

Old Dominion University

ODU Digital Commons

Mechanical & Aerospace Engineering Theses & Dissertations

Mechanical & Aerospace Engineering

Spring 2016

Long-Term Quantitative Study of Wind-Tunnel Balance Calibrations Across Multiple Calibration Systems

Ryan James Callahan

Old Dominion University, rcall017@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/mae_etds



Part of the [Aerospace Engineering Commons](#)

Recommended Citation

Callahan, Ryan J.. "Long-Term Quantitative Study of Wind-Tunnel Balance Calibrations Across Multiple Calibration Systems" (2016). Master of Science (MS), Thesis, Mechanical & Aerospace Engineering, Old Dominion University, DOI: 10.25777/09b8-cm59
https://digitalcommons.odu.edu/mae_etds/3

This Thesis is brought to you for free and open access by the Mechanical & Aerospace Engineering at ODU Digital Commons. It has been accepted for inclusion in Mechanical & Aerospace Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**LONG-TERM QUANTITATIVE STUDY OF WIND-TUNNEL BALANCE
CALIBRATIONS ACROSS MULTIPLE CALIBRATION SYSTEMS**

by

Ryan James Callahan
B.S.A.E. May 2014, Embry-Riddle Aeronautical University

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

AEROSPACE ENGINEERING

OLD DOMINION UNIVERSITY
May 2016

Approved by:

Drew Landman (Director)

Colin Britcher (Member)

Brett Newman (Member)

ABSTRACT

LONG-TERM QUANTITATIVE STUDY OF WIND-TUNNEL BALANCE CALIBRATIONS ACROSS MULTIPLE CALIBRATION SYSTEMS

Ryan James Callahan
Old Dominion University, 2016
Director: Dr. Drew Landman

Internal strain-gauge balances are multi-dimensional force transducers widely employed in wind tunnels to obtain accurate aerodynamic force and moment data on an aircraft model. Balances are calibrated periodically to assure measurement accuracy. The calibration provides a mathematical model relating the forces applied to the balance and the output signal from the balance. Currently, there are multiple types of internal strain-gauge balances used in wind-tunnel testing as well as multiple methods by which balances are calibrated. Because of the differences in the hardware and methodologies used to calibrate a wind-tunnel balance, it is hard to quantify the differences seen between the calibration results. The purpose of this study is to understand how the calibration of a balance is affected by the calibration system. Additionally, the study examines how the performance of a balance changes over time from a calibration perspective. These differences are quantified across different types of balances and different types of calibration systems. The long-term study employs multiple rigorous statistical methods as well as statistical process control techniques to provide insight into the stability of a process over time.

Copyright, 2016, by Ryan James Callahan, All Rights Reserved.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Drew Landman, who has been integral in developing this research and helping me understand the intricacies of wind-tunnel balances. Additionally, his knowledge of DOE, RSM and general ground-test knowledge made learning and applying the skills for this research nearly painless. I would also like to thank my mentor at NASA Langley and the benefactor for this research, Dr. Sean Commo. His knowledge of the calibration capabilities at NASA Langley and his experience with the calibration process proved invaluable. Finally, I would like to thank my parents, who supported me every step of the way.

