \bar{Y}_{q\cdot} - \bar{B}_{I'} \right)^2}$$

Example of consistent bias for probe #2362 used to measure resistivity of silicon wafers

The table below comes from the table of resistivity measurements from a type A analysis of random effects with the average for each wafer subtracted from each measurement. The differences, as shown, represent the biases for each probe with respect to the other probes. Probe #2362 has an average bias, over the five wafers, of -0.02724 ohm.cm. If measurements made with this probe are corrected for this bias, the standard deviation of the correction is a type A uncertainty.

Table of biases for probes and silicon wafers (ohm.cm)

	Wafers				
Probe	138	139	140	141	142
1	0.02476	-0.00356	0.04002	0.03938	0.00620
181	0.01076	0.03944	0.01871	-0.01072	0.03761
182	0.01926	0.00574	-0.02008	0.02458	-0.00439
2062	-0.01754	-0.03226	-0.01258	-0.02802	-0.00110
2362	-0.03725	-0.00936	-0.02608	-0.02522	-0.03830

Average bias for probe #2362 = - 0.02724

Standard deviation of bias = 0.01171 with 4 degrees of freedom

Standard deviation of correction = 0.01171/sqrt(5) = 0.00523

Note on different approaches to instrument bias

The analysis on this page considers the case where only one instrument is used to make the certification measurements; namely probe #2362, and the certified values are corrected for bias due to this probe. The analysis in the section on type A analysis of random effects considers the case where any one of the probes could be used to make the certification measurements.

2.5.3.3.3. Bias with sparse data

Strategy for dealing with limited data

The purpose of this discussion is to outline methods for dealing with biases that may be real but which cannot be estimated reliably because of the sparsity of the data. For example, a test between two, of many possible, configurations of the measurement process cannot produce a reliable enough estimate of bias to permit a correction, but it can reveal problems with the measurement process. The strategy for a significant bias is to apply a 'zero' correction. The type A uncertainty component is the standard deviation of the correction, and the calculation depends on whether the bias is

- inconsistent
- consistent

The analyses in this section can be produced using both Dataplot code and R code.

Example of differences among wiring settings

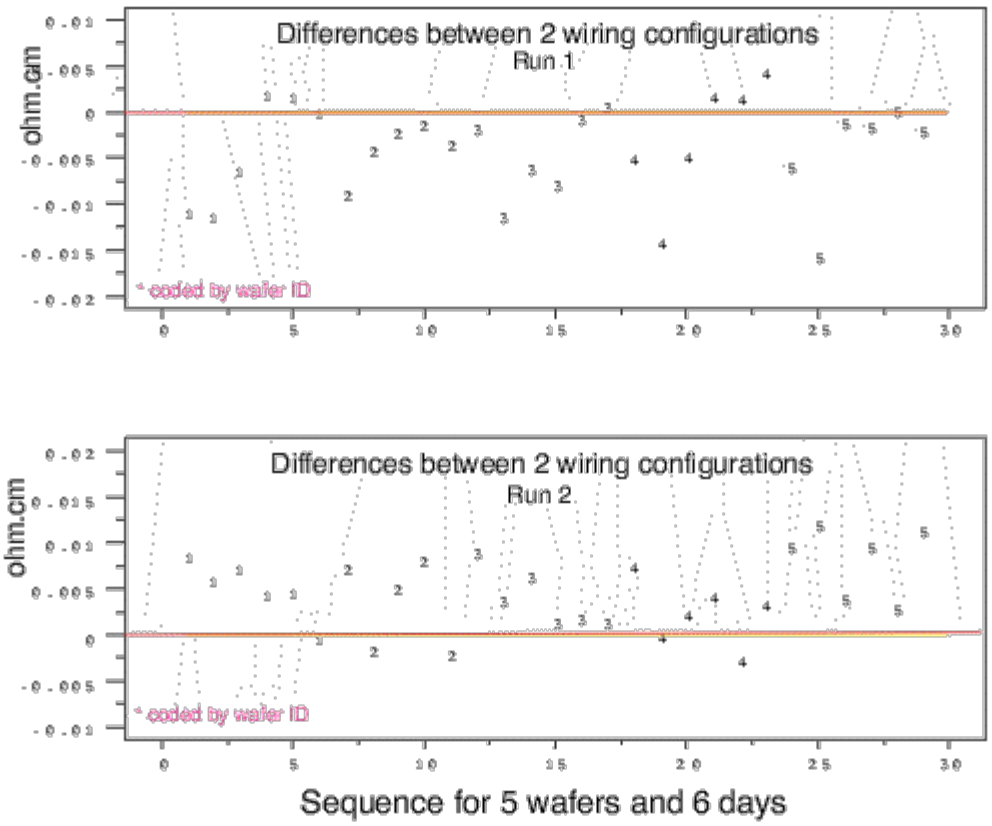
An example is given of a study of wiring settings for a single gauge. The gauge, a 4-point probe for measuring resistivity of silicon wafers, can be wired in several ways. Because it was not possible to test all wiring configurations during the gauge study, measurements were made in only two configurations as a way of identifying possible problems.

Data on wiring configurations

Measurements were made on six wafers over six days (except for 5 measurements on wafer 39) with probe #2062 wired in two configurations. This sequence of measurements was repeated after about a month resulting in two runs. A database of differences between measurements in the two configurations on the same day are analyzed for significance.

Plot the differences between the two wiring configurations

A plot of the differences between the two configurations shows that the differences for run 1 are, for the most part, less than zero, and the differences for run 2 are greater than zero.



Statistical test for difference between two configurations

A *t*-statistic is used as an approximate test where we are assuming the differences are approximately normal. The average difference and standard deviation of the difference are required for this test. If

$$t = \left| \frac{\sqrt{N}}{s_{diff}} Avg_{diff} \right| > 2$$

the difference between the two configurations is statistically significant.

The average and standard deviation computed from the N = 29 differences in each run from the table above are shown along with corresponding *t*-values which confirm that the differences are significant, but in opposite directions, for both runs.

Average differences between wiring configurations

Run	Probe	Average	Std dev	N	t
1	2062	- 0.00383	0.00514	29	- 4.0
2	2062	+ 0.00489	0.00400	29	+ 6.6

Case of inconsistent bias

The data reveal a significant wiring bias for both runs that changes direction between runs. Because of this inconsistency, a 'zero' correction is applied to the results, and the type A uncertainty is taken to be

$$s_{correction} = \frac{1}{\sqrt{3}} MaxBias$$

For this study, the type A uncertainty for wiring bias is

$$s_{correction} = \frac{1}{\sqrt{3}} 0.00489\, ohm.cm = 0.0028\, ohm.cm$$

Case of consistent bias

Even if the bias is consistent over time, a 'zero' correction is applied to the results, and for a single run, the estimated standard deviation of the correction is

$$s_{correction} = \frac{1}{\sqrt{N}} s_{diff}$$

For two runs (1 and 2), the estimated standard deviation of the correction is

$$s_{correction} = \frac{1}{\sqrt{2N}} \sqrt{s_{diff_1}^2 + s_{diff_2}^2}$$

2.5.4. Type B evaluations

Type B evaluations apply to both error and bias

Type B evaluations can apply to both random error and bias. The distinguishing feature is that the calculation of the uncertainty component is not based on a statistical analysis of data. The distinction to keep in mind with regard to random error and bias is that:

- random errors cannot be corrected
- biases can, theoretically at least, be corrected or eliminated from the result.

Sources of type B evaluations

Some examples of sources of uncertainty that lead to type B evaluations are:

- Reference standards calibrated by another laboratory
- Physical constants used in the calculation of the reported value
- Environmental effects that cannot be sampled
- Possible configuration/geometry misalignment in the instrument
- Lack of resolution of the instrument

Documented sources of uncertainty from other processes

Documented sources of uncertainty, such as calibration reports for reference standards or published reports of uncertainties for physical constants, pose no difficulties in the analysis. The uncertainty will usually be reported as an expanded uncertainty, U, which is converted to the standard uncertainty,

$$u = U/k$$

If the k factor is not known or documented, it is probably conservative to assume that $k = 2$.

Sources of uncertainty that are local to the measurement process

Sources of uncertainty that are local to the measurement process but which cannot be adequately sampled to allow a statistical analysis require type B evaluations. One technique, which is widely used, is to estimate the worst-case effect, a, for the source of interest, from

- experience
- scientific judgment

- scant data

A standard deviation, assuming that the effect is two-sided, can then be computed based on a uniform, triangular, or normal distribution of possible effects.

Following the Guide to the Expression of Uncertainty of Measurement (GUM), the convention is to assign **infinite degrees of freedom** to standard deviations derived in this manner.

2.5.4.1. Standard deviations from assumed distributions

Difficulty of obtaining reliable uncertainty estimates

The methods described on this page attempt to avoid the difficulty of allowing for sources of error for which reliable estimates of uncertainty do not exist. The methods are based on assumptions that may, or may not, be valid and require the experimenter to consider the effect of the assumptions on the final uncertainty.

Difficulty of obtaining reliable uncertainty estimates

The ISO guidelines do not allow worst-case estimates of bias to be added to the other components, but require they in some way be converted to equivalent standard deviations. The approach is to consider that any error or bias, for the situation at hand, is a random draw from a known statistical distribution. Then the standard deviation is calculated from known (or assumed) characteristics of the distribution. Distributions that can be considered are:

- Uniform
- Triangular
- Normal (Gaussian)

Standard deviation for a uniform distribution

The uniform distribution leads to the most conservative estimate of uncertainty; i.e., it gives the largest standard deviation. The calculation of the standard deviation is based on the assumption that the end-points, ± a, of the distribution are known. It also embodies the assumption that all effects on the reported value, between -a and +a, are equally likely for the particular source of uncertainty.

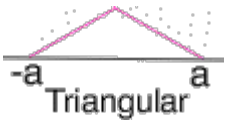
$$s_{source} = \frac{1}{\sqrt{3}} a$$

Standard deviation for a triangular

The triangular distribution leads to a less conservative estimate of uncertainty; i.e., it gives a smaller standard deviation than the uniform distribution. The calculation of the standard deviation is based on the assumption that the end-

distribution

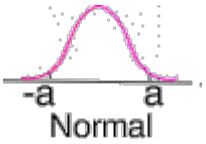
points, ± a, of the distribution are known and the mode of the triangular distribution occurs at zero.



$$s_{source} = \frac{1}{\sqrt{6}} a$$

Standard deviation for a normal distribution

The normal distribution leads to the least conservative estimate of uncertainty; i.e., it gives the smallest standard deviation. The calculation of the standard deviation is based on the assumption that the end-points, ± a, encompass 99.7 percent of the distribution.



$$s_{source} = \frac{1}{3} a$$

Degrees of freedom

In the context of using the Welch-Saitterthwaite formula with the above distributions, the degrees of freedom is assumed to be infinite.

2.5.5. Propagation of error considerations

Top-down approach consists of estimating the uncertainty from direct repetitions of the measurement result

The approach to uncertainty analysis that has been followed up to this point in the discussion has been what is called a top-down approach. Uncertainty components are estimated from direct repetitions of the measurement result. To contrast this with a propagation of error approach, consider the simple example where we estimate the area of a rectangle from replicate measurements of length and width. The area

$$area = length \times width$$

can be computed from each replicate. The standard deviation of the reported area is estimated directly from the replicates of area.

Advantages of top-down approach

This approach has the following advantages:

- proper treatment of covariances between measurements of length and width
- proper treatment of unsuspected sources of error that would emerge if measurements covered a range of operating conditions and a sufficiently long time period
- independence from propagation of error model

Propagation of error approach combines estimates from individual auxiliary measurements

The formal propagation of error approach is to compute:

1. standard deviation from the length measurements
2. standard deviation from the width measurements

and combine the two into a standard deviation for area using the approximation for products of two variables (ignoring a possible covariance between length and width),

$$s_{area} = \sqrt{width^2 s^2_{length} + length^2 s^2_{width}}$$

Exact formula

Goodman (1960) derived an exact formula for the variance between two products. Given two random variables, x and y (correspond to width and length in the above approximate formula), the exact formula for the variance is:

$$V(\bar{x}\bar{y}) = [X^2V(y) + Y^2V(x) + 2XYE_{11} + 2X\frac{E_{12}}{n} + 2Y\frac{E_{21}}{n} + \frac{V(x)V(y)}{n} + \frac{Cov((\Delta x)^2,(\Delta y)^2) - E_{11}^2}{n^2}]/n$$

with

- X = E(x) and Y = E(y) (corresponds to width and length, respectively, in the approximate formula)
- V(x) = variance of x and V(y) = variance Y (corresponds to s^2 for width and length, respectively, in the approximate formula)
- $E_{ij} = \{(\Delta x)^i, (\Delta y)^j\}$ where $\Delta x = x$ - X and $\Delta y = y$ - Y
- $Cov((\Delta x)^2, (\Delta y)^2) = E_{22} - V(x)V(y)$

To obtain the standard deviation, simply take the square root of the above formula. Also, an estimate of the statistic is obtained by substituting sample estimates for the corresponding population values on the right hand side of the equation.

Approximate formula assumes indpendence

The approximate formula assumes that length and width are independent. The exact formula assumes that length and width are not independent.

Disadvantages of propagation of error approach

In the ideal case, the propagation of error estimate above will not differ from the estimate made directly from the area measurements. However, in complicated scenarios, they may differ because of:

- unsuspected covariances
- disturbances that affect the reported value and not the elementary measurements (usually a result of mis-specification of the model)
- mistakes in propagating the error through the defining formulas

Propagation of error formula

Sometimes the measurement of interest cannot be replicated directly and it is necessary to estimate its uncertainty via propagation of error formulas (Ku). The propagation of error formula for

$$Y = f(X, Z, \ldots)$$

a function of one or more variables with measurements, X, Z, ... gives the following estimate for the standard deviation of Y:

$$s_y = \sqrt{\left(\frac{\partial Y}{\partial X}\right)^2 s_x^2 + \left(\frac{\partial Y}{\partial Z}\right)^2 s_z^2 + \cdots + \left(\frac{\partial Y}{\partial X}\right)\left(\frac{\partial Y}{\partial Z}\right) s_{xz}^2 + \cdots}$$

where

- s_x is the standard deviation of the X measurements
- s_z is the standard deviation of Z measurements
- s_y is the standard deviation of Y measurements
- $\partial Y / \partial X$ is the partial derivative of the function Y with respect to X, etc.

s_{xz} is the estimated covariance between the *X*,*Z* measurements

Treatment of covariance terms

Covariance terms can be difficult to estimate if measurements are not made in pairs. Sometimes, these terms are omitted from the formula. Guidance on when this is acceptable practice is given below:

1. If the measurements of *X, Z* are independent, the associated covariance term is *zero*.
2. Generally, reported values of test items from calibration designs have *non-zero* covariances that must be taken into account if *Y* is a summation such as the mass of two weights, or the length of two gage blocks end-to-end, etc.
3. Practically speaking, covariance terms should be included in the computation only if they have been estimated from sufficient data. See Ku (1966) for guidance on what constitutes sufficient data.

Sensitivity coefficients

The partial derivatives are the sensitivity coefficients for the associated components.

Examples of propagation of error analyses

Examples of propagation of error that are shown in this chapter are:

- Case study of propagation of error for resistivity measurements
- Comparison of check standard analysis and propagation of error for linear calibration
- Propagation of error for quadratic calibration showing effect of covariance terms

Specific formulas

Formulas for specific functions can be found in the following sections:

- functions of a single variable
- functions of two variables
- functions of many variables

2.5.5.1. Formulas for functions of one variable

Case: Y=f(X,Z)

Standard deviations of reported values that are functions of a single variable are reproduced from a paper by H. Ku (Ku).

The reported value, *Y*, is a function of the average of *N* measurements on a single variable.

Notes	***Function*** Y ***of*** $\bar{X}$ $\bar{X}$ *is an average of N measurements*	***Standard deviation of*** Y s_x *= standard deviation of X.*
	$Y = \bar{X}$	$\frac{1}{\sqrt{N}} s_x$
	$Y = \frac{\bar{X}}{1+\bar{X}}$	$\frac{s_x}{\sqrt{N}\left(1+\bar{X}\right)^2}$
	$Y = \left(\bar{X}\right)^2$	$\frac{2\bar{X}}{\sqrt{N}} s_x$
	$Y = \sqrt{\bar{X}}$	$\frac{s_x}{2\sqrt{N\bar{X}}}$
	$Y = \ln \bar{X}$	$\frac{s_x}{\sqrt{N}\,\bar{X}}$
Approximation could be seriously in error if n is small--	$Y = e^{\bar{X}}$	$\frac{e^{\bar{X}}}{\sqrt{N}} s_x$
Not directly derived from		$\frac{Y}{\sqrt{2(N-1)}}$

the formulas	$Y = \frac{100 s_x}{\bar{X}}$	Note: we need to assume that the original data follow an approximately normal distribution.

2.5.5.2. Formulas for functions of two variables

Case: Y=f(X,Z)

Standard deviations of reported values that are functions of measurements on two variables are reproduced from a paper by H. Ku (Ku).

The reported value, Y is a function of averages of N measurements on two variables.

Function Y ***of*** $\bar{X}, \bar{Z}$ $\bar{X}$ *and* $\bar{Z}$ *are averages of N measurements*	***Standard deviation of*** Y s_X = *standard dev of X;* s_Z = *standard dev of Z;* s_{XZ}^2 = *covariance of X,Z* ***Note: Covariance term is to be included only if there is a reliable estimate***
$Y = A\bar{X} + B\bar{Z}$	$\frac{1}{\sqrt{N}}\sqrt{A^2 s_X^2 + B^2 s_Z^2 + 2AB s_{XZ}^2}$
$Y = \frac{\bar{X}}{\bar{Z}}$	$\frac{1}{\sqrt{N}}\frac{\bar{X}}{\bar{Z}}\sqrt{\frac{s_X^2}{\left(\bar{X}\right)^2} + \frac{s_Z^2}{\left(\bar{Z}\right)^2} - 2\frac{s_{XZ}^2}{\bar{X}\,\bar{Z}}}$
$Y = \frac{\bar{X}}{\bar{X} + \bar{Z}}$	$\left(\frac{Y}{\bar{X}}\right)^2 \frac{1}{\sqrt{N}}\sqrt{\left(\bar{X}\right)^2 s_Z^2 + \left(\bar{Z}\right)^2 s_X^2 - 2\bar{X}\,\bar{Z}\,s_{XZ}^2}$
$Y = \bar{X}\,\bar{Z}$	$\frac{XZ}{\sqrt{N}}\sqrt{\frac{s_X^2}{\bar{X}^2} + \frac{s_Z^2}{Z^2} + 2\frac{s_{XZ}^2}{\bar{X}\bar{Z}}}$
$Y = c\left(\bar{X}\right)^a\left(\bar{Z}\right)^b$	$\frac{Y}{\sqrt{N}}\sqrt{a^2\frac{s_X^2}{\bar{X}^2} + b^2\frac{s_Z^2}{Z^2} + 2ab\frac{s_{XZ}^2}{\bar{X}\bar{Z}}}$ Note: this is an approximation. The exact result could be obtained starting from the exact formula for the standard deviation of a product derived by

	Goodman (1960).

2. Measurement Process Characterization
2.5. Uncertainty analysis
2.5.5. Propagation of error considerations

2.5.5.3. Propagation of error for many variables

Example from fluid flow with a nonlinear function

Computing uncertainty for measurands based on more complicated functions can be done using basic propagation of errors principles. For example, suppose we want to compute the uncertainty of the discharge coefficient for fluid flow (Whetstone et al.). The measurement equation is

$$C_d = \frac{\dot{m}\sqrt{1-\left(\frac{d}{D}\right)^4}}{Kd^2F\sqrt{\rho}\sqrt{\Delta P}}$$

where

$$\begin{aligned} C_d &= \text{discharge coefficient} \\ \dot{m} &= \text{mass flow rate} \\ d &= \text{orifice diameter} \\ D &= \text{pipe diameter} \\ \rho &= \text{fluid density} \\ \Delta P &= \text{differential pressure} \\ K &= \text{constant} \\ F &= \text{thermal expansion factor (constant)} \end{aligned}$$

Assuming the variables in the equation are uncorrelated, the squared uncertainty of the discharge coefficient is

$$s^2_{Cd} = \left[\frac{\partial C_d}{\partial \dot{m}}\right]^2 s^2_{\dot{m}} + \left[\frac{\partial C_d}{\partial d}\right]^2 s^2_d + \left[\frac{\partial C_d}{\partial D}\right]^2 s^2_D + \left[\frac{\partial C_d}{\partial \rho}\right]^2 s^2_\rho + \left[\frac{\partial C_d}{\partial \Delta P}\right]^2 s^2_{\Delta P}$$

and the partial derivatives are the following.

$$\frac{\partial C_d}{\partial \dot{m}} = \frac{\sqrt{1-\left(\frac{d}{D}\right)^4}}{Kd^2F\sqrt{\rho}\sqrt{\Delta P}}$$

$$\frac{\partial C_d}{\partial d} = \frac{-2\dot{m}d}{KFD^4\sqrt{\rho}\sqrt{\Delta P}\sqrt{1-\left(\frac{d}{D}\right)^4}} - \frac{2\dot{m}\sqrt{1-\left(\frac{d}{D}\right)^4}}{Kd^3F\sqrt{\rho}\sqrt{\Delta P}}$$

$$\frac{\partial C_d}{\partial D} = \frac{2\dot{m}d^2}{KFD^5\sqrt{\rho}\sqrt{\Delta P}\sqrt{1-\left(\frac{d}{D}\right)^4}}$$

$$\frac{\partial C_d}{\partial \rho} = \frac{-\dot{m}\sqrt{1-\left(\frac{d}{D}\right)^4}}{2Kd^2F\rho^{\frac{3}{2}}\sqrt{\Delta P}}$$

$$\frac{\partial C_d}{\partial \Delta P} = \frac{-\dot{m}\sqrt{1-\left(\frac{d}{D}\right)^4}}{2Kd^2F\sqrt{\rho}(\Delta P)^{\frac{3}{2}}}$$

Software can simplify propagation of error

Propagation of error for more complicated functions can be done reliably with software capable of symbolic computations or algebraic representations.

Symbolic computation software can also be used to combine the partial derivatives with the appropriate standard deviations, and then the standard deviation for the discharge coefficient can be evaluated and plotted for specific values of the secondary variables, as shown in the comparison of check standard analysis and propagation of error.

Simplification for dealing with multiplicative variables

Propagation of error for several variables can be simplified considerably for the special case where:

- the function, *Y*, is a simple multiplicative function of secondary variables, and
- uncertainty is evaluated as a percentage.

For three variables, *X*, *Z*, *W*, the function

$$Y = X \cdot Z \cdot W$$

has a standard deviation in absolute units of

$$s_Y = \sqrt{(ZW)^2 s_x^2 + (XW)^2 s_z^2 + (XZ)^2 s_w^2}$$
$$= Y\sqrt{\frac{s_x^2}{X^2} + \frac{s_z^2}{Z^2} + \frac{s_w^2}{W^2}}$$

In percent units, the standard deviation can be written as

$$\frac{s_Y}{Y} = \sqrt{\frac{s_x^2}{X^2} + \frac{s_z^2}{Z^2} + \frac{s_w^2}{W^2}}$$

if all covariances are negligible. These formulas are easily extended to more than three variables.

2.5.6. Uncertainty budgets and sensitivity coefficients

Case study showing uncertainty budget

Uncertainty components are listed in a table along with their corresponding sensitivity coefficients, standard deviations and degrees of freedom. A table of typical entries illustrates the concept.

Typical budget of type A and type B uncertainty components

Type A components	Sensitivity coefficient	Standard deviation	Degrees freedom
1. Time (repeatability)	a_1	s_1	*v1*
2. Time (reproducibility)	a_2	s_2	*v2*
3. Time (long-term)	a_3	s_3	*v3*
Type B components			
5. Reference standard	(nominal test / nominal ref)	s_4	*v4*

Sensitivity coefficients show how components are related to result

The sensitivity coefficient shows the relationship of the individual uncertainty component to the standard deviation of the reported value for a test item. The sensitivity coefficient relates to the result that is being reported and not to the method of estimating uncertainty components where the uncertainty, *u*, is

$$u = \sqrt{\sum_{i=1}^{R} a_i^2 s_i^2}$$

Sensitivity coefficients for type A components of uncertainty

This section defines sensitivity coefficients that are appropriate for type A components estimated from repeated measurements. The pages on type A evaluations, particularly the pages related to estimation of repeatability and reproducibility components, should be reviewed before continuing on this page. The convention for the notation for sensitivity coefficients for this section is that:

1. a_1 refers to the sensitivity coefficient for the

repeatability standard deviation, s_1
2. a_2 refers to the sensitivity coefficient for the reproducibility standard deviation, s_2
3. a_3 refers to the sensitivity coefficient for the stability standard deviation, s_3

with some of the coefficients possibly equal to zero.

