NASA Contractor Report 4303

Data Quality Analysis at the National Transonic Facility

Pamela N. Stewart

NASA

Data Quality Analysis at the National Transonic Facility

Pamela N. Stewart
Wyle Laboratories
Hampton, Virginia

NASA

DATA QUALITY ANALYSIS

AT THE NATIONAL TRANSONIC FACILITY

by Pamela N. Stewart

Summary

The data quality analysis program was developed to satisfy the need
for a computer driven systematic analysis of data taken during
calibrations of the high speed digital data acquisition system at the
National Transonic Facility (NTF). The end result of the data quality
analysis program is a comprehensive report which identifies those data
channels which have the capability of producing poor quality data.
This report is generated each time a calibration of the system is
performed.

The data quality analysis program performs five distinct checks on the
calibration data. The five checks are for non-linearity, noise, short
term drift, long term drift, and the proper functioning of the
calibrator. The tests are performed to identify problems prior to the
collecting and recording of data.

The data quality analysis program has eliminated individual interpretation of data quality and has established a standard set of evaluation guidelines. The Test Directors can now make more informed decisions regarding tunnel operations. These decisions may be made rapidly which is very important due to the high costs of cryogenic facility operations. Finally, since the institution of the data quality report, the solutions to instrumentation problems that previously were difficult to detect until after post test processing was completed are found with fewer delays.

Introduction

The NTF is a cryogenic wind tunnel at the NASA Langley Research Center in Hampton, Virginia. The NTF has a variety of data systems. The system addressed here is a high speed digital data acquisition system that is used to obtain data from thermocouples, pressure transducers, platinum resistance thermometers, and strain gages.

This is an analog-to-digital converter system with an independent preamplifier for each data channel. The NTF tunnel system is normally configured for 192 channels, and the Model Preparation Area (MPA) system for 128 channels. However, if the systems were fully expanded they could contain up to 2048 channels. The analog-to-digital converters are of the successive approximation type with a resolution of 14 bits plus sign ($\pm 16,384$ counts), and a total system throughput rate of 50,000 samples per second.

2

Incoming signals are amplified by the preamplifier gain which is x500 or x100 for low level cards and x10 or x1 for high level cards. Following preamplification, the signal passes through a one Hertz filter where the majority of extraneous noise is eliminated. The signal is then further amplified by setting an internal Programmable Gain Amplifier (PGA). In the current data acquisition system, PGA gains of 4, 2, and 1 are available. This provides a measurement range of 5.12 mv to 10.24 v fullscale. This results in a system resolution of approximately 0.3 microvolts per count. It is clear then that with such resolution even a small amount of electrical noise can result in significant data contamination. Figure 1 gives an overview of the NTF data acquisition system.
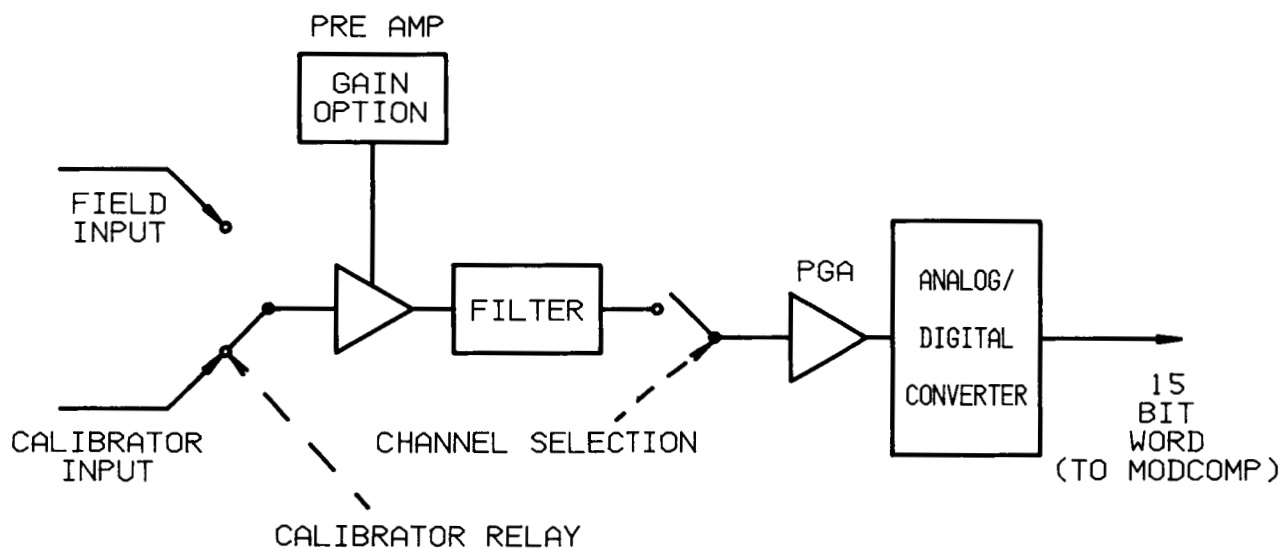
PRE AMP

