AIAA 2000-2201

Langley Wind Tunnel Data Quality Assurance - Check Standard Results (Invited)

M. Hemsch
NASA Langley Research Center
Hampton, VA

J. Grubb and W. Krieger
Lockheed Martin
Hampton, VA

D. Cler
NASA Langley Research Center
Hampton, VA

21st AIAA Advanced Measurement Technology and Ground Testing Conference

19-22 June 2000 / Denver, CO

Langley Wind Tunnel Data Quality Assurance - Check Standard Results

Michael J. Hemsch [*]
NASA Langley Research Center
Hampton, VA

John P. Grubb [†] and William B. Krieger [⊥]
Lockheed Martin Technology Services Group
Hampton, VA

Daniel L. Cler [□]
NASA Langley Research Center
Hampton, VA

Abstract

A framework for statistical evaluation, control and improvement of wind tunnel measurement processes is presented. The methodology is adapted from elements of the Measurement Assurance Plans developed by the National Bureau of Standards (now the National Institute of Standards and Technology) for standards and calibration laboratories. The present methodology is based on the notions of statistical quality control (SQC) together with check standard testing and a small number of customer repeat-run sets. The results of check standard and customer repeat-run sets are analyzed using the statistical control chart methods of Walter A. Shewhart long familiar to the SQC community. Control chart results are presented for various measurement processes in five facilities at Langley Research Center. The processes include test section calibration, force and moment measurements with a balance, and instrument calibration.

Nomenclature

A	proportionality coefficient (see Eq. 20)
AF	uncorrected axial-force in balance axis system
AF_{max}	balance axial force limit
AOA	angle of attack
NF	uncorrected normal force in balance axis system
NF_{max}	balance normal force limit
PM	uncorrected pitching moment in balance axis system
PM_{max}	balance pitching moment limit
A_2	coefficient of $\overline{R}$ for computing lower and upper control limits for $\overline{x}_j$
B	proportionality coefficient (see Eq. 21)
C	proportionality coefficient (see Eq. 27)
C_A	uncorrected axial-force coefficient in balance axis system
C_{Mach}	Mach number ratio (see Eq. 12)
C_N	uncorrected normal-force coefficient in balance axis system
C_m	uncorrected pitching-moment coefficient in balance axis system
C'	calibration coefficient for low-speed tunnels (see Eq. 8)
d_2	coefficient used to remove bias when using $\overline{R}$ or $\overline{mR}$ as an estimator for σ
D_3	coefficient of $\overline{R}$ for computing lower control limits for R_j
D_4	coefficient of $\overline{R}$ for computing upper control limits for R_j
$\overline{mR}$	average of moving ranges
mR_j	two-point moving range for interval j

[*] Aerospace Technologist, Associate Fellow AIAA
[†] Senior Aeronautical Engineer
[⊥] Aeronautical Engineering Specialist
[□] Aerospace Technologist

$M_{reference}$	Mach number inferred from tunnel indicated stagnation pressure and measured nozzle block position
$M_{wall\ tap}$	Mach number inferred from wall static pressure and tunnel indicated stagnation pressure
M_∞	Mach number on the test section centerline
n	number of observations in a group
p_{static}	static pressure
p_{total}	stagnation pressure
q_∞	dynamic pressure on the test section centerline
R_j	range of group j
$\overline{R}$	average of group ranges
SF	scale factor for balance force coefficients (see Eq. 32)
S_{ref}	reference area for balance force and moment coefficients
x	value of an individual measurement
$\overline{x}$	average value of a set of individual measurements
$\overline{x}_j$	average value of measurements in group j
$\overline{\overline{x}}$	grand average of group averages
μ	population mean
$\hat{\mu}$	estimate of the population mean
ρ	correlation coefficient
σ	population standard deviation
$\hat{\sigma}$	estimate of the population standard deviation
$\hat{\sigma}_{within-test}$	estimated population standard deviation for individual values of measurements made during a single test
$\hat{\sigma}_{single\ meas.}$	estimated population standard deviation for individual values of measurements made during identical tests
bg	between-group
wg	within-group
CL	centerline
DQA	data quality assurance
LaRC	Langley Research Center (NASA)
LCL	lower control limit
LWTE	Langley Wind Tunnel Enterprise
MAP	Measurement Assurance Plan
NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
SQC	statistical quality control
UCL	upper control limit

Introduction

At the 19th AIAA Advanced Measurement Technology and Ground Testing Conference in 1996, the first author presented an outline of the data quality assurance program that has been adopted by the Langley Wind Tunnel Enterprise (LWTE).[1] The key features of the program as presented in the 1996 paper are

1. Measurement assurance based on statistical quality control (SQC),[2-5]

2. Pre-test measurement uncertainty prediction to insure that the data quality objectives of a test can actually be met,[6-7]

3. National standard corrections and procedures, most notably, correction of wall and support interference to the reference condition known as "free-air".[8-9] Round-robin testing would be used to remove remaining residuals between facilities.[4]

The present paper is divided into two main parts. In the first part, we present the general statistical framework and its rational for the LWTE data quality assurance program. In the second part, we present the procedures that the nine subscribing facilities have adopted for obtaining and analyzing the statistical data, together with examples from five facilities. We close with some final remarks on implementation of such a program and interpretation of the data. An outline of the rest of the paper is presented for convenience below. A description of the present work on the third program feature listed above is given elsewhere.[8]

Statistical Framework

Inference, Deduction and Doubt
Simple Statistical Control and Shewhart's Charts
Charts for Grouped Data
Charts for Individuals
Eisenhart and Complex Statistical Control
Cameron and Check Standard Testing
Schumacher, Customer Repeat-Sets, and Scaling

LWTE Methodology and Examples

Selection and Care of the Check Standards
Test Condition Check Standard
Airframe Check Standard
Selection of the Check Standard Test Matrices
Test Condition Matrix
Airframe Matrix
Analysis of the Test-Condition Data

Statistical Framework

Inference, Deduction and Doubt

Experimentalists have long recognized that variation in supposedly identical repetitions of a precision measurement is to be expected.[5] Eventually, it was recognized that the results of repeated experiments could be considered to be the elements of a virtual population, the characteristics of which could be inferred by sampling.[3,5,10] Statistical inference is the process of estimating the parameters of that population and statistical deduction is the process of deducing the properties of samples from the population.[11] This relationship is shown diagrammatically in Figure 1 taken from Bury.[11]

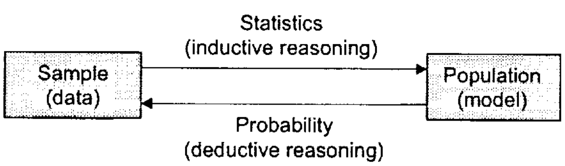

Figure 1 - The relationship between statistical inference and deduction (Bury[11]).

Both inference and deduction are essential for our statistical framework. But an additional notion is required which is that of "doubt" and the removal of it.[3]

The definition of doubt which we have found to be closest to our intended meaning was given by Webster[12] in his first dictionary (1828) as follows:

> *A fluctuation of mind respecting truth or propriety, arising from defect of knowledge or evidence; uncertainty of mind; suspense; unsettled state of opinion...*

Following the above definition, we mean doubt about the stability and meaning of the measurement process itself in a statistical sense. ***If the measurement process is not stable in both mean and dispersion, there is no population to talk about or from which to deduce the properties of samples, including any measurement sample of interest to the customer.*** It is this doubt about whether there exists a population in the statistical sense[3,5,10] that makes it difficult to assign risk to a decision based on the measurement.

Note that by doubt we do not mean that we are ignorant of the *type* of distribution. The type of distribution is relatively unimportant in ground testing. Rather, by doubt, we mean that we are unable to say, with any quantitative degree of certainty, that there is a population at all, Gaussian or otherwise.[†]

Simple Statistical Control and Shewhart's Charts

The key population parameters for measurement are the mean, μ, and the standard deviation, σ. We consider the mean to be the *right answer*, subject to corrections, and the standard deviation to be the *standard uncertainty* (for a single measurement) for whatever process variation is being expressed. For a population well-described by a Gaussian distribution, they are the only parameters. It also turns out that they are the only parameters of interest for almost any ground testing measurement population.

Sample characteristics can be used to make estimates of the population parameters if and only if it is known and confirmed that there *is* a population. For example, the sample average, $\bar{x}$, is an unbiased estimator[‡] for the value of μ and the sample variance, S^2, is an unbiased estimator for the value of σ^2. Of course, both of these estimates will be "fuzzy" for finite sample sizes. Fortunately, the Measurement Postulate states that the fuzziness for both will be reduced to arbitrarily small levels as the sample size is increased.[5]

In 1924, Walter A. Shewhart wrote an internal Western Electric memo describing a visual statistical

† Our statistical framework for evaluation, control and improvement refers only to Type A evaluation[4] of uncertainty and is based on *frequentist* notions of probability.[3,10,13] All other evaluations of uncertainty (Type B) are developed from an assumed probability distribution based on the degree of belief that an event will occur.[14] It would be improper to call such a notion of probability *frequentist*.[13]

‡ An unbiased estimator converges to the corresponding population parameter as more observations are made *provided* that there is a population. A biased estimator converges to a value that is proportional to the corresponding population parameter.

tool that can be used to check the validity of the Measurement Postulate for any measurement process, i.e. that there actually is or is not a population that has meaning for statistics and therefore measurement. His methods, as applied to manufacturing processes, are described in any text on statistical quality control.[15-19] He himself described the application to precision measurement in 1939.[3] It is beyond the scope of this paper to give the full justification for and details of his methods. However, we will briefly describe some of those methods that are particularly useful to ground testing measurement assurance and attempt to demonstrate their value through various examples.

If a stable population of values from a repeated measurement process can be said to exist, the process is said to be in a state of *statistical control*.[3-5] A process for which a stable population does not exist is said to be out-of-control. To be specific,

> *"A measurement process is in a state of statistical control if the resulting observations from the process, when collected under any fixed experimental conditions within the scope of the a priori well-defined conditions of the measurement process, behave like random drawings from some fixed distribution with fixed location [mean] and fixed scale [standard deviation] parameters."*[20]

Charts for Grouped Data

Shewhart used what he called "control charts" to determine whether a process is in statistical control. The following description of control charts is taken from Wheeler.[15] An example is provided in Figure 2 to illustrate their features. The data for the example charts of Figure 2 were obtained from a repeatability study of model leveling in one of the preparation bays in the National Transonic Facility. Twenty-seven engineers and technicians were divided into 9 groups of 3 observers each. Each group entered the preparation bay and placed the leveling plate on the model. Each individual then set the support apparatus to zero relative to gravity using a 20" bubble. The mounting system was moved off zero between measurements. The bubble was removed and replaced between each measurement. Hence, each group made three readings of the mounting system accelerometer when the model was ostensibly level with respect to gravity. The objective of the experiment was to determine if the observers/operators were effectively equivalent in a statistical sense and to determine their measurement capability if they were.