Note on long-term errors

Even if no day-to-day nor run-to-run measurements were made in determining the reported value, the sensitivity coefficient is non-zero if that standard deviation proved to be significant in the analysis of data.

Sensitivity coefficients for other type A components of random error

Procedures for estimating differences among instruments, operators, etc., which are treated as random components of uncertainty in the laboratory, show how to estimate the standard deviations so that the sensitivity coefficients = 1.

Sensitivity coefficients for type A components for bias

This Handbook follows the ISO guidelines in that biases are corrected (correction may be zero), and the uncertainty component is the standard deviation of the correction. Procedures for dealing with biases show how to estimate the standard deviation of the correction so that the sensitivity coefficients are equal to one.

Sensitivity coefficients for specific applications

The following pages outline methods for computing sensitivity coefficients where the components of uncertainty are derived in the following manner:

1. From measurements on the test item itself
2. From measurements on a check standard
3. From measurements in a 2-level design
4. From measurements in a 3-level design

and give an example of an uncertainty budget with sensitivity coefficients from a 3-level design.

Sensitivity coefficients for type B evaluations

The majority of sensitivity coefficients for type B evaluations will be one with a few exceptions. The sensitivity coefficient for the uncertainty of a reference standard is the nominal value of the test item divided by the nominal value of the reference standard.

Case study- sensitivity coefficients for propagation of error

If the uncertainty of the reported value is calculated from propagation of error, the sensitivity coefficients are the multipliers of the individual variance terms in the propagation of error formula. Formulas are given for selected functions of:

1. functions of a single variable
2. functions of two variables
3. several variables

2.5.6.1. Sensitivity coefficients for measurements on the test item

From data on the test item itself

If the temporal component is estimated from N short-term readings on the test item itself

$$Y_1, Y_2, ..., Y_N$$

and

$$s_1 = \frac{1}{\sqrt{N-1}}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}$$

and the reported value is the average, the standard deviation of the reported value is

$$s_{reported\ value} = \frac{1}{\sqrt{N}} s_1$$

with degrees of freedom $\nu_1 = N - 1$.

Sensitivity coefficients

The sensitivity coefficient is $a_1 = \sqrt{1/N}$. The risk in using this method is that it may seriously underestimate the uncertainty.

To improve the reliability of the uncertainty calculation

If possible, the measurements on the test item should be repeated over M days and averaged to estimate the reported value. The standard deviation for the reported value is computed from the daily averages>, and the standard deviation for the temporal component is:

$$s_{reported\ value} = \frac{1}{\sqrt{M}}\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(\bar{Y}_{m\bullet} - \bar{Y}_{\bullet\bullet}\right)^2}$$

with degrees of freedom $\nu_2 = M - 1$ where $\bar{Y}_{m\bullet}$ are the daily averages and $\bar{Y}_{\bullet\bullet}$ is the grand average.

The sensitivity coefficients are: $a_1 = 0$; $a_2 = \sqrt{1/M}$.

Note on

Even if no day-to-day nor run-to-run measurements were

long-term errors	made in determining the reported value, the sensitivity coefficient is non-zero if that standard deviation proved to be significant in the analysis of data.

2.5.6.2. Sensitivity coefficients for measurements on a check standard

From measurements on check standards

If the temporal component of the measurement process is evaluated from measurements on a check standard and there are M days ($M = 1$ is permissible) of measurements on the test item that are **structured in the same manner as the measurements on the check standard**, the standard deviation for the reported value is

$$s_{reported\ value} = \frac{1}{\sqrt{M}} s_2$$

with degrees of freedom $\nu_2 = K - 1$ from the K entries in the check standard database.

Standard deviation from check standard measurements

The computation of the standard deviation from the check standard values and its relationship to components of instrument precision and day-to-day variability of the process are explained in the section on two-level nested designs using check standards.

Sensitivity coefficients

The sensitivity coefficients are: a_1; $a_2 = \sqrt{1/M}$.

2.5.6.3. Sensitivity coefficients for measurements from a 2-level design

Sensitivity coefficients from a 2-level design

If the temporal components are estimated from a 2-level nested design, and the reported value for a test item is an average over

- N short-term repetitions
- M ($M = 1$ is permissible) days

of measurements on the test item, the standard deviation for the reported value is:

$$s_{reported\ value} = \sqrt{\frac{1}{M}s_{days}^2 + \frac{1}{MN}s_1^2}$$

See the relationships in the section on 2-level nested design for definitions of the standard deviations and their respective degrees of freedom.

Problem with estimating degrees of freedom

If degrees of freedom are required for the uncertainty of the reported value, the formula above cannot be used directly and must be rewritten in terms of the standard deviations, s_1 and s_2.

$$s_{reported\ value} = \sqrt{\frac{1}{M}s_2^2 + \frac{J-N}{MNJ}s_1^2}$$

Sensitivity coefficients

The sensitivity coefficients are: $a_1 = \sqrt{(J-N)/PMNJ}$; $a_2 = \sqrt{1/M}$.

Specific sensitivity coefficients are shown in the table below for selections of N, M.

Sensitivity coefficients for two components of uncertainty

Number	Number	Short-term	Day-to-

short-term N	day-to-day M	sensitivity coefficient a_1	day sensitivity coefficient a_2
1	1	$\sqrt{(J-1)/J}$	1
N	1	$\sqrt{(J-N)/NJ}$	1
N	M	$\sqrt{(J-N)/MNJ}$	$\sqrt{1/M}$

2.5.6.4. Sensitivity coefficients for measurements from a 3-level design

Sensitivity coefficients from a 3-level design

If the temporal components are estimated from a 3-level nested design and the reported value is an average over

- *N* short-term repetitions
- *M* days
- *P* runs

of measurements on the test item, the standard deviation for the reported value is:

$$s_{reported\ value} = \sqrt{\frac{1}{P}s_{runs}^2 + \frac{1}{PM}s_{days}^2 + \frac{1}{PMN}s_1^2}$$

See the section on analysis of variability for definitions and relationships among the standard deviations shown in the equation above.

Problem with estimating degrees of freedom

If degrees of freedom are required for the uncertainty, the formula above cannot be used directly and must be rewritten in terms of the standard deviations s_1, s_2, and s_3.

$$s_{reported\ value} = \sqrt{\frac{1}{P}s_3^2 + \frac{K-M}{PMK}s_2^2 + \frac{J-N}{PMNJ}s_1^2}$$

Sensitivity coefficients

The sensitivity coefficients are:

$a_1 = \sqrt{(J-N)/PMNJ}$; $a_2 = \sqrt{(K-M)/PMK}$;
$a_3 = \sqrt{1/P}$.

Specific sensitivity coefficients are shown in the table below for selections of *N*, *M*, *P*. In addition, the following constraints must be observed:

J must be > or = N and K must be > or = M

Sensitivity coefficients for three components of uncertainty

Number short-term N	Number day-to-day M	Number run-to-run P	Short-term sensitivity coefficient a_1	Day-to-day sensitivity coefficient a_2	Run-to-run sensitivity coefficient a_3
1	1	1	$\sqrt{(J-1)/J}$	$\sqrt{(K-1)/K}$	1
N	1	1	$\sqrt{(J-N)/NJ}$	$\sqrt{(K-1)/K}$	1
N	M	1	$\sqrt{(J-N)/MNJ}$	$\sqrt{(K-M)/MK}$	1
N	M	P	$\sqrt{(J-N)/PMNJ}$	$\sqrt{(K-M)/MPK}$	$\sqrt{1/P}$

2.5.6.5. Example of uncertainty budget

Example of uncertainty budget for three components of temporal uncertainty

An uncertainty budget that illustrates several principles of uncertainty analysis is shown below. The reported value for a test item is the average of N short-term measurements where the temporal components of uncertainty were estimated from a 3-level nested design with J short-term repetitions over K days.

The number of measurements made on the test item is the same as the number of short-term measurements in the design; i.e., $N = J$. Because there were no repetitions over days or runs on the test item, $M = 1$; $P = 1$. The sensitivity coefficients for this design are shown on the foregoing page.

Example of instrument bias

This example also illustrates the case where the measuring instrument is biased relative to the other instruments in the laboratory, with a bias correction applied accordingly. The sensitivity coefficient, given that the bias correction is based on measurements on Q artifacts, is defined as $a_4 = 1$, and the standard deviation, s_4, is the standard deviation of the correction.

Example of error budget for type A and type B uncertainties

Type A components	Sensitivity coefficient	Standard deviation	Degrees freedom
1. Repeatability	$a_1 = 0$	s_1	$J - 1$
2. Reproducibility	$a_2 = \sqrt{K - 1 / K}$	s_2	$K - 1$
2. Stability	$a_3 = 1$	s_3	$L - 1$
3. Instrument bias	$a_4 = 1$	s_4	$Q - 1$

2.5.7. Standard and expanded uncertainties

Definition of standard uncertainty

The sensitivity coefficients and standard deviations are combined by root sum of squares to obtain a 'standard uncertainty'. Given R components, the standard uncertainty is:

$$u = \sqrt{\sum_{i=1}^{R} a_i^2 s_i^2}$$

Expanded uncertainty assures a high level of confidence

If the purpose of the uncertainty statement is to provide coverage with a high level of confidence, an expanded uncertainty is computed as

$$U = ku$$

where k is chosen to be the $t_{1-\alpha/2,\nu}$ critical value from the *t*-table with ν degrees of freedom. For large degrees of freedom, $k = 2$ approximates 95 % coverage.

Interpretation of uncertainty statement

The expanded uncertainty defined above is assumed to provide a high level of coverage for the unknown true value of the measurement of interest so that for any measurement result, Y,

$$Y - U \le TrueValue \le Y + U$$

2.5.7.1. Degrees of freedom

Degrees of freedom for individual components of uncertainty

Degrees of freedom for type A uncertainties are the degrees of freedom for the respective standard deviations. Degrees of freedom for Type B evaluations may be available from published reports or calibration certificates. Special cases where the standard deviation must be estimated from fragmentary data or scientific judgment are assumed to have infinite degrees of freedom; for example,

- Worst-case estimate based on a robustness study or other evidence
- Estimate based on an assumed distribution of possible errors
- Type B uncertainty component for which degrees of freedom are not documented

Degrees of freedom for the standard uncertainty

Degrees of freedom for the standard uncertainty, u, which may be a combination of many standard deviations, is not generally known. This is particularly troublesome if there are large components of uncertainty with small degrees of freedom. In this case, the degrees of freedom is approximated by the Welch-Satterthwaite formula (Brownlee).

$$\nu = \frac{u^4}{\sum_{i=1}^{R} \frac{a_i^4 s_i^4}{\nu_i}}$$

Case study: Uncertainty and degrees of freedom

A case study of type A uncertainty analysis shows the computations of temporal components of uncertainty; instrument bias; geometrical bias; standard uncertainty; degrees of freedom; and expanded uncertainty.

2.5.8. Treatment of uncorrected bias

Background

The ISO Guide (ISO) for expressing measurement uncertainties assumes that all biases are corrected and that the uncertainty applies to the corrected result. For measurements at the factory floor level, this approach has several disadvantages. It may not be practical, may be expensive and may not be economically sound to correct for biases that do not impact the commercial value of the product (Turgel and Vecchia).

Reasons for not correcting for bias

Corrections may be expensive to implement if they require modifications to existing software and "paper and pencil" corrections can be both time consuming and prone to error. In the scientific or metrology laboratory, biases may be documented in certain situations, but the mechanism that causes the bias may not be fully understood, or repeatable, which makes it difficult to argue for correction. In these cases, the best course of action is to report the measurement as taken and adjust the uncertainty to account for the "bias".

The question is how to adjust the uncertainty

A method needs to be developed which assures that the resulting uncertainty has the following properties (Phillips and Eberhardt):

1. The final uncertainty must be greater than or equal to the uncertainty that would be quoted if the bias were corrected.
2. The final uncertainty must reduce to the same uncertainty given that the bias correction is applied.
3. The level of coverage that is achieved by the final uncertainty statement should be at least the level obtained for the case of corrected bias.
4. The method should be transferable so that both the uncertainty and the bias can be used as components of uncertainty in another uncertainty statement.
5. The method should be easy to implement.

2.5.8.1. Computation of revised uncertainty

Definition of the bias and corrected measurement

If the bias is δ and the corrected measurement is defined by

$$Y_{cor} = Y - \delta ,$$

the corrected value of Y has the usual expanded uncertainty interval which is symmetric around the unknown true value for the measurement process and is of the following type:

$$Y_{cor} - U \le TrueValue \le Y_{cor} + U$$

Definition of asymmetric uncertainty interval to account for uncorrected measurement

If no correction is made for the bias, the uncertainty interval is contaminated by the effect of the bias term as follows:

$$Y - (U + \delta) \le TrueValue \le Y + (U - \delta)$$

and can be rewritten in terms of upper and lower endpoints that are asymmetric around the true value; namely,

$$Y - U_{-} \le TrueValue \le Y + U_{+}$$

Conditions on the relationship between the bias and U

The definition above can lead to a negative uncertainty limit; e.g., if the bias is positive and greater than U, the upper endpoint becomes negative. The requirement that the uncertainty limits be greater than or equal to zero for all values of the bias guarantees non-negative uncertainty limits and is accepted at the cost of somewhat wider uncertainty intervals. This leads to the following set of restrictions on the uncertainty limits:

$$U_{-} = \begin{cases} U + \delta & \text{if} \quad U + \delta > 0 \\ 0 & \text{if} \quad U + \delta \le 0 \end{cases}$$

$$U_{+} = \begin{cases} U - \delta & \text{if} \quad U - \delta > 0 \\ 0 & \text{if} \quad U - \delta \le 0 \end{cases}$$

Situation where bias is not known exactly but must be

If the bias is not known exactly, its magnitude is estimated from repeated measurements, from sparse data or from theoretical considerations, and the standard deviation is estimated from repeated measurements or from an assumed distribution. The standard deviation of the bias becomes a

estimated component in the uncertainty analysis with the standard uncertainty restructured to be:

$$u_C = \sqrt{u^2 + u_{bias}^2}$$

and the expanded uncertainty limits become:

$$Limits = \begin{cases} U_- = ku_C + \delta \\ U_+ = ku_C - \delta \end{cases}$$

.

Interpretation The uncertainty intervals described above have the desirable properties outlined on a previous page. For more information on theory and industrial examples, the reader should consult the paper by the authors of this technique (Phillips and Eberhardt).

2. Measurement Process Characterization

2.6. Case studies

Contents The purpose of this section is to illustrate the planning, procedures, and analyses outlined in the various sections of this chapter with data taken from measurement processes at the National Institute of Standards and Technology.

1. **Gauge study of resistivity probes**
2. **Check standard study for resistivity measurements**
3. **Type A uncertainty analysis**
4. **Type B uncertainty analysis and propagation of error**

2.6.1. Gauge study of resistivity probes

Purpose The purpose of this case study is to outline the analysis of a gauge study that was undertaken to identify the sources of uncertainty in resistivity measurements of silicon wafers.

Outline

1. Background and data
2. Analysis and interpretation
3. Graphs showing repeatability standard deviations
4. Graphs showing day-to-day variability
5. Graphs showing differences among gauges
6. Run this example yourself with Dataplot
7. Dataplot macros

2.6.1.1. Background and data

Description of measurements

Measurements of resistivity on 100 ohm.cm wafers were made according to an ASTM Standard Test Method (ASTM F84) to assess the sources of uncertainty in the measurement system. Resistivity measurements have been studied over the years, and it is clear from those data that there are sources of variability affecting the process beyond the basic imprecision of the gauges. Changes in measurement results have been noted over days and over months and the data in this study are structured to quantify these time-dependent changes in the measurement process.

Gauges

The gauges for the study were five probes used to measure resistivity of silicon wafers. The five gauges are assumed to represent a random sample of typical 4-point gauges for making resistivity measurements. There is a question of whether or not the gauges are essentially equivalent or whether biases among them are possible.

Check standards

The check standards for the study were five wafers selected at random from the batch of 100 ohm.cm wafers.

Operators

The effect of operator was not considered to be significant for this study.

Database of measurements

The 3-level nested design consisted of:

- $J = 6$ measurements at the center of each wafer per day
- $K = 6$ days
- $L = 2$ runs

To characterize the probes and the influence of wafers on the measurements, the design was repeated over:

- $Q = 5$ wafers (check standards 138, 139, 140, 141, 142)
- $I = 5$ probes (1, 281, 283, 2062, 2362)

The runs were separated by about one month in time. The $J = 6$ measurements at the center of each wafer are reduced to an average and repeatability standard deviation and recorded in a database with identifications for wafer, probe,

and day.

Software The analyses used in this case study can be generated using both Dataplot code and R code.

2.6.1.1.1. Database of resistivity measurements

The check standards are five wafers chosen at random from a batch of wafers

Measurements of resistivity (ohm.cm) were made according to an ASTM Standard Test Method (F4) at NIST to assess the sources of uncertainty in the measurement system. The gauges for the study were five probes owned by NIST; the check standards for the study were five wafers selected at random from a batch of wafers cut from one silicon crystal doped with phosphorous to give a nominal resistivity of 100 ohm.cm.

Measurements on the check standards are used to estimate repeatability, day effect, and run effect

The effect of operator was not considered to be significant for this study; therefore, 'day' replaces 'operator' as a factor in the nested design. Averages and standard deviations from $J = 6$ measurements at the center of each wafer are shown in the table.

- $J = 6$ measurements at the center of the wafer per day
- $K = 6$ days (one operator) per repetition
- $L = 2$ runs (complete)
- $Q = 5$ wafers (check standards 138, 139, 140, 141, 142)
- $R = 5$ probes (1, 281, 283, 2062, 2362)