NOMENCLATURE

α	=	Level of Significance
$b, c, d(n)i, j, k$	=	Least squares estimators of model coefficients
$\beta(n)$	=	True regression coefficients
$CL_{X,\bar{X}}$	=	Center line for X or \bar{X} chart
$CL_{mR,R}$	=	Center line for mR or R chart
n	=	Number of experimental factors
ν	=	Degrees of freedom
$LCL_{X,\bar{X}}$	=	Lower control limit
$UCL_{X,\bar{X}}$	=	Upper control limit
$UCL_{mR,R}$	=	Upper control limit
$\overline{m\bar{R}}$	=	Average moving range
\bar{R}	=	Average subgroup range
\bar{X}	=	Subgroup average
$\bar{\bar{X}}$	=	Grand average
$F_{j,k}$	=	Load parameters for $i^{\text{th}}/j^{\text{th}}$ term
R_t	=	Measured response
X_M	=	Model matrix
y_i	=	i^{th} measurement (response)
\hat{y}_i	=	Predicted value of the i^{th} measurement (response)
SS_T	=	Total sums of squares
SS_R	=	Regression model sums of squares
SS_E	=	Residual error sums of squares
SS_{lof}	=	Sums of squares for lack of fit
SS_{PE}	=	Sums of squares for pure error
$SS_{Reduced}$	=	Reduced sum of squares
MS_R	=	Regression model mean squares
MS_E	=	Residual error mean squares
$MS_{Partial}$	=	Partial mean squares
lof	=	Lack of fit
PE	=	Pure error
k	=	Total number of factors included in the model
σ^2	=	Error variance
$\widehat{\sigma^2}$	=	Predicted error variance
F_0	=	F test statistic
$F_{Critical}$	=	Critical value for F_0 comparison, based on confidence level
n_{Rep}	=	Number of replicates
m	=	Factor level

p	=	Number of model parameters, including the mean
H	=	Hat matrix
h_{ii}	=	Diagonal elements of hat matrix
R^2	=	R squared statistic
R_{adj}^2	=	R squared adjusted statistic
R_{pred}^2	=	R squared predicted statistic
$PRESS$	=	Prediction error sums of squares
e	=	Residual error
$\mu_{\hat{y}(x_0)}$	=	Actual mean response (Population)
$\hat{y}(x_0)$	=	Predicted mean response at design point x_0
$t_{\frac{\alpha}{2}, n-p}$	=	t test statistic for two tailed test at 1-confidence level for n-p degrees of freedom
V	=	Voltage
$(X_M^T X_M)^{-1}$	=	Variance/Covariance Matrix
C_{ii}	=	Diagonal element of the Variance/Covariance matrix
x_0	=	Design point of interest
r_i	=	Studentized residual
$SENS_{i,comp}$	=	Primary Sensitivity for an individual balance component for an individual balance calibration
\overline{SENS}_{comp}	=	Average primary sensitivity per component over all calibrations
$RMSE$	=	Root mean square error
X_m	=	Model matrix

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vi
 Chapter	
1. INTRODUCTION AND LITERATURE REVIEW	1
1.1. Project Objective Statement.....	3
1.2. Literature Study	4
2. BACKGROUND, BALANCES AND CALIBRATION CAPABILITIES.....	18
2.1. Internal Strain-Gauge Balances	18
2.2. Balance Calibration Methods and Methodologies	34
2.3. Rigorous Load Schedule Design.....	39
3. MODELING AND UNDERSTANDING WIND-TUNNEL BALANCES	52
3.1. Wind Tunnel Balance Characterization	52
3.2. Statistical Characterization of Data	53
4. ANALYSIS OF BALANCE CALIBRATION DATA.....	75
4.1. Introduction of Results and Initial Remarks	75
4.2. Comparison of Calibration Methods across a Common Balance	76
4.3. Long-Term Repeatability of a Balance on a Single System	96
4.4. Historical Summary of Balance Calibrations	141
5. CONCLUSIONS AND FUTURE WORK	145
5.1. Conclusions.....	145
5.2. Future Work	149
REFERENCES	151
APPENDICES	154
A. First Order Interaction Coefficients NTF-113C	154
B. First Order Interaction Coefficients MC-60E	170
VITA.....	186

LIST OF TABLES

Table	Page
1. Rated Loads for the Task MK-29 and Associated Accuracy's.....	25
2. Rated Loads for the MC-60E and Associated Accuracy's	29
3. Rated Loads for the Single-Piece Balances and Associated Accuracy's.....	34
4. Automatic Balance Calibration Load Ranges.....	35
5. Table of ABCS Load Schedule Excerpt	42
6. Excerpt from Standard Load Schedule for Single Vector System.....	46
7. Excerpt from Standard Load Schedule for Manual Stand with 5 Load Levels	48
8. Assorted Statistics for Design Comparisons.....	51
9. Sample ANOVA Table	56
10. Tabulated Statistics from a Common Balance Calibration (Single-Piece Balance – SVS)....	76
11. NF Load Prediction from NF Replicates (NTF-113C).....	116
12. NF Load Prediction from Normal Force Replicates (MC-60E)	136
13. Summary Table of Calibrations for Each Balance on Each System.....	142
14. In-depth Table of Balance History Included in this Thesis	143