Typically, for Shewhart's charts, short-term repeat sets are obtained, separated significantly over time. These repeat sets are called "groups". The ranges[‡] and averages of the groups are computed and plotted as time series, one for each as shown in Figure 2. These two plots together are called $\bar{x}, R$ charts. We could just as easily have used $\bar{x}, S$ charts but we have found that the range is more intuitive for most people and it is easily estimated without calculation.

Once the two time series plots are made, a central line is added to each plot. For the $\bar{x}$ plot, the central line is given by the grand average of the group averages, $\bar{\bar{x}}$. For the R plot, the central line is given by the arithmetic average of the group ranges, $\bar{R}$. The central lines provide a visual reference for detecting shifts or trends in either the mean or the dispersion.

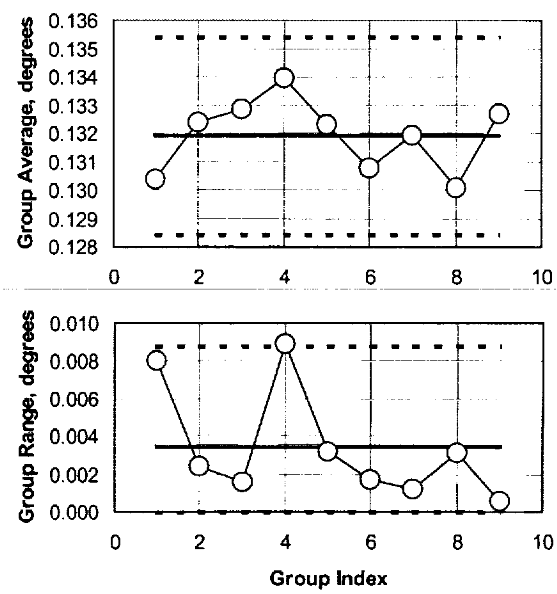

Figure 2 - $\bar{x}, R$ control charts for NTF leveling study.

Shewhart also added upper and lower limits for the group properties based on roughly 300:1 odds.[15-19] For such odds, for a process in statistical control, it would be unlikely that we would find any points beyond the limits. If we find any points beyond the limits or if we find any unusual patterns whose likelihood of occurrence is beyond 300:1 odds, ***then***

‡ The range is defined as the absolute value of the difference between the maximum and minimum values in the group.

we agree to declare that the process is not in control and there is no stable population. Conversely, if we have taken enough data, the points are within the limits, and no unusual, unrandom-like behavior is evident, we can safely say that the process is in control and, therefore, the properties of future samples are *predictable*. Of course, periodic sampling must continue forever to verify that the process stays in control.

The limits for the group ranges are given by

$$LCL(R) = D_3 \bar{R} \qquad UCL(R) = D_4 \bar{R} \tag{1}$$

and the limits for the group averages are given by

$$LCL(\bar{x}) = \bar{\bar{x}} - A_2 \bar{R} \qquad UCL(\bar{x}) = \bar{\bar{x}} + A_2 \bar{R} \tag{2}$$

where the coefficients of $\bar{R}$ are given in any SQC textbook and are dependent upon the number of observations in a group. The coefficients are given in Table 1 for the group size examples used in this paper.

n	A_2	D_3	D_4	d_2
2	1.880	0	3.267	1.128
3	1.023	0	2.575	1.693
10	0.308	0.223	1.777	3.078

Table 1 - Chart coefficients for grouped data.[18-19]

Although the group averages of the leveling-plate example of Figure 2 are within the limits, the group ranges are not, so we must conclude that the process is not in control and that inferences about the population parameters for this system of observers is not meaningful. Typically, 4-5 groups can be sufficient to say that a process is out-of-control, but 20-25 groups are needed to say that the process is in control.[15-17] Additional study and perhaps training of the observers/operators is needed to bring the process into control. ***Once the process is in control and it is clear that it is staying in control, we would have an unequivocal demonstration that there is a measurement population and a meaningful statement of measurement uncertainty can be made in the frequentist (Type A) sense.***

Charts for Individuals

Many times in measurement processes, there is no obvious grouping and it is more appropriate to use Charts for Individuals. For this type of chart, we plot the individual values separated in time and make a two-point moving range chart of the differences between adjacent point values as shown in Figure 3. For the moving range chart, the centerline and limits are given by[18-19]

$$CL(mR) = \overline{mR} \qquad LCL(mR) = 0 \qquad UCL(mR) = 3.267\ \overline{mR} \tag{3}$$

For the individual values chart, the centerline and limits are given by[18-19]

$$CL(x) = \bar{x} \qquad LCL(x) = \bar{x} - 2.66\,\overline{mR} \qquad UCL(x) = \bar{x} + 2.66\,\overline{mR} \tag{4}$$

Figure 3 shows the residuals from a linear fit to a set of calibration data for the mounting system roll potentiometer in the 16-Foot Transonic Tunnel using the Charts for Individuals for statistical analysis. For this particular case, the run of eight residuals in a row below the centerline is unlikely at 300:1 odds[16-19,24], so we'd have to say that the residuals don't constitute a random sample as required by the least-squares fitting procedure. A higher-order fit seems to be required.

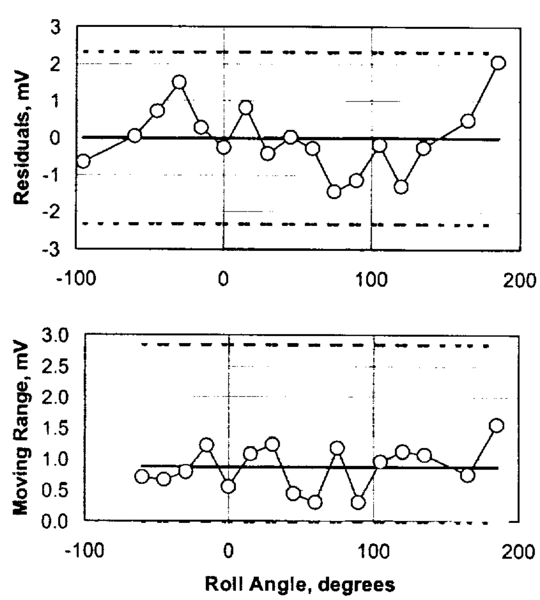

Figure 3 - x, mR charts for residuals for calibration of a roll potentiometer.

Eisenhart and Complex Statistical Control

Shewhart[3] established the basic statistical framework for measurement evaluation and control, but it was Churchill Eisenhart[5] who established the methodology as appropriate and practical for precision measurement in standards and calibration laboratories like the National Bureau of Standards (NBS), now the National Institute for Standards and Technology (NIST). Furthermore, Eisenhart recognized that the SQC methodology would have to deal with the fact that measurement processes have sources of variation whose expression depends upon the time frames involved. This notion is specifically recognized in the NIST[21] and ANSI/NCSL/ISO[14] measurement uncertainty guides as "repeatability" and "reproducibility". The definition for *repeatability* is given by them as:

> *... closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement ...*
>
> *Repeatability conditions include:*
>
> - *the same measurement procedure*
> - *the same measuring instrument[s], used under the same conditions*
> - *the same location*
> - *repetition over a short period of time*

Similarly, the definition for *reproducibility* is given as:

> *... closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement ... A valid statement of reproducibility requires specification of the conditions changed ...*
>
> *The changed conditions may include:*
>
> - *principle of measurement*
> - *method of measurement*
> - *observer*
> - *measuring instrument*
> - *reference standard*
> - *location*
> - *conditions of use*
> - *time*

To deal with the different time frames, Eisenhart proposed evaluating the measurement process for two time frames. He called the resulting evaluation "complex statistical control" as compared to the one-level variation control addressed by Shewhart which Eisenhart called "simple statistical control". In effect, one keeps track of two populations --- one short-term and the other long-term. Shewhart's range chart for grouped data is used to track the short-term dispersion. The Charts for Individuals are used to track the group means and their dispersion.

The combined charts are called 3-Way Charts by Wheeler.[16-17] The limits given in Eq. 1 are used for the group range data and the limits given in Eqs. 3 and 4 are used for the moving range and group average data respectively. When Eqs. 3 and 4 are used in 3-Way Chart analyses, the group averages are treated as individual values. A typical analysis is shown in Figures 4 and 5. Figure 4 shows individual repeat lift-coefficient values[‡] obtained on a single-element 2-D airfoil in the Low-Turbulence Pressure Tunnel. The measurements were made using the external balance. Groups of ten measurements each were made at 2° angle of attack at $M_\infty = 0.2$. The total pressure was 150 psia. The objective of this test was to determine the short-term (within-group) variation and the medium-term (within-test/between-group) variation for use in designing a future test with very stringent measurement uncertainty requirements.

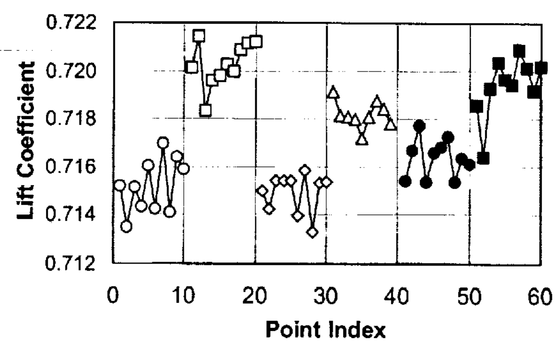

Figure 4 - Six groups of 10 individual values each of airfoil lift coefficient obtained over several months in the Low-Turbulence Pressure Tunnel.

The individual groups were obtained by simply pressing the data acquisition button ten times in a row. The groups were spaced out over time, separated by at least a few hours and a change in test conditions. In addition, the first four groups were separated from the last two by several months. The model was not removed from the tunnel during the interim nor were any changes made to the tunnel. It is clear from the data that the group-to-group variation is significantly larger than the within-group variation. In the terminology of two-level analysis, this means that the "between-group" variation is not negligible.

‡ We recommend plotting the individual repeat values for possible insight that might be lost if only group averages and ranges are examined.

The two-level statistical control analysis using 3-Way Charts is shown in Figure 5. The top plot shows the group ranges over time. The middle plot shows the group averages over time with limits derived from the moving-range chart. The bottom plot shows the moving ranges obtained from the group averages. The ranges, averages and moving ranges are all within their limits for this test. It is perhaps easiest to think loosely of the group range chart as tracking the stability of the short-term/repeatability/within-group dispersion. Similarly, the group average chart tracks the stability of the overall mean with respect to drift and intermittancy. The moving range chart then tracks the stability of the between-group dispersion. Note that for this example, the across-test stability cannot be considered checked because the model was not removed from the tunnel between groups.