Run	Wafer	Probe	Month	Day	Op	Temp	Average	Std Dev
1	138.	1.	3.	15.	1.	22.98	95.1772	0.1191
1	138.	1.	3.	17.	1.	23.02	95.1567	0.0183
1	138.	1.	3.	18.	1.	22.79	95.1937	0.1282
1	138.	1.	3.	21.	1.	23.17	95.1959	0.0398
1	138.	1.	3.	23.	2.	23.25	95.1442	0.0346
1	138.	1.	3.	23.	1.	23.20	95.0610	0.1539
1	138.	281.	3.	16.	1.	22.99	95.1591	0.0963
1	138.	281.	3.	17.	1.	22.97	95.1195	0.0606
1	138.	281.	3.	18.	1.	22.83	95.1065	0.0842
1	138.	281.	3.	21.	1.	23.28	95.0925	0.0973
1	138.	281.	3.	23.	2.	23.14	95.1990	0.1062

```
1  138.    281.      3.     23.      1.   23.16   95.1682
0.1090
1  138.    283.      3.     16.      1.   22.95   95.1252
0.0531
1  138.    283.      3.     17.      1.   23.08   95.1600
0.0998
1  138.    283.      3.     18.      1.   23.13   95.0818
0.1108
1  138.    283.      3.     21.      1.   23.28   95.1620
0.0408
1  138.    283.      3.     22.      1.   23.36   95.1735
0.0501
1  138.    283.      3.     24.      2.   22.97   95.1932
0.0287
1  138.   2062.      3.     16.      1.   22.97   95.1311
0.1066
1  138.   2062.      3.     17.      1.   22.98   95.1132
0.0415
1  138.   2062.      3.     18.      1.   23.16   95.0432
0.0491
1  138.   2062.      3.     21.      1.   23.16   95.1254
0.0603
1  138.   2062.      3.     22.      1.   23.28   95.1322
0.0561
1  138.   2062.      3.     24.      2.   23.19   95.1299
0.0349
1  138.   2362.      3.     15.      1.   23.08   95.1162
0.0480
1  138.   2362.      3.     17.      1.   23.01   95.0569
0.0577
1  138.   2362.      3.     18.      1.   22.97   95.0598
0.0516
1  138.   2362.      3.     22.      1.   23.23   95.1487
0.0386
1  138.   2362.      3.     23.      2.   23.28   95.0743
0.0256
1  138.   2362.      3.     24.      2.   23.10   95.1010
0.0420
1  139.      1.      3.     15.      1.   23.01   99.3528
0.1424
1  139.      1.      3.     17.      1.   23.00   99.2940
0.0660
1  139.      1.      3.     17.      1.   23.01   99.2340
0.1179
1  139.      1.      3.     21.      1.   23.20   99.3489
0.0506
1  139.      1.      3.     23.      2.   23.22   99.2625
0.1111
1  139.      1.      3.     23.      1.   23.22   99.3787
0.1103
1  139.    281.      3.     16.      1.   22.95   99.3244
0.1134
1  139.    281.      3.     17.      1.   22.98   99.3378
0.0949
1  139.    281.      3.     18.      1.   22.86   99.3424
0.0847
1  139.    281.      3.     22.      1.   23.17   99.4033
0.0801
1  139.    281.      3.     23.      2.   23.10   99.3717
0.0630
1  139.    281.      3.     23.      1.   23.14   99.3493
0.1157
1  139.    283.      3.     16.      1.   22.94   99.3065
0.0381
1  139.    283.      3.     17.      1.   23.09   99.3280
0.1153
1  139.    283.      3.     18.      1.   23.11   99.3000
0.0818
1  139.    283.      3.     21.      1.   23.25   99.3347
0.0972
1  139.    283.      3.     22.      1.   23.36   99.3929
0.1189
1  139.    283.      3.     23.      1.   23.18   99.2644
0.0622
1  139.   2062.      3.     16.      1.   22.94   99.3324
0.1531
1  139.   2062.      3.     17.      1.   23.08   99.3254
0.0543
1  139.   2062.      3.     18.      1.   23.15   99.2555
0.1024
1  139.   2062.      3.     18.      1.   23.18   99.1946
```

```
0.0851
1  139.   2062.      3.     22.      1.   23.27   99.3542
0.1227
1  139.   2062.      3.     24.      2.   23.23   99.2365
0.1218
1  139.   2362.      3.     15.      1.   23.08   99.2939
0.0818
1  139.   2362.      3.     17.      1.   23.02   99.3234
0.0723
1  139.   2362.      3.     18.      1.   22.93   99.2748
0.0756
1  139.   2362.      3.     22.      1.   23.29   99.3512
0.0475
1  139.   2362.      3.     23.      2.   23.25   99.2350
0.0517
1  139.   2362.      3.     24.      2.   23.05   99.3574
0.0485
1  140.      1.      3.     15.      1.   23.07   96.1334
0.1052
1  140.      1.      3.     17.      1.   23.08   96.1250
0.0916
1  140.      1.      3.     18.      1.   22.77   96.0665
0.0836
1  140.      1.      3.     21.      1.   23.18   96.0725
0.0620
1  140.      1.      3.     23.      2.   23.20   96.1006
0.0582
1  140.      1.      3.     23.      1.   23.21   96.1131
0.1757
1  140.    281.      3.     16.      1.   22.94   96.0467
0.0565
1  140.    281.      3.     17.      1.   22.99   96.1081
0.1293
1  140.    281.      3.     18.      1.   22.91   96.0578
0.1148
1  140.    281.      3.     22.      1.   23.15   96.0700
0.0495
1  140.    281.      3.     22.      1.   23.33   96.1052
0.1722
1  140.    281.      3.     23.      1.   23.19   96.0952
0.1786
1  140.    283.      3.     16.      1.   22.89   96.0650
0.1301
1  140.    283.      3.     17.      1.   23.07   96.0870
0.0881
1  140.    283.      3.     18.      1.   23.07   95.8906
0.1842
1  140.    283.      3.     21.      1.   23.24   96.0842
0.1008
1  140.    283.      3.     22.      1.   23.34   96.0189
0.0865
1  140.    283.      3.     23.      1.   23.19   96.1047
0.0923
1  140.   2062.      3.     16.      1.   22.95   96.0379
0.2190
1  140.   2062.      3.     17.      1.   22.97   96.0671
0.0991
1  140.   2062.      3.     18.      1.   23.15   96.0206
0.0648
1  140.   2062.      3.     21.      1.   23.14   96.0207
0.1410
1  140.   2062.      3.     22.      1.   23.32   96.0587
0.1634
1  140.   2062.      3.     24.      2.   23.17   96.0903
0.0406
1  140.   2362.      3.     15.      1.   23.08   96.0771
0.1024
1  140.   2362.      3.     17.      1.   23.00   95.9976
0.0943
1  140.   2362.      3.     18.      1.   23.01   96.0148
0.0622
1  140.   2362.      3.     22.      1.   23.27   96.0397
0.0702
1  140.   2362.      3.     23.      2.   23.24   96.0407
0.0627
1  140.   2362.      3.     24.      2.   23.13   96.0445
0.0622
1  141.      1.      3.     15.      1.   23.01  101.2124
0.0900
1  141.      1.      3.     17.      1.   23.08  101.1018
0.0820
```

```
1  141.      1.      3.     18.      1.   22.75  101.1119
0.0500
1  141.      1.      3.     21.      1.   23.21  101.1072
0.0641
1  141.      1.      3.     23.      2.   23.25  101.0802
0.0704
1  141.      1.      3.     23.      1.   23.19  101.1350
0.0699
1  141.    281.      3.     16.      1.   22.93  101.0287
0.0520
1  141.    281.      3.     17.      1.   23.00  101.0131
0.0710
1  141.    281.      3.     18.      1.   22.90  101.1329
0.0800
1  141.    281.      3.     22.      1.   23.19  101.0562
0.1594
1  141.    281.      3.     23.      2.   23.18  101.0891
0.1252
1  141.    281.      3.     23.      1.   23.17  101.1283
0.1151
1  141.    283.      3.     16.      1.   22.85  101.1597
0.0990
1  141.    283.      3.     17.      1.   23.09  101.0784
0.0810
1  141.    283.      3.     18.      1.   23.08  101.0715
0.0460
1  141.    283.      3.     21.      1.   23.27  101.0910
0.0880
1  141.    283.      3.     22.      1.   23.34  101.0967
0.0901
1  141.    283.      3.     24.      2.   23.00  101.1627
0.0888
1  141.   2062.      3.     16.      1.   22.97  101.1077
0.0970
1  141.   2062.      3.     17.      1.   22.96  101.0245
0.1210
1  141.   2062.      3.     18.      1.   23.19  100.9650
0.0700
1  141.   2062.      3.     18.      1.   23.18  101.0319
0.1070
1  141.   2062.      3.     22.      1.   23.34  101.0849
0.0960
1  141.   2062.      3.     24.      2.   23.21  101.1302
0.0505
1  141.   2362.      3.     15.      1.   23.08  101.0471
0.0320
1  141.   2362.      3.     17.      1.   23.01  101.0224
0.1020
1  141.   2362.      3.     18.      1.   23.05  101.0702
0.0580
1  141.   2362.      3.     22.      1.   23.22  101.0904
0.1049
1  141.   2362.      3.     23.      2.   23.29  101.0626
0.0702
1  141.   2362.      3.     24.      2.   23.15  101.0686
0.0661
1  142.      1.      3.     15.      1.   23.02   94.3160
0.1372
1  142.      1.      3.     17.      1.   23.04   94.2808
0.0999
1  142.      1.      3.     18.      1.   22.73   94.2478
0.0803
1  142.      1.      3.     21.      1.   23.19   94.2862
0.0700
1  142.      1.      3.     23.      2.   23.25   94.1859
0.0899
1  142.      1.      3.     23.      1.   23.21   94.2389
0.0686
1  142.    281.      3.     16.      1.   22.98   94.2640
0.0862
1  142.    281.      3.     17.      1.   23.00   94.3333
0.1330
1  142.    281.      3.     18.      1.   22.88   94.2994
0.0908
1  142.    281.      3.     21.      1.   23.28   94.2873
0.0846
1  142.    281.      3.     23.      2.   23.07   94.2576
0.0795
1  142.    281.      3.     23.      1.   23.12   94.3027
0.0389
1  142.    283.      3.     16.      1.   22.92   94.2846
```

```
0.1021
1  142.    283.      3.     17.      1.   23.08   94.2197
0.0627
1  142.    283.      3.     18.      1.   23.09   94.2119
0.0785
1  142.    283.      3.     21.      1.   23.29   94.2536
0.0712
1  142.    283.      3.     22.      1.   23.34   94.2280
0.0692
1  142.    283.      3.     24.      2.   22.92   94.2944
0.0958
1  142.   2062.      3.     16.      1.   22.96   94.2238
0.0492
1  142.   2062.      3.     17.      1.   22.95   94.3061
0.2194
1  142.   2062.      3.     18.      1.   23.16   94.1868
0.0474
1  142.   2062.      3.     21.      1.   23.11   94.2645
0.0697
1  142.   2062.      3.     22.      1.   23.31   94.3101
0.0532
1  142.   2062.      3.     24.      2.   23.24   94.2204
0.1023
1  142.   2362.      3.     15.      1.   23.08   94.2437
0.0503
1  142.   2362.      3.     17.      1.   23.00   94.2115
0.0919
1  142.   2362.      3.     18.      1.   22.99   94.2348
0.0282
1  142.   2362.      3.     22.      1.   23.26   94.2124
0.0513
1  142.   2362.      3.     23.      2.   23.27   94.2214
0.0627
1  142.   2362.      3.     24.      2.   23.08   94.1651
0.1010
2  138.      1.      4.     13.      1.   23.12   95.1996
0.0645
2  138.      1.      4.     15.      1.   22.73   95.1315
0.1192
2  138.      1.      4.     18.      2.   22.76   95.1845
0.0452
2  138.      1.      4.     19.      1.   22.73   95.1359
0.1498
2  138.      1.      4.     20.      2.   22.73   95.1435
0.0629
2  138.      1.      4.     21.      2.   22.93   95.1839
0.0563
2  138.    281.      4.     14.      2.   22.46   95.2106
0.1049
2  138.    281.      4.     18.      2.   22.80   95.2505
0.0771
2  138.    281.      4.     18.      2.   22.77   95.2648
0.1046
2  138.    281.      4.     20.      2.   22.80   95.2197
0.1779
2  138.    281.      4.     20.      2.   22.87   95.2003
0.1376
2  138.    281.      4.     21.      2.   22.95   95.0982
0.1611
2  138.    283.      4.     18.      2.   22.83   95.1211
0.0794
2  138.    283.      4.     13.      1.   23.17   95.1327
0.0409
2  138.    283.      4.     18.      1.   22.67   95.2053
0.1525
2  138.    283.      4.     19.      2.   23.00   95.1292
0.0655
2  138.    283.      4.     21.      2.   22.91   95.1669
0.0619
2  138.    283.      4.     21.      2.   22.96   95.1401
0.0831
2  138.   2062.      4.     15.      1.   22.64   95.2479
0.2867
2  138.   2062.      4.     15.      1.   22.67   95.2224
0.1945
2  138.   2062.      4.     19.      2.   22.99   95.2810
0.1960
2  138.   2062.      4.     19.      1.   22.75   95.1869
0.1571
2  138.   2062.      4.     21.      2.   22.84   95.3053
0.2012
```

```
2  138.   2062.      4.     21.      2.   22.92   95.1432
0.1532
2  138.   2362.      4.     12.      1.   22.74   95.1687
0.0785
2  138.   2362.      4.     18.      2.   22.75   95.1564
0.0430
2  138.   2362.      4.     19.      2.   22.88   95.1354
0.0983
2  138.   2362.      4.     19.      1.   22.73   95.0422
0.0773
2  138.   2362.      4.     20.      2.   22.86   95.1354
0.0587
2  138.   2362.      4.     21.      2.   22.94   95.1075
0.0776
2  139.      1.      4.     13.      2.   23.14   99.3274
0.0220
2  139.      1.      4.     15.      2.   22.77   99.5020
0.0997
2  139.      1.      4.     18.      2.   22.80   99.4016
0.0704
2  139.      1.      4.     19.      1.   22.68   99.3181
0.1245
2  139.      1.      4.     20.      2.   22.78   99.3858
0.0903
2  139.      1.      4.     21.      2.   22.93   99.3141
0.0255
2  139.    281.      4.     14.      2.   23.05   99.2915
0.0859
2  139.    281.      4.     15.      2.   22.71   99.4032
0.1322
2  139.    281.      4.     18.      2.   22.79   99.4612
0.1765
2  139.    281.      4.     20.      2.   22.74   99.4001
0.0889
2  139.    281.      4.     20.      2.   22.91   99.3765
0.1041
2  139.    281.      4.     21.      2.   22.92   99.3507
0.0717
2  139.    283.      4.     13.      2.   23.11   99.3848
0.0792
2  139.    283.      4.     18.      2.   22.84   99.4952
0.1122
2  139.    283.      4.     18.      2.   22.76   99.3220
0.0915
2  139.    283.      4.     19.      2.   23.03   99.4165
0.0503
2  139.    283.      4.     21.      2.   22.87   99.3791
0.1138
2  139.    283.      4.     21.      2.   22.98   99.3985
0.0661
2  139.   2062.      4.     14.      2.   22.43   99.4283
0.0891
2  139.   2062.      4.     15.      2.   22.70   99.4139
0.2147
2  139.   2062.      4.     19.      2.   22.97   99.3813
0.1143
2  139.   2062.      4.     19.      1.   22.77   99.4314
0.1685
2  139.   2062.      4.     21.      2.   22.79   99.4166
0.2080
2  139.   2062.      4.     21.      2.   22.94   99.4052
0.2400
2  139.   2362.      4.     12.      1.   22.82   99.3408
0.1279
2  139.   2362.      4.     18.      2.   22.77   99.3116
0.1131
2  139.   2362.      4.     19.      2.   22.82   99.3241
0.0519
2  139.   2362.      4.     19.      1.   22.74   99.2991
0.0903
2  139.   2362.      4.     20.      2.   22.88   99.3049
0.0783
2  139.   2362.      4.     21.      2.   22.94   99.2782
0.0718
2  140.      1.      4.     13.      1.   23.10   96.0811
0.0463
2  140.      1.      4.     15.      2.   22.75   96.1460
0.0725
2  140.      1.      4.     18.      2.   22.78   96.1582
0.1428
2  140.      1.      4.     19.      1.   22.70   96.1039
```

```
0.1056
2  140.      1.      4.     20.      2.   22.75   96.1262
0.0672
2  140.      1.      4.     21.      2.   22.93   96.1478
0.0562
2  140.    281.      4.     15.      2.   22.71   96.1153
0.1097
2  140.    281.      4.     14.      2.   22.49   96.1297
0.1202
2  140.    281.      4.     18.      2.   22.81   96.1233
0.1331
2  140.    281.      4.     20.      2.   22.78   96.1731
0.1484
2  140.    281.      4.     20.      2.   22.89   96.0872
0.0857
2  140.    281.      4.     21.      2.   22.91   96.1331
0.0944
2  140.    283.      4.     13.      2.   23.22   96.1135
0.0983
2  140.    283.      4.     18.      2.   22.85   96.1111
0.1210
2  140.    283.      4.     18.      2.   22.78   96.1221
0.0644
2  140.    283.      4.     19.      2.   23.01   96.1063
0.0921
2  140.    283.      4.     21.      2.   22.91   96.1155
0.0704
2  140.    283.      4.     21.      2.   22.94   96.1308
0.0258
2  140.   2062.      4.     15.      2.   22.60   95.9767
0.2225
2  140.   2062.      4.     15.      2.   22.66   96.1277
0.1792
2  140.   2062.      4.     19.      2.   22.96   96.1858
0.1312
2  140.   2062.      4.     19.      1.   22.75   96.1912
0.1936
2  140.   2062.      4.     21.      2.   22.82   96.1650
0.1902
2  140.   2062.      4.     21.      2.   22.92   96.1603
0.1777
2  140.   2362.      4.     12.      1.   22.88   96.0793
0.0996
2  140.   2362.      4.     18.      2.   22.76   96.1115
0.0533
2  140.   2362.      4.     19.      2.   22.79   96.0803
0.0364
2  140.   2362.      4.     19.      1.   22.71   96.0411
0.0768
2  140.   2362.      4.     20.      2.   22.84   96.0988
0.1042
2  140.   2362.      4.     21.      1.   22.94   96.0482
0.0868
2  141.      1.      4.     13.      1.   23.07  101.1984
0.0803
2  141.      1.      4.     15.      2.   22.72  101.1645
0.0914
2  141.      1.      4.     18.      2.   22.75  101.2454
0.1109
2  141.      1.      4.     19.      1.   22.69  101.1096
0.1376
2  141.      1.      4.     20.      2.   22.83  101.2066
0.0717
2  141.      1.      4.     21.      2.   22.93  101.0645
0.1205
2  141.    281.      4.     15.      2.   22.72  101.1615
0.1272
2  141.    281.      4.     14.      2.   22.40  101.1650
0.0595
2  141.    281.      4.     18.      2.   22.78  101.1815
0.1393
2  141.    281.      4.     20.      2.   22.73  101.1106
0.1189
2  141.    281.      4.     20.      2.   22.86  101.1420
0.0713
2  141.    281.      4.     21.      2.   22.94  101.0116
0.1088
2  141.    283.      4.     13.      2.   23.26  101.1554
0.0429
2  141.    283.      4.     18.      2.   22.85  101.1267
0.0751
```

```
2  141.    283.      4.     18.      2.   22.76  101.1227
0.0826
2  141.    283.      4.     19.      2.   22.82  101.0635
0.1715
2  141.    283.      4.     21.      2.   22.89  101.1264
0.1447
2  141.    283.      4.     21.      2.   22.96  101.0853
0.1189
2  141.   2062.      4.     15.      2.   22.65  101.1332
0.2532
2  141.   2062.      4.     15.      1.   22.68  101.1487
0.1413
2  141.   2062.      4.     19.      2.   22.95  101.1778
0.1772
2  141.   2062.      4.     19.      1.   22.77  101.0988
0.0884
2  141.   2062.      4.     21.      2.   22.87  101.1686
0.2940
2  141.   2062.      4.     21.      2.   22.94  101.3289
0.2072
2  141.   2362.      4.     12.      1.   22.83  101.1353
0.0585
2  141.   2362.      4.     18.      2.   22.83  101.1201
0.0868
2  141.   2362.      4.     19.      2.   22.91  101.0946
0.0855
2  141.   2362.      4.     19.      1.   22.71  100.9977
0.0645
2  141.   2362.      4.     20.      2.   22.87  101.0963
0.0638
2  141.   2362.      4.     21.      2.   22.94  101.0300
0.0549
2  142.      1.      4.     13.      1.   23.07   94.3049
0.1197
2  142.      1.      4.     15.      2.   22.73   94.3153
0.0566
2  142.      1.      4.     18.      2.   22.77   94.3073
0.0875
2  142.      1.      4.     19.      1.   22.67   94.2803
0.0376
2  142.      1.      4.     20.      2.   22.80   94.3008
0.0703
2  142.      1.      4.     21.      2.   22.93   94.2916
0.0604
2  142.    281.      4.     14.      2.   22.90   94.2557
0.0619
2  142.    281.      4.     18.      2.   22.83   94.3542
0.1027
2  142.    281.      4.     18.      2.   22.80   94.3007
0.1492
2  142.    281.      4.     20.      2.   22.76   94.3351
0.1059
2  142.    281.      4.     20.      2.   22.88   94.3406
0.1508
2  142.    281.      4.     21.      2.   22.92   94.2621
0.0946
2  142.    283.      4.     13.      2.   23.25   94.3124
0.0534
2  142.    283.      4.     18.      2.   22.85   94.3680
0.1643
2  142.    283.      4.     18.      1.   22.67   94.3442
0.0346
2  142.    283.      4.     19.      2.   22.80   94.3391
0.0616
2  142.    283.      4.     21.      2.   22.91   94.2238
0.0721
2  142.    283.      4.     21.      2.   22.95   94.2721
0.0998
2  142.   2062.      4.     14.      2.   22.49   94.2915
0.2189
2  142.   2062.      4.     15.      2.   22.69   94.2803
0.0690
2  142.   2062.      4.     19.      2.   22.94   94.2818
0.0987
2  142.   2062.      4.     19.      1.   22.76   94.2227
0.2628
2  142.   2062.      4.     21.      2.   22.74   94.4109
0.1230
2  142.   2062.      4.     21.      2.   22.94   94.2616
0.0929
2  142.   2362.      4.     12.      1.   22.86   94.2052
```

```
0.0813
2  142.   2362.      4.     18.      2.   22.83   94.2824
0.0605
2  142.   2362.      4.     19.      2.   22.85   94.2396
0.0882
2  142.   2362.      4.     19.      1.   22.75   94.2087
0.0702
2  142.   2362.      4.     20.      2.   22.86   94.2937
0.0591
2  142.   2362.      4.     21.      1.   22.93   94.2330
0.0556
```

2.6.1.2. Analysis and interpretation

Graphs of probe effect on repeatability

A graphical analysis shows repeatability standard deviations plotted by wafer and probe. Probes are coded by numbers with probe #2362 coded as #5. The plots show that for both runs the precision of this probe is better than for the other probes.

Probe #2362, because of its superior precision, was chosen as the tool for measuring all 100 ohm.cm resistivity wafers at NIST. Therefore, the remainder of the analysis focuses on this probe.

Plot of repeatability standard deviations for probe #2362 from the nested design over days, wafers, runs

The precision of probe #2362 is first checked for consistency by plotting the repeatability standard deviations over days, wafers and runs. Days are coded by letter. The plots verify that, for both runs, probe repeatability is not dependent on wafers or days although the standard deviations on days D, E, and F of run 2 are larger in some instances than for the other days. This is not surprising because repeated probing on the wafer surfaces can cause slight degradation. Then the repeatability standard deviations are pooled over:

- $K = 6$ days for $K(J - 1) = 30$ degrees of freedom
- $L = 2$ runs for $LK(J - 1) = 60$ degrees of freedom
- $Q = 5$ wafers for $QLK(J - 1) = 300$ degrees of freedom

The results of pooling are shown below. Intermediate steps are not shown, but the section on repeatability standard deviations shows an example of pooling over wafers.

Pooled level-1 standard deviations (ohm.cm)

Probe	Run 1	DF	Run 2	DF	Pooled DF
2362.	0.0658	150	0.0758	150	0.0710 300

Graphs of reproducibility and stability for

Averages of the 6 center measurements on each wafer are plotted on a single graph for each wafer. The points (connected by lines) on the left side of each graph are

probe #2362

averages at the wafer center plotted over 5 days; the points on the right are the same measurements repeated after one month as a check on the stability of the measurement process. The plots show day-to-day variability as well as slight variability from run-to-run.

Earlier work discounts long-term drift in the gauge as the cause of these changes. A reasonable conclusion is that day-to-day and run-to-run variations come from random fluctuations in the measurement process.

Level-2 (reproducibility) standard deviations computed from day averages and pooled over wafers and runs

Level-2 standard deviations (with $K - 1 = 5$ degrees of freedom each) are computed from the daily averages that are recorded in the database. Then the level-2 standard deviations are pooled over:

- $L = 2$ runs for $L(K - 1) = 10$ degrees of freedom
- $Q = 5$ wafers for $QL(K - 1) = 50$ degrees of freedom

as shown in the table below. The table shows that the level-2 standard deviations are consistent over wafers and runs.

Level-2 standard deviations (ohm.cm) for 5 wafers

		Run 1			Run 2		
Wafer	Probe	Average	Stddev	DF	Average	Stddev	DF
138.	2362.	95.0928	0.0359	5	95.1243	0.0453	5
139.	2362.	99.3060	0.0472	5	99.3098	0.0215	5
140.	2362.	96.0357	0.0273	5	96.0765	0.0276	5
141.	2362.	101.0602	0.0232	5	101.0790	0.0537	5
142.	2362.	94.2148	0.0274	5	94.2438	0.0370	5
	2362.	Pooled	0.0333	25		0.0388	25
		(over 2 runs)				0.0362	50

Level-3 (stability) standard deviations computed from run averages and pooled

Level-3 standard deviations are computed from the averages of the two runs. Then the level-3 standard deviations are pooled over the five wafers to obtain a standard deviation with 5 degrees of freedom as shown in the table below.

over
wafers

Level-3 standard deviations (ohm.cm) for 5 wafers

Wafer	Probe	Run 1 Average	Run 2 Average	Diff	Stddev	DF
138.	2362.	95.0928	95.1243	-0.0315	0.0223	1
139.	2362.	99.3060	99.3098	-0.0038	0.0027	1
140.	2362.	96.0357	96.0765	-0.0408	0.0289	1
141.	2362.	101.0602	101.0790	-0.0188	0.0133	1
142.	2362.	94.2148	94.2438	-0.0290	0.0205	1
	2362.			Pooled	0.0197	5

[*Graphs of probe biases*](#)

A graphical analysis shows the relative biases among the 5 probes. For each wafer, differences from the wafer average by probe are plotted versus wafer number. The graphs verify that probe #2362 (coded as 5) is biased low relative to the other probes. The bias shows up more strongly after the probes have been in use (run 2).

[*Formulas for computation of biases for probe #2362*](#)

Biases by probe are shown in the following table.

Differences from the mean for each wafer

Wafer	Probe	Run 1	Run 2
138.	1.	0.0248	-0.0119
138.	281.	0.0108	0.0323
138.	283.	0.0193	-0.0258
138.	2062.	-0.0175	0.0561
138.	2362.	-0.0372	-0.0507
139.	1.	-0.0036	-0.0007
139.	281.	0.0394	0.0050
139.	283.	0.0057	0.0239
139.	2062.	-0.0323	0.0373
139.	2362.	-0.0094	-0.0657
140.	1.	0.0400	0.0109
140.	281.	0.0187	0.0106
140.	283.	-0.0201	0.0003
140.	2062.	-0.0126	0.0182
140.	2362.	-0.0261	-0.0398
141.	1.	0.0394	0.0324
141.	281.	-0.0107	-0.0037
141.	283.	0.0246	-0.0191
141.	2062.	-0.0280	0.0436
141.	2362.	-0.0252	-0.0534

```
142.    1.    0.0062    0.0093
142.  281.    0.0376    0.0174
142.  283.   -0.0044    0.0192
142. 2062.   -0.0011    0.0008
142. 2362.   -0.0383   -0.0469
```

How to deal with bias due to the probe

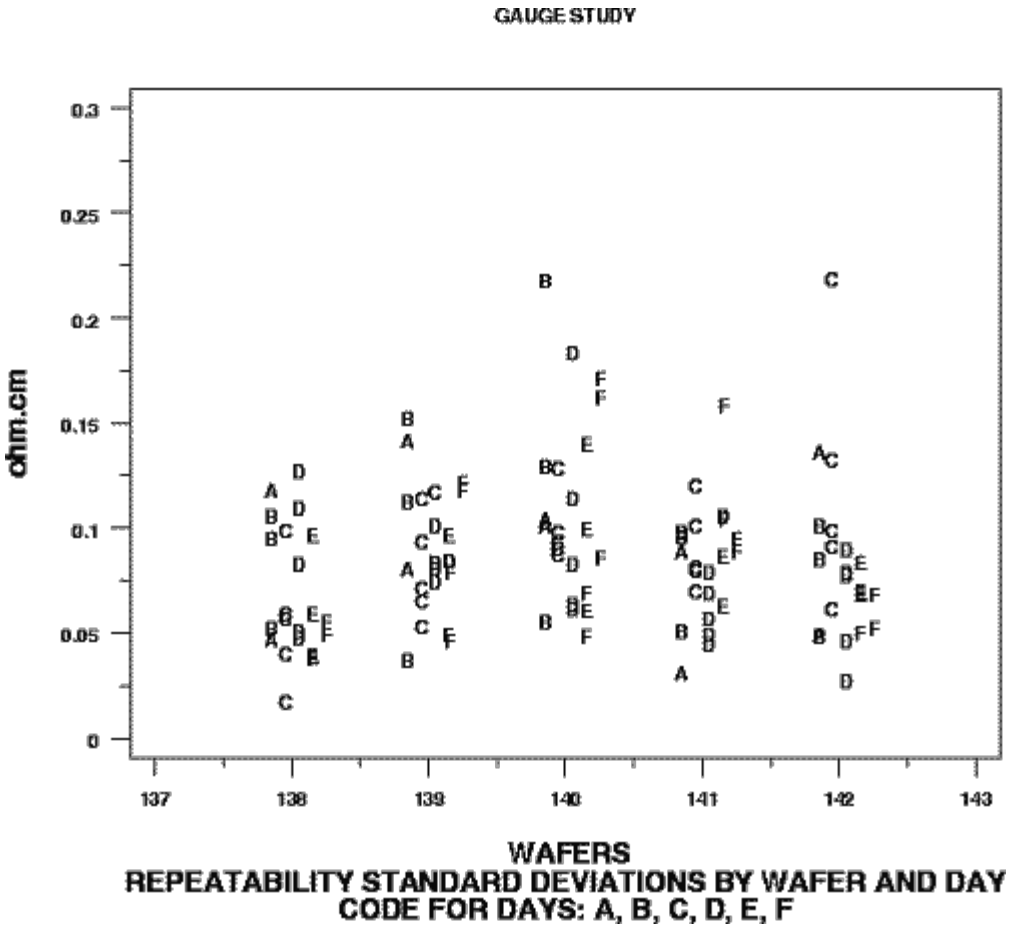
Probe #2362 was chosen for the certification process because of its superior precision, but its bias relative to the other probes creates a problem. There are two possibilities for handling this problem:

1. Correct all measurements made with probe #2362 to the average of the probes.
2. Include the standard deviation for the difference among probes in the uncertainty budget.