GAIN OPTION

FIELD INPUT

FILTER

PGA

ANALOG/ DIGITAL CONVERTER

15 BIT WORD (TO MODCOMP)

CALIBRATOR INPUT

CHANNEL SELECTION

CALIBRATOR RELAY

FIGURE 1, NTF DIGITAL DATA AQUISITION SYSTEM OVERVIEW

The data system is calibrated by the computer controlled switching between input data and a programmable precision voltage source used as a calibration standard. Under computer control the voltage standard is set to a specific output voltage, a bank of sixteen calibrate relays is engaged, and, after a delay, the values are scanned for the sixteen channels. The process is repeated five times, at a different percentage of the full scale voltage each time. The five voltage levels that were chosen for the NTF data system are +3/4 and -3/4 of full scale, +3/8 and -3/8 of full scale, and zero. The choice of voltage levels is software controlled and may be easily changed if necessary. Due to the utilization of the above mentioned one Hertz filters, it is necessary to allow for a 2 second filter charge time for each calibration step. After all channels have been scanned at each of the five voltage levels a least squares linear fit is performed. This calibration process is normally completed within twenty minutes or less, depending on the number of data channels that need to be calibrated.

The data quality analysis program was the final step in a series of modifications that were made in the calibration procedures at NTF. Previous modifications, which included the elimination of thermal effects in the data due to the prolonged closure of calibrator relays, significantly improved the quality of acquired data.

However, prior to the creation of the data quaility analysis program there was no mechanism available for objective evalution of system performance. When a calibration was performed, the quality of data was a subjective determination based on the opinions of various individuals. The lack of concrete analysis procedures required errors in the data to be of significant magnitude to be detected. Once an error was detected, finding the source of the problem and solving it was frequently tedious and time consuming.

The data quality report was developed with several goals in mind. The first goal was to establish a computer driven systematic method of analysis of calibration data. Another objective was to provide test directors with a concise, comprehensive report so that decisions could be made without long time delays. Finally, the report would provide ample information about data so that solutions to instrumentation problems would be easier and quicker to find.

The data analysis methods that will be outlined in this paper are for a wind tunnel application, however, they may be applied to data systems in other environments.

Overview of Analysis Methods

The data quality analysis program consists of five checks that are performed on the data acquired during a calibration of the system. Each check uses a statistical method to determine whether the data obtained during a calibration of the data system is within predetermined specifications. If two means are being compared a one tailed t test is used. If two standard deviations are being compared a Chi-Square test is used. The five checks that are performed on the data are for non-linearity, noise, short and long term drift, and the proper functioning of the calibrator.