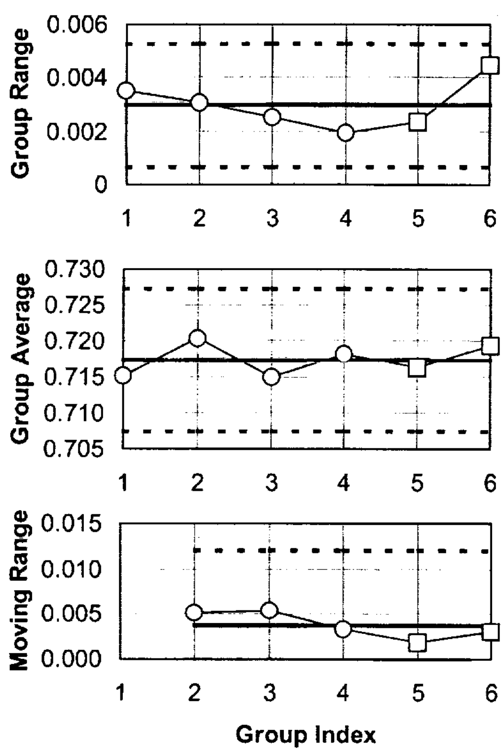

Figure 5 - Three-Way Charts for the LTPT lift-coefficient data shown in Figure 4.

Formulas for estimating the population standard deviations for the two levels of variation are given by[5]

$$\begin{aligned}\hat{\sigma}_{wg} &= \overline{R}/d_2 \\ &= 0.00298/3.078 \\ &= 0.00097\end{aligned} \tag{5}$$

and

$$\begin{aligned}\hat{\sigma}_{bg} &= \sqrt{\hat{\sigma}_{\bar{x}}^2 - \hat{\sigma}_{wg}^2/n} \\ &= \sqrt{\left(\overline{mR}/d_2\right)^2 - \hat{\sigma}_{wg}^2/n} \\ &= \sqrt{(0.00371/1.128)^2 - (0.00097)^2/10} \\ &= 0.0033\end{aligned} \tag{6}$$

where $\hat{\sigma}_{wg}$ is the estimate for the repeatability (within-group) and $\hat{\sigma}_{bg}$ is the estimate for the between-group variation. The estimated within-test reproducibility for a single measurement for the procedures followed is then

$$\hat{\sigma}_{within\text{-}test} = \sqrt{\hat{\sigma}_{wg}^2 + \hat{\sigma}_{bg}^2} = 0.0034 \tag{7}$$

The range and moving range are not unbiased estimators of the population standard deviation. Hence, in order to obtain $\hat{\sigma}$, it is necessary to divide them by d_2 which is a function only of the number of observations in a group. Note that the between-group variation is dominant even for this single test. This is typical of precision measurement processes.[5]

Cameron and Check Standard Testing

Consideration of the requirements for SQC in measurement quickly leads to the realization that some sort of surrogate process is needed as a stand-in for customer measurements. Such a process was introduced and perfected for the NBS standards and calibration laboratories by Joseph F. Cameron and coworkers.[22] The surrogate process consists of testing a stable artifact called a *check standard*[4,20,23] regularly.

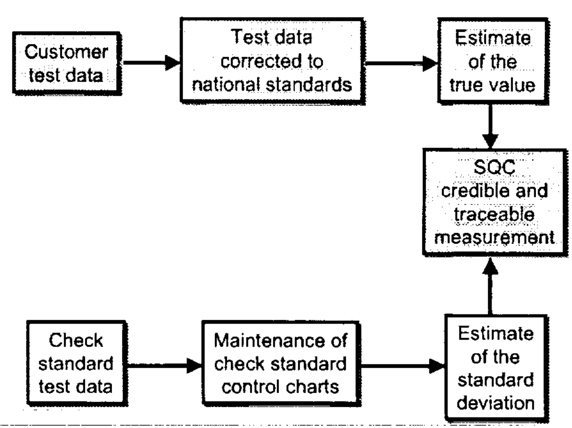

Figure 6 - Diagram for SQC measurement assurance using a check standard.

It is important to recognize the difference between reference and working standards and check standards. *Reference and working standards exist to assign* μ

to a measurement process. Check standards exist to assign σ. Reference and working standards are used for calibration. Check standards are used to determine the measurement uncertainty and to remove any doubt that the measurement process is stable and meaningful in a statistical (frequentist) sense.

With the addition of Cameron's contribution, the measurement assurance process consists of the familiar traceability through working and reference standards to national standards for the measurement of μ and the additional traceability for the measurement of σ through the regular testing of check standards. The process is shown diagrammatically in Figure 6.

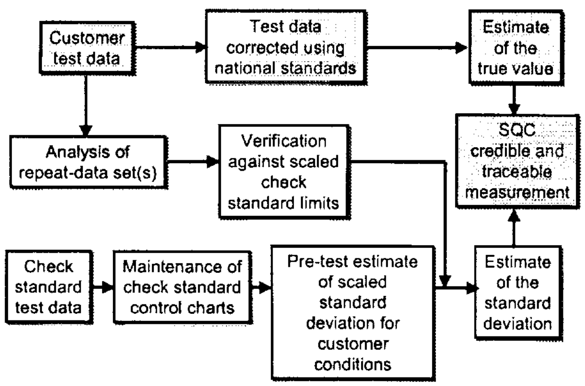

Figure 7 - Additions to SQC measurement assurance process of Figure 6 to include scaling and customer repeat-data verification.

Schumacher, Customer Repeat-Sets, and Scaling

Two more ideas are needed to complete the basic statistical framework for the Langley WTE data quality assurance methodogy. Both are due to Rolf B. F. Schumacher.[24-25] The first idea recognizes the need to verify that the measurement process remains in statistical control between check standard tests and during customer tests. Schumacher dealt with this issue in his standards and calibration laboratory at Rockwell International by obtaining a repeat-data set during calibration of a customer's measurement article.[24-25] Schumacher was dealing only with simple statistical control. Hence, he could verify that the measurement process was still in control by checking to see that the group range of the repeat-data set fell within the lower and upper control limits of the range chart established with check standard testing. We will have to modify his method somewhat for use with any process exhibiting two-level variation.

The other idea needed is that of scaling. If the customer's measurement article is not closely similar to that of the check standard, then the dispersion inferred from the check standard testing must be somehow scaled to the conditions of the customer test. For example, if the calibration measurement process for a resistor uses a check standard of 100 ohms, estimation of the variation associated with calibration of a resistor of nominally 50 ohms may require some scaling.

Figure 7 shows the SQC measurement assurance process of Figure 6 with the additional notions of customer repeat-data set verification and scaling.

LWTE Methodology and Examples

From the previous section, we have three elements to incorporate in our methodology for data quality assurance in the LWTE ground testing facilities: (1) Eisenhart's approach to statistical analysis using control charts as described above, together with Cameron's addition of check standard testing, (2) Schumacher's addition of customer repeat sets for checking the continuing stability of the measurement process and (3) Schumacher's addition of scaling to convert the check standard results to the customer's conditions. Our approach for the three elements is given in the rest of this main section.

We first consider the statistical control/check standard testing element. After roughly five years of experience with the SQC process in the LWTE facilities, we have found that there are three key activities associated with each check standard:

1. Selection and care of the standard;
2. Selection of the test matrix for check standard testing, including how often the tests will be conducted each year;
3. Analysis of the test data.

The LWTE procedures for carrying out those three activities for its two primary types of check standards are described in the next four sections. Our procedures for elements 2 and 3 --- customer repeat sets and scaling --- are described in the two sections following the check standard discussion.

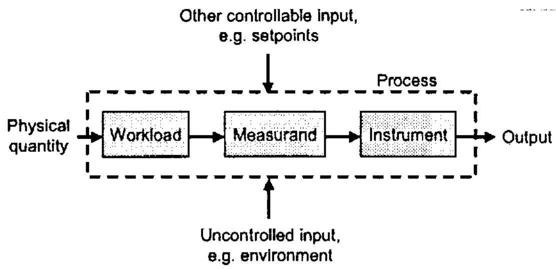

Figure 8 - Simple schematic of a measurement process.

Facility	Test Condition Check Standard
14- by 22-Foot Subsonic Tunnel[26]	Stand-mounted pitot-static probe on the centerline at the usual axial location of the model
Unitary Plan Wind Tunnel[27]	Wall tap upstream of the model nose (interim)
National Transonic Facility[28]	Sting-mounted pitot-static probe on the centerline at the model center of rotation
16-Foot Transonic Tunnel[29]	Sting-mounted pitot-static probe on the centerline at the model center of rotation
Low-Turbulence Pressure Tunnel[30]	Sting-mounted pitot-static probe on the centerline near the model center of rotation
0.3-Meter Transonic Cryogenic Tunnel[31]	Pitot probe on the wake rake and an array of wall taps
8-Foot High Temperature Tunnel[32]	Pitot and static pressure probes and a temperature probe on a rake in the test section
Jet Exit Test Facility[33]	Standard nozzles with pitot and static pressure probes on rakes
Transonic Dynamics Tunnel[34]	Sting-mounted pitot-static probe on the centerline at the model center of rotation

Table 2 - Check standards used for the test condition measurement process in the subscribing WTE facilities.

Selection and Care of the Check Standards

As noted above, the SQC approach to data quality assurance uses check standards as surrogates in the measurement processes of interest.[4] In fact, it is useful to think of any artifact used as a check standard as simply a means of applying a work load to the instrument(s) of interest in the presence of the "environment" (i.e. the tunnel, etc.) as shown in Figure 8. Consequently, it is important to select check standard artifacts which (1) represent the measurement process of interest adequately and (2) can be maintained with sufficient care that the population mean does not change significantly. At present, for the subscribing facilities, the LWTE uses the artifacts listed in Tables 2 and 3.*

Test Condition Check Standard

The artifacts listed in Table 2 are used to evaluate the measurement process for the test conditions, e.g. q_∞ and M_∞, for the test facilities. All of these artifacts use static pressure tap orifices. Since the bias error associated with imperfect static pressure orifices can easily exceed the variation due to the measurement process[35] and since it is extremely easy to affect those orifice bias errors by nicking, denting, etc., the orifice[36], it is crucially important to protect the orifices. In fact, we recommend that the check standard artifacts for test conditions not be used for any other purpose and that they be used and stored with extreme care.

Facility	Airframe Check Standard
14- by 22-Foot Subsonic Tunnel[26]	High-wing subsonic transport (Elliptical Wing)
Unitary Plan Wind Tunnel[27]	Generic fighter (previously supersonic transport HSR TCA 2A and 2B)
National Transonic Facility[28]	Low-wing subsonic transport (Pathfinder I)
16-Foot Transonic Tunnel[29]	Supersonic transport (HSR TCA 2A)
Low-Turbulence Pressure Tunnel[30]	EA-6B single-element airfoil
0.3-Meter Transonic Cryogenic Tunnel[31]	NACA 0012 airfoil
8-Foot High Temperature Tunnel[32]	No model
Jet Exit Test Facility[33]	3 calibration nozzles
Transonic Dynamics Tunnel[34]	Generic fighter model (static testing)

Table 3 - Check standards used for force- and moment-coefficient measurements in the subscribing WTE facilities.

Airframe Check Standard

Selection of a check standard airframe (see Table 3) for application of the test environment workload to a balance requires attention similar to that for the test condition standard. Although our experience with this particular kind of check standard is rather brief

* The statistical control verification process for test conditions could use an actual calibration system rather than a separate surrogate. However, this is not always feasible because of the long installation time requirements for some calibration systems which can be unacceptable for a process that must be repeated several times a year.