The better choice is (1) if we can assume that the probes in the study represent a random sample of probes of this type. This is particularly true when the unit (resistivity) is defined by a test method.

2. Measurement Process Characterization
2.6. Case studies
2.6.1. Gauge study of resistivity probes

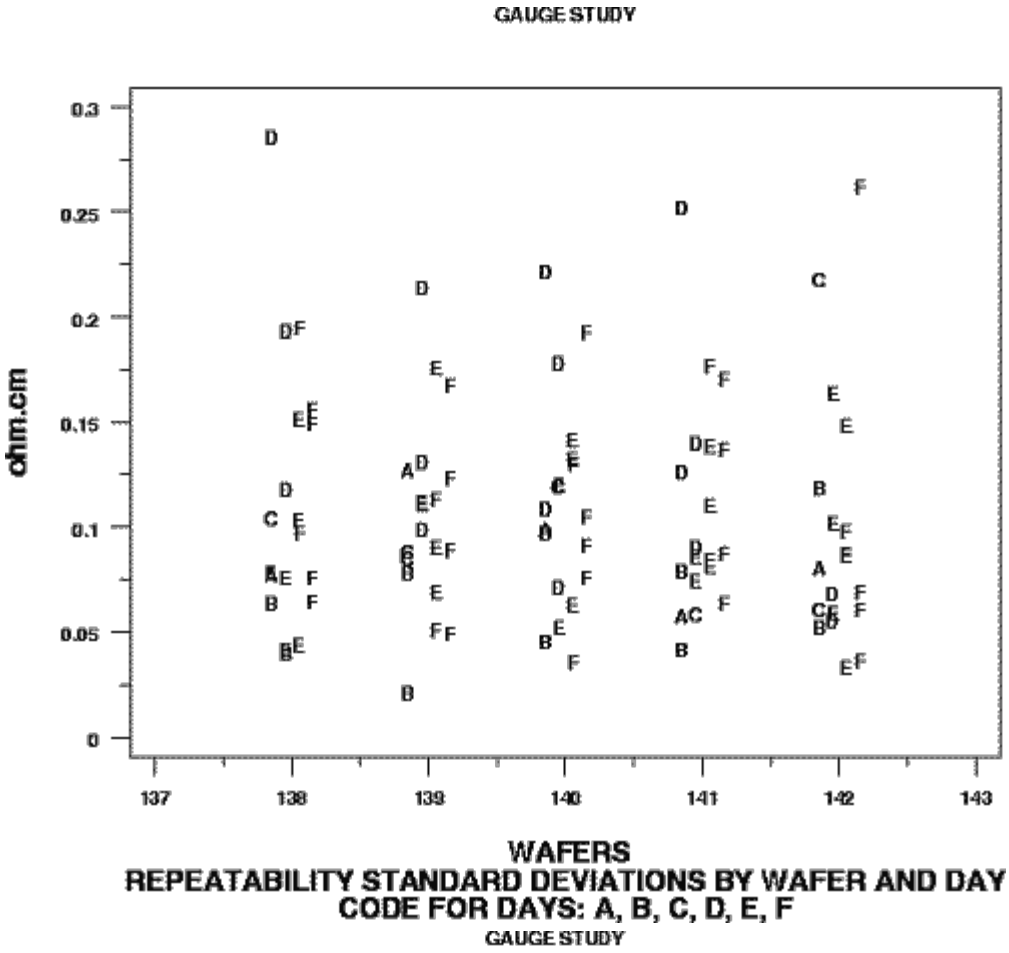
2.6.1.3. Repeatability standard deviations



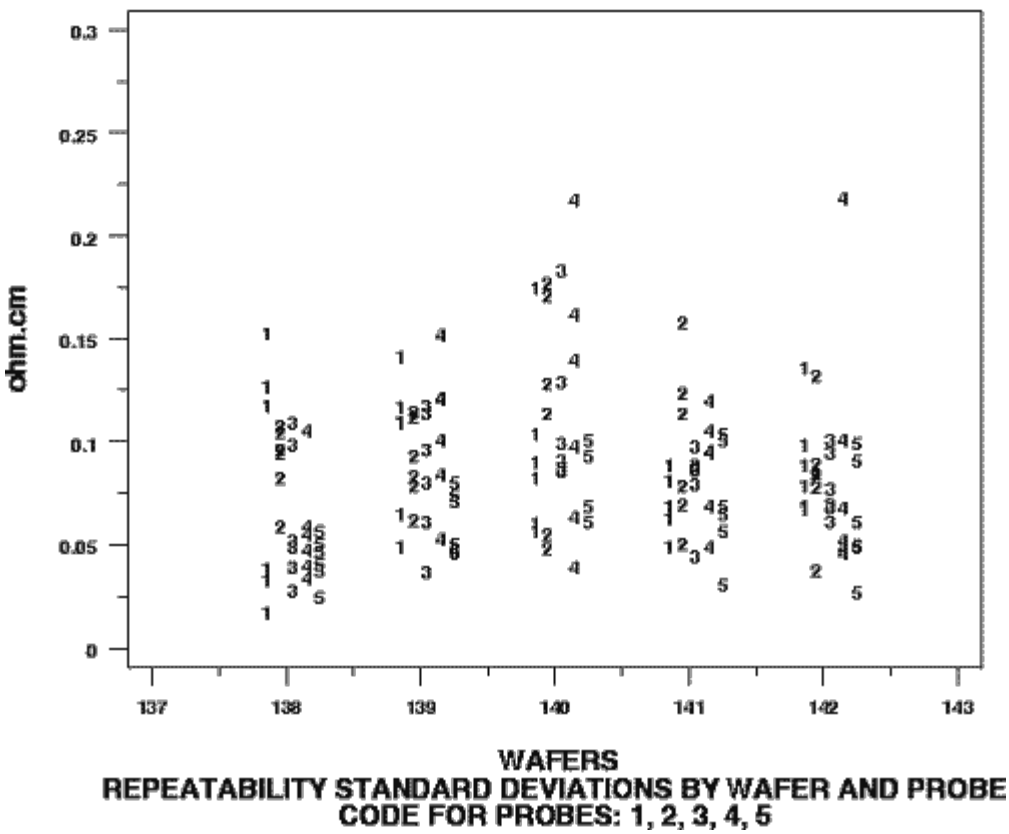
Run 1 - *Graph of repeatability standard deviations for probe #2362 -- 6 days and 5 wafers showing that repeatability is constant across wafers and days*

Run 2 - *Graph of repeatability standard deviations for probe #2362 -- 6 days and 5 wafers showing that*

repeatability is constant across wafers and days

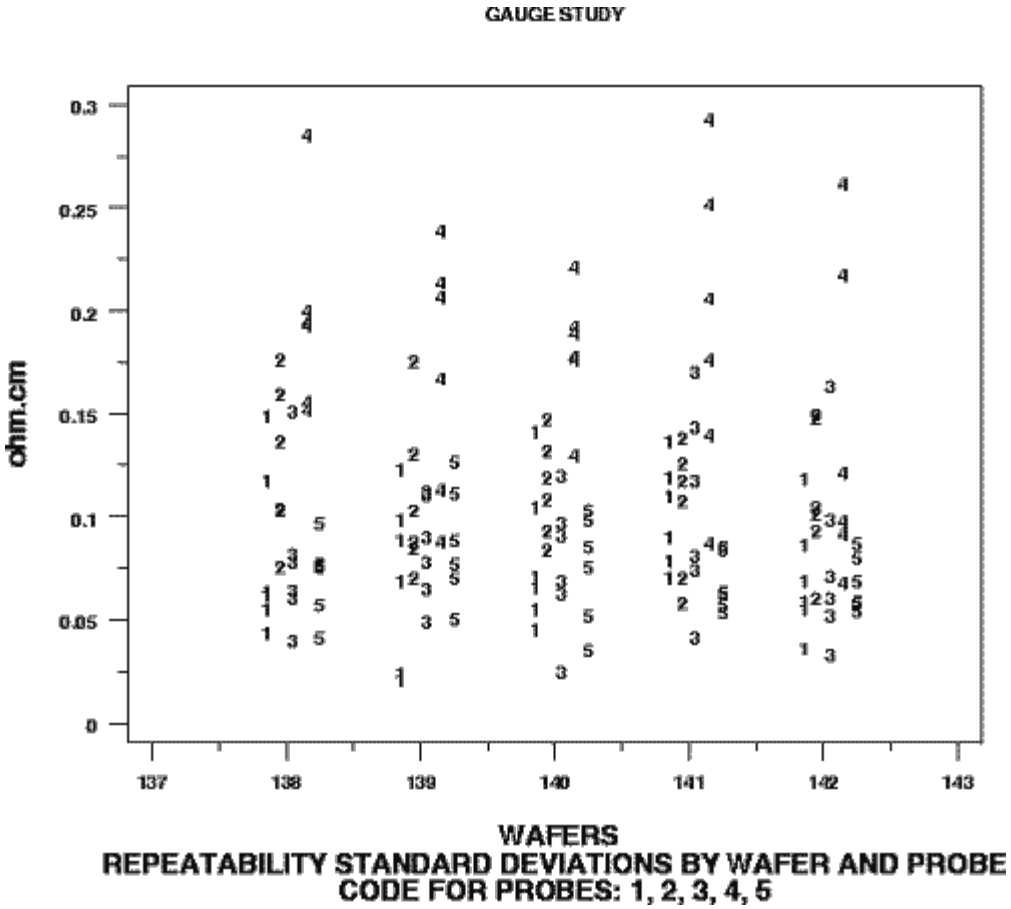


Run 1 *- Graph showing repeatability standard deviations for five probes as a function of wafers and probes*



Symbols for codes: 1 = #1; 2 = #281; 3 = #283; 4 = #2062; 5 = #2362

Run 2 - *Graph showing repeatability standard deviations for 5 probes as a function of wafers and probes*



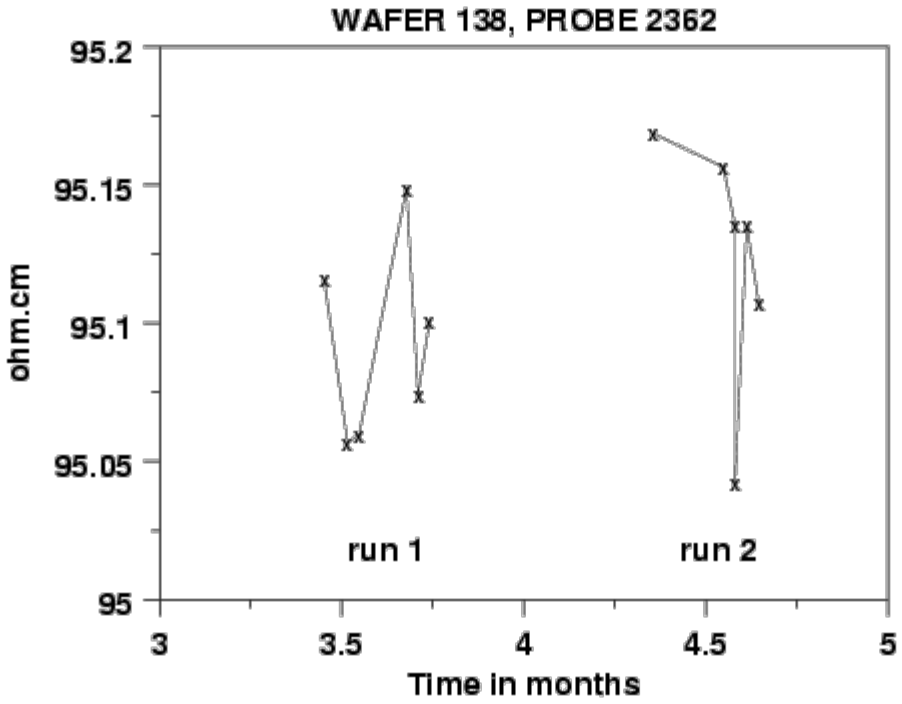
Symbols for probes: 1 = #1; 2 = #281; 3 = #283; 4 = #2062; 5 = #2362

2.6.1.4. Effects of days and long-term stability

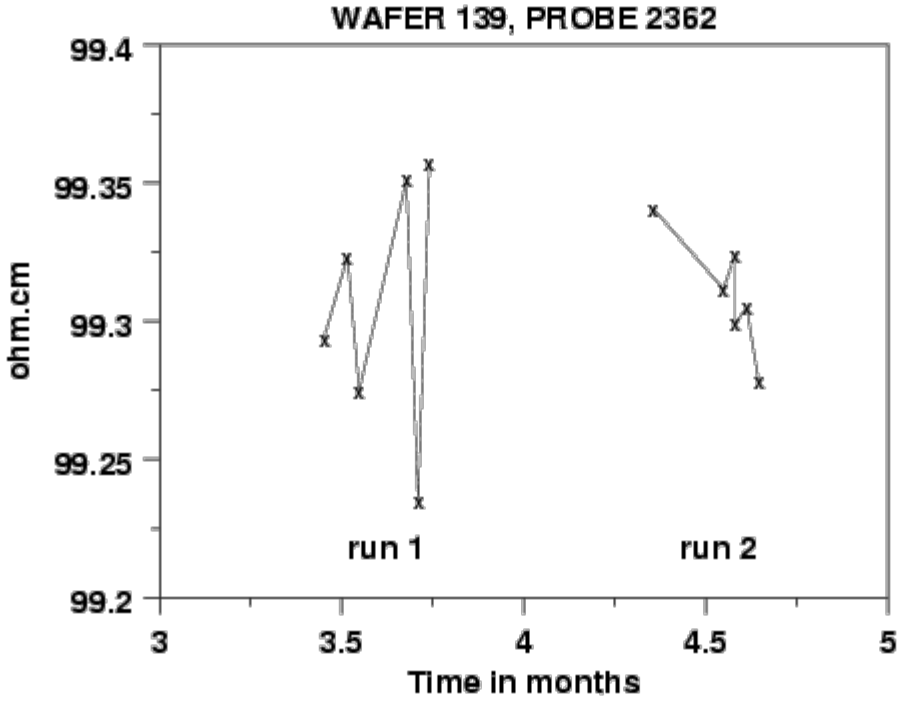
Effects of days and long-term stability on the measurements

The data points that are plotted in the five graphs shown below are averages of resistivity measurements at the center of each wafer for wafers #138, 139, 140, 141, 142. Data for each of two runs are shown on each graph. The six days of measurements for each run are separated by approximately one month and show, with the exception of wafer #139, that there is a very slight shift upwards between run 1 and run 2. The size of the effect is estimated as a level-3 standard deviation in the analysis of the data.

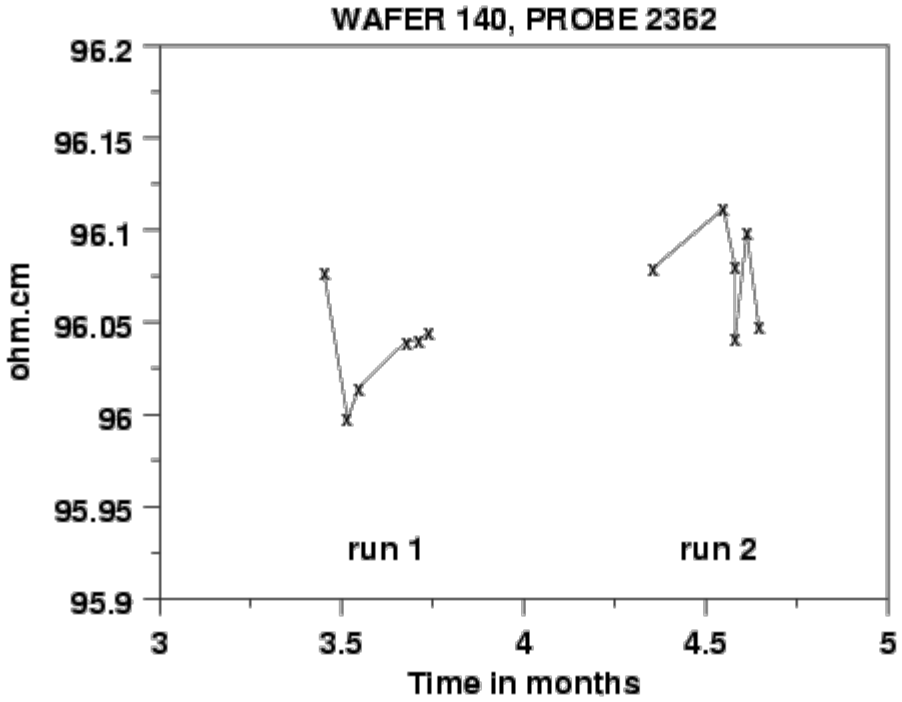
Wafer 138



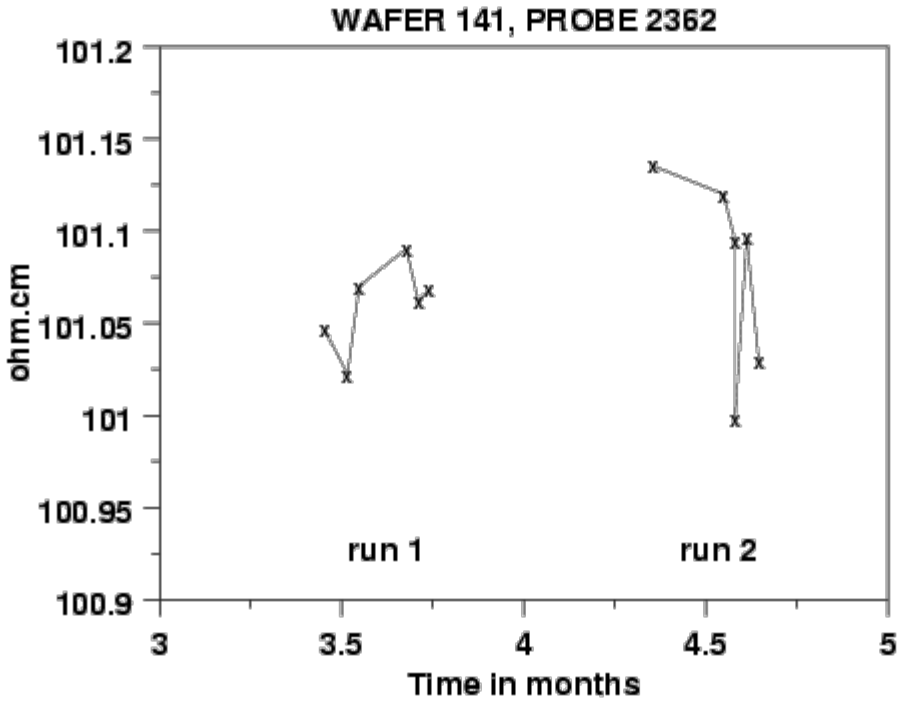
Wafer 139



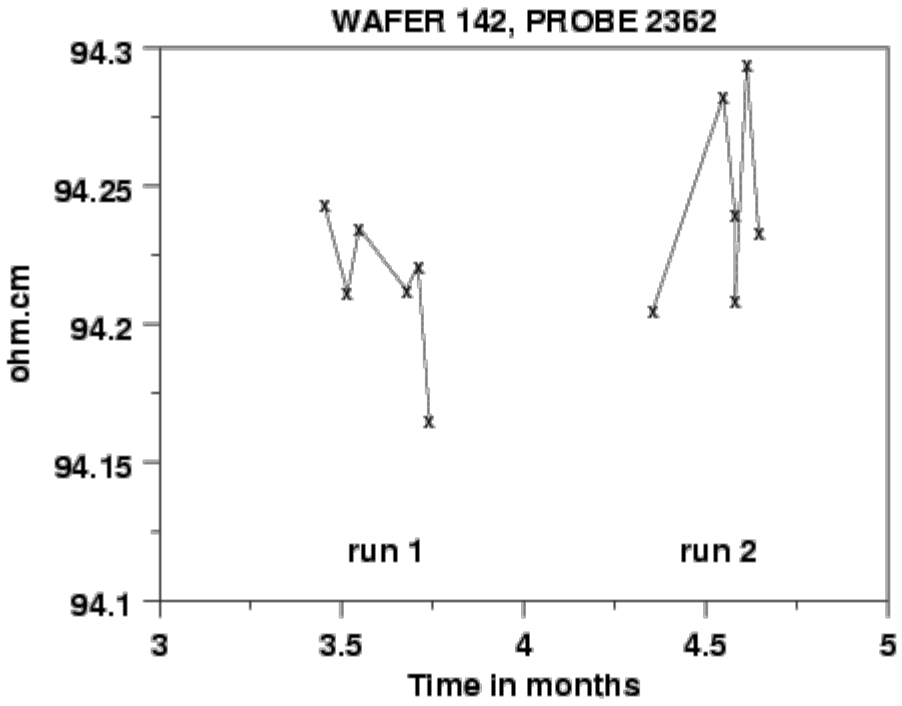
Wafer 140



Wafer 141

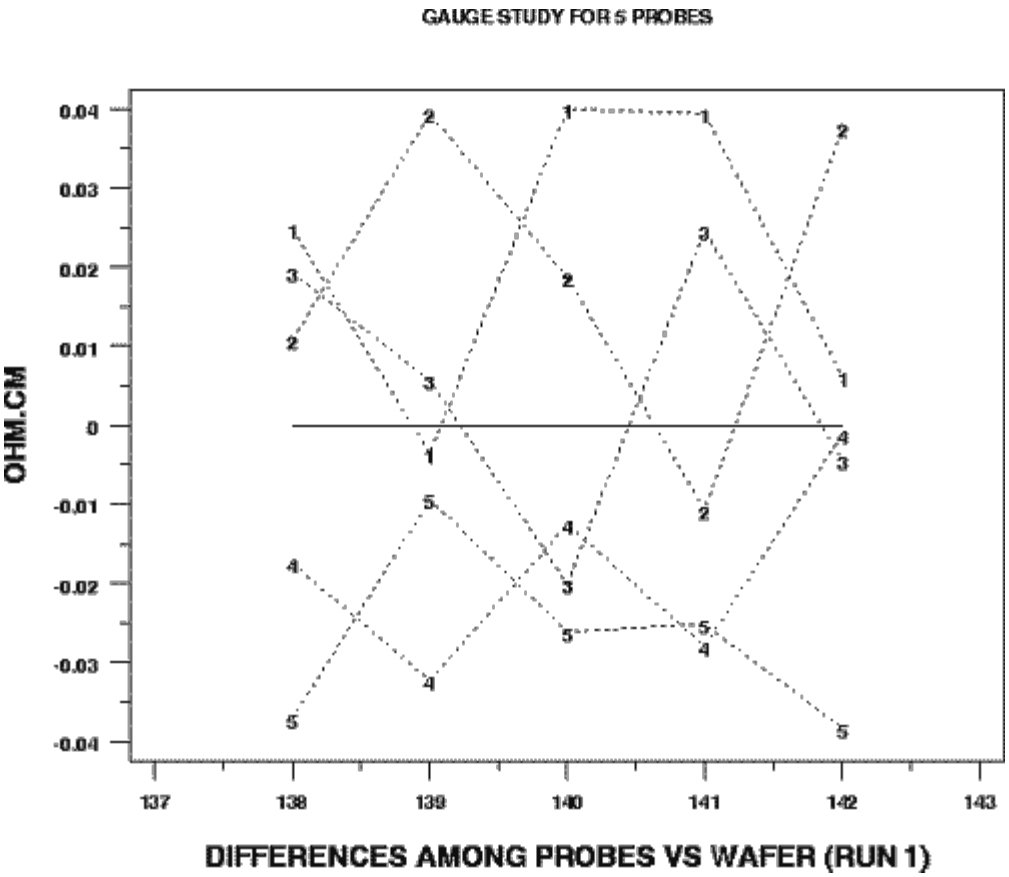

WAFER 141, PROBE 2362
101.2
101.15
101.1
101.05
101
100.95
100.9
ohm.cm
run 1
run 2
3
3.5
4
4.5
5
Time in months

Wafer 142


WAFER 142, PROBE 2362
94.3
94.25
94.2
94.15
94.1
ohm.cm
run 1
run 2
3
3.5
4
4.5
5
Time in months

2.6.1.5. Differences among 5 probes

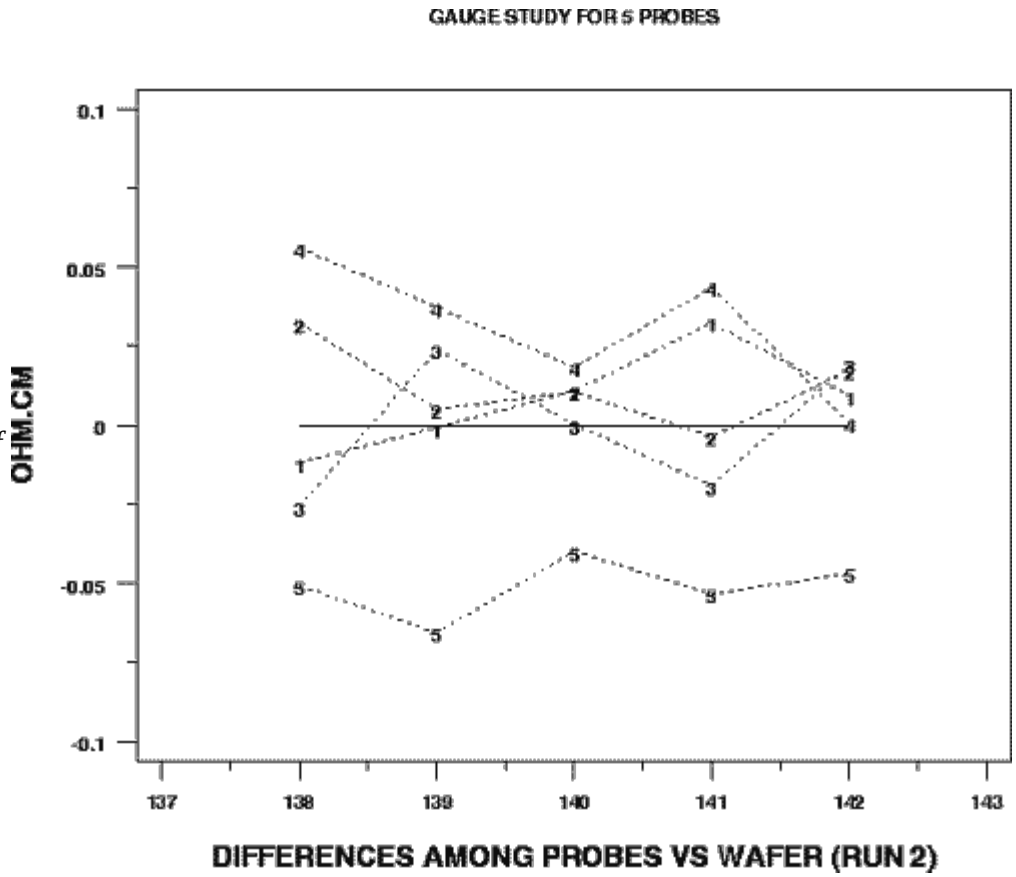
Run 1 - *Graph of differences from wafer averages for each of 5 probes showing that probes #2062 and #2362 are biased low relative to the other probes*



Symbols for probes: 1 = #1; 2 = #281; 3 = #283; 4 = #2062; 5 = #2362

Run 2 - *Graph of differences from wafer averages*

for each of 5 probes showing that probe #2362 continues to be biased low relative to the other probes



Symbols for probes: 1 = #1; 2 = #281; 3 = #283; 4 = #2062; 5 = #2362

2.6.1.6. Run gauge study example using Dataplot

View of Dataplot macros for this case study

This page allows you to repeat the analysis outlined in the case study description on the previous page using Dataplot . It is required that you have already downloaded and installed Dataplot and configured your browser. to run Dataplot. Output from each analysis step below will be displayed in one or more of the Dataplot windows. The four main windows are the Output Window, the Graphics window, the Command History window, and the data sheet window. Across the top of the main windows there are menus for executing Dataplot commands. Across the bottom is a command entry window where commands can be typed in.