LIST OF FIGURES

Figure	Page
1. Example of Calibration Quality Monitoring Chart [4]	7
2. Example of Bad Calibration [4].....	8
3. Example of Statistical Process Control Charts [6].....	10
4. Test Condition Mach Number Coefficient (<i>CMach</i>) Check Standard Example	12
5. Example of Within-Group Variation/Repeatability (left) and Between-Group Variation/Repeatability (right) for NF and AF [10]	14
6. Estimated Standard Deviations for Within-Group Variation/Repeatability (left) and Between Group Variation/Repeatability (right) for NF and AF [10]	15
7. Diagram of Wind Tunnel Model with Balance.....	18
8. Diagram of Common Wind Tunnel Balance (Single-Piece)	20
9. Example of Bidirectional Behavior	21
10. Multi-Piece Balance – Task Mk-29B	24
11. Internal Structure of Task Mk-29	25
12. Load Rhombus for the Mk-29.....	26
13. Unitized Multi-Piece Balance	28
14. Internal Diagram of Unitized Balance	28
15. Simple Diagram of Unitized Balance Flexures	29
16. Load Rhombus for MC-60E	30
17. Single-Piece Balance	32
18. Single-Piece Balance Diagram	32
19. Isometric of a Single-Piece Balance	33

Figure	Page
20. Automatic Balance Calibration System	35
21. NASA Langley Single Vector System.....	37
22. Single Vector System Diagram.....	37
23. Small Stand Manual Calibration - Close Up.....	38
24. Large Stand Manual Calibration	38
25. Plot of Load Series Excerpt from ABCS Calibration	41
26. Graphic showing CCD Design.....	44
27. Excerpt from Standard Load Schedule for Single Vector System.....	45
28. Excerpt from Standard Load Schedule for Manual Stand with 5 Load Levels	47
29. Box-Behnken Design	50
30. Normal Plot of Residuals	67
31. Residuals vs. Predicted	68
32. Residuals vs. Run.....	68
33. Example of Control Charts	71
34. Cross-System Comparisons, NF Primary Sensitivity (NTF-113C).....	79
35. Cross-System Comparisons, AF Primary Sensitivity (NTF-113C).....	79
36. Cross-System Comparisons, PM Primary Sensitivity (NTF-113C)	80
37. Cross-System Comparisons, RM Primary Sensitivity (NTF-113C).....	80
38. Cross-System Comparisons, YM Primary Sensitivity (NTF-113C).....	81
39. Cross-System Comparisons, SF Primary Sensitivity (NTF-113C)	81
40. Cross-System Comparisons, NF Primary Sensitivity (MC-60E)	83
41. Cross-System Comparisons, AF Primary Sensitivity (MC-60E)	83