(roughly two years), we have found that the following considerations are important:

1. Since separated flow would likely add variation that is model (artifact) dependent, it is recommended that the check standard model be used in conditions producing attached flow only. Scaling for conditions of separated flow in a customer test will be discussed below.
2. Since the model may also be used as a check standard for measuring test section flow angularity, it is recommended that the slope of the model normal-force coefficient be comparable to those normally tested in the tunnel. For subsonic and transonic tunnels, a compromise on a fighter model may be appropriate.
3. It should be possible to build up the model in a prep bay so that test section occupancy time is kept to a minimum.
4. The model should be able to accommodate a balance that is reasonably representative of the balances normally used in the tunnel.
5. As with the test condition check standard, extreme care should be observed for the model check standard. Of special concern are the leading and trailing edges, the surface finish and boundary-layer trips.

Selection of the Check Standard Test Matrices

The issues associated with choosing a test matrix for check standard testing seem to be: (1) economy, (2) operating regime coverage and (3) use of the resulting estimates of population dispersion in statements of measurement uncertainty. After struggling with these issues for several years, we have recently settled on measuring two levels of variation: (1) back-to-back repeatability and (2) test-to-test reproducibility.‡ We are attempting to carry out four tests of each kind of artifact each year with the only rule being that the check standard tests must be separated by at least one customer test.

Test Condition Matrix

Each test using a test condition artifact of Table 2 is designed to require no more than one day of test section occupancy. The test matrix consists of 6-10 test conditions. At each test condition, three data points are obtained back-to-back to form a group. Only one group is obtained in a given test at each condition. While it is important to cover a reasonable portion of the test condition envelope, it is often impractical, especially in pressure and cryogenic tunnels, to cover the entire range.

We have recently adopted the practice of covering a fraction of the envelope so that we can relatively easily meet our economy goals and still check a fairly wide range of conditions. In addition, we recommend to the facility that, perhaps once a year, it obtain groups at other important conditions to see if the variation observed is consistent with the normally covered range. For example, we are presently conducting check standard testing in the NTF at two total pressures and one temperature (120 F) for a wide range of dynamic pressure as shown in Table 4 below. We will check the variation at cryogenic temperatures and higher Mach numbers less frequently.

p_{total}, psia	q_∞, psf	M_∞
18	15.0	0.091
18	34.0	0.138
18	77.2	0.209
50	175.0	0.189
50	396.8	0.289
50	900.0	0.454
50	1422.7	0.600

Table 4 - Test conditions for check standard testing in the National Transonic Facility.

Airframe Matrix

The check standard tests using the airframe artifacts of Table 3 are designed to require no more than two days of test section occupancy. The test matrix consists of groups of five back-to-back pitch-only (AOA) runs taken at the same conditions used for the Table 2 artifacts. The five runs are taken as follows:

1. Inverted
2. Upright
3. Upright
4. Upright
5. Inverted

This grouping presumes a remote roll capability. If such a capability is not available, we obtain only the upright runs. Each run consists of 8-10 data points covering the linear, attached-flow region. The first two runs are combined to obtain a measurement of the flow angularity. The last two runs are used similarly. Together the two measurements form a flow-angularity group. The three upright runs form a group for measurement of the balance normal-force,

‡ Initially, we did not appreciate the difficulty of sorting out the various levels of variation without carefully and narrowly defining what we mean by repeatability and reproducibility. The difficulty is exacerbated when $\sigma^2_{bg} >> \sigma^2_{wg}$. Most of the data used in the examples presented in this paper were *not* obtained according to the procedures listed above. The actual procedures used are described in the text.

axial-force and pitching moment coefficients. To avoid mixing results from different instruments, we now compute the coefficients in the balance axis system and no base, cavity or wall corrections are applied.

Analysis of the Test Condition Data

We are currently using two types of statistical analyses for the test condition data: (1) Shewhart's control charts and (2) simple time series plots of individual values. Both of these analyses result in easily interpreted visual displays. The first, of course, is an integral part of the approach that has been adopted by the LWTE. We have found that the second type of charts is helpful in diagnosing measurement system troubles during a test and in explaining the control chart results to those who are not familiar with them. We'll first describe our use of control charts and then describe the other type of display.

Control Charts

Each group of test condition check standard data is processed and added to the historical statistical control charts *for that test condition* as follows:

1. The group range and average are computed from the three back-to-back points for the properties of interest.
2. The group range and average are then added to the historical control charts.
3. If the previous 10 groups indicate that the measurement process is in control, the points are added to the charts without changing the chart centerlines and limits.
4. If fewer than 10 groups have been taken previously or if the previous 10 groups do not show that the measurement process is in control, the new group range and average are included with the previous groups to recompute the chart centerline and limits.
5. The charts are checked to see if the process indicates that it is in control. If it does not so indicate, then action must be taken to correct the problem.[15-19]

A set of 3-Way Charts is maintained for each property of interest and each condition in the test matrix. An example is given in Figure 9 for the probe used in the 14- by 22-Foot Subsonic Tunnel. For check standard testing, the probe is mounted on a test stand on the centerline of the test section at one of the most-used model locations. The only test condition property analyzed for that tunnel is the calibration coefficient defined by

$$C' = \frac{(p_{total} - p_{static})_{probe}}{(p_{total} - p_{static})_{reference}} \qquad (8)$$

The quantity $\sigma_{C'}$ can be shown to be roughly equal to $\sigma_{q_\infty} / q_\infty$. For the data of Figure 9, the test section walls, ceiling and floor were in the closed position and the boundary-layer removal system was off. The data were obtained in five separate tests over roughly 3½ years.

As noted above, the present procedure of obtaining back-to-back data points for a group was adopted recently and the groups for the tests of Figure 9 were obtained differently. For those five tests, the data points in a group were obtained as part of a set of repeat dynamic-pressure sweeps. Between each sweep, the fan was turned off and a wind-off zero was obtained. Nevertheless, the data and the charts provide a good example of the method of analysis used for test condition check standard testing.

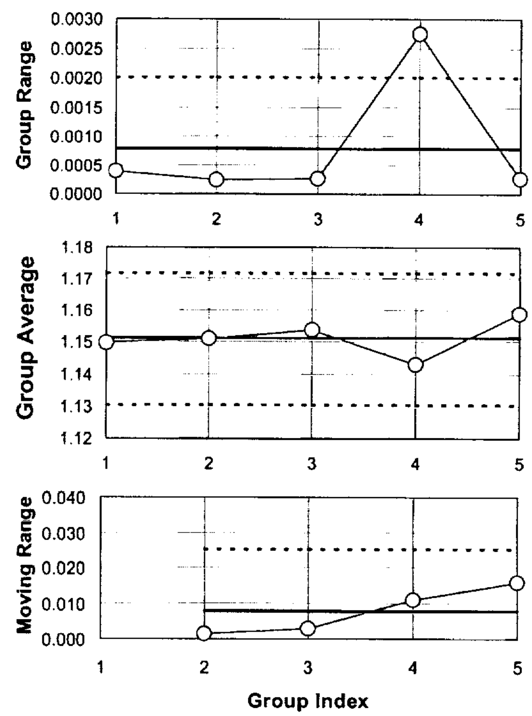

Figure 9 - Three-Way Charts for test condition check standard data, parameter C', obtained in the 14- by 22-Foot Subsonic Tunnel at $q_\infty = 60$ psf.

Reading Figure 9, we see immediately that the measurement process is not in control because the range of the group obtained in the fourth test is above the upper limit. We do expect, however, that the charts will exhibit control from group 5 on because the variation experienced in group 4 is not likely to happen with back-to-back points. (We suspect that the excessive within-group variation of group 4 is due to problems with sealing the test section and the plenum.)

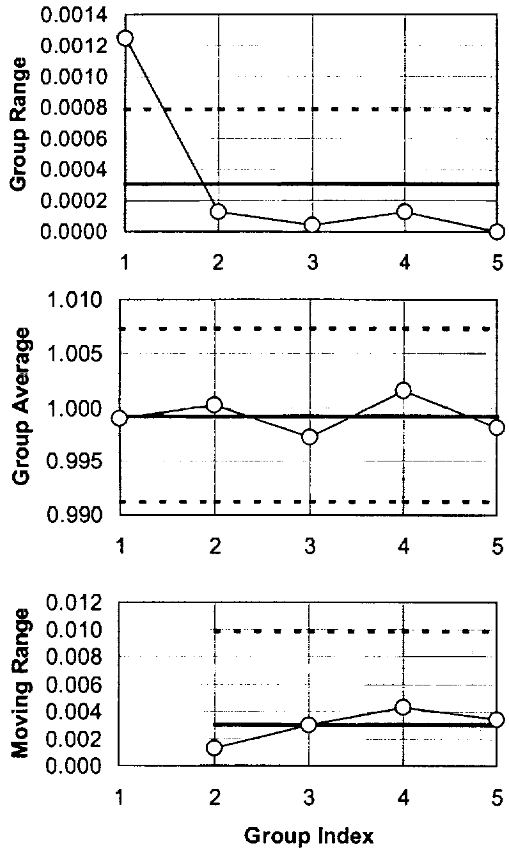

Figure 10 - Three-Way Charts for test condition check standard data, parameter C_{Mach}, obtained in the Unitary Plan Wind Tunnel at $M_\infty = 2.4$.

Even though the measurement process is not in statistical control, we will compute the within-group and between-group estimates of $\hat{\sigma}_{wg}$ and $\hat{\sigma}_{bg}$ as before for illustration purposes:

$$\hat{\sigma}_{wg} = \bar{R}/d_2 = 0.00299/1.693 = 0.00177 \quad (9)$$

and

$$\hat{\sigma}_{bg} = \sqrt{\left(\overline{mR}/d_2\right)^2 - \hat{\sigma}_{wg}^2/n} = \sqrt{(0.00882/1.128)^2 - (0.00177)^2/3} = 0.00775 \quad (10)$$

so that

$$\hat{\sigma}_{single\ measurement} = \sqrt{\hat{\sigma}_{wg}^2 + \hat{\sigma}_{bg}^2} = 0.0079 \quad (11)$$

Another example is provided from test condition data obtained in Test Section 2 of the Unitary Plan Wind Tunnel at $M_\infty = 2.4$. These data were obtained during back-to-back angle-of-attack sweeps of the check standard model. Since the flow in the test section is supersonic, the wall tap should not be affected by the presence of the model. The data for this example are shown in Figure 10. They were taken at one of the canonical angles of attack and, consequently, were not obtained back-to-back as prescribed in the present WTE procedures noted above. The test condition property analyzed for this example is the Mach number ratio defined as

$$C_{Mach} = \frac{M_{wall\ tap}}{M_{reference}} \quad (12)$$

The quantity $\sigma_{C_{Mach}}$ can be shown to be roughly equal to $\sigma_{M_\infty}/M_\infty$.