Data Analysis Steps	Results and Conclusions
Click on the links below to start Dataplot and run this case study yourself. Each step may use results from previous steps, so please be patient. Wait until the software verifies that the current step is complete before clicking on the next step.	*The links in this column will connect you with more detailed information about each analysis step from the case study description.*
Graphical analyses of variability *Graphs to test for:* 1. Wafer/day effect on repeatability (run 1) 2. Wafer/day effect on repeatability (run 2) 3. Probe effect on repeatability (run 1) 4. Probe effect on repeatability (run 2) 5. Reproducibility and stability	1. and 2. Interpretation: The plots verify that, for both runs, the repeatability of probe #2362 is not dependent on wafers or days, although the standard deviations on days D, E, and F of run 2 are larger in some instances than for the other days. 3. and 4. Interpretation: Probe #2362 appears as #5 in the plots which show that, for both runs, the precision of this probe is better than for the other probes. 5. Interpretation: There is a separate plot for each wafer. The points on the left side of each plot are averages at the wafer center plotted over 5 days; the points on

	the right are the same measurements repeated after one month to check on the stability of the measurement process. The plots show day-to-day variability as well as slight variability from run-to-run.
Table of estimates for probe #2362 1. Level-1 (repeatability) 2. Level-2 (reproducibility) 3. Level-3 (stability)	1., 2. and 3.: Interpretation: The repeatability of the gauge (level-1 standard deviation) dominates the imprecision associated with measurements and days and runs are less important contributors. Of course, even if the gauge has high precision, biases may contribute substantially to the uncertainty of measurement.
Bias estimates 1. Differences among probes - run 1 2. Differences among probes - run 2	1. and 2. Interpretation: The graphs show the relative biases among the 5 probes. For each wafer, differences from the wafer average by probe are plotted versus wafer number. The graphs verify that probe #2362 (coded as 5) is biased low relative to the other probes. The bias shows up more strongly after the probes have been in use (run 2).

2.6.1.7. Dataplot macros

Plot of wafer and day effect on repeatability standard deviations for run 1

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
read mpc61.dat run wafer probe mo day op hum y
sw
y1label ohm.cm
title GAUGE STUDY
lines blank all
let z = pattern 1 2 3 4 5 6 for I = 1 1 300
let z2 = wafer + z/10 -0.25
characters a b c d e f
X1LABEL WAFERS
X2LABEL REPEATABILITY STANDARD DEVIATIONS BY
WAFER AND DAY
X3LABEL CODE FOR DAYS: A, B, C, D, E, F
TITLE RUN 1
plot sw z2 day subset run 1
```

Plot of wafer and day effect on repeatability standard deviations for run 2

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
read mpc61.dat run wafer probe mo day op hum y
sw
y1label ohm.cm
title GAUGE STUDY
lines blank all
let z = pattern 1 2 3 4 5 6 for I = 1 1 300
let z2 = wafer + z/10 -0.25
characters a b c d e f
X1LABEL WAFERS
X2LABEL REPEATABILITY STANDARD DEVIATIONS BY
WAFER AND DAY
X3LABEL CODE FOR DAYS: A, B, C, D, E, F
TITLE RUN 2
plot sw z2 day subset run 2
```

Plot of repeatability standard deviations for 5 probes - run 1

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
read mpc61.dat run wafer probe mo day op hum y
sw
y1label ohm.cm
title GAUGE STUDY
lines blank all
let z = pattern 1 2 3 4 5 6 for I = 1 1 300
let z2 = wafer + z/10 -0.25
characters 1 2 3 4 5
X1LABEL WAFERS
X2LABEL REPEATABILITY STANDARD DEVIATIONS BY
WAFER AND PROBE
X3LABEL CODE FOR PROBES: 1= SRM1; 2= 281; 3=283;
4=2062; 5=2362
TITLE RUN 1
plot sw z2 probe subset run 1
```

Plot of repeatability standard deviations for 5 probes - run 2

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
read mpc61.dat run wafer probe mo day op hum y
sw
y1label ohm.cm
title GAUGE STUDY
lines blank all
let z = pattern 1 2 3 4 5 6 for I = 1 1 300
let z2 = wafer + z/10 -0.25
characters 1 2 3 4 5
X1LABEL WAFERS
X2LABEL REPEATABILITY STANDARD DEVIATIONS BY
WAFER AND PROBE
X3LABEL CODE FOR PROBES: 1= SRM1; 2= 281; 3=283;
4=2062; 5=2362
TITLE RUN 2
plot sw z2 probe subset run 2
```

Plot of differences from the wafer mean for 5 probes - run 1

```
reset data
reset plot control
reset i/o
dimension 500 30
read mpc61a.dat wafer probe d1 d2
let biasrun1 = mean d1 subset probe 2362
print biasrun1
title GAUGE STUDY FOR 5 PROBES
Y1LABEL OHM.CM
lines dotted dotted dotted dotted dotted solid
characters 1 2 3 4 5 blank
xlimits 137 143
let zero = pattern 0 for I = 1 1 30
x1label DIFFERENCES AMONG PROBES VS WAFER (RUN
1)
plot d1 wafer probe and
plot zero wafer
```

Plot of differences from the wafer mean for 5 probes - run 2

```
reset data
reset plot control
reset i/o
dimension 500 30
read mpc61a.dat wafer probe d1 d2
let biasrun2 = mean d2 subset probe 2362
print biasrun2
title GAUGE STUDY FOR 5 PROBES
Y1LABEL OHM.CM
lines dotted dotted dotted dotted dotted solid
characters 1 2 3 4 5 blank
xlimits 137 143
let zero = pattern 0 for I = 1 1 30
x1label DIFFERENCES AMONG PROBES VS WAFER (RUN
2)
plot d2 wafer probe and
plot zero wafer
```

Plot of averages by day showing reproducibility and stability for measurements made with probe #2362 on 5 wafers

```
reset data
reset plot control
reset i/o
dimension 300 50
label size 3
read mcp61b.dat  wafer probe mo1 day1 y1 mo2
day2 y2 diff
let t = mo1+(day1-1)/31.
let t2= mo2+(day2-1)/31.
x3label WAFER 138
multiplot 3 2
plot y1 t subset wafer 138 and
plot y2 t2 subset wafer 138
x3label wafer 139
plot y1 t subset wafer 139 and
plot y2 t2 subset wafer 139
x3label WAFER 140
plot y1 t subset wafer 140 and
plot y2 t2 subset wafer 140
x3label WAFER 140
plot y1 t subset wafer 141 and
plot y2 t2 subset wafer 141
```

```
x3label WAFER 142
plot y1 t subset wafer 142 and
plot y2 t2 subset wafer 142
```

2.6.2. Check standard for resistivity measurements

Purpose The purpose of this page is to outline the analysis of check standard data with respect to controlling the precision and long-term variability of the process.

Outline

1. Background and data
2. Analysis and interpretation
3. Run this example yourself using Dataplot

2.6.2.1. Background and data

Explanation of check standard measurements	The process involves the measurement of resistivity (ohm.cm) of individual silicon wafers cut from a single crystal (# 51939). The wafers were doped with phosphorous to give a nominal resistivity of 100 ohm.cm. A single wafer (#137), chosen at random from a batch of 130 wafers, was designated as the check standard for this process.
Design of data collection and Database	The measurements were carried out according to an ASTM Test Method (F84) with NIST probe #2362. The measurements on the check standard duplicate certification measurements that were being made, during the same time period, on individual wafers from crystal #51939. For the check standard there were: • $J = 6$ repetitions at the center of the wafer on each day • $K = 25$ days The $K = 25$ days cover the time during which the individual wafers were being certified at the National Institute of Standards and Technology.
Software	The analyses used in this case study can be generated using both Dataplot code and R code.

2.6.2.1.1. Database for resistivity check standard

Description of check standard

A single wafer (#137), chosen at random from a batch of 130 wafers, is the check standard for resistivity measurements at the 100 ohm.cm level at the National Institute of Standards and Technology. The average of six measurements at the center of the wafer is the check standard value for one occasion, and the standard deviation of the six measurements is the short-term standard deviation. The columns of the database contain the following:

1. Crystal ID
2. Check standard ID
3. Month
4. Day
5. Hour
6. Minute
7. Operator
8. Humidity
9. Probe ID
10. Temperature
11. Check standard value
12. Short-term standard deviation
13. Degrees of freedom

Database of measurements on check standard

Crystal	Waf	Mo	Da	Hr	Mn	Op	Hum	Probe	Temp	Avg	Stddev	DF
51939	137	03	24	18	01	drr	42	2362	23.003	97.070	0.085	5
51939	137	03	25	12	41	drr	35	2362	23.115	97.049	0.052	5
51939	137	03	25	15	57	drr	33	2362	23.196	97.048	0.038	5
51939	137	03	28	10	10	JMT	47	2362	23.383	97.084	0.036	5
51939	137	03	28	13	31	JMT	44	2362	23.491	97.106	0.049	5
51939	137	03	28	17	33	drr	43	2362	23.352	97.014	0.036	5

```
51939 137 03 29 14 40 drr 36 2362 23.202 97.047
                     0.052 5

51939 137 03 29 16 33 drr 35 2362 23.222 97.078
                     0.117 5

51939 137 03 30 05 45 JMT 32 2362 23.337 97.065
                     0.085 5

51939 137 03 30 09 26 JMT 33 2362 23.321 97.061
                     0.052 5

51939 137 03 25 14 59 drr 34 2362 22.993 97.060
                     0.060 5

51939 137 03 31 10 10 JMT 37 2362 23.164 97.102
                     0.048 5

51939 137 03 31 13 00 JMT 37 2362 23.169 97.096
                     0.026 5

51939 137 03 31 15 32 JMT 35 2362 23.156 97.035
                     0.088 5

51939 137 04 01 13 05 JMT 34 2362 23.097 97.114
                     0.031 5

51939 137 04 01 15 32 JMT 34 2362 23.127 97.069
                     0.037 5

51939 137 04 01 10 32 JMT 48 2362 22.963 97.095
                     0.032 5

51939 137 04 06 14 38 JMT 49 2362 23.454 97.088
                     0.056 5

51939 137 04 07 10 50 JMT 34 2362 23.285 97.079
                     0.067 5

51939 137 04 07 15 46 JMT 33 2362 23.123 97.016
                     0.116 5

51939 137 04 08 09 37 JMT 33 2362 23.373 97.051
                     0.046 5

51939 137 04 08 12 53 JMT 33 2362 23.296 97.070
                     0.078 5

51939 137 04 08 15 03 JMT 33 2362 23.218 97.065
                     0.040 5

51939 137 04 11 09 30 JMT 36 2362 23.415 97.111
                     0.038 5

51939 137 04 11 11 34 JMT 35 2362 23.395 97.073
                     0.039 5
```

2.6.2.2. Analysis and interpretation

Estimates of the repeatability standard deviation and level-2 standard deviation

The level-1 standard deviations (with $J - 1 = 5$ degrees of freedom each) from the database are pooled over the $K = 25$ days to obtain a reliable estimate of repeatability. This pooled value is

$$s_1 = 0.06139 \text{ ohm.cm}$$

with $K(J - 1) = 125$ degrees of freedom. The level-2 standard deviation is computed from the daily averages to be

$$s_2 = 0.02680 \text{ ohm.cm}$$

with $K - 1 = 24$ degrees of freedom.

Relationship to uncertainty calculations

These standard deviations are appropriate for estimating the uncertainty of the average of six measurements on a wafer that is of the same material and construction as the check standard. The computations are explained in the section on sensitivity coefficients for check standard measurements. For other numbers of measurements on the test wafer, the computations are explained in the section on sensitivity coefficients for level-2 designs.

Illustrative table showing computations of repeatability and level-2 standard deviations

A tabular presentation of a subset of check standard data ($J = 6$ repetitions and $K = 6$ days) illustrates the computations. The pooled repeatability standard deviation with $K(J - 1) = 30$ degrees of freedom from this limited database is shown in the next to last row of the table. A level-2 standard deviation with $K - 1 = 5$ degrees of freedom is computed from the center averages and is shown in the last row of the table.

Control chart for probe #2362

The control chart for monitoring the precision of probe #2362 is constructed as discussed in the section on control charts for standard deviations. The upper control limit (UCL) for testing for degradation of the probe is computed using the critical value from the F table with numerator degrees of freedom $J - 1 = 5$ and denominator degrees of freedom $K(J - 1) = 125$. For a 0.05 significance level,

$$F_{0.05,5,125} = 2.29$$

$$UCL = s_1\sqrt{F_{0.05,5,125}} = 0.09238\ ohm \cdot cm$$

Interpretation of control chart for probe #2362

The control chart shows two points exceeding the upper control limit. We expect 5 % of the standard deviations to exceed the UCL for a measurement process that is in-control. Two outliers are not indicative of significant problems with the repeatability for the probe, but the probe should be monitored closely in the future.

Control chart for bias and variability

The control limits for monitoring the bias and long-term variability of resistivity with a Shewhart control chart are given by

$$UCL = Average + 2*s_2 = 97.1234\ ohm.cm$$
$$Centerline = Average = 97.0698\ ohm.cm$$
$$LCL = Average - 2*s_2 = 97.0162\ ohm.cm$$

Interpretation of control chart for bias

The control chart shows that the points scatter randomly about the center line with no serious problems, although one point exceeds the upper control limit and one point exceeds the lower control limit by a small amount. The conclusion is that there is:

- No evidence of bias, change or drift in the measurement process.
- No evidence of long-term lack of control.

Future measurements that exceed the control limits must be evaluated for long-term changes in bias and/or variability.

2.6.2.2.1. Repeatability and level-2 standard deviations

Example

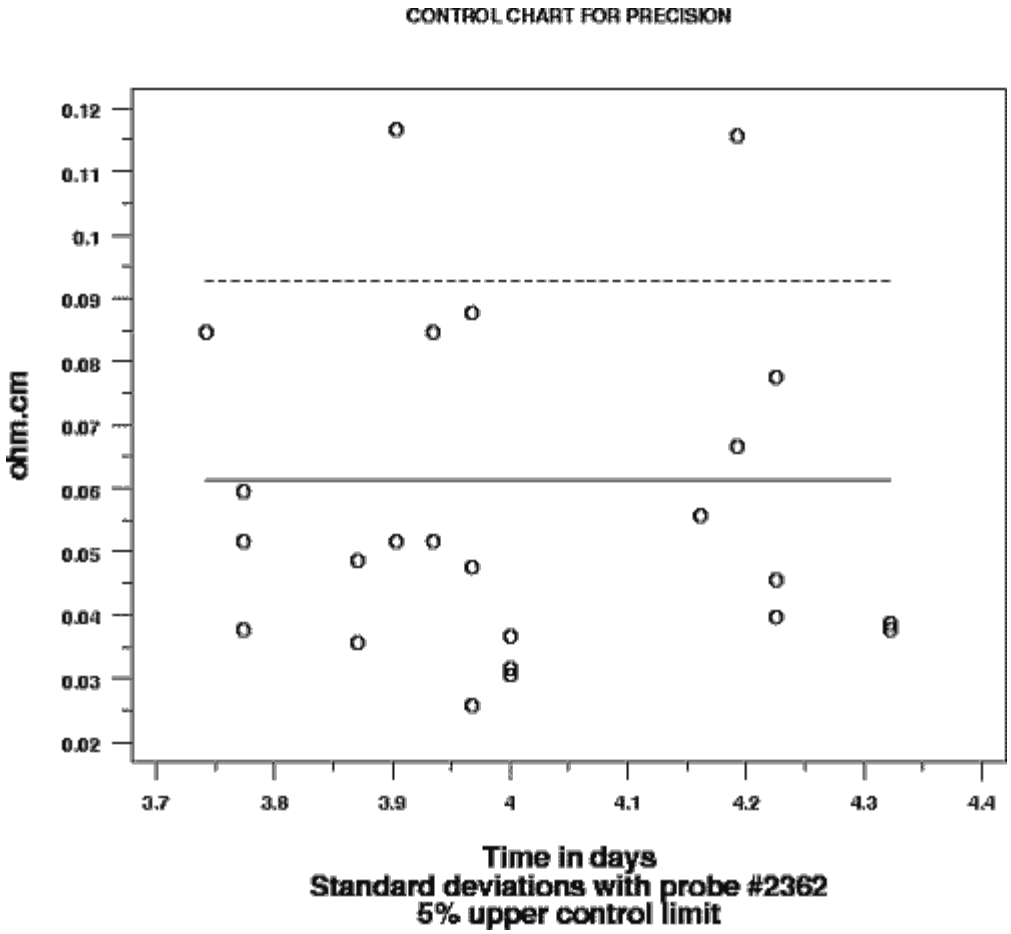
The table below illustrates the computation of repeatability and level-2 standard deviations from measurements on a check standard. The check standard measurements are resistivities at the center of a 100 ohm.cm wafer. There are $J = 6$ repetitions per day and $K = 5$ days for this example.

Table of data, averages, and repeatability standard deviations

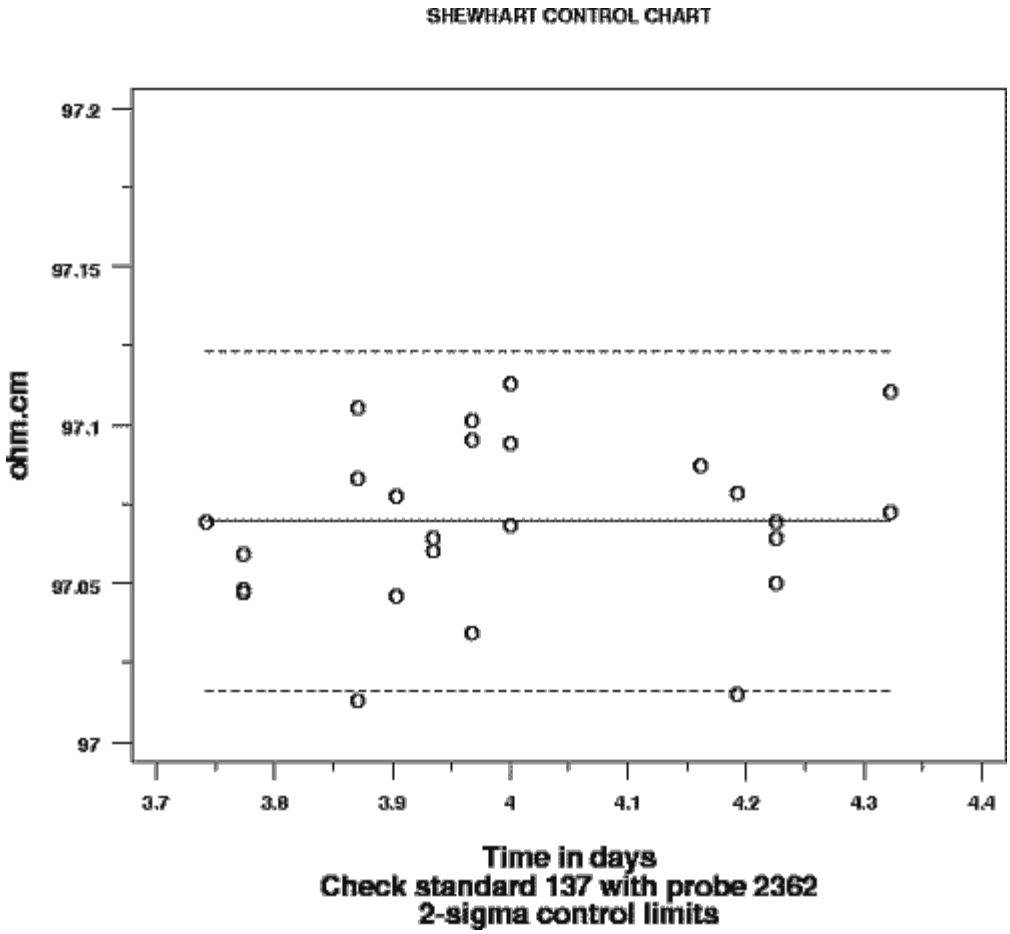
Measurements on check standard #137 Repetitions per day						
Days	**1**	**2**	**3**	**4**	**5**	**6**
1	96.920	97.054	97.057	97.035	97.189	96.965
2	97.118	96.947	97.110	97.047	96.945	97.013
3	97.034	97.084	97.023	97.045	97.061	97.074
4	97.047	97.099	97.087	97.076	97.117	97.070
5	97.127	97.067	97.106	96.995	97.052	97.121
6	96.995	96.984	97.053	97.065	96.976	96.997
Averages	97.040	97.039	97.073	97.044	97.057	97.037
Repeatability Standard Deviations	0.0777	0.0602	0.0341	0.0281	0.0896	0.0614
Pooled Repeatability Standard Deviation	**0.0625 30 df**					
Level-2 Standard Deviation	**0.0139 5 df**					

2.6.2.3. Control chart for probe precision

Control chart for probe #2362 showing violations of the control limits -- all standard deviations are based on 6 repetitions and the control limits are 95% limits



2.6.2.4. Control chart for bias and long-term variability



Shewhart control chart for measurements on a resistivity check standard showing that the process is in-control -- all measurements are averages of 6 repetitions

2.6.2.5. Run check standard example yourself

View of Dataplot macros for this case study

This page allows you to repeat the analysis outlined in the case study description on the previous page using Dataplot. It is required that you have already downloaded and installed Dataplot and configured your browser to run Dataplot. Output from each analysis step below will be displayed in one or more of the Dataplot windows. The four main windows are the Output Window, the Graphics window, the Command History window, and the data sheet window. Across the top of the main windows there are menus for executing Dataplot commands. Across the bottom is a command entry window where commands can be typed in.