The non-linearity check performs a linear regression analysis on the five calibration points to determine the degree of non-linearity. This check is used to detect bad channels or calibrator relays, ADC problems, or bit dropping or setting in the data system. The noise check uses a Chi-Square test to compare the standard deviation of the data samples to the manufacturer's specifications to determine if the noise level is excessive.

Excessive noise may be caused by interface problems or improper grounding or shielding. The long and short term drift checks compare current data to previous data to detect excessive drifting. Drift problems may be caused by temperature variations, grounding or

shielding problems, or a PGA failure. Finally, a check for the proper

functioning of the calibrator is made to identify any calibrator zero

offset.

Non-Linearity Check

Non-linearity is a measure of how the actual input-to-output

performance for a device deviates from an ideal linear

relationship.[2]

The concept of an ideal linear relationship implies that there are

certain data values that should be received every time a calibration

of the data system is performed. In the data quality analysis program

these values are known as expected values. As Table 2 shows, the

expected values for the non-linear example shown in Figure 2 are 0 mV,

100 mV, and 200 mV. For illustration purposes, this example is based

on a three point calibration. At NTF, a five point calibration is

performed because five points provide a more accurate check of

linearity. The numbers in the Actual Values column represent the

values that were actually scanned when the calibration was performed.

These values are 100 mV, 220 mV, and 280 mV. As Figure 2 illustrates,

there is a zero offset of 100mV. This means that each of the points

in the Actual Values column are inaccurate by a value of 100 mV. The

numbers in the Predicted Values column represent the data points that

would have been obtained if there had not been a zero offset of 100

mV. As stated above, the values that were expected if the calibration had yielded results that were perfectly linear were 100mV and 200mV. This indicates that the system exhibited an error in linearity of 20 mV.
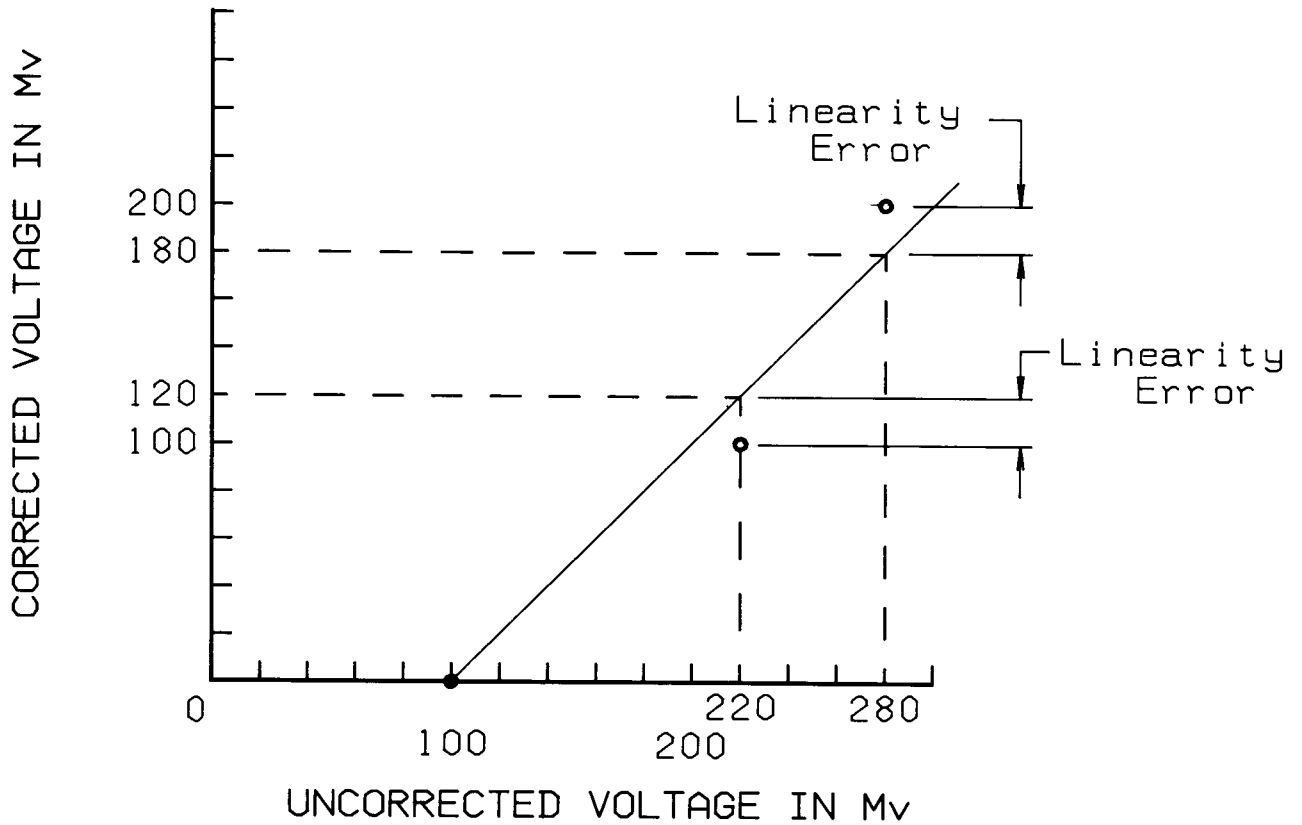