As with the results of Figure 9, one of the groups on the range chart of Figure 10 is above the upper limit. One of the advantages of using the average range (and the average moving range) to make estimates of $\hat{\sigma}_{wg}$ and $\hat{\sigma}_{bg}$ versus a global estimate using the pooled variances is that the range estimates are more robust for out-of-control processes like those of Figures 9 and 10. This does not mean that we don't have to work to get the processes into control. It just means that estimates made early in the effort are more likely to reflect the future in-control values.

Again, we estimate $\hat{\sigma}_{wg}$ and $\hat{\sigma}_{bg}$ as before:

$$\hat{\sigma}_{wg} = \bar{R}/d_2 = 0.000308/1.693 = 0.000182 \quad (13)$$

and

$$\hat{\sigma}_{bg} = \sqrt{\left(\overline{mR}/d_2\right)^2 - \hat{\sigma}_{wg}^2/n} = \sqrt{(0.00301/1.128)^2 - (0.000182)^2/3} = 0.00267 \quad (14)$$

so that

$$\hat{\sigma}_{single\ measurement} = \sqrt{\hat{\sigma}_{wg}^2 + \hat{\sigma}_{bg}^2} = 0.0027 \qquad (15)$$

Note that we would have one set of 3-Way Charts for each test condition and test property to be tracked. This is normally not onerous if only 6-10 test conditions and one or two test properties are used. So far we have been tracking the calibration coefficient, C', for subsonic tunnels and q_{probe} / q_{ref} and M_{probe} / M_{ref} for transonic and supersonic tunnels.

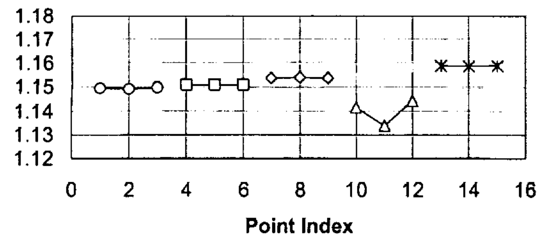

Figure 11 - Individual data points for five test condition check standard tests in the 14- by 22-Foot Subsonic Tunnel at $q_\infty = 60$ psf. The test condition property analyzed is C'.

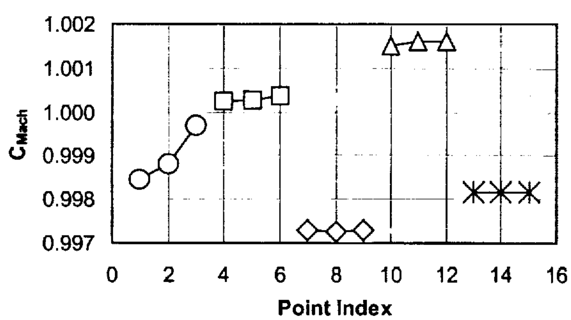

Figure 12 - Individual data points for five test condition check standard tests in Test Section 2 of the Unitary Plan Wind Tunnel at $M_\infty = 2.4$.

Plots of Individual Values

In addition to the control charts described in the previous subsection, we have found that time series plots of individual values are helpful for interpretation when a process is not in control. Furthermore, since most people are not yet used to the methods of analysis described in this paper, it is often easier to talk about the issues using both sets of charts. The individual values corresponding to the charts of Figures 9 and 10 are presented in Figures 11 and 12 respectively.

Finding that a test condition measurement process is out of control after the check standard test is complete is not nearly as helpful as finding it out during the test. So we recommend that the charts be updated during the test, by hand if necessary. This can be accomplished by simply marking the new points on the charts without recalculating the centerlines and limits.

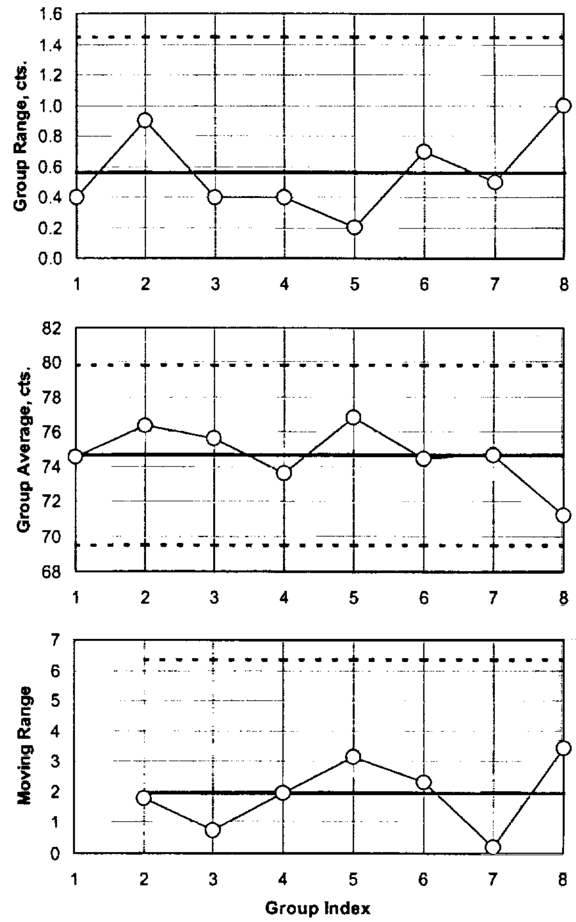

Figure 13 - 3-Way Charts for HSR check standard model in the 16-Foot Transonic Tunnel. Axial-Force Coefficient, $M_\infty = 0.9, \alpha = 2°$.

Analysis of the Airframe Balance Data

Statistical analysis of the airframe data is considerably more complicated than is normally found in the SQC measurement assurance literature for calibration laboratories because of the sheer volume of data. For example, the NTF test matrix has 3 test properties (C_N, C_m, C_A), 7 test conditions and 10 angles of attack. If we follow the same scheme given above for test condition check standard testing, we would have 210 sets of 3-way Charts for the NTF, not including the charts for flow angularity. Such a huge amount of data would make it very difficult to keep track of the statistical control state of the tunnel's balance measurement process.

It is possible to avoid dealing with such a huge amount of data (replicated for each tunnel!) by taking advantage of our procedural design described earlier. Since we are only concerned with attached-flow conditions, we expect that the variation to be observed in that angle of attack region will be, *on the average*, independent of the angle of attack. Hence, we need keep track of the results at only a single angle of attack. The number of sets of control charts is thus reduced to one for each test property and test condition. In addition, each time a group is obtained, we create a set of 3 quick-look plots to (1) review the individual data points for gross blunders/surprises, (2) review the *scatter* of the repeat points in the group also for gross blunders/surprises, and (3) check once again that the (within-group) variation is independent of angle of attack in the attached-flow region.

In the next two subsections, we describe our analysis procedures together with several examples.

Control Charts

For each tunnel and test condition, a nominal angle of attack must be chosen for the control chart analysis. So far we have used an angle in the middle of the attached-flow region.[††] After a group of three back-to-back runs has been obtained at a given test condition during a check standard test, linear interpolation is performed in each run to obtain the values of C_N, C_m, C_A at the chosen angle of attack. The interpolation takes out any scatter due to set point variation. The three values of each test property obtained this way form a group and are analyzed in the same way discussed above for test condition check standard data.

The first example for this subsection is the control chart analysis for the 16-Foot Transonic Tunnel's check standard model. For the tests shown here, the runs were *not* obtained back-to-back but rather were separated by a single run at another test condition. The results for the uncorrected axial-force coefficient in body coordinates are given in Figure 13. The population standard deviations for the two levels of variation are given by[5]

$$\begin{aligned}\hat{\sigma}_{wg} &= \bar{R}/d_2 \\ &= 0.562/1.693 \\ &= 0.332 \text{ counts}\end{aligned} \qquad (16)$$

and

$$\begin{aligned}\hat{\sigma}_{bg} &= \sqrt{\hat{\sigma}_{\bar{x}}^2 - \hat{\sigma}_{wg}^2/n} \\ &= \sqrt{\left(\overline{mR}/d_2\right)^2 - \hat{\sigma}_{wg}^2/n} \\ &= \sqrt{(1.95/1.128)^2 - (0.332)^2/3} \\ &= 1.72 \text{ counts}\end{aligned} \qquad (17)$$

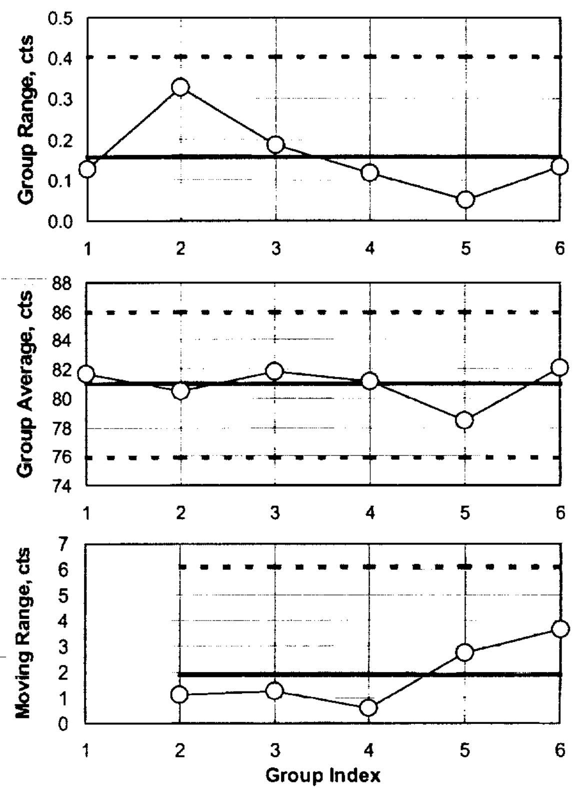

Figure 14 - 3-Way Charts for HSR check standard model in Test Section 2 of the Unitary Plan Wind Tunnel. Axial-Force Coefficient, $M_\infty = 2.4, \alpha = 2.5°$.

The second example for this subsection is the control chart analysis for the Unitary Plan Wind Tunnel's check standard model in Test Section 2. The results for the uncorrected axial-force coefficient in body coordinates are given in Figure 14. The population standard deviations for the two levels of variation are given by[5]

†† We have assumed that the attached flow region corresponds to the region of minimum scatter.

$$
\begin{aligned}
\hat{\sigma}_{wg} &= \bar{R}/d_2 \\
&= 0.157/1.693 \\
&= 0.093 \text{ counts}
\end{aligned}
\tag{18}
$$

and

$$
\begin{aligned}
\hat{\sigma}_{bg} &= \sqrt{\hat{\sigma}_{\bar{x}}^2 - \hat{\sigma}_{wg}^2 / n} \\
&= \sqrt{\left(\overline{mR}/d_2\right)^2 - \hat{\sigma}_{wg}^2 / n} \\
&= \sqrt{(1.89/1.128)^2 - (0.093)^2/3} \\
&= 1.67 \text{ counts}
\end{aligned}
\tag{19}
$$

Both examples (Figures 13 and 14) show no evidence of a lack of statistical control for either repeatability or reproducibility. However, there are too few groups as yet for either tunnel to be able to claim definitively that the balance measurement process is in statistical control.