Data Analysis Steps	Results and Conclusions
Click on the links below to start Dataplot and run this case study yourself. Each step may use results from previous steps, so please be patient. Wait until the software verifies that the current step is complete before clicking on the next step.	*The links in this column will connect you with more detailed information about each analysis step from the case study description.*
Graphical tests of assumptions Histogram Normal probability plot	The histogram and normal probability plots show no evidence of non-normality.
Control chart for precision Control chart for probe #2362 Computations: 1. Pooled repeatability standard deviation 2. Control limit	The precision control chart shows two points exceeding the upper control limit. We expect 5% of the standard deviations to exceed the UCL even when the measurement process is in-control.
Control chart for check standard Control chart for check standard #137	The Shewhart control chart shows that the points scatter randomly about the center

Computations: 1. Average check standard value 2. Process standard deviation 3. Upper and lower control limits	line with no serious problems, although one point exceeds the upper control limit and one point exceeds the lower control limit by a small amount. The conclusion is that there is no evidence of bias or lack of long-term control.

2.6.2.6. Dataplot macros

Histogram for check standard #137 to test assumption of normality

```
reset data
reset plot control
reset i/o
dimension 500 30
skip 14
read mpc62.dat crystal wafer mo day hour min op
hum probe temp y sw df
histogram y
```

Normal probability plot for check standard #137 to test assumption of normality

```
reset data
reset plot control
reset i/o
dimension 500 30
skip 14
read mpc62.dat crystal wafer mo day hour min op
hum probe temp y sw df
normal probabilty plot y
```

Control chart for precision of probe #2372 and computation of control parameter estimates

```
reset data
reset plot control
reset i/o
dimension 500 30
skip 14
read mpc62.dat crystal wafer mo day hour min op
hum probe temp y sw df
let time = mo +(day-1)/31.
let s = sw*sw
let spool = mean s
let spool = spool**.5
print spool
let f = fppf(.95, 5, 125)
let ucl = spool*(f)**.5
print ucl
title Control chart for precision
characters blank blank O
lines solid dashed blank
y1label ohm.cm
x1label Time in days
x2label Standard deviations with probe #2362
x3label 5% upper control limit
let center = sw - sw + spool
let cl = sw - sw + ucl
plot center cl sw vs time
```

Shewhart control chart for check standard #137 with computation of control

```
reset data
reset plot control
reset i/o
dimension 500 30
skip 14
read mpc62.dat crystal wafer mo day hour min op
hum probe temp y sw df
let time = mo +(day-1)/31.
let avg = mean y
let sprocess = standard deviation y
let ucl = avg + 2*sprocess
let lcl = avg - 2*sprocess
```

chart parameters

```
print avg
print sprocess
print ucl lcl
title Shewhart control chart
characters O blank blank blank
lines blank dashed solid dashed
y1label ohm.cm
x1label Time in days
x2label Check standard 137 with probe 2362
x3label 2-sigma control limits
let ybar = y - y + avg
let lc1 = y - y + lcl
let lc2 = y - y + ucl
plot y lc1 ybar lc2 vs time
```

2.6.3. Evaluation of type A uncertainty

Purpose The purpose of this case study is to demonstrate the computation of uncertainty for a measurement process with several sources of uncertainty from data taken during a gauge study.

Outline

1. Background and data for the study
2. Graphical and quantitative analyses and interpretations
3. Run this example yourself with Dataplot

2.6.3.1. Background and data

Description of measurements	The measurements in question are resistivities (ohm.cm) of silicon wafers. The intent is to calculate an uncertainty associated with the resistivity measurements of approximately 100 silicon wafers that were certified with probe #2362 in wiring configuration A, according to ASTM Method F84 (ASTM F84) which is the defined reference for this measurement. The reported value for each wafer is the average of six measurements made at the center of the wafer on a single day. Probe #2362 is one of five probes owned by the National Institute of Standards and Technology that is capable of making the measurements.
Sources of uncertainty in NIST measurements	The uncertainty analysis takes into account the following sources of variability: • Repeatability of measurements at the center of the wafer • Day-to-day effects • Run-to-run effects • Bias due to probe #2362 • Bias due to wiring configuration
Database of 3-level nested design -- for estimating time-dependent sources of uncertainty	The certification measurements themselves are not the primary source for estimating uncertainty components because they do not yield information on day-to-day effects and long-term effects. The standard deviations for the three time-dependent sources of uncertainty are estimated from a 3-level nested design. The design was replicated on each of $Q = 5$ wafers which were chosen at random, for this purpose, from the lot of wafers. The certification measurements were made between the two runs in order to check on the long-term stability of the process. The data consist of repeatability standard deviations (with $J - 1 = 5$ degrees of freedom each) from measurements at the wafer center.
Software	The analyses used in this case study can be generated using both Dataplot code and R code.

2.6.3.1.1. Database of resistivity measurements

Check standards are five wafers chosen at random from a batch of wafers

Measurements of resistivity (ohm.cm) were made according to an ASTM Standard Test Method (F4) at the National Institute of Standards and Technology to assess the sources of uncertainty in the measurement system. The gauges for the study were five probes owned by NIST; the check standards for the study were five wafers selected at random from a batch of wafers cut from one silicon crystal doped with phosphorous to give a nominal resistivity of 100 ohm.cm.

Measurements on the check standards are used to estimate repeatability, day effect, run effect

The effect of operator was not considered to be significant for this study. Averages and standard deviations from $J = 6$ measurements at the center of each wafer are shown in the table.

- $J = 6$ measurements at the center of the wafer per day
- $K = 6$ days (one operator) per repetition
- $L = 2$ runs (complete)
- $Q = 5$ wafers (check standards 138, 139, 140, 141, 142)
- $I = 5$ probes (1, 281, 283, 2062, 2362)

Run	Wafer	Probe	Month	Day	Operator	Temp	Average	Standard Deviation
1	138.	1.	3.	15.	1.	22.98	95.1772	0.1191
1	138.	1.	3.	17.	1.	23.02	95.1567	0.0183
1	138.	1.	3.	18.	1.	22.79	95.1937	0.1282
1	138.	1.	3.	21.	1.	23.17	95.1959	0.0398
1	138.	1.	3.	23.	2.	23.25	95.1442	0.0346
1	138.	1.	3.	23.	1.	23.20	95.0610	0.1539
1	138.	281.	3.	16.	1.	22.99	95.1591	0.0963
1	138.	281.	3.	17.	1.	22.97	95.1195	0.0606
1	138.	281.	3.	18.	1.	22.83	95.1065	0.0842
1	138.	281.	3.	21.	1.	23.28	95.0925	0.0973

```
1  138.    281.      3.     23.      2.   23.14   95.1990
0.1062
1  138.    281.      3.     23.      1.   23.16   95.1682
0.1090
1  138.    283.      3.     16.      1.   22.95   95.1252
0.0531
1  138.    283.      3.     17.      1.   23.08   95.1600
0.0998
1  138.    283.      3.     18.      1.   23.13   95.0818
0.1108
1  138.    283.      3.     21.      1.   23.28   95.1620
0.0408
1  138.    283.      3.     22.      1.   23.36   95.1735
0.0501
1  138.    283.      3.     24.      2.   22.97   95.1932
0.0287
1  138.   2062.      3.     16.      1.   22.97   95.1311
0.1066
1  138.   2062.      3.     17.      1.   22.98   95.1132
0.0415
1  138.   2062.      3.     18.      1.   23.16   95.0432
0.0491
1  138.   2062.      3.     21.      1.   23.16   95.1254
0.0603
1  138.   2062.      3.     22.      1.   23.28   95.1322
0.0561
1  138.   2062.      3.     24.      2.   23.19   95.1299
0.0349
1  138.   2362.      3.     15.      1.   23.08   95.1162
0.0480
1  138.   2362.      3.     17.      1.   23.01   95.0569
0.0577
1  138.   2362.      3.     18.      1.   22.97   95.0598
0.0516
1  138.   2362.      3.     22.      1.   23.23   95.1487
0.0386
1  138.   2362.      3.     23.      2.   23.28   95.0743
0.0256
1  138.   2362.      3.     24.      2.   23.10   95.1010
0.0420
1  139.      1.      3.     15.      1.   23.01   99.3528
0.1424
1  139.      1.      3.     17.      1.   23.00   99.2940
0.0660
1  139.      1.      3.     17.      1.   23.01   99.2340
0.1179
1  139.      1.      3.     21.      1.   23.20   99.3489
0.0506
1  139.      1.      3.     23.      2.   23.22   99.2625
0.1111
1  139.      1.      3.     23.      1.   23.22   99.3787
0.1103
1  139.    281.      3.     16.      1.   22.95   99.3244
0.1134
1  139.    281.      3.     17.      1.   22.98   99.3378
0.0949
1  139.    281.      3.     18.      1.   22.86   99.3424
0.0847
1  139.    281.      3.     22.      1.   23.17   99.4033
0.0801
1  139.    281.      3.     23.      2.   23.10   99.3717
0.0630
1  139.    281.      3.     23.      1.   23.14   99.3493
0.1157
1  139.    283.      3.     16.      1.   22.94   99.3065
0.0381
1  139.    283.      3.     17.      1.   23.09   99.3280
0.1153
1  139.    283.      3.     18.      1.   23.11   99.3000
0.0818
1  139.    283.      3.     21.      1.   23.25   99.3347
0.0972
1  139.    283.      3.     22.      1.   23.36   99.3929
0.1189
1  139.    283.      3.     23.      1.   23.18   99.2644
0.0622
1  139.   2062.      3.     16.      1.   22.94   99.3324
0.1531
1  139.   2062.      3.     17.      1.   23.08   99.3254
0.0543
1  139.   2062.      3.     18.      1.   23.15   99.2555
```

```
0.1024
1  139.   2062.      3.     18.      1.   23.18   99.1946
0.0851
1  139.   2062.      3.     22.      1.   23.27   99.3542
0.1227
1  139.   2062.      3.     24.      2.   23.23   99.2365
0.1218
1  139.   2362.      3.     15.      1.   23.08   99.2939
0.0818
1  139.   2362.      3.     17.      1.   23.02   99.3234
0.0723
1  139.   2362.      3.     18.      1.   22.93   99.2748
0.0756
1  139.   2362.      3.     22.      1.   23.29   99.3512
0.0475
1  139.   2362.      3.     23.      2.   23.25   99.2350
0.0517
1  139.   2362.      3.     24.      2.   23.05   99.3574
0.0485
1  140.      1.      3.     15.      1.   23.07   96.1334
0.1052
1  140.      1.      3.     17.      1.   23.08   96.1250
0.0916
1  140.      1.      3.     18.      1.   22.77   96.0665
0.0836
1  140.      1.      3.     21.      1.   23.18   96.0725
0.0620
1  140.      1.      3.     23.      2.   23.20   96.1006
0.0582
1  140.      1.      3.     23.      1.   23.21   96.1131
0.1757
1  140.    281.      3.     16.      1.   22.94   96.0467
0.0565
1  140.    281.      3.     17.      1.   22.99   96.1081
0.1293
1  140.    281.      3.     18.      1.   22.91   96.0578
0.1148
1  140.    281.      3.     22.      1.   23.15   96.0700
0.0495
1  140.    281.      3.     22.      1.   23.33   96.1052
0.1722
1  140.    281.      3.     23.      1.   23.19   96.0952
0.1786
1  140.    283.      3.     16.      1.   22.89   96.0650
0.1301
1  140.    283.      3.     17.      1.   23.07   96.0870
0.0881
1  140.    283.      3.     18.      1.   23.07   95.8906
0.1842
1  140.    283.      3.     21.      1.   23.24   96.0842
0.1008
1  140.    283.      3.     22.      1.   23.34   96.0189
0.0865
1  140.    283.      3.     23.      1.   23.19   96.1047
0.0923
1  140.   2062.      3.     16.      1.   22.95   96.0379
0.2190
1  140.   2062.      3.     17.      1.   22.97   96.0671
0.0991
1  140.   2062.      3.     18.      1.   23.15   96.0206
0.0648
1  140.   2062.      3.     21.      1.   23.14   96.0207
0.1410
1  140.   2062.      3.     22.      1.   23.32   96.0587
0.1634
1  140.   2062.      3.     24.      2.   23.17   96.0903
0.0406
1  140.   2362.      3.     15.      1.   23.08   96.0771
0.1024
1  140.   2362.      3.     17.      1.   23.00   95.9976
0.0943
1  140.   2362.      3.     18.      1.   23.01   96.0148
0.0622
1  140.   2362.      3.     22.      1.   23.27   96.0397
0.0702
1  140.   2362.      3.     23.      2.   23.24   96.0407
0.0627
1  140.   2362.      3.     24.      2.   23.13   96.0445
0.0622
1  141.      1.      3.     15.      1.   23.01  101.2124
0.0900
```

```
1  141.      1.      3.     17.      1.   23.08  101.1018
0.0820
1  141.      1.      3.     18.      1.   22.75  101.1119
0.0500
1  141.      1.      3.     21.      1.   23.21  101.1072
0.0641
1  141.      1.      3.     23.      2.   23.25  101.0802
0.0704
1  141.      1.      3.     23.      1.   23.19  101.1350
0.0699
1  141.    281.      3.     16.      1.   22.93  101.0287
0.0520
1  141.    281.      3.     17.      1.   23.00  101.0131
0.0710
1  141.    281.      3.     18.      1.   22.90  101.1329
0.0800
1  141.    281.      3.     22.      1.   23.19  101.0562
0.1594
1  141.    281.      3.     23.      2.   23.18  101.0891
0.1252
1  141.    281.      3.     23.      1.   23.17  101.1283
0.1151
1  141.    283.      3.     16.      1.   22.85  101.1597
0.0990
1  141.    283.      3.     17.      1.   23.09  101.0784
0.0810
1  141.    283.      3.     18.      1.   23.08  101.0715
0.0460
1  141.    283.      3.     21.      1.   23.27  101.0910
0.0880
1  141.    283.      3.     22.      1.   23.34  101.0967
0.0901
1  141.    283.      3.     24.      2.   23.00  101.1627
0.0888
1  141.   2062.      3.     16.      1.   22.97  101.1077
0.0970
1  141.   2062.      3.     17.      1.   22.96  101.0245
0.1210
1  141.   2062.      3.     18.      1.   23.19  100.9650
0.0700
1  141.   2062.      3.     18.      1.   23.18  101.0319
0.1070
1  141.   2062.      3.     22.      1.   23.34  101.0849
0.0960
1  141.   2062.      3.     24.      2.   23.21  101.1302
0.0505
1  141.   2362.      3.     15.      1.   23.08  101.0471
0.0320
1  141.   2362.      3.     17.      1.   23.01  101.0224
0.1020
1  141.   2362.      3.     18.      1.   23.05  101.0702
0.0580
1  141.   2362.      3.     22.      1.   23.22  101.0904
0.1049
1  141.   2362.      3.     23.      2.   23.29  101.0626
0.0702
1  141.   2362.      3.     24.      2.   23.15  101.0686
0.0661
1  142.      1.      3.     15.      1.   23.02   94.3160
0.1372
1  142.      1.      3.     17.      1.   23.04   94.2808
0.0999
1  142.      1.      3.     18.      1.   22.73   94.2478
0.0803
1  142.      1.      3.     21.      1.   23.19   94.2862
0.0700
1  142.      1.      3.     23.      2.   23.25   94.1859
0.0899
1  142.      1.      3.     23.      1.   23.21   94.2389
0.0686
1  142.    281.      3.     16.      1.   22.98   94.2640
0.0862
1  142.    281.      3.     17.      1.   23.00   94.3333
0.1330
1  142.    281.      3.     18.      1.   22.88   94.2994
0.0908
1  142.    281.      3.     21.      1.   23.28   94.2873
0.0846
1  142.    281.      3.     23.      2.   23.07   94.2576
0.0795
1  142.    281.      3.     23.      1.   23.12   94.3027
```

```
0.0389
1  142.    283.      3.     16.      1.   22.92   94.2846
0.1021
1  142.    283.      3.     17.      1.   23.08   94.2197
0.0627
1  142.    283.      3.     18.      1.   23.09   94.2119
0.0785
1  142.    283.      3.     21.      1.   23.29   94.2536
0.0712
1  142.    283.      3.     22.      1.   23.34   94.2280
0.0692
1  142.    283.      3.     24.      2.   22.92   94.2944
0.0958
1  142.   2062.      3.     16.      1.   22.96   94.2238
0.0492
1  142.   2062.      3.     17.      1.   22.95   94.3061
0.2194
1  142.   2062.      3.     18.      1.   23.16   94.1868
0.0474
1  142.   2062.      3.     21.      1.   23.11   94.2645
0.0697
1  142.   2062.      3.     22.      1.   23.31   94.3101
0.0532
1  142.   2062.      3.     24.      2.   23.24   94.2204
0.1023
1  142.   2362.      3.     15.      1.   23.08   94.2437
0.0503
1  142.   2362.      3.     17.      1.   23.00   94.2115
0.0919
1  142.   2362.      3.     18.      1.   22.99   94.2348
0.0282
1  142.   2362.      3.     22.      1.   23.26   94.2124
0.0513
1  142.   2362.      3.     23.      2.   23.27   94.2214
0.0627
1  142.   2362.      3.     24.      2.   23.08   94.1651
0.1010
2  138.      1.      4.     13.      1.   23.12   95.1996
0.0645
2  138.      1.      4.     15.      1.   22.73   95.1315
0.1192
2  138.      1.      4.     18.      2.   22.76   95.1845
0.0452
2  138.      1.      4.     19.      1.   22.73   95.1359
0.1498
2  138.      1.      4.     20.      2.   22.73   95.1435
0.0629
2  138.      1.      4.     21.      2.   22.93   95.1839
0.0563
2  138.    281.      4.     14.      2.   22.46   95.2106
0.1049
2  138.    281.      4.     18.      2.   22.80   95.2505
0.0771
2  138.    281.      4.     18.      2.   22.77   95.2648
0.1046
2  138.    281.      4.     20.      2.   22.80   95.2197
0.1779
2  138.    281.      4.     20.      2.   22.87   95.2003
0.1376
2  138.    281.      4.     21.      2.   22.95   95.0982
0.1611
2  138.    283.      4.     18.      2.   22.83   95.1211
0.0794
2  138.    283.      4.     13.      1.   23.17   95.1327
0.0409
2  138.    283.      4.     18.      1.   22.67   95.2053
0.1525
2  138.    283.      4.     19.      2.   23.00   95.1292
0.0655
2  138.    283.      4.     21.      2.   22.91   95.1669
0.0619
2  138.    283.      4.     21.      2.   22.96   95.1401
0.0831
2  138.   2062.      4.     15.      1.   22.64   95.2479
0.2867
2  138.   2062.      4.     15.      1.   22.67   95.2224
0.1945
2  138.   2062.      4.     19.      2.   22.99   95.2810
0.1960
2  138.   2062.      4.     19.      1.   22.75   95.1869
0.1571
```

```
2  138.   2062.      4.     21.      2.   22.84   95.3053
0.2012
2  138.   2062.      4.     21.      2.   22.92   95.1432
0.1532
2  138.   2362.      4.     12.      1.   22.74   95.1687
0.0785
2  138.   2362.      4.     18.      2.   22.75   95.1564
0.0430
2  138.   2362.      4.     19.      2.   22.88   95.1354
0.0983
2  138.   2362.      4.     19.      1.   22.73   95.0422
0.0773
2  138.   2362.      4.     20.      2.   22.86   95.1354
0.0587
2  138.   2362.      4.     21.      2.   22.94   95.1075
0.0776
2  139.      1.      4.     13.      2.   23.14   99.3274
0.0220
2  139.      1.      4.     15.      2.   22.77   99.5020
0.0997
2  139.      1.      4.     18.      2.   22.80   99.4016
0.0704
2  139.      1.      4.     19.      1.   22.68   99.3181
0.1245
2  139.      1.      4.     20.      2.   22.78   99.3858
0.0903
2  139.      1.      4.     21.      2.   22.93   99.3141
0.0255
2  139.    281.      4.     14.      2.   23.05   99.2915
0.0859
2  139.    281.      4.     15.      2.   22.71   99.4032
0.1322
2  139.    281.      4.     18.      2.   22.79   99.4612
0.1765
2  139.    281.      4.     20.      2.   22.74   99.4001
0.0889
2  139.    281.      4.     20.      2.   22.91   99.3765
0.1041
2  139.    281.      4.     21.      2.   22.92   99.3507
0.0717
2  139.    283.      4.     13.      2.   23.11   99.3848
0.0792
2  139.    283.      4.     18.      2.   22.84   99.4952
0.1122
2  139.    283.      4.     18.      2.   22.76   99.3220
0.0915
2  139.    283.      4.     19.      2.   23.03   99.4165
0.0503
2  139.    283.      4.     21.      2.   22.87   99.3791
0.1138
2  139.    283.      4.     21.      2.   22.98   99.3985
0.0661
2  139.   2062.      4.     14.      2.   22.43   99.4283
0.0891
2  139.   2062.      4.     15.      2.   22.70   99.4139
0.2147
2  139.   2062.      4.     19.      2.   22.97   99.3813
0.1143
2  139.   2062.      4.     19.      1.   22.77   99.4314
0.1685
2  139.   2062.      4.     21.      2.   22.79   99.4166
0.2080
2  139.   2062.      4.     21.      2.   22.94   99.4052
0.2400
2  139.   2362.      4.     12.      1.   22.82   99.3408
0.1279
2  139.   2362.      4.     18.      2.   22.77   99.3116
0.1131
2  139.   2362.      4.     19.      2.   22.82   99.3241
0.0519
2  139.   2362.      4.     19.      1.   22.74   99.2991
0.0903
2  139.   2362.      4.     20.      2.   22.88   99.3049
0.0783
2  139.   2362.      4.     21.      2.   22.94   99.2782
0.0718
2  140.      1.      4.     13.      1.   23.10   96.0811
0.0463
2  140.      1.      4.     15.      2.   22.75   96.1460
0.0725
2  140.      1.      4.     18.      2.   22.78   96.1582
```

```
0.1428
2  140.      1.      4.     19.      1.   22.70   96.1039
0.1056
2  140.      1.      4.     20.      2.   22.75   96.1262
0.0672
2  140.      1.      4.     21.      2.   22.93   96.1478
0.0562
2  140.    281.      4.     15.      2.   22.71   96.1153
0.1097
2  140.    281.      4.     14.      2.   22.49   96.1297
0.1202
2  140.    281.      4.     18.      2.   22.81   96.1233
0.1331
2  140.    281.      4.     20.      2.   22.78   96.1731
0.1484
2  140.    281.      4.     20.      2.   22.89   96.0872
0.0857
2  140.    281.      4.     21.      2.   22.91   96.1331
0.0944
2  140.    283.      4.     13.      2.   23.22   96.1135
0.0983
2  140.    283.      4.     18.      2.   22.85   96.1111
0.1210
2  140.    283.      4.     18.      2.   22.78   96.1221
0.0644
2  140.    283.      4.     19.      2.   23.01   96.1063
0.0921
2  140.    283.      4.     21.      2.   22.91   96.1155
0.0704
2  140.    283.      4.     21.      2.   22.94   96.1308
0.0258
2  140.   2062.      4.     15.      2.   22.60   95.9767
0.2225
2  140.   2062.      4.     15.      2.   22.66   96.1277
0.1792
2  140.   2062.      4.     19.      2.   22.96   96.1858
0.1312
2  140.   2062.      4.     19.      1.   22.75   96.1912
0.1936
2  140.   2062.      4.     21.      2.   22.82   96.1650
0.1902
2  140.   2062.      4.     21.      2.   22.92   96.1603
0.1777
2  140.   2362.      4.     12.      1.   22.88   96.0793
0.0996
2  140.   2362.      4.     18.      2.   22.76   96.1115
0.0533
2  140.   2362.      4.     19.      2.   22.79   96.0803
0.0364
2  140.   2362.      4.     19.      1.   22.71   96.0411
0.0768
2  140.   2362.      4.     20.      2.   22.84   96.0988
0.1042
2  140.   2362.      4.     21.      1.   22.94   96.0482
0.0868
2  141.      1.      4.     13.      1.   23.07  101.1984
0.0803
2  141.      1.      4.     15.      2.   22.72  101.1645
0.0914
2  141.      1.      4.     18.      2.   22.75  101.2454
0.1109
2  141.      1.      4.     19.      1.   22.69  101.1096
0.1376
2  141.      1.      4.     20.      2.   22.83  101.2066
0.0717
2  141.      1.      4.     21.      2.   22.93  101.0645
0.1205
2  141.    281.      4.     15.      2.   22.72  101.1615
0.1272
2  141.    281.      4.     14.      2.   22.40  101.1650
0.0595
2  141.    281.      4.     18.      2.   22.78  101.1815
0.1393
2  141.    281.      4.     20.      2.   22.73  101.1106
0.1189
2  141.    281.      4.     20.      2.   22.86  101.1420
0.0713
2  141.    281.      4.     21.      2.   22.94  101.0116
0.1088
2  141.    283.      4.     13.      2.   23.26  101.1554
0.0429
```

```
2  141.    283.      4.     18.      2.   22.85  101.1267
0.0751
2  141.    283.      4.     18.      2.   22.76  101.1227
0.0826
2  141.    283.      4.     19.      2.   22.82  101.0635
0.1715
2  141.    283.      4.     21.      2.   22.89  101.1264
0.1447
2  141.    283.      4.     21.      2.   22.96  101.0853
0.1189
2  141.   2062.      4.     15.      2.   22.65  101.1332
0.2532
2  141.   2062.      4.     15.      1.   22.68  101.1487
0.1413
2  141.   2062.      4.     19.      2.   22.95  101.1778
0.1772
2  141.   2062.      4.     19.      1.   22.77  101.0988
0.0884
2  141.   2062.      4.     21.      2.   22.87  101.1686
0.2940
2  141.   2062.      4.     21.      2.   22.94  101.3289
0.2072
2  141.   2362.      4.     12.      1.   22.83  101.1353
0.0585
2  141.   2362.      4.     18.      2.   22.83  101.1201
0.0868
2  141.   2362.      4.     19.      2.   22.91  101.0946
0.0855
2  141.   2362.      4.     19.      1.   22.71  100.9977
0.0645
2  141.   2362.      4.     20.      2.   22.87  101.0963
0.0638
2  141.   2362.      4.     21.      2.   22.94  101.0300
0.0549
2  142.      1.      4.     13.      1.   23.07   94.3049
0.1197
2  142.      1.      4.     15.      2.   22.73   94.3153
0.0566
2  142.      1.      4.     18.      2.   22.77   94.3073
0.0875
2  142.      1.      4.     19.      1.   22.67   94.2803
0.0376
2  142.      1.      4.     20.      2.   22.80   94.3008
0.0703
2  142.      1.      4.     21.      2.   22.93   94.2916
0.0604
2  142.    281.      4.     14.      2.   22.90   94.2557
0.0619
2  142.    281.      4.     18.      2.   22.83   94.3542
0.1027
2  142.    281.      4.     18.      2.   22.80   94.3007
0.1492
2  142.    281.      4.     20.      2.   22.76   94.3351
0.1059
2  142.    281.      4.     20.      2.   22.88   94.3406
0.1508
2  142.    281.      4.     21.      2.   22.92   94.2621
0.0946
2  142.    283.      4.     13.      2.   23.25   94.3124
0.0534
2  142.    283.      4.     18.      2.   22.85   94.3680
0.1643
2  142.    283.      4.     18.      1.   22.67   94.3442
0.0346
2  142.    283.      4.     19.      2.   22.80   94.3391
0.0616
2  142.    283.      4.     21.      2.   22.91   94.2238
0.0721
2  142.    283.      4.     21.      2.   22.95   94.2721
0.0998
2  142.   2062.      4.     14.      2.   22.49   94.2915
0.2189
2  142.   2062.      4.     15.      2.   22.69   94.2803
0.0690
2  142.   2062.      4.     19.      2.   22.94   94.2818
0.0987
2  142.   2062.      4.     19.      1.   22.76   94.2227
0.2628
2  142.   2062.      4.     21.      2.   22.74   94.4109
0.1230
2  142.   2062.      4.     21.      2.   22.94   94.2616
```

```
0.0929
2  142.   2362.      4.     12.      1.   22.86   94.2052
0.0813
2  142.   2362.      4.     18.      2.   22.83   94.2824
0.0605
2  142.   2362.      4.     19.      2.   22.85   94.2396
0.0882
2  142.   2362.      4.     19.      1.   22.75   94.2087
0.0702
2  142.   2362.      4.     20.      2.   22.86   94.2937
0.0591
2  142.   2362.      4.     21.      1.   22.93   94.2330
0.0556
```