Figure	Page
42. Cross-System Comparisons, PM Primary Sensitivity (MC-60E).....	84
43. Cross-System Comparisons, RM Primary Sensitivity (MC-60E)	84
44. Cross-System Comparisons, YM Primary Sensitivity (MC-60E)	85
45. Cross-System Comparisons, SF Primary Sensitivity (MC-60E)	85
46. Cross-System Comparisons, NF Primary Sensitivity (MK-29B)	87
47. Cross-System Comparisons, AF Primary Sensitivity (MK-29B)	87
48. Cross-System Comparisons, PM Primary Sensitivity (MK-29B)	88
49. Cross-System Comparisons, RM Primary Sensitivity (MK-29B).....	88
50. Cross-System Comparisons, YM Primary Sensitivity (MK-29B).....	89
51. Cross-System Comparisons, SF Primary Sensitivity (MK-29B).....	89
52. Cross-System Comparisons, NF Primary Sensitivity (NTF-118B).....	91
53. Cross-System Comparisons, AF Primary Sensitivity (NTF-118B).....	92
54. Cross-System Comparisons, PM Primary Sensitivity (NTF-118B)	92
55. Cross-System Comparisons, RM Primary Sensitivity (NTF-118B).....	93
56. Cross-System Comparisons, YM Primary Sensitivity (NTF-118B).....	93
57. Cross-System Comparisons, SF Primary Sensitivity (NTF-118B)	94
58. X-mR Control Chart for NF Primary Sensitivity (NTF-113C)	98
59. X-mR Control Chart for AF Primary Sensitivity (NTF-113C)	98
60. X-mR Control Chart for PM Primary Sensitivity (NTF-113C).....	99
61. X-mR Control Chart for RM Primary Sensitivity (NTF-113C)	99
62. X-mR Control Chart for YM Primary Sensitivity (NTF-113C)	100
63. X-mR Control Chart for SF Primary Sensitivity (NTF-113C)	100

Figure	Page
64. X-mR Control Chart of RMSE on NF Confirmation Point Residuals (NTF-113C)	104
65. X-mR Control Chart of RMSE on AF Confirmation Point Residuals (NTF-113C)	105
66. X-mR Control Chart of RMSE on PM Confirmation Point Residuals (NTF-113C).....	105
67. X-mR Control Chart of RMSE on RM Confirmation Point Residuals (NTF-113C)	106
68. X-mR Control Chart of RMSE on YM Confirmation Point Residuals (NTF-113C)	106
69. X-mR Control Chart of RMSE on SF Confirmation Point Residuals (NTF-113C).....	107
70. X-mR Control Chart of RMSE on NF Model Point Residuals (NTF-113C)	108
71. X-mR Control Chart of RMSE on AF Model Point Residuals (NTF-113C)	109
72. X-mR Chart of RMSE on PM Model Point Residuals (NTF-113C).....	109
73. X-mR Control Chart of RMSE on RM Model Point Residuals (NTF-113C)	110
74. X-mR Control Chart of RMSE on YM Model Point Residuals (NTF-113C)	110
75. X-mR Control Chart of RMSE on SF Model Point Residuals (NTF-113C).....	111
76. \bar{X} -R Control Chart for N1 Bridge, NF Replicates (NTF-113C).....	113
77. \bar{X} -R Control Chart for N2 Bridge, NF Replicates (NTF-113C).....	114
78. Change in Predicted Load from Average NF Replicated Points (NTF-113C)	117
79. Phase II Control Chart for NF Primary Sensitivity (NTF-113C)	119
80. Phase II Control Chart for AF Primary Sensitivity (NTF-113C)	119
81. Phase II Control Chart for PM Primary Sensitivity (NTF-113C).....	120
82. Phase II Control Chart for RM Primary Sensitivity (NTF-113C)	120
83. Phase II Control Chart for YM Primary Sensitivity (NTF-113C)	121
84. Phase II Control Chart for SF Primary Sensitivity (NTF-113C)	121
85. X-mR Control Chart for Normal Force Primary Sensitivity (MC-60)	123