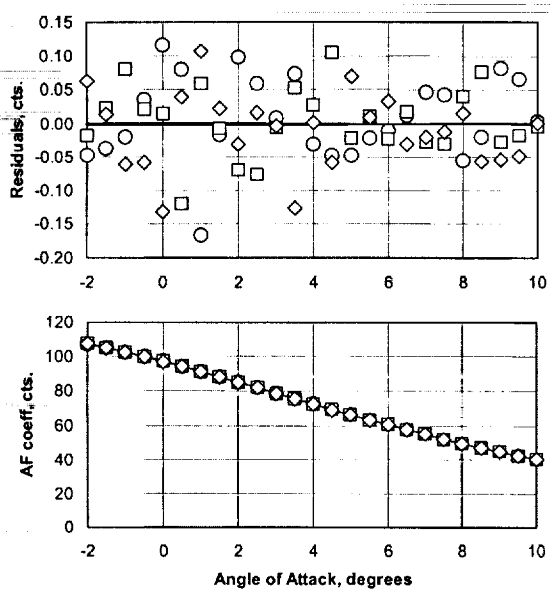

Figure 15 - Quick-look data for check standard model, Unitary Plan Wind Tunnel, Group 6, Axial-Force Coefficient, M=2.4.

It is interesting to note that the Unitary Plan Wind Tunnel's repeatability is about 3 times better than the repeatability of the 16-Foot Transonic Tunnel, although the two tunnels have comparable levels of reproducibility. It is likely, in this case, that the difference in repeatability is due to the different procedures used to obtain a group rather than anything inherent in the tunnels, ***once again illustrating the need to rigorously define and follow set procedures.***

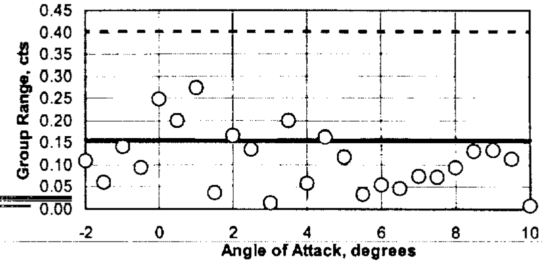

Figure 16 - Quick-look range plot for check standard model group. Data for Unitary Plan Wind Tunnel model at M=2.4 (Figure 15).

Quick-look plots

In the discussions above, we pointed out the importance of looking at the individual data as well as the aggregate group range and average data in the control charts (e.g. Figures 4, 11, 12). For check standard model testing, we prepare "quick-look plots" for each group for that purpose. We have added an additional feature to the quick-look plots for the airframe check standard testing. This feature allows us to check our assumption that the variation at the chosen angle of attack is representative of all of the angles of attack for which attached flow is to be expected. Two examples are given in Figures 15-18.

The first example is for the uncorrected axial-force coefficient for the Unitary Plan Wind Tunnel's check standard model. The data in the upper plot in Figure 15 and in Figure 16 are obtained in the following way:

1. Linear interpolation is performed for each run to obtained data estimates at the exact nominal angles of attack. This procedure is performed to remove the effects of set point variation.
2. The interpolation results at each of the nominal angles of attack are averaged.
3. The resulting averages at each nominal angle of attack are subtracted from the original data estimates to obtain the residuals shown in the top plot of Figure 15 and the ranges shown in Figure 16.

Figure 15 shows the original coefficient data and the residuals about the averages at each nominal angle of attack. Figure 16 shows the ranges of the residuals at the nominal angles of attack. The data of Figure 16 behave as expected from the previous 5 tests for the attached flow region (roughly 0 to 4 degrees) and show no special behavior for the angle of attack chosen for the statistical control charts.

The second example is for the uncorrected axial-force coefficient for the 16-Foot Transonic Tunnel's check standard model. The groups were not obtained with back-to-back runs as noted above for the UPWT check standard testing. Rather, at least one other run at a different Mach number was obtained between each check standard run in the group.

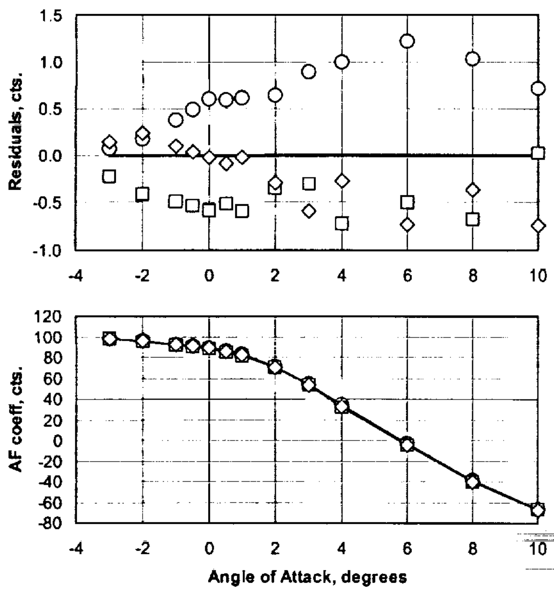

Figure 17 - Quick-look data for check standard model, 16-Foot Transonic Tunnel, Group 8, Axial-Force Coefficient, M=0.9.

The lower plot in Figure 17 shows the actual data obtained in the 3-run set with the upper plot in Figure 17 and Figure 18 showing the residuals and the range respectively at each nominal angle of attack. As noted above, one element of the procedure for creating the airframe control charts is to choose a single nominal angle of attack in the middle of the attached flow region. Typically, the scatter is smallest for that region. That procedure was followed in choosing $\alpha = 2°$ for the control charts for the 16-Foot Transonic Tunnel (Figure 13). Hence, the fact that the ranges for $\alpha \geq 3°$ are at or above the upper control limit suggests a change in the process. For this particular test, two on-board accelerometers were installed for the measurement of α and the offsets to α due to model and mounting system vibration. Unfortunately, the instrument cables were rather large and stiff and the tunnel staff were concerned that the axial force measurements would be affected. The results shown in Figure 18 seem to confirm their fears.

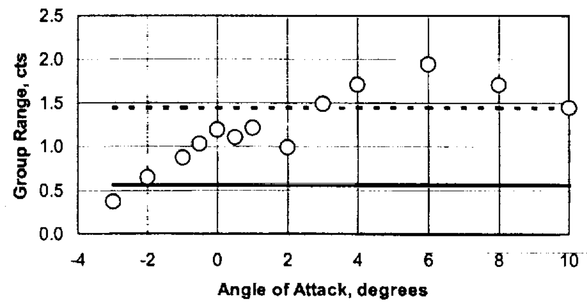

Figure 18 - Quick-look range plot for check standard model group. Data for 16-Foot Transonic Tunnel model at M=0.9 (Figure 17).

Customer Repeat Sets and Scaling

As noted above, Schumacher[24] pointed out that it is necessary to obtain statistical information *during customer tests* as well as during check standard tests. And we need a method for comparing the statistical information from the two kinds of tests (scaling). This information is also needed for preparing the customer reproducibility statement. In this section, we will first describe the problem of scaling and our interim approach to it. Then, we will show how we check the within-group variation (repeatability) in a customer test followed by our method for checking the within-test variation (reproducibility). Finally, we will describe how we check the scaling method and how we report the measurement reproducibility to the customer, including the dispersion for the separated-flow regions.

Interim Method for Scaling

Proper scaling requires that we know the dependence of σ_{wg} and σ_{bg} on tunnel conditions and instrument properties for each of our tunnels. Unfortunately, it is beyond the scope of this paper to present a complete discussion of scaling and σ dependence and we have not yet acquired sufficient data to be definitive about the dependence. Therefore, in order to do scaling for the interim, we will have to make some assumptions and the information that we do have suggests that σ_{wg} and σ_{bg} depend primarily on the tunnel dynamic pressure and the full-scale limit of the instrument of concern (e.g. axial-force beam of a balance, individual port of a multi-port pressure transducer, etc.). In fact, the data suggest two different types of dependence.

For the lowest portion of the dynamic pressure range in a given tunnel, both σ_{wg} and σ_{bg} seem to be proportional to the full-scale limit of the instrument, i.e.

$$\sigma = A \tag{20}$$

where A is a constant. Furthermore, the value of A_{wg} seems to be reasonably close to the combined repeatability, nonlinearity and hysteresis that is typically reported for each class of instrument. For example, A_{wg} might be on the order of 0.05% of the full-scale limit of a balance beam. We do not have enough data to draw conclusions about A_{bg}.

For most of the dynamic pressure range (middle and high), both σ_{wg} and σ_{bg} seem to be proportional to the dynamic pressure, i.e.

$$\sigma = B q_\infty \tag{21}$$

The relationship, if any, of B to the instrument being used is not known. We do know that typically

$$B_{bg}^2 >> B_{wg}^2 \tag{22}$$

For the interim, we are assuming that B is proportional to A. This assumption then leads to scaling both σ_{wg} and σ_{bg} by the full-scale limit of the instrument.

Since typical balance and pressure data are presented as coefficients with q_∞ in the denominator, we expect to see force, moment and pressure *coefficient* dispersion that is inversely proportional to q_∞ at low dynamic pressures and independent of q_∞ at the middle and higher dynamic pressures.

Although the evidence is sparse, the two terms for σ seem to be somewhat correlated so that a reasonable representation of the dynamic pressure dependence for balance and pressure coefficients could be

$$\sigma = \sqrt{\left(\frac{A}{q_\infty}\right)^2 + 2\rho\frac{AB}{q_\infty} + B^2} \tag{23}$$

where ρ is the correlation coefficient.[6] Thus, for no correlation, we would have

$$\sigma = \sqrt{\left(\frac{A}{q_\infty}\right)^2 + B^2} \tag{24}$$

and for perfect correlation, we would have

$$\sigma = \frac{A}{q_\infty} + B \tag{25}$$

Parameter	Airframe Check Standard (UPWT)	Customer (Test 1729)
Balance limit	60 lbf	85 lbf
Reference area	2.385 ft^2	0.3277 ft^2
Dynamic pressure	840 psf	540 psf
Mach	2.4	3.0
$\hat{\sigma}_{wg}$	0.093 counts	0.96 counts
$\hat{\sigma}_{bg}$	1.67 counts	17.2 counts
CL for R	0.157 counts	1.62 counts
LCL for R	0	0
UCL for R	0.404 counts	4.17 counts
CL for mR	1.89 counts	19.5 counts
LCL for mR	0	0
UCL for mR	6.17 counts	63.7 counts

Table 5 - Parameters for comparison of axial force measurements in check standard and example customer test in the Unitary Plan Wind Tunnel.

To illustrate the interim scaling approach, we will scale the airframe check standard results of the Unitary Plan Wind Tunnel (Figure 14) for the axial-force coefficient to the conditions/instruments used in Test 1729 and use those results in the next two subsections to compare the scaled limits to the repeat-run results from that test. A comparison of the axial force limits, reference areas and tunnel conditions is given in Table 5. For this case we assume that the "A" contribution to the variation is negligible and derive the scale factor (SF) as shown next.