FIGURE 2, NON-LINEARITY CHECK

8

	Actual	Predicted
Expected	(Scanned)	(Calculated)
Values	Values	Values
0 mV	100 mV	0 mV
100 mV	220 mV	120 mV
200 mV	280 mV	180 mV

The values in the predicted column are computed by performing a linear regression analysis. The slope (b) for the line that is formed using the actual values that were obtained in the calibration was calculated using the following formula:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (x)^2} \qquad (1)$$

where x = Expected Value

y = Actual Value

The intercept (a) for the line that is formed from the values obtained from the calibration, was calculated using the formula:

$$a = \frac{\sum y - b \sum x}{n}$$

where y = Actual Value

n = 50 samples of data   (2)

b = Slope [from (1)]

The predicted values (x') are computed using the formula:

$$x' = \frac{y-a}{b}$$

where y = Actual Value   (3)

b = Slope [from (1)]

a = Intercept [from (2)]

The non-linearity error is obtained by subtracting the expected value from the predicted value.

For a channel to pass the non-linearity check, any non-linearity error must fall within a tolerance that was computed using the three components in the following formula:

Tolerance = (FSV(Error)) + Uncertainty of Discrete System

The first component is the full scale value (FSV) which is ±16,384 counts. The second factor is the non-linearity error (Error) listed in the specifications for this system. The final component represents

10

the uncertainty imposed by the use of a discrete system. Each data point contains an uncertainty of one count because the resolution of the system is one count. However, a variation for one of the five points will produce a shift in the offset of 1/5. The error for that point is its value minus the offset which is 1-1/5. This error is added to the product of the full scale value and the non-linearity error listed in the specifications to get the final non-linearity tolerance.

Noise Check

Noise is defined as any extraneous or unwanted signal which contaminates measurement.[2] In the data quality analysis program, the noise check detects any external noise that is affecting the data. There are two potential sources of noise in the NTF data acquisition system, the noise generated by the calibrator and the noise generated within the data acquisition system itself. As previously stated, each time a calibration of the system is performed, fifty samples of data are read and averaged at +3/4 and -3/4 of full scale, +3/8 and -3/8 of full scale and at zero. A standard deviation is then computed for each of the five, fifty sample readings. For a channel to pass the noise check, each of the five standard deviations must fall within a tolerance that is acceptable for this system. A precision error index was used to compute the maximum system standard deviation that is acceptable for this system. This index, combined with the noise

specification listed by the calibrator manufacturer represents the tolerance value used in the noise test. A Chi-Square distribution was used to establish a confidence interval for the standard deviations that are routinely calculated each time a calibration of the system is performed.