Figure	Page
86. X-mR Control Chart for Axial Force Primary Sensitivity (MC-60).....	123
87. X-mR Control Chart for Pitching Moment Primary Sensitivity (MC-60).....	124
88. X-mR Control Chart for Rolling Moment Primary Sensitivity (MC-60).....	124
89. X-mR Control Chart for Yawing Moment Primary Sensitivity (MC-60)	125
90. X-mR Control Chart for Side Force Primary Sensitivity (MC-60E).....	125
91. X-mR Control Chart of RMSE on NF Confirmation Point Residuals (MC-60E).....	127
92. X-mR Control Chart of RMSE on AF Confirmation Point Residuals (MC-60E).....	127
93. X-mR Control Chart of RMSE on PM Confirmation Point Residuals (MC-60E)	128
94. X-mR Control Chart of RMSE on RM Confirmation Point Residuals (MC-60E).....	128
95. X-mR Control Chart of RMSE on YM Confirmation Point Residuals (MC-60E).....	129
96. X-mR Control Chart of RMSE on SF Confirmation Point Residuals (MC-60E).....	129
97. X-mR Control Chart of RMSE on Normal Force Model Point Residuals (MC-60E).....	131
98. X-mR Control Chart of RMSE on Axial Force Model Point Residuals (MC-60E)	131
99. X-mR Control Chart of RMSE on Pitching Moment Model Point Residuals (MC-60E)	132
100. X-mR Control Chart of RMSE on Rolling Moment Model Point Residuals (MC-60E)....	132
101. X-mR Control Chart of RMSE on Yawing Moment Model Point Residuals (MC-60E)...	133
102. X-mR Control Chart of RMSE on Side Force Model Point Residuals (MC-60E).....	133
103. \bar{X} -R Control Chart for N1 Bridge, NF Replicates (MC-60E)	135
104. \bar{X} -R Control Chart for N2 Bridge, NF Replicates (MC-60E)	135
105. Change in Predicted Load from Average NF Replicated Points (MC-60E).....	137
106. Phase II Control Chart for NF Primary Sensitivity (MC-60E).....	138
107. Phase II Control Chart for AF Primary Sensitivity (MC-60E).....	139

Figure	Page
108. Phase II Control Chart for PM Primary Sensitivity (MC-60E)	139
109. Phase II Control Chart for RM Primary Sensitivity (MC-60E).....	140
110. Phase II Control Chart for YM Primary Sensitivity (MC-60E).....	140
111. Phase II Control Chart for SF Primary Sensitivity (MC-60E).....	141
112. Normal Force Model - AF First Order Coefficient (NTF-113C)	154
113. Normal Force Model - PM First Order Coefficient (NTF-113C).....	155
114. Normal Force Model - RM First Order Coefficient (NTF-113C)	155
115. Normal Force Model - YM First Order Coefficient (NTF-113C)	156
116. Normal Force Model - SF First Order Coefficient (NTF-113C)	156
117. Axial Force Model - NF First Order Coefficient (NTF-113C).....	157
118. Axial Force Model - PM First Order Coefficient (NTF-113C)	157
119. Axial Force Model - RM First Order Coefficient (NTF-113C).....	158
120. Axial Force Model - YM First Order Coefficient (NTF-113C)	158
121. Axial Force Model - SF First Order Coefficient (NTF-113C)	159
122. Pitching Moment Model - NF First Order Coefficient (NTF-113C).....	159
123. Pitching Moment Model - AF First Order Coefficient (NTF-113C).....	160
124. Pitching Moment Model - RM First Order Coefficient (NTF-113C).....	160
125. Pitching Moment Model - YM First Order Coefficient (NTF-113C).....	161
126. Pitching Moment Model - SF First Order Coefficient (NTF-113C).....	161
127. Rolling Moment Model - NF First Order Coefficient (NTF-113C)	162
128. Rolling Moment Model - AF First Order Coefficient (NTF-113C)	162
129. Rolling Moment Model - PM First Order Coefficient (NTF-113C).....	163