Neglecting the "A" factor contribution, we have from Eq. 21

$$\sigma_{AF} = B q_\infty \tag{26}$$

Assuming also that B is proportional to the balance limit, we have

$$B = C\, AF_{max} \tag{27}$$

where AF_{max} is the axial force limit and C is a constant for axial force for the balance class of interest in the tunnel of interest. Substituting Eq. 27 into Eq. 26 and rearranging gives

$$C = \frac{\sigma_{AF}}{AF_{max}\, q_\infty} \tag{28}$$

But the dispersion of the axial force is related to the dispersion of the axial-force coefficent by[#]

[#] Note that we actually handle the very short-term variation of q_∞ (frame-to-frame) by computing and storing C_A for a data point.

$$\sigma_{AF} = \sigma_{C_A} S_{ref} q_\infty \quad (29)$$

Substituting Eq. 29 into Eq. 28 gives an expression for relating the check standard dispersion of the axial-force coefficient to that of the customer test:

$$C = \frac{\sigma_{C_A} S_{ref}}{AF_{max}} \quad (30)$$

Since C is assumed to be independent of conditions and instruments of the same class, we can use the above expression to write

$$\frac{\sigma_{C_A,customer} S_{ref,customer}}{AF_{max,customer}} = \frac{\sigma_{C_A,standard} S_{ref,standard}}{AF_{max,standard}} \quad (31)$$

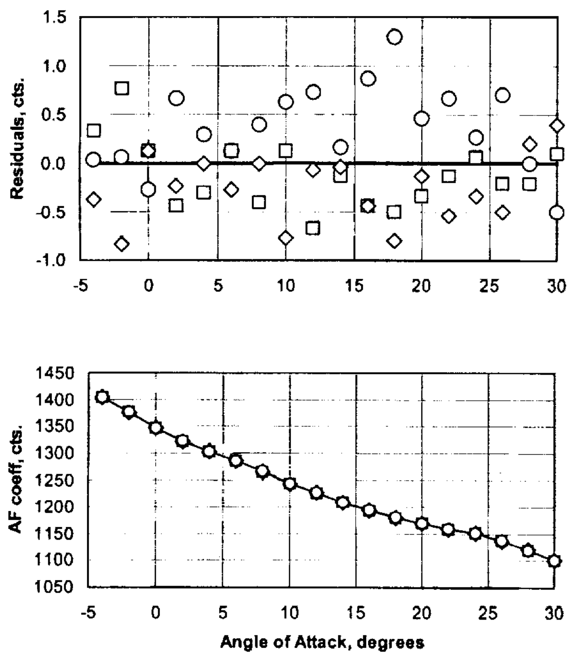

Figure 19 - Quick-look plots of original data and residuals for example customer test of Table 5. Repeat run-set 1, beginning of the test.

Rewriting Eq. 31 gives

$$\sigma_{C_A,customer} = \sigma_{C_A,standard} \left(\frac{S_{ref,standard} AF_{max,customer}}{S_{ref,customer} AF_{max,standard}} \right) \quad (32)$$
$$= \sigma_{C_A,standard} SF$$

For the comparison of Table 5, the scale factor SF is

$$SF = \frac{(2.385\, ft^2)(85\, lbf)}{(0.3277\, ft^2)(60\, lbf)} = 10.3 \quad (33)$$

It is easy to show that this scale factor also applies to the centerlines and limits of the control charts.

Within-Group Checks

Our within-group checks during a customer test are conducted as follows:

1. Prior to the test, pre-test estimates of σ_{wg} and σ_{bg} for the uncorrected normal- and axial-force and pitching-moment coefficients in balance coordinates are made using the scaling described in the previous subsection (see Table 5).
2. One repeat-run set at one test condition is obtained at the onset of the test. The data are analyzed and displayed as described above for a check standard test with the exception that the range plot uses scaled centerlines and control limits based on the value of $\hat{\sigma}_{wg}$ found in Step 1.
3. An identical repeat-run set is obtained and analysis conducted at the conclusion of the test for the same configuration and test condition.

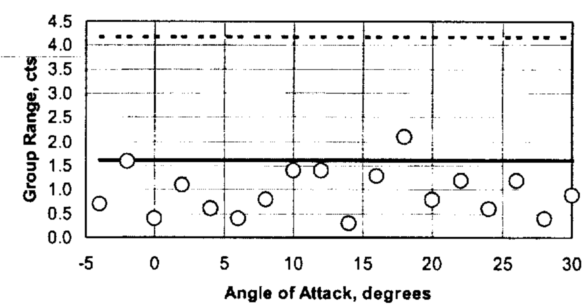

Figure 20 - Quick-look within-group range plot for example customer test of Table 5. Repeat run-set 1, beginning of the test.

An example within-group check analysis for a customer test is given in Figures 19-22. The scaling information is given in Table 5 and was derived in the previous subsection. We remind the reader that there is nothing significant about the dispersion results lying below the scaled centerline. Such a result is to be expected for roughly half of the samples from any measurement process. If *all* of the customer results over many tests lay below the centerline (or all above it), then we would have to reconsider the scaling process. We will discuss that subject in more depth two subsections below.

For Test 1729, then, the repeat-run checks show no evidence that the balance measurement process is out-of-control for the within-group variation compared to the scaled check standard data.

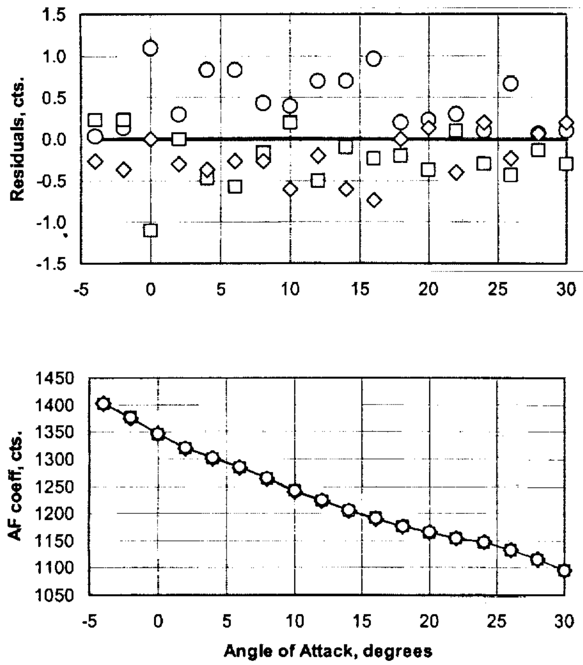

Figure 21 - Quick-look plots of original data and residuals for example customer test of Table 5. Repeat run-set 2, end of the test.

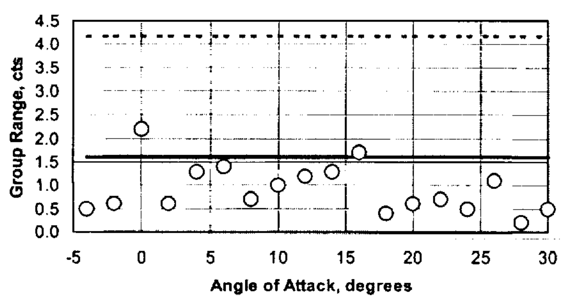

Figure 22 - Quick-look within-group range plot for example customer test of Table 5. Repeat run-set 2, end of the test.

Within-Test Checks

Within-test checks for a customer test are carried out as follows: The averages of the data for the two run sets, at the nominal angles of attack, are subtracted. The absolute values of those differences are then displayed in the same manner as for individual repeat-run sets with the exception that the values are compared to the scaled centerlines and limits for the moving range. The within-test check for the example of Table 5 is given in Figures 23 and 24.

An objection could be raised that it not appropriate to compare within-test dispersion to across-test dispersion. Before we adopted the present procedures we obtained several groups in a typical check standard test. Although the results are sparse, it seems reasonable to assume for the interim that the within-test variation might account for roughly 80% of the across-test variation. We will suggest a way of examining this estimate in the next subsection.

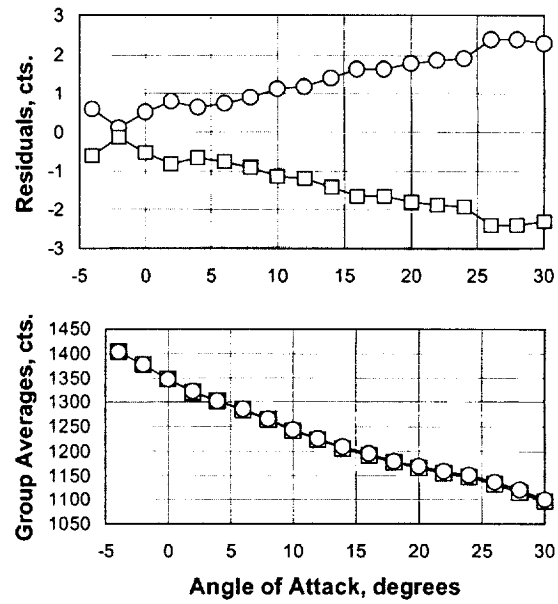

Figure 23 - Quick-look plots of averages and residuals about the grand averages for example customer test of Table 5.

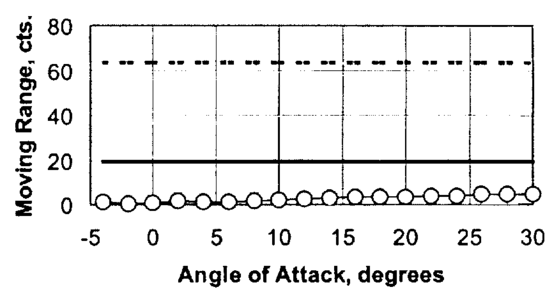

Figure 24 - Quick-look within-test moving-range plot for example customer test of Table 5.

As above for the within-group variation, there is no evidence that the between-group variation is out-of-control.

Control-Chart Checks of Scaling Method

The central problem of scaling is determining whether it is satisfactory in the face of the wide confidence limits to be expected for sampling for dispersion. For example, the confidence limits for the population standard deviation (simple statistical control) are given in Figure 25 as a function of the number of observations. It is obvious from the figure that small numbers of observations are not sufficient and that we need roughly 20 observations (19 degrees of freedom) to capture the dispersion appropriately for measurement uncertainty statements. Furthermore, since the across-test reproducibility is by far the dominant level of variation, it is clear that we need 20 check standard tests to evaluate it satisfactorily.

At present, of course, we are also lacking definitive information regarding the dependence of the dispersion on the instrument limits. Our approach to dealing with this problem is to use control charts to track the actual dispersion results for the customer tests scaled to the conditions of the check standard tests. However, we have not yet compiled enough data using the recently adopted procedures to confidently interpret the results.

Reporting Measurement Reproducibility

In our view the ANSI/NCSL/ISO Guide is definitive in the matter of reporting measurement uncertainty. We quote from section 7.1.4 of the Guide[14]

Although in practice the amount of information necessary to document a measurement result depends on its intended use, the basic principle of what is required remains unchanged: when reporting the result of a measurement and its uncertainty, it is preferable to err on the side of providing too much information rather than too little. For example, one should

a) describe clearly the methods used to calculate the measurement result and its uncertainty from the experimental observations and input data;

b) list all uncertainty components and document fully how they were evaluated;

c) present the data analysis in such a way that each of its important steps can be readily followed and the calculation of the reported result can be independently repeated if necessary;

d) give all corrections and constants used in the analysis and their sources.