Precision errors are random errors caused primarily by noise. Noise specifications for elements which can operate at different gain levels may be reported as Relative to Input (RTI), Relative to Output (RTO) or a combination of these two. The equation for the precision index (S) for this particular system is as follows:

$$S = \sqrt{(e_1 G_1 G_2)^2 + (e_2 G_2)^2 + (e_3 G_2)^2 + (e_4)^2} \qquad (4)$$

where e_1 = Noise (RTI) for Input Preamplifier

e_2 = Noise (RTO) for Input Preamplifier

e_3 = Noise (RTI) for Programmable Gain Amplifier

e_4 = Noise (RTO) for Programmable Gain Amplifier

G_1 = Preamp Gain

G_2 = PGA Gain

As shown in Figure 3, the term e_1 represents the noise present at the input to the preamplifier and e_2 represents the noise generated at the output of the preamplifier. The term e_3 represents the noise

12

generated at the input to the programmable gain amplifier and e_4 represents the noise generated at the output of the programmable gain amplifier.

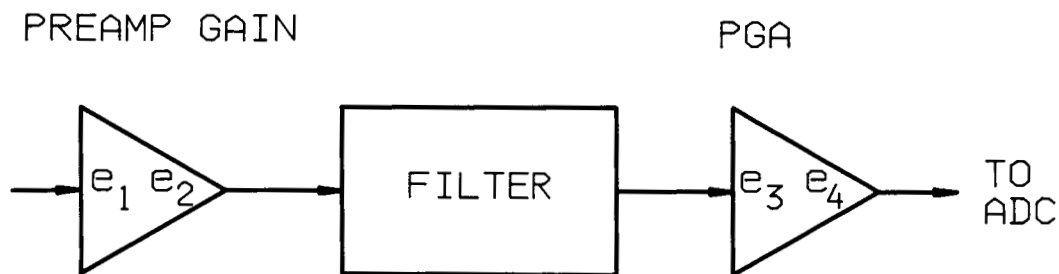PREAMP GAIN   PGA

FIGURE 3, PRECISION ERRORS FOR NOISE CHECK

As stated above, the other component which introduces noise into the data acquisition system is the programmable precision voltage source used as a calibrator. The manufacturer's specifications for the calibrator used lists 2.75 uv of wide band noise on the range used to calibrate low level channels and 27.5 uv of wide band noise on the range used to calibrate high level channels.

In order to compute a confidence interval for the standard deviations that are routinely calculated each time a calibration of the system is performed, the following Chi-Square statistic was used:

(5)

$$X^2 \leq \frac{(n-1)S^2}{\sigma_2}$$

where n = 50 samples
of data

This may be transformed into the following:

(6)

$$S \leq \sigma\sqrt{\frac{x^2}{n-1}}$$

where n = 50

The value for the variable X^2 was computed from a Chi-Square distribution table for forty nine degrees of freedom at a confidence level of .995. The variable σ represents the system standard deviation which was computed using a precision error index as outlined above.

The total system noise is the square root of the sum of the squares of the noise from the calibrator and the noise from the precision index calculated above. This value is then multiplied by the Chi-Square distribution statistic to obtain the overall tolerance for the noise check. See Appendices C and D for a list of the manufacturer's specifications and their applications in computing tolerances for the noise check.

14

Short Term Drift Check

The short term drift check detects shifts in the data during short periods of time. In order to detect short term drift, calibrations are normally performed at the beginning of each eight hour shift. A portion of the data from the current calibration is then compared to identical data from the previous calibration. If the difference between the two sets of data does not fall within a predetermined tolerance range, the channel has failed the short term drift check.

The data used for comparison of short term drift is the averaged value obtained from the fifty samples of data read when the calibrator is set to zero. Each time a calibration is performed, this set of data is retained on disk to be compared with the data obtained from the next calibration.