2.6.3.1.2. Measurements on wiring configurations

Check wafers were measured with the probe wired in two configurations

Measurements of resistivity (ohm.cm) were made according to an ASTM Standard Test Method (F4) to identify differences between 2 wiring configurations for probe #2362. The check standards for the study were five wafers selected at random from a batch of wafers cut from one silicon crystal doped with phosphorous to give a nominal resistivity of 100 ohm.cm.

Description of database

The data are averages of $K = 6$ days' measurements and $J = 6$ repetitions at the center of each wafer. There are $L = 2$ complete runs, separated by two months time, on each wafer.

The data recorded in the 10 columns are:

1. Wafer
2. Probe
3. Average - configuration A; run 1
4. Standard deviation - configuration A; run 1
5. Average - configuration B; run 1
6. Standard deviation - configuration B; run 1
7. Average - configuration A; run 2
8. Standard deviation - configuration A; run 2
9. Average - configuration B; run 2
10. Standard deviation - configuration B; run 2

```
Wafer Probe Config A-run1   Config B-run1   Config A-run2
Config B-run2.

138. 2362.  95.1162 0.0480  95.0993 0.0466  95.1687 0.0785
95.1589 0.0642
138. 2362.  95.0569 0.0577  95.0657 0.0450  95.1564 0.0430
95.1705 0.0730
138. 2362.  95.0598 0.0516  95.0622 0.0664  95.1354 0.0983
95.1221 0.0695
138. 2362.  95.1487 0.0386  95.1625 0.0311  95.0422 0.0773
95.0513 0.0840
138. 2362.  95.0743 0.0256  95.0599 0.0488  95.1354 0.0587
95.1531 0.0482
138. 2362.  95.1010 0.0420  95.0944 0.0393  95.1075 0.0776
95.1537 0.0230
139. 2362.  99.2939 0.0818  99.3018 0.0905  99.3408 0.1279
99.3637 0.1025
139. 2362.  99.3234 0.0723  99.3488 0.0350  99.3116 0.1131
99.3881 0.0451
```

```
139. 2362.  99.2748 0.0756  99.3571 0.1993  99.3241 0.0519
99.3737 0.0699
139. 2362.  99.3512 0.0475  99.3512 0.1286  99.2991 0.0903
99.3066 0.0709
139. 2362.  99.2350 0.0517  99.2255 0.0738  99.3049 0.0783
99.3040 0.0744
139. 2362.  99.3574 0.0485  99.3605 0.0459  99.2782 0.0718
99.3680 0.0470
140. 2362.  96.0771 0.1024  96.0915 0.1257  96.0793 0.0996
96.1041 0.0890
140. 2362.  95.9976 0.0943  96.0057 0.0806  96.1115 0.0533
96.0774 0.0983
140. 2362.  96.0148 0.0622  96.0244 0.0833  96.0803 0.0364
96.1004 0.0758
140. 2362.  96.0397 0.0702  96.0422 0.0738  96.0411 0.0768
96.0677 0.0663
140. 2362.  96.0407 0.0627  96.0738 0.0800  96.0988 0.1042
96.0585 0.0960
140. 2362.  96.0445 0.0622  96.0557 0.1129  96.0482 0.0868
96.0062 0.0895
141. 2362. 101.0471 0.0320 101.0241 0.0670 101.1353 0.0585
101.1156 0.1027
141. 2362. 101.0224 0.1020 101.0660 0.1030 101.1201 0.0868
101.1077 0.1141
141. 2362. 101.0702 0.0580 101.0509 0.0710 101.0946 0.0855
101.0455 0.1070
141. 2362. 101.0904 0.1049 101.0983 0.0894 100.9977 0.0645
101.0274 0.0666
141. 2362. 101.0626 0.0702 101.0614 0.0849 101.0963 0.0638
101.1106 0.0788
141. 2362. 101.0686 0.0661 101.0811 0.0490 101.0300 0.0549
101.1073 0.0663
142. 2362.  94.2437 0.0503  94.2088 0.0815  94.2052 0.0813
94.2487 0.0719
142. 2362.  94.2115 0.0919  94.2043 0.1176  94.2824 0.0605
94.2886 0.0499
142. 2362.  94.2348 0.0282  94.2324 0.0519  94.2396 0.0882
94.2739 0.1075
142. 2362.  94.2124 0.0513  94.2347 0.0694  94.2087 0.0702
94.2023 0.0416
142. 2362.  94.2214 0.0627  94.2416 0.0757  94.2937 0.0591
94.2600 0.0731
142. 2362.  94.1651 0.1010  94.2287 0.0919  94.2330 0.0556
94.2406 0.0651
```

2.6.3.2. Analysis and interpretation

Purpose of this page

The purpose of this page is to outline an analysis of data taken during a gauge study to quantify the type A uncertainty component for resistivity (ohm.cm) measurements on silicon wafers made with a gauge that was part of the initial study.

Summary of standard deviations at three levels

The level-1, level-2, and level-3 standard deviations for the uncertainty analysis are summarized in the table below from the gauge case study.

Standard deviations for probe #2362

Level	Symbol	Estimate	DF
Level-1	s_1	0.0729	300
Level-2	s_2	0.0362	50
Level-3	s_3	0.0197	5

Calculation of individual components for days and runs

The standard deviation that estimates the day effect is

$$s_{days} = \sqrt{s_2^2 - \frac{1}{6}s_1^2} = 0.0217 \ \ ohm.cm$$

The standard deviation that estimates the run effect is

$$s_{runs} = \sqrt{s_3^2 - \frac{1}{6}s_2^2} = 0.0130 \ \ ohm.cm$$

Calculation of the standard deviation of the certified value showing sensitivity coefficients

The certified value for each wafer is the average of $N = 6$ repeatability measurements at the center of the wafer on $M = 1$ days and over $P = 1$ runs. Notice that N, M and P are not necessarily the same as the number of measurements in the gauge study per wafer; namely, J, K and L. The standard deviation of a certified value (for time-dependent sources of error), is

$$s = \sqrt{s_{runs}^2 + s_{days}^2 - \frac{1}{6}s_1^2}$$

Standard deviations for days and runs are included in this calculation, even though there were no replications over days or runs for the certification measurements. These factors contribute

to the overall uncertainty of the measurement process even though they are not sampled for the particular measurements of interest.

The equation must be rewritten to calculate degrees of freedom

Degrees of freedom cannot be calculated from the equation above because the calculations for the individual components involve differences among variances. The table of sensitivity coefficients for a 3-level design shows that for

$$N = J,\ M = 1,\ P = 1$$

the equation above can be rewritten in the form

$$s = \sqrt{\frac{5}{6}s_2^2 + s_3^2}$$

Then the degrees of freedom can be approximated using the Welch-Satterthwaite method.

Probe bias - Graphs of probe biases

A graphical analysis shows the relative biases among the 5 probes. For each wafer, differences from the wafer average by probe are plotted versus wafer number. The graphs verify that probe #2362 (coded as 5) is biased low relative to the other probes. The bias shows up more strongly after the probes have been in use (run 2).

How to deal with bias due to the probe

Probe #2362 was chosen for the certification process because of its superior precision, but its bias relative to the other probes creates a problem. There are two possibilities for handling this problem:

1. Correct all measurements made with probe #2362 to the average of the probes.
2. Include the standard deviation for the difference among probes in the uncertainty budget.

The best strategy, as followed in the certification process, is to correct all measurements for the average bias of probe #2362 and take the standard deviation of the correction as a type A component of uncertainty.

Correction for bias or probe #2362 and uncertainty

Biases by probe and wafer are shown in the gauge case study. Biases for probe #2362 are summarized in table below for the two runs. The correction is taken to be the negative of the average bias. The standard deviation of the correction is the standard deviation of the average of the ten biases.

Estimated biases for probe #2362

Wafer	Probe	Run 1	Run 2	All
138	2362	-0.0372	-0.0507	
139	2362	-0.0094	-0.0657	
140	2362	-0.0261	-0.0398	
141	2362	-0.0252	-0.0534	

```
142   2362   -0.0383   -0.0469

Average      -0.0272   -0.0513   -0.0393
Standard deviation                0.0162
(10 values)
```

Configurations Database and plot of differences

Measurements on the check wafers were made with the probe wired in two different configurations (A, B). A plot of differences between configuration A and configuration B shows no bias between the two configurations.

Test for difference between configurations

This finding is consistent over runs 1 and 2 and is confirmed by the t-statistics in the table below where the average differences and standard deviations are computed from 6 days of measurements on 5 wafers. A t-statistic < 2 indicates no significant difference. The conclusion is that there is no bias due to wiring configuration and no contribution to uncertainty from this source.

Differences between configurations

Status	Average	Std dev	DF	t
Pre	-0.00858	0.0242	29	1.9
Post	-0.0110	0.0354	29	1.7

Error budget showing sensitivity coefficients, standard deviations and degrees of freedom

The error budget showing sensitivity coefficients for computing the standard uncertainty and degrees of freedom is outlined below.

Error budget for resistivity (ohm.cm)

Source	Type	Sensitivity	Standard Deviation	DF
Repeatability	A	$a_1 = 0$	0.0729	300
Reproducibility	A	$a_2 = \sqrt{5/6}$	0.0362	50
Run-to-run	A	$a_3 = 1$	0.0197	5
Probe #2362	A	$a_4 = \sqrt{1/10}$	0.0162	5
Wiring Configuration A	A	$a_5 = 1$	0	--

Standard uncertainty includes components for repeatability, days, runs and probe

The standard uncertainty is computed from the error budget as

$$u = \sqrt{\sum_i a_i^2 s_i^2} = \sqrt{\frac{5}{6}s_2^2 + s_3^2 + \frac{1}{10}s_{probe}^2} = 0.0388 ohm.cm$$

Approximate

The degrees of freedom associated with *u* are approximated by

degrees of freedom and expanded uncertainty

the Welch-Satterthwaite formula as:

$$\nu = \frac{u^4}{\sum_{i=1}^{5} \frac{a_i^4 s_i^4}{\nu_i}} = 42$$

where the ν_i are the degrees of freedom given in the rightmost column of the table.

The critical value at the 0.05 significance level with 42 degrees of freedom, from the t-table, is 2.018 so the expanded uncertainty is

$$U = 2.018\ u = 0.078\ ohm.cm$$

2.6.3.2.1. Difference between 2 wiring configurations

Measurements with the probe configured in two ways

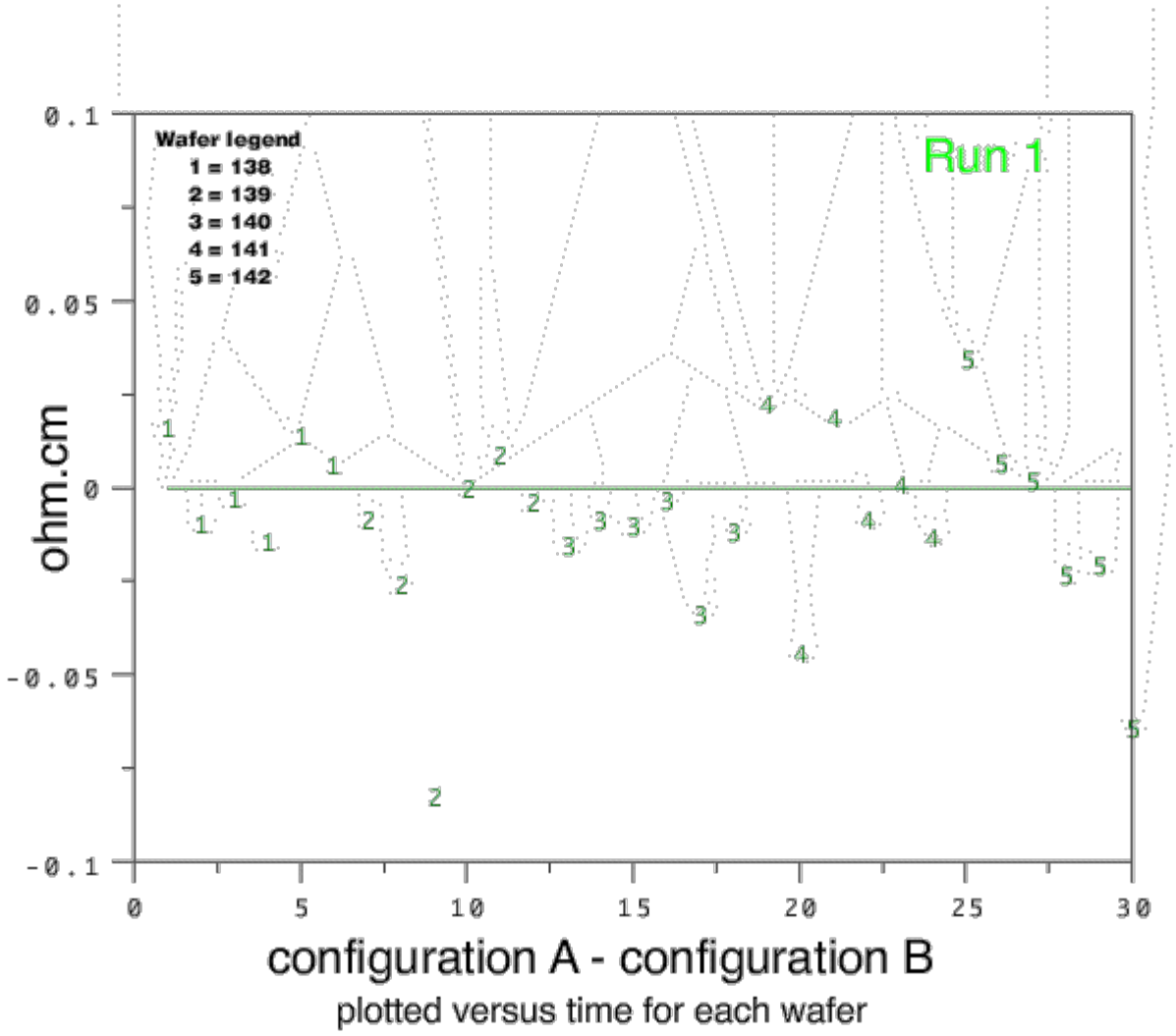
The graphs below are constructed from resistivity measurements (ohm.cm) on five wafers where the probe (#2362) was wired in two different configurations, A and B. The probe is a 4-point probe with many possible wiring configurations. For this experiment, only two configurations were tested as a means of identifying large discrepancies.

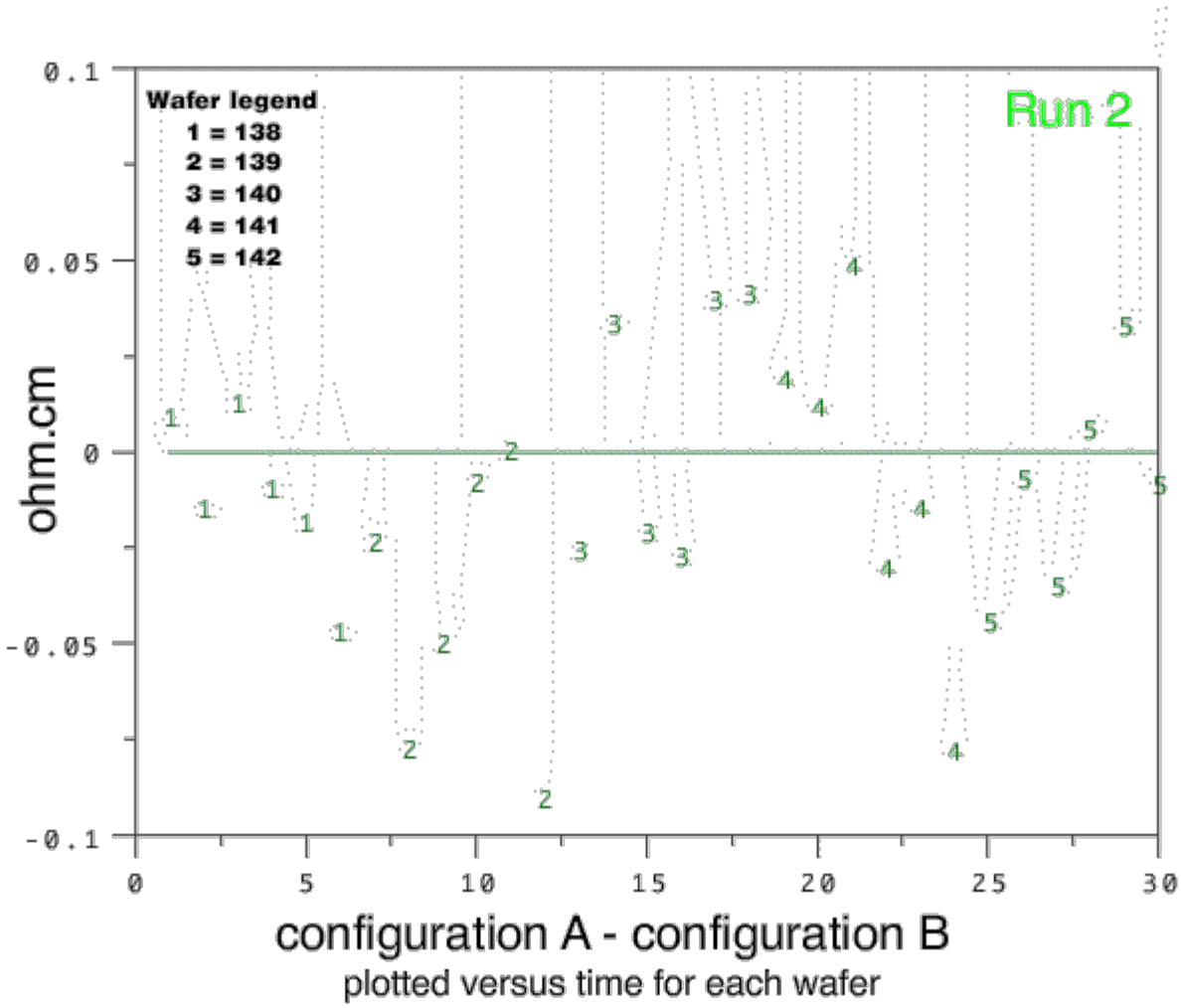
Artifacts for the study

The five wafers; namely, #138, #139, #140, #141, and #142 are coded 1, 2, 3, 4, 5, respectively, in the graphs. These wafers were chosen at random from a batch of approximately 100 wafers that were being certified for resistivity.

Interpretation

Differences between measurements in configurations A and B, made on the same day, are plotted over six days for each wafer. The two graphs represent two runs separated by approximately two months time. The dotted line in the center is the zero line. The pattern of data points scatters fairly randomly above and below the zero line -- indicating no difference between configurations for probe #2362. The conclusion applies to probe #2362 and cannot be extended to all probes of this type.


Wafer legend
1 = 138
2 = 139
3 = 140
4 = 141
5 = 142
Run 1
0.1
0.05
0
-0.05
-0.1
ohm.cm
0
5
10
15
20
25
30
configuration A - configuration B
plotted versus time for each wafer


Wafer legend
1 = 138
2 = 139
3 = 140
4 = 141
5 = 142
Run 2
ohm.cm
0.1
0.05
0
-0.05
-0.1
0
5
10
15
20
25
30
configuration A - configuration B
plotted versus time for each wafer

2.6.3.3. Run the type A uncertainty analysis using Dataplot

View of Dataplot macros for this case study

This page allows you to repeat the analysis outlined in the case study description on the previous page using Dataplot . It is required that you have already downloaded and installed Dataplot and configured your browser. to run Dataplot. Output from each analysis step below will be displayed in one or more of the Dataplot windows. The four main windows are the Output Window, the Graphics window, the Command History window, and the data sheet window. Across the top of the main windows there are menus for executing Dataplot commands. Across the bottom is a command entry window where commands can be typed in.