A test of the foregoing list is to ask oneself "Have I provided enough information in a sufficiently clear manner that my result can be updated in the future if new information or data become available?"

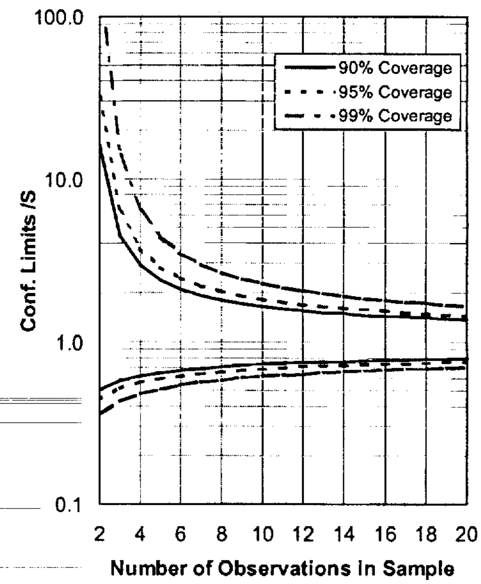

Figure 25 - Confidence limits for sample standard deviation for simple statistical control for sampling from a process governed by the Gaussian distribution.[37]

For the complex testing carried out in wind-tunnels and other types of ground testing facilities, carrying out the prescriptions of the Guide may seem excessively burdensome. However, not carrying them out essentially negates the value of the uncertainty effort. We have adopted, or are about to adopt, several procedures to reduce the burden to reasonable levels. First, we are committing all of our data quality assurance efforts to management by the ISO 9000 standard.** Second, we are placing our test documentation and analysis results on special web sites for access by our customers.[38] These sites not only make the information readily available to the customer in a timely (and secure) manner at whatever level of detail they may desire, but also allow us to archive them electronically to a high degree of safety for further analysis if required.

We believe that the following information should be included in any wind-tunnel measurement assurance statement that is based on statistical quality control and the methods described in this paper:

** At Langley Research Center, the ISO 9000 management process is known as the Langley Management System (LMS).

1. An accessible reference to the overall measurement assurance scheme for the tunnel of interest. The reference should include the following:
    a. A description of the check standard testing process;
    b. The control charts and quick-look plots for the check standard testing process;
    c. The scaling methodology in use;
    d. The control-chart checks on the scaling methodology.
2. A description of the scaled predictions for the customer test of interest. This should be provided soon enough prior to the test that the customer can change the test matrix or experimental design if necessary.
3. The quick-look plots for the repeatability (range) and reproducibility (moving range), including comments on whether the measurement dispersion results are within the process limits.
4. If the measurement dispersion results are within the process limits, then a clear statement should be made that the scaled values of $\hat{\sigma}_{wg}$ and $\hat{\sigma}_x = \sqrt{\hat{\sigma}_{wg}^2 + \hat{\sigma}_{bg}^2}$ are the tunnel staff's best estimate of the repeatability and reproducibility respectively.
5. For separated flow regions, it is recommended to the customer that they multiply the reported values for $\sigma_{wg}, \sigma_{bg}, \sigma_x$ for attached flow by the ratio of the separated-flow scatter to the attached-flow scatter as observed in the repeat-run checks.

Final Remarks

In this paper we argue, having adopted the methods of Shewhart, Eisenhart, Schumacher and NIST, that

1. The precision measurement processes of ground testing must be predictable and have a "right" answer.
2. To be predictable and have a right answer, the measurement processes must be in statistical control.
3. Any estimate of the right answer is meaningless without a credible estimate of the process standard deviation together with the degrees of freedom associated with that estimate.
4. Using statistical control charts together with a continuing data base of check standard results and customer checks will serve to establish that the process is stable and give unequivocal estimates of the measurement process standard deviations and degrees of freedom.

We have also described the specific data gathering and analysis procedures that we have adopted for carrying out the above prescription for credible measurement.

Finally, we remark that, while we believe we have followed the essential prescriptions of Shewhart, Eisenhart, Schumacher and NIST, they are not responsible for any errors of application or interpretation that we might have made.

Acknowledgements

The authors wish to thank the staffs of the facilities of the Wind Tunnel Enterprise at Langley Research Center who answered our questions, followed our prescriptions, shared their insights, and generously provided their data for this paper. Funding for the second and third authors was provided through NASA Contract NAS1-96014.

References

1. Hemsch, Michael J., "Development and Status of Data Quality Assurance Program at NASA Langley Research Center --- Toward National Standards", AIAA Paper 96-2214, June 1996.
2. Shewhart, Walter A., *Economic Control of Manufactured Product*, American Society of Quality, 1980 (originally published in 1931 by Van Nostrand).
3. Shewhart, Walter A., *Statistical Method from the Viewpoint of Quality Control,* 1939, reprinted by Dover 1986.
4. Croarkin, Carroll, and Tobias, Paul (Eds.), *NIST/Sematech Engineering Statistics Handbook.* http://www.nist.gov/stat.handbook Accessed April 21, 1999.
5. Eisenhart, Churchill, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems", *Precision Measurement and Calibration: Statistical Concepts and Procedures*, NBS SP 300 - Vol. 1, February 1969 (Harry H. Ku, editor), pp. 21-47.
6. Coleman, Hugh W., and Steele, W. Glenn, *Experimentation and Uncertainty Analysis for Engineers,* 2nd Ed., Wiley, 1999.
7. Moffat, Robert J., "Using Uncertainty Analysis in the Planning of an Experiment", Journal of Fluids Engineering, Vol. 107, June 1985, pp. 173-178.
8. Iyer, Venkit, Everhart, Joel L., Bir, Pamela J., Ulbrich, Norbert, "Implementation of the WICS Wall-Interference Correction System at the National

Transonic Facility", AIAA Paper 2000-2383, June 2000.

9. Ewald, B. F. R., (Ed.), *Wind Tunnel Wall Correction*, AGARD-AG-336, October 1998.
10. Von Mises, Richard, *Probability, Statistics and Truth*, Dover 1981 (originally published in 1928).
11. Bury, Karl, *Statistical Distributions in Engineering*, Cambridge University Press, 1999.
12. Webster, Noah, *American Dictionary of the English Language*, 1828 Edition, reprinted by the Foundation for American Christian Education, San Francisco, CA, 1989.
13. Mayo, Deborah G., *Error and the Growth of Experimental Knowledge*, U. of Chicago Press, 1996.
14. Anon., *U.S. Guide to the Expression of Uncertainty in Measurement*, ANSI/NCSL Z540-2-1997.
15. Wheeler, Donald J. *Understanding Variation: The Key to Managing Chaos*, SPC Press, 1993.
16. Wheeler, Donald J., and Chambers, David S., *Understanding Statistical Process Control*, 2nd Edition, SPC Press, 1992.
17. Wheeler, Donald J., *Advanced Topics in Statistical Process Control: The Power of Shewhart's Charts*, SPC Press, 1995.
18. Montgomery, Douglas C., *Introduction to Statistical Quality Control*, 3rd Edition, Wiley, 1996.
19. Grant, Eugene L., and Leavenworth, Richard S., *Statistical Quality Control*, 7th Edition, McGraw-Hill, 1996.
20. Belanger, Brian, *Measurement Assurance Programs: Part I - Introduction*, NBS SP 676-I, May 1984.
21. Taylor, Barry N., and Kuyatt, Chris E., *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST TN 1297, 1994 Edition, (see also Rev. July 1999 http://www.physics.nist.gov/Pubs /guidelines/ Accessed April 21, 2000.
22. Pontius, Paul E., and Cameron, Joseph M., "Realistic Uncertainties and the Mass Measurement Process", *Precision Measurement and Calibration: Statistical Concepts and Procedures*, NBS SP 300 - Vol. 1, February 1969 (Harry H. Ku, editor), pp. 2-20.
23. Croarkin, Carroll, *Measurement Assurance Programs: Part II - Development and Implementation*, NBS SP 676-II, April 1984.
24. Schumacher, Rolf B. F., *Measurement Uncertainty - Measurement Assurance Handbook*, 1996.
25. Schumacher, Rolf B. F., "Statistical Control in a Standards Laboratory", *Measurements and Data*, Vol. 3, No. 15, 1969, pp. 58-64.
26. Gentry, Garl L., Jr., Quinto, P. Frank, Gatlin, Gregory M., and Applin, Zachary T., *The Langley 14- by 22-Foot Subsonic Tunnel: Description, Flow Characteristics, and Guide for Users*, NASA TP 3008, September 1990.
27. Jackson, Charlie M., Corlett, William A., and Monta, William J., *Description and Calibration of the Langley Unitary Plan Wind Tunnel*, NASA TP 1905, November 1981.
28. Fuller, Dennis E., *Guide for Users of the National Transonic Facility*, NASA TM-83124, 1981.
29. Capone, Francis J., Bangert, Linda S., Asbury, Scott C., Mills, Charles T. L., and Bare, E. Ann, *The NASA Langley 16-Foot Transonic Tunnel: Historical Overview, Facility Description, Calibration, Flow Characteristics, and Test Capabilities*, NASA TP 3521, September 1995.
30. McGhee, Robert J., Beasley, William D., and Foster, Jean M., *Recent Modifications and Calibration of the Langley Low-Turbulence Pressure Tunnel*, NASA TP 2328, July 1984.
31. Mineck, Raymond E., and Hill, Acquilla S., *Calibration of the 13- by 13-Inch Adaptive Wall Test Section for the Langley 0.3-Meter Transonic Cryogenic Tunnel*, NASA TP 3049, December 1990.
32. Huebner, L. D., Rock, K. E., Voland, R. T., Wieting, A. R., *Calibration of the 8-Foot High-Temperature Tunnel for Hypersonic Airbreathing Propulsion Testing*, AIAA-CP-96-2197, June 1996.
33. Staff, *A User's Guide to the Langley 16-Foot Transonic Tunnel Complex (Revision 1)*, NASA TM 102750, September 1990.
34. Staff, *The Langley Transonic Dynamics Tunnel*, Langley Working Paper LWP-799, September 1969.
35. Bryer, D. W., and Pankhurst, R. C., *Pressure Probe Methods for Determining Wind Speed and Flow Direction*, National Physical Laboratory. 1971.
36. Chue, S. H., "Pressure Probes for Fluid Measurement", *Progress in Aerospace Sciences*, Vol. 16, No. 2, 1975, pp. 147-223.
37. Anon., *Manual on Presentation of Data and Control Chart Analysis: 6th Edition*, ASTM Manual Series MNL 7, Revision of Special Publication (STP) 15D, 1995.
38. Jones, Kennie, *Improving Productivity and Value of Experimental Processes through Application of Information Technology*, AIAA 98-2718, June 1998.