The only specification for drift listed by the manufacturer of the NTF data acquisition system was for a sixty day period. Therefore, the tolerance value used in the short term drift check was developed by observing and evaluating this particular system over a three week time period. The maximum amount of drift observed on a daily basis during this time period was about 3.75uv on a 5mv full scale range. For a channel to pass the short term drift check, the averaged value obtained from the fifty samples of data read when the calibrator is set to zero must not differ from the previous day's reading by more than 3.75uv.

Long Term Drift Check

The long term drift check detects shifts in the data during a sixty day time period. A sixty day time period was chosen because the long term drift specification listed by the manufacturer of the data acquisition system is for a period of sixty days.

Before a model test is begun, all data acquisition channels are manually reset to zero. As shown in Figure 4, a one-tailed t test was used to determine a confidence interval for long term drift. The solid line rising through the middle of the curve represents the manufacturer's specifications for long term drift. The values read with the calibrator set to zero must fall within three standard deviations of the manufacturer's specifications for the data acquisition system to remain within tolerance. The tolerances for the long term drift check were determined by using the manufacturer's specifications to calculate a bias error. Bias errors are fixed errors that contribute to the difference between the true value and the average value of many repeated readings.
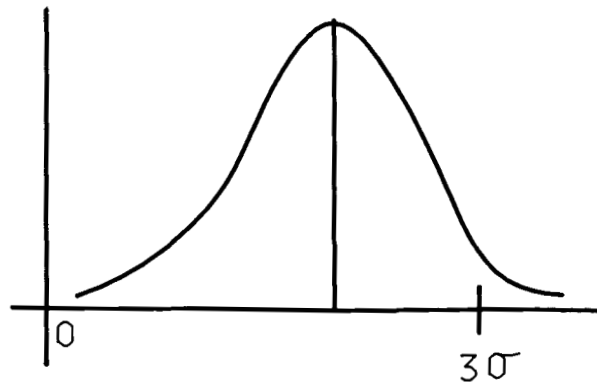

FIGURE 4, ONE-TAILED T TEST FOR LONG TERM DRIFT

16

A bias limit (B) is defined as the square root of the sum of the squares of all known elemental bias errors.

$$B = \sqrt{b_1^2 + b_2^2 + b_3^2 + b_4^2 + b_5^2} \qquad (7)$$

where b_1 = Gain Stability

b_2 = Linearity

b_3 = Zero Stability

b_4 = Zero Drift

b_5 = Common Mode Rejection

As shown in the equation above, there are five known elemental bias errors for this data acquisition system. These errors are gain stability, non-linearity, zero stability, zero drift, and common mode rejection. It is important to note that these five errors are just for the data acquisition system and the calibrator. Other possible sources of error have been measured and do not appear to appreciably contribute. Therefore, these other sources of error were not taken into account.

The error introduced by gain stability is the amount of error that will occur after amplification, due primarily to fluctuations in temperature. Non-linearity is a measure of how the actual input-to-output performance for a device deviates from an ideal linear relationship.[2] There are two sources of linearity error, the error from the preamplifier, and the analog-to-digital converter. This

linearity component represents the manufacturer's specification for linearity and is unrelated to the non-linearity check that was outlined previously. Zero stability and zero drift refer to the amount of drift away from an established zero that will occur primarily due to temperature induced fluctuations in the preamplifier and programmable gain amplifier. Inaccuracies due to temperature changes are the errors listed in the manufacturer's specifications for the data acquisition system.

Common Mode Rejection is a measure of the ability of a differential device to discriminate against voltages common to both input leads.[2] Effectively, Common Mode Rejection refers to the amount of system noise that a data system can inherently compensate for and remove.