Figure	Page
130. Rolling Moment Model - YM First Order Coefficient (NTF-113C).....	163
131. Rolling Moment Model - SF First Order Coefficient (NTF-113C).....	164
132. Yawing Moment Model - NF First Order Coefficient (NTF-113C).....	164
133. Yawing Moment Model - AF First Order Coefficient (NTF-113C).....	165
134. Yawing Moment Model - PM First Order Coefficient (NTF-113C).....	165
135. Yawing Moment Model - RM First Order Coefficient (NTF-113C)	166
136. Yawing Moment Model - SF First Order Coefficient (NTF-113C)	166
137. Side Force Model - NF First Order Coefficient (NTF-113C).....	167
138. Side Force Model - AF First Order Coefficient (NTF-113C).....	167
139. Side Force Model - PM First Order Coefficient (NTF-113C).....	168
140. Side Force Model - RM First Order Coefficient (NTF-113C)	168
141. Side Force Model - YM First Order Coefficient (NTF-113C)	169
142. Normal Force Model - AF First Order Coefficient (MC-60E).....	170
143. Normal Force Model - PM First Order Coefficient (MC-60E)	171
144. Normal Force Model - RM First Order Coefficient (MC-60E).....	171
145. Normal Force Model - YM First Order Coefficient (MC-60E).....	172
146. Normal Force Model - SF First Order Coefficient (MC-60E).....	172
147. Axial Force Model - NF First Order Coefficient (MC-60E)	173
148. Axial Force Model - PM First Order Coefficient (MC-60E).....	173
149. Axial Force Model - RM First Order Coefficient (MC-60E)	174
150. Axial Force Model - YM First Order Coefficient (MC-60E).....	174
151. Axial Force Model - SF First Order Coefficient (MC-60E).....	175

Figure	Page
152. Pitching Moment Model - NF First Order Coefficient (MC-60E)	175
153. Pitching Moment Model - AF First Order Coefficient (MC-60E)	176
154. Pitching Moment Model - RM First Order Coefficient (MC-60E)	176
155. Pitching Moment Model - YM First Order Coefficient (MC-60E)	177
156. Pitching Moment Model - SF First Order Coefficient (MC-60E)	177
157. Rolling Moment Model - NF First Order Coefficient (MC-60E)	178
158. Rolling Moment Model - AF First Order Coefficient (MC-60E)	178
159. Rolling Moment Model - PM First Order Coefficient (MC-60E)	179
160. Rolling Moment Model - YM First Order Coefficient (MC-60E)	179
161. Rolling Moment Model - SF First Order Coefficient (MC-60E)	180
162. Yawing Moment Model - NF First Order Coefficient (MC-60E)	180
163. Yawing Moment Model - AF First Order Coefficient (MC-60E)	181
164. Yawing Moment Model - PM First Order Coefficient (MC-60E)	181
165. Yawing Moment Model - RM First Order Coefficient (MC-60E)	182
166. Yawing Moment Model - SF First Order Coefficient (MC-60E)	182
167. Side Force Model - NF First Order Coefficient (MC-60E)	183
168. Side Force Model - AF First Order Coefficient (MC-60E)	183
169. Side Force Model - PM First Order Coefficient (MC-60E)	184
170. Side Force Model - RM First Order Coefficient (MC-60E)	184
171. Side Force Model - YM First Order Coefficient (MC-60E)	185

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Wind-tunnel balances are the primary source for aerodynamic force and moment data in ground-based aeronautics research. Unlike most metrology applications, there is no reference standard for a balance traceable to any standards organization, such as the National Institute of Standards and Technology (NIST) [1]. As a result, balances are characterized utilizing a variety of systems and philosophies. Calibration systems used by NASA include manual gravity loaded systems, where loads are applied to a balance using calibrated weights by a calibration technician, and automatic systems, where loads are applied and measured through programmable actuators and load cells. Due to the differences in these systems and the methodologies employed during calibration, discrepancies have been observed in the calibration data. Furthermore, understanding the causes of the discrepancies is important to furthering the confidence in the data collected during a wind-tunnel test.

The balance itself is only part of the total uncertainty of any wind-tunnel test. The balance calibration, the response model, the wind-tunnel model, the model setup, and tunnel uncertainties all play into the total uncertainty and, by extension, the total confidence in the data taken from a wind-tunnel test [1]. To eliminate or reduce most of the uncertainties associated with the wind-tunnel, the balance is typically calibrated in a laboratory environment using calibrated weights which are traceable to a NIST standard. The purpose of the calibration is to develop a relationship between applied load and measured output from the balance, as well as reduce bias in instrument readings. Each model set produced by a calibration directly correlates to a primary load component that the balance can measure. Six models, one for each of the six components, provide a full understanding of the balance mechanics and forces on a wind-tunnel

model. These components are as follows: normal force (NF), axial force (AF), side force (SF), pitching moment (PM), yawing moment (YM), and rolling moment (RM). Unfortunately, the coefficients derived from a single calibration do not expose the breadth of random error or unstable biases within the calibration process itself.