Data Analysis Steps	Results and Conclusions
Click on the links below to start Dataplot and run this case study yourself. Each step may use results from previous steps, so please be patient. Wait until the software verifies that the current step is complete before clicking on the next step.	*The links in this column will connect you with more detailed information about each analysis step from the case study description.*
Time-dependent components from 3-level nested design Pool repeatability standard deviations for: 1. Run 1 2. Run 2 Compute level-2 standard deviations for: 3. Run 1 4. Run 2 5. Pool level-2 standard deviations	*Database of measurements with probe #2362* 1. The repeatability standard deviation is 0.0658 ohm.cm for run 1 and 0.0758 ohm.cm for run 2. This represents the basic precision of the measuring instrument. 2. The level-2 standard deviation pooled over 5 wafers and 2 runs is 0.0362 ohm.cm. This is significant in the calculation of uncertainty. 3. The level-3 standard deviation pooled over 5 wafers is 0.0197 ohm.cm. This is small compared to the other components but is

6. Compute level-3 standard deviations	included in the uncertainty calculation for completeness.
Bias due to probe #2362 1. Plot biases for 5 NIST probes 2. Compute wafer bias and average bias for probe #2362 3. Correction for bias and standard deviation	*Database of measurements with 5 probes* 1. The plot shows that probe #2362 is biased low relative to the other probes and that this bias is consistent over 5 wafers. 2. The bias correction is the average bias = 0.0393 ohm.cm over the 5 wafers. The correction is to be subtracted from all measurements made with probe #2362. 3. The uncertainty of the bias correction = 0.0051 ohm.cm is computed from the standard deviation of the biases for the 5 wafers.
Bias due to wiring configuration A 1. Plot differences between wiring configurations 2. Averages, standard deviations and t-statistics	*Database of wiring configurations A and B* 1. The plot of measurements in wiring configurations A and B shows no difference between A and B. 2. The statistical test confirms that there is no difference between the wiring configurations.
Uncertainty 1. Standard uncertainty, df, t-value and expanded uncertainty	*Elements of error budget* 1. The uncertainty is computed from the error budget. The uncertainty for an average of 6 measurements on one day with probe #2362 is 0.078 with 42 degrees of freedom.

2.6.3.4. Dataplot macros

Reads data and plots the repeatability standard deviations for probe #2362 and pools standard deviations over days, wafers -- run 1

```
reset data
reset plot control
reset i/o
dimension 500 rows
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
retain run wafer probe y sr    subset probe =
2362
let df = sr - sr + 5.
y1label ohm.cm
characters * all
lines blank all
x2label Repeatability standard deviations for
probe 2362 - run 1
plot sr subset run 1
let var = sr*sr
let df11 = sum df subset run 1
let s11 = sum var subset run 1
. repeatability standard deviation for run 1
let s11 = (5.*s11/df11)**(1/2)
print s11 df11
. end of calculations
```

Reads data and plots repeatability standard deviations for probe #2362 and pools standard deviations over days, wafers -- run 2

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
retain run wafer probe y sr subset probe 2362
let df = sr - sr + 5.
y1label ohm.cm
characters * all
lines blank all
x2label Repeatability standard deviations for
probe 2362 - run 2
plot sr subset run 2
let var = sr*sr
let df11 = sum df subset run 1
let df12 = sum df subset run 2
let s11 = sum var subset run 1
let s12 = sum var subset run 2
let s11 = (5.*s11/df11)**(1/2)
let s12 = (5.*s12/df12)**(1/2)
print s11 df11
print s12 df12
let s1 = ((s11**2 + s12**2)/2.)**(1/2)
let df1=df11+df12
. repeatability standard deviation and df for
run 2
print s1 df1
. end of calculations
```

Computes level-2 standard deviations

```
reset data
reset plot control
reset i/o
dimension 500 rows
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
```

from daily averages and pools over wafers -- run 1

```
read mpc633a.dat run wafer probe y sr
retain run wafer probe y sr subset probe 2362
sd plot y wafer subset run 1
let s21 = yplot
let wafer1 = xplot
retain s21 wafer1 subset tagplot = 1
let nwaf = size s21
let df21 = 5 for i = 1 1 nwaf
. level-2 standard deviations and df for 5
wafers - run 1
print wafer1 s21 df21
. end of calculations
```

Computes level-2 standard deviations from daily averages and pools over wafers -- run 2

```
reset data
reset plot control
reset i/o
dimension 500 rows
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
retain run wafer probe y sr subset probe 2362
sd plot y wafer subset run 2
let s22 = yplot
let wafer1 = xplot
retain s22 wafer1 subset tagplot = 1
let nwaf = size s22
let df22 = 5 for i = 1 1 nwaf
. level-2 standard deviations and df for 5
wafers - run 1
print wafer1 s22 df22
. end of calculations
```

Pools level-2 standard deviations over wafers and runs

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
retain run wafer probe y sr subset probe 2362
sd plot y wafer subset run 1
let s21 = yplot
let wafer1 = xplot
sd plot y wafer subset run 2
let s22 = yplot
retain s21 s22 wafer1 subset tagplot = 1
let nwaf = size wafer1
let df21 = 5 for i = 1 1 nwaf
let df22 = 5 for i = 1 1 nwaf
let s2a = (s21**2)/5 + (s22**2)/5
let s2 = sum s2a
let s2 = sqrt(s2/2)
let df2a = df21 + df22
let df2 = sum df2a
. pooled level-2 standard deviation and df
across wafers and runs
print s2 df2
. end of calculations
```

Computes level-3standard deviations from run averages and pools over wafers

```
reset data
reset plot control
reset i/o
dimension 500 rows
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
retain run wafer probe y sr subset probe 2362
.
mean plot y wafer subset run 1
let m31 = yplot
let wafer1 = xplot
mean plot y wafer subset run 2
let m32 = yplot
retain m31 m32 wafer1 subset tagplot = 1
let nwaf = size m31
let s31 =(((m31-m32)**2)/2.)**(1/2)
let df31 = 1 for i = 1 1 nwaf
. level-3 standard deviations and df for 5
wafers
```

```
print wafer1 s31 df31
let s31 = (s31**2)/5
let s3 = sum s31
let s3 = sqrt(s3)
let df3=sum df31
. pooled level-3 std deviation and df over 5
wafers
print s3 df3
. end of calculations
```

Plot differences from the average wafer value for each probe showing bias for probe #2362

```
reset data
reset plot control
reset i/o
dimension 500 30
read mpc61a.dat wafer probe d1 d2
let biasrun1 = mean d1 subset probe 2362
let biasrun2 = mean d2 subset probe 2362
print biasrun1 biasrun2
title GAUGE STUDY FOR 5 PROBES
Y1LABEL OHM.CM
lines dotted dotted dotted dotted dotted solid
characters 1 2 3 4 5 blank
xlimits 137 143
let zero = pattern 0 for I = 1 1 30
x1label DIFFERENCES AMONG PROBES VS WAFER (RUN
1)
plot d1 wafer probe and
plot zero wafer
let biasrun2 = mean d2 subset probe 2362
print biasrun2
title GAUGE STUDY FOR 5 PROBES
Y1LABEL OHM.CM
lines dotted dotted dotted dotted dotted solid
characters 1 2 3 4 5 blank
xlimits 137 143
let zero = pattern 0 for I = 1 1 30
x1label DIFFERENCES AMONG PROBES VS WAFER (RUN
2)
plot d2 wafer probe and
plot zero wafer
. end of calculations
```

Compute bias for probe #2362 by wafer

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
set read format
.
cross tabulate mean y run wafer
retain run wafer probe y sr subset probe 2362
skip 1
read dpst1f.dat runid wafid ybar
print runid wafid ybar
let ngroups = size ybar
skip 0
.
let m3 = y - y
feedback off
loop for k = 1 1 ngroups
    let runa = runid(k)
    let wafera = wafid(k)
    let ytemp = ybar(k)
    let m3 = ytemp subset run = runa subset
wafer = wafera
end of loop
feedback on
.
let d = y - m3
let bias1 = average d subset run 1
let bias2 = average d subset run 2
.
mean plot d wafer subset run 1
let b1 = yplot
let wafer1 = xplot
mean plot d wafer subset run 2
let b2 = yplot
```

```
retain b1 b2 wafer1 subset tagplot = 1
let nwaf = size b1
. biases for run 1 and run 2 by wafers
print wafer1 b1 b2
. average biases over wafers for run 1 and run 2
print bias1 bias2
. end of calculations
```

Compute correction for bias for measurements with probe #2362 and the standard deviation of the correction

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
set read format f1.0,f6.0,f8.0,32x,f10.4,f10.4
read mpc633a.dat run wafer probe y sr
set read format
.
cross tabulate mean y run wafer
retain run wafer probe y sr subset probe 2362
skip 1
read dpst1f.dat runid wafid ybar
let ngroups = size ybar
skip 0
.
let m3 = y - y
feedback off
loop for k = 1 1 ngroups
    let runa = runid(k)
    let wafera = wafid(k)
    let ytemp = ybar(k)
    let m3 = ytemp subset run = runa subset
wafer = wafera
end of loop
feedback on
.
let d = y - m3
let bias1 = average d subset run 1
let bias2 = average d subset run 2
.
mean plot d wafer subset run 1
let b1 = yplot
let wafer1 = xplot
mean plot d wafer subset run 2
let b2 = yplot
retain b1 b2 wafer1 subset tagplot = 1
.
extend b1 b2
let sd = standard deviation b1
let sdcorr = sd/(10**(1/2))
let correct = -(bias1+bias2)/2.
. correction for probe #2362, standard dev, and
standard dev of corr
print correct sd sdcorr
. end of calculations
```

Plot differences between wiring configurations A and B

```
reset data
reset plot control
reset i/o
dimension 500 30
label size 3
read mpc633k.dat wafer probe a1 s1 b1 s2 a2 s3
b2 s4
let diff1 = a1 - b1
let diff2 = a2 - b2
let t = sequence 1 1 30
lines blank all
characters 1 2 3 4 5
y1label ohm.cm
x1label Config A - Config B -- Run 1
x2label over 6 days and 5 wafers
x3label legend for wafers 138, 139, 140, 141,
142: 1, 2, 3, 4, 5
plot diff1 t wafer
x1label Config A - Config B -- Run 2
plot diff2 t wafer
. end of calculations
```

```
reset data
```

Compute average differences between configuration A and B; standard deviations and t-statistics for testing significance

```
reset plot control
reset i/o
separator character @
dimension 500 rows
label size 3
read mpc633k.dat wafer probe a1 s1 b1 s2 a2 s3
b2 s4
let diff1 = a1 - b1
let diff2 = a2 - b2
let d1 = average diff1
let d2 = average diff2
let s1 = standard deviation diff1
let s2 = standard deviation diff2
let t1 = (30.)**(1/2)*(d1/s1)
let t2 = (30.)**(1/2)*(d2/s2)
. Average config A-config B; std dev difference;
t-statistic for run 1
print d1 s1 t1
. Average config A-config B; std dev difference;
t-statistic for run 2
print d2 s2 t2
separator character ;
. end of calculations
```

Compute standard uncertainty, effective degrees of freedom, t value and expanded uncertainty

```
reset data
reset plot control
reset i/o
dimension 500 rows
label size 3
read mpc633m.dat sz a df
let c = a*sz*sz
let d = c*c
let e = d/(df)
let sume = sum e
let u = sum c
let u = u**(1/2)
let effdf=(u**4)/sume
let tvalue=tppf(.975,effdf)
let expu=tvalue*u
.
. uncertainty, effective degrees of freedom,
tvalue and
. expanded uncertainty
print u effdf tvalue expu
. end of calculations
```

2.6.4. Evaluation of type B uncertainty and propagation of error

Focus of this case study

The purpose of this case study is to demonstrate uncertainty analysis using statistical techniques coupled with type B analyses and propagation of error. It is a continuation of the case study of type A uncertainties.

Background - description of measurements and constraints

The measurements in question are volume resistivities (ohm.cm) of silicon wafers which have the following definition:

$$\rho = Xo \cdot K_a \cdot F_t \cdot t \cdot F_{t/s}$$

with explanations of the quantities and their nominal values shown below:

ρ = *resistivity* = 0.00128 *ohm.cm*
X = *voltage/current* (*ohm*)
t = $thickness_{wafer}$(*cm*) = 0.628 *cm*
K_a = $factor_{electrical}$ = 4.50 *ohm.cm*
F_F = $correction_{temp} \approx 1^{\circ} C$
$F_{t/s}$ = $factor_{thickness/separation} \approx 1.0$

Type A evaluations

The resistivity measurements, discussed in the case study of type A evaluations, were replicated to cover the following sources of uncertainty in the measurement process, and the associated uncertainties are reported in units of resistivity (ohm.cm).

- Repeatability of measurements at the center of the wafer
- Day-to-day effects
- Run-to-run effects
- Bias due to probe #2362
- Bias due to wiring configuration

Need for propagation of error

Not all factors could be replicated during the gauge experiment. Wafer thickness and measurements required for the scale corrections were measured off-line. Thus, the type B evaluation of uncertainty is computed using propagation of error. The propagation of error formula in units of resistivity is as follows:

$$s_\rho = \rho\sqrt{\frac{s_Y^2}{X^2} + \frac{s_l^2}{t^2} + \frac{s_{Ka}^2}{K_a^2} + \frac{s_{FT}^2}{F_T} + \frac{s_{F(t/S)}^2}{F_{t/S}}}$$

Standard deviations for type B evaluations

Standard deviations for the type B components are summarized here. For a complete explanation, see the publication (Ehrstein and Croarkin).

Electrical measurements

There are two basic sources of uncertainty for the electrical measurements. The first is the least-count of the digital volt meter in the measurement of X with a maximum bound of

$$a = 0.0000534 \text{ ohm}$$

which is assumed to be the half-width of a uniform distribution. The second is the uncertainty of the electrical scale factor. This has two sources of uncertainty:

1. error in the solution of the transcendental equation for determining the factor
2. errors in measured voltages

The maximum bounds to these errors are assumed to be half-widths of

$$a = 0.0001 \text{ ohm.cm} \text{ and } a = 0.00038 \text{ ohm.cm}$$

respectively, from uniform distributions. The corresponding standard deviations are shown below.

$$s_X = 0.0000534/\sqrt{3} = 0.0000308 \text{ ohm}$$

$$s_{Ks} = \sqrt{\frac{0.0001^2}{3} + \frac{0.00038^2}{3}} = 0.000227 \text{ ohm.cm}$$

Thickness

The standard deviation for thickness, t, accounts for two sources of uncertainty:

1. calibration of the thickness measuring tool with precision gauge blocks
2. variation in thicknesses of the silicon wafers

The maximum bounds to these errors are assumed to be half-widths of

$$a = 0.000015 \text{ cm} \text{ and } a = 0.000001 \text{ cm}$$

respectively, from uniform distributions. Thus, the standard deviation for thickness is

$$s_t = \sqrt{\frac{.000015^2}{3} + \frac{.000001^2}{3}} = .00000868 \ cm$$

Temperature correction

The standard deviation for the temperature correction is calculated from its defining equation as shown below. Thus, the standard deviation for the correction is the standard deviation associated with the measurement of temperature multiplied by the temperature coefficient, $C(t) = 0.0083$. The maximum bound to the error of the temperature measurement is assumed to be the half-width

$$a = 0.13 \ ^\circ C$$

of a triangular distribution. Thus the standard deviation of the correction for

$$F_T = 1 - C_T(T - 23^\circ C)$$

is

$$s_{F(T)} = C_T s_T = 0.0083\sqrt{\frac{0.13^2}{6}} = 0.000441^\circ C$$

Thickness scale factor

The standard deviation for the thickness scale factor is negligible.

Associated sensitivity coefficients

Sensitivity coefficients for translating the standard deviations for the type B components into units of resistivity (ohm.cm) from the propagation of error equation are listed below and in the error budget. The sensitivity coefficient for a source is the multiplicative factor associated with the standard deviation in the formula above; i.e., the partial derivative with respect to that variable from the propagation of error equation.

$a_6 = (\rho/X) = 100/0.111 = 900.901$
$a_7 = (\rho/K_a) = 100/4.50 = 22.222$
$a_8 = (\rho/t) = 100/0.628 = 159.24$
$a_9 = (\rho/F_T) = 100$
$a_{10} = (\rho/F_{t/S}) = 100$

Sensitivity coefficients and degrees of freedom

Sensitivity coefficients for the type A components are shown in the case study of type A uncertainty analysis and repeated below. Degrees of freedom for type B uncertainties based on assumed distributions, according to the convention, are assumed to be infinite.

Error budget showing

The error budget showing sensitivity coefficients for computing the relative standard uncertainty of volume resistivity (ohm.cm)

sensitivity coefficients, standard deviations and degrees of freedom

with degrees of freedom is outlined below.

Error budget for volume resistivity (ohm.cm)

Source	Type	Sensitivity	Standard Deviation	DF
Repeatability	A	$a_1 = 0$	0.0729	300
Reproducibility	A	$a_2 = \sqrt{5/6}$	0.0362	50
Run-to-run	A	$a_3 = 1$	0.0197	5
Probe #2362	A	$a_4 = \sqrt{1/10}$	0.0162	5
Wiring Configuration A	A	$a_5 = 1$	0	--
Resistance ratio	B	$a_6 = 900.901$	0.0000308	∞
Electrical scale	B	$a_7 = 22.222$	0.000227	∞
Thickness	B	$a_8 = 159.20$	0.00000868	∞
Temperature correction	B	$a_9 = 100$	0.000441	∞
Thickness scale	B	$a_{10} = 100$	0	--

Standard uncertainty

The standard uncertainty is computed as:

$$u = \sqrt{\sum_{i=1}^{10} a_i^2 s_i^2} = 0.065 \ \ ohm.cm$$

Approximate degrees of freedom and expanded uncertainty

The degrees of freedom associated with *u* are approximated by the Welch-Satterthwaite formula as:

$$\nu = \frac{u^4}{\sum_{i=1}^{5} \frac{a_i^4 s_i^4}{\nu_i}} = 42$$

This calculation is not affected by components with infinite degrees of freedom, and therefore, the degrees of freedom for the standard uncertainty is the same as the degrees of freedom for the type A uncertainty. The critical value at the 0.05 significance level with 42 degrees of freedom, from the t-table, is 2.018 so the expanded uncertainty is

$$U = 2.018 \ u = 0.13 \ ohm.cm$$

2.7. References

Degrees of freedom	K. A. Brownlee (1960). ***Statistical Theory and Methodology in Science and Engineering,*** John Wiley & Sons, Inc., New York, p. 236.
Calibration designs	J. M. Cameron, M. C. Croarkin and R. C. Raybold (1977). ***Designs for the Calibration of Standards of Mass,*** NBS Technical Note 952, U.S. Dept. Commerce, 58 pages.
Calibration designs for eliminating drift	J. M. Cameron and G. E. Hailes (1974). ***Designs for the Calibration of Small Groups of Standards in the Presence of Drift***, Technical Note 844, U.S. Dept. Commerce, 31 pages.
Measurement assurance for measurements on ICs	Carroll Croarkin and Ruth Varner (1982). ***Measurement Assurance for Dimensional Measurements on Integrated-circuit Photomasks,*** NBS Technical Note 1164, U.S. Dept. Commerce, 44 pages.
Calibration designs for gauge blocks	Ted Doiron (1993). ***Drift Eliminating Designs for Non-Simultaneous Comparison Calibrations***, *J Research National Institute of Standards and Technology, 98,* pp. 217-224.
Type A & B uncertainty analyses for resistivities	J. R. Ehrstein and M. C. Croarkin (1998). ***Standard Reference Materials: The Certification of 100 mm Diameter Silicon Resistivity SRMs 2541 through 2547 Using Dual-Configuration Four-Point Probe Measurements,*** NIST Special Publication 260-131, Revised, 84 pages.
Calibration designs for electrical standards	W. G. Eicke and J. M. Cameron (1967). ***Designs for Surveillance of the Volt Maintained By a Group of Saturated Standard Cells,*** NBS Technical Note 430, U.S. Dept. Commerce 19 pages.
Theory of uncertainty analysis	Churchill Eisenhart (1962). ***Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems****J Research National Bureau of Standards-C. Engineering and Instrumentation, Vol. 67C, No.2*, p. 161-187.
Confidence, prediction, and	Gerald J. Hahn and William Q. Meeker (1991). ***Statistical Intervals: A Guide for Practitioners***, John Wiley & Sons,

tolerance intervals	Inc., New York.
Original calibration designs for weighings	J. A. Hayford (1893). ***On the Least Square Adjustment of Weighings,*** U.S. Coast and Geodetic Survey Appendix 10, Report for 1892.
Uncertainties for values from a calibration curve	Thomas E. Hockersmith and Harry H. Ku (1993). ***Uncertainties associated with proving ring calibrations***, NBS Special Publication 300: Precision Measurement and Calibration, Statistical Concepts and Procedures, Vol. 1, pp. 257-263, H. H. Ku, editor.
EWMA control charts	J. Stuart Hunter (1986). ***The Exponentially Weighted Moving Average***, *J Quality Technology, Vol. 18, No. 4*, pp. 203-207.
Fundamentals of mass metrology	K. B. Jaeger and R. S. Davis (1984). ***A Primer for Mass Metrology***, NBS Special Publication 700-1, 85 pages.
Fundamentals of propagation of error	Harry Ku (1966). ***Notes on the Use of Propagation of Error Formulas,*** *J Research of National Bureau of Standards-C. Engineering and Instrumentation, Vol. 70C, No.4, pp. 263-273.*
Handbook of statistical methods	Mary Gibbons Natrella (1963). ***Experimental Statistics,*** NBS Handbook 91, US Deptartment of Commerce.
Omnitab	Sally T. Peavy, Shirley G. Bremer, Ruth N. Varner, David Hogben (1986). ***OMNITAB 80: An Interpretive System for Statistical and Numerical Data Analysis,*** NBS Special Publication 701, US Deptartment of Commerce.
Uncertainties for uncorrected bias	Steve D. Phillips and Keith R. Eberhardt (1997). ***Guidelines for Expressing the Uncertainty of Measurement Results Containing Uncorrected Bias***, *NIST Journal of Research, Vol. 102, No. 5*.
Calibration of roundness artifacts	Charles P. Reeve (1979). ***Calibration designs for roundness standards***, NBSIR 79-1758, 21 pages.
Calibration designs for angle blocks	Charles P. Reeve (1967). ***The Calibration of Angle Blocks by Comparison***, NBSIR 80-19767, 24 pages.
SI units	Barry N. Taylor (1991). ***Interpretation of the SI for the United States and Metric Conversion Policy for Federal Agencies,*** NIST Special Publication 841, U.S.

Deptartment of Commerce.

Uncertainties for calibrated values — Raymond Turgel and Dominic Vecchia (1987). ***Precision Calibration of Phase Meters,*** *IEEE Transactions on Instrumentation and Measurement, Vol. IM-36, No. 4.*, pp. 918-922.

Example of propagation of error for flow measurements — James R. Whetstone et al. (1989). ***Measurements of Coefficients of Discharge for Concentric Flange-Tapped Square-Edged Orifice Meters in Water Over the Reynolds Number Range 600 to 2,700,000***, NIST Technical Note 1264. pp. 97.

Mathematica software — Stephen Wolfram (1993). ***Mathematica, A System of Doing Mathematics by Computer, 2nd edition,*** Addison-Wesley Publishing Co., New York.

Restrained least squares — Marvin Zelen (1962). ***"Linear Estimation and Related Topics"*** in *Survey of Numerical Analysis* edited by John Todd, McGraw-Hill Book Co. Inc., New York, pp. 558-577.

ASTM F84 for resistivity — ***ASTM Method F84-93, Standard Test Method for Measuring Resistivity of Silicon Wafers With an In-line Four-Point Probe.*** Annual Book of ASTM Standards, 10.05, West Conshohocken, PA 19428.

ASTM E691 for interlaboratory testing — ***ASTM Method E691-92, Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method.*** Annual Book of ASTM Standards, 10.05, West Conshohocken, PA 19428.

Guide to uncertainty analysis — ***Guide to the Expression of Uncertainty of Measurement*** (1993). ISBN 91-67-10188-9, 1st ed. ISO, Case postale 56, CH-1211, Genève 20, Switzerland, 101 pages.

ISO 5725 for interlaboratory testing — ***ISO 5725: 1997. Accuracy (trueness and precision) of measurement results, Part 2: Basic method for repeatability and reproducibility of a standard measurement method,*** ISO, Case postale 56, CH-1211, Genève 20, Switzerland.

ISO 11095 on linear calibration — ***ISO 11095: 1997. Linear Calibration using Reference Materials,*** ISO, Case postale 56, CH-1211, Genève 20, Switzerland.

MSA gauge studies manual — ***Measurement Systems Analysis Reference Manual, 2nd ed.,*** (1995). Chrysler Corp., Ford Motor Corp., General Motors Corp., 120 pages.

NCSL RP on uncertainty — ***Determining and Reporting Measurement Uncertainties,*** National Conference of Standards Laboratories RP-12,

analysis	(1994), Suite 305B, 1800 30th St., Boulder, CO 80301.
ISO Vocabulary for metrology	***International Vocabulary of Basic and General Terms in Metrology, 2nd ed.,*** (1993). ISO, Case postale 56, CH-1211, Genève 20, Switzerland, 59 pages.
Exact variance for length and width	Leo Goodman (1960). ***"On the Exact Variance of Products"*** in *Journal of the American Statistical Association*, December, 1960, pp. 708-713.