At the beginning of each model test a zero baseline is established by adjusting the zero offset for each card. During the lifetime of the test, a zero offset check is performed and compared to the established baseline. The zero offset is obtained from a fifty sample average taken with the calibrator set to zero. For a channel to pass the long term drift check, the averaged value with the calibrator set to zero must not deviate by a value larger than the calculated tolerance for that channel. See Appendices A and B for a list of the manufacturer's specifications for this particular data acquisition system and their applications in calculating bias errors for the long term drift check.

Proper Functioning of the Calibrator

The final check performed is to ensure that the calibrator is
operating correctly. The data for this check is provided via eight
data channels, two per preamplifier gain, with the input to each
channel having signal high, signal low, and guard shorted together at
the amplifier. These channels are known as reference channels. Data
is read from the reference channel once with the calibrator turned off
and then again with the calibrator set to zero. Theoretically, due to
the presence of the shorted inputs, a value of zero should be read on
each channel when the calibrator is turned off. A calibrator offset
is obtained by subtracting this value from the value read when the
calibrator was set to zero. This offset must be within a
predetermined tolerance to ensure that the calibrator is operating
correctly.

The tolerance for the calibrator was calculated using a confidence
limit. The following formula was used:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{\frac{1}{2}} \tag{8}$$

In the formula above σ_1 represents the noise introduced by the data
acquisition system and σ_2 represents the wide band noise introduced by
the calibrator. The variable $z_{\frac{\alpha}{2}}$ has a value of three. The number
three was chosen because the confidence limit implies that the
calibrator offsets should lie within three standard deviations of the

19

tolerance value calculated. The term $(\bar{x}_1 - \bar{x}_2)$ represents the noise specifications supplied by the manufacturer of the calibrator.

Summary of Results

The data quality analysis program has had many positive effects in the area of data acquisition at NTF. The generated report is retained on a test-by-test basis and provides a history of data quality from the system. A well defined set of criteria has been established and is now utilized on a daily basis to identify data channels which would produce poor quality data. The institution of a strict set of guidelines to evaluate data has also eliminated individual interpretation. The method for evaluating data quality has proven to be extremely effective in decision making and identifying problems. The data acquisition system has remained within the calculated tolerances except for periods when legitimate problems were present. Test directors at NTF are now able to make decisions regarding tunnel operations in a more informed and expedient manner.

Due to the data quality analysis program, problems with instrumentation are much easier to solve. For example, malfunctions due to bad calibrator relays are much easier to identify. Thus, bad relays are now replaced before negatively affecting the quality of the data. Other problems that the program was able to identify include defective cables which caused high noise levels, and a high ambient humidity level which induced linearity and drift errors.

20

Finally, the data quality program has pointed out inaccuracies in the calibrator that is currently being used. It has been discovered that the data acquisition system is now as accurate as the calibrator itself. As a result, it has been determined that a new calibrator is necessary. Therefore, numerous calibrators from several different manufacturer's are being evaluated.

Appendix A

Bias Error Specifications for Data Acquisition System

Gain Stability - b_1

Specification: $\pm0.01\% \pm 0.002\%$ / C

Temperature $\pm3^\circ$C

$b_1 = 0.01\% + 3(0.002\%)$

   $= 0.016\%$

Linearity - b_2

Amplifier Error: 0.02%

ADC Error 0.006%

$b_2 = 0.02\% + 0.006\%$

   $= 0.026\%$

<u>Zero</u> <u>Stability</u> - b_3

Low Level: 4uv RTI + 15uv RTO

High Level: 150uv RTO

<u>Zero</u> <u>Drift</u> - b_4

Low Level: 1uv + 15uv RTO /$^{\circ}$C

High Level: 25uv RTO /$^{\circ}$C

<u>Common</u> <u>Mode</u> <u>Voltage</u> - b_5

CMV = 1 Volt

CMR = 66db + gain

    = 66db + 40db

    = 106db

Derivation of Long Term Drift Test Tolerance for 5mv Channel

$$B = \sqrt{b_1^2 + b_2^2 + b_3^2 + b_4^2 + b_5^2}$$

Gain Stability (b_1) = 0.01% + 3(0.002%) = 0.016%

Linearity (b_2) = 0.02% + 0.006% = 0.026%

Zero Stability (b_3)

V_{in} = 5.12mv

V_{out} = 10V

$$b_3 = \frac{4 \times 10^{-6}}{5.12 \times 10^{-3}} + \frac{15 \times 10^{-6}}{10}$$

= .078%

Zero Drift (b_4)

V_{in} = 5.12mv

V_{out} = 10V

$$b_4 = \frac{1 \times 10^{-6}}{5.12 \times 10^{-3}} + 3 \frac{15 \times 10^{-6}}{10}$$

= 0.020%

Common Mode Rejection (b_5)

$$b_5 = \frac{1 \times 10^{-6}}{5.12 \times 10^{-3}}$$

$$= 0.020\%$$

$$B = \sqrt{(.016)^2 + (.026)_2 + (.078)_2 + (.020)_2 + (.020)^2}$$

Appendix C

Precision Error Specifications for Data Acquisition System

$$Error = \sqrt{S_1^2 + S_2^2 + S_3^2 + S_4^2}$$

$$S_1 = e_1 G_1 G_2$$

$$S_2 = e_2 G_2$$

$$S_3 = e_3 G_2$$

$$S_4 = e_4$$

e_1 = Noise Related to Input (preamp) 1uv

e_2 = Noise Related to Output (preamp) 50uv

e_3 = Noise Related to Input (PGA) 20uv

e_4 = Noise Related to Output (PGA) 1.25mV

Derivation of Noise Test Tolerances for 5mV Channel

Determine :

$$Error = \sqrt{S_1^2 + S_2^2 + S_3^2 + S_4^2}$$

$S_1 = e_1 G_1 G_2$

$S_2 = e_2 G_2$

$S_3 = e_3 G_2$

$S_4 = e_4$

$G_1 = 500$ (PreAmp Gain)

$G_2 = 4$ (PGA)

$e_1 = 1uv$

$e_2 = 50uv$

$e_3 = 20uv$

$e_4 = 1250uv$

$$Error = \sqrt{(1x500x4)^2 + (50x4)^2 + (20x4)^2 + (1250)^2}$$

References

(1) McPhee, J.R.: Electrical Noise Reduction Techniques
     Contributing to Improved Data Quality at the National
     Transonic Facility. NASA CR-4193, 1988.

(2) Taylor, James L.; Fundamentals of Measurement Error. Neff
     Instrument Corporation,c.1988.

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-4303		

4. Title and Subtitle		5. Report Date
Data Quality Analysis at the National Transonic Facility		July 1990
		6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Pamela N. Stewart	
	10. Work Unit No.
	505-61-01-01

9. Performing Organization Name and Address	11. Contract or Grant No.
Wyle Laboratories 3200 Magruder Blvd. Hampton, VA 23665-5225	NAS1-18304
	13. Type of Report and Period Covered

12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Langley Research Center Hampton, VA 23665-5225	14. Sponsoring Agency Code

15. Supplementary Notes

Langley Technical Monitor: W. E. Bruce, Jr.

16. Abstract

This report describes the data quality analysis program that was developed at Langley Research Center's National Transonic Facility. The program provides a computer driven systematic analysis of data taken during calibrations of the high speed digital data acquisition system. The report outlines five distinct checks that are performed on the calibration data. The five checks are for non-linearity, noise, short term drift, long term drift, and the proper functioning of the calibrator. The program has eliminated individual interpretations of data quality and has established a standard set of evaluation guidelines.

Although the analysis methods described in the report are used for wind tunnel data acquisition, they may be applied to data systems in other environments.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
high speed digital data acquisition noise non-linearity short term drift calibration	Unclassified - Unlimited Subject Category 61

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	27	A03