In the past, many of the differences seen between calibrations have often been blamed on changes in the balance electrical systems or structure due to use. These alleged problems are not founded in fact as no conclusive study of balance repeatability has been attempted. In all likelihood, some of the differences between calibrations could be attributed to human error during the calibration process.

The three calibration systems chosen for the study are the Automatic Balance Calibration System (ABCS), the Single-Vector System (SVS) and the Manual Calibration Stand. The three types of balances chosen for the study are a multi-piece balance designated the Task Mk-29B, a unitized multi-piece balance designated the Triumph MC-60E and two NASA Langley single-piece balances designated the NTF-113C and the NTF-118B.

This thesis presents preliminary results from the first two years of the balance calibration study and focuses on system and balance comparisons using actual balance data from calibrations on each system. The study also takes the first look at long-term calibration repeatability on a single balance for a single system. Statistical process control (SPC) is used extensively to provide well-founded insights into the long-term performance of a given balance type for a given calibration system. Other performance metrics are presented in the thesis and lend further insight into the current state-of-the-art balance calibration capabilities available to NASA and the aerospace industry.

1.1. Project Objective Statement

The long-term balance calibration study is an assessment of the balance calibration systems currently used by NASA. It is designed to help guide customers to select calibration services to meet the requirements of their test needs, and to guide the evolution of calibration services and balances available to NASA and its partners. The results of this study may be used to identify and help fix potential weaknesses in existing calibration approaches, hardware, and load schedule designs as well as recommend investments in new balances and new balance designs. Using statically-rigorous and mathematically-defendable methods, the project seeks to evaluate three types of balances across three common types of calibration systems currently used by NASA. The primary source of data for comparison is calibration data from actual wind-tunnel balances that have been calibrated on the three systems. Historical as well as recent calibration data are considered. Subsequently, the project seeks to quantify the differences between calibration systems and balances and to determine just how repeatable a balance and calibration system together can be by repeating multiple calibrations over a long-term period.

The research is split into two major types of analyses. Comparing a single balance across systems and analysis of long-term repeated calibrations. By looking at a single balance on multiple systems, some of the effects a system have on the balance itself may be understood. This comparison will yield reliable results as the experiment takes data directly from the balances. Conversely, by looking at multiple balances on a single system, the study can attempt to discern differences in balances as calibrated on a single system. This comparison is less sound as the data for the study is taken from the balance itself. Using balance data from three different balances to inform on a single system introduces obvious biases and correlations. Therefore, this type of

comparison will be omitted from this study. Finally, using data from repeated calibrations, the long-term repeatability of a balance-calibration system will be evaluated.

1.2. Literature Study

The variance associated with repeated calibrations has seen few studies and is not well understood. Some limited studies have been conducted and offer some limited conclusions on the subject. The following sections will provide relevant information on the history, evolution and challenges associated with operating and calibrating a wind-tunnel balance. Additionally, the groundwork for the application of statistical methods applied to wind-tunnel balances will be explored. Finally, previous efforts to understand balance calibration stability will be reviewed.

1.2.1. Previous Balance Calibration Studies

The NASA Langley Research Center (LaRC) has previously undertaken a similar balance calibration study where two balance types were evaluated on two calibration systems. A single-piece balance and multi-piece balance were studied using the SVS and the Manual Stand [1]. For descriptions of these balances and systems, see Chapter 2. The analyses focused on the math model generated for each calibration. This was chosen as the primary focus as it represents the deliverable values for the customer. The math model is provided as a 6x27 matrix of fitted coefficients that models a wind-tunnel balance with a quadratic fit. The study found acceptable results for a single-piece balance. The standard deviation of back-computed residuals were within 0.1% full-scale when compared across the two systems [1]. Balance output uncertainty is usually quoted minimally at 0.05% [2]. Between two different systems, with two different methodologies, residuals of 0.1% are thought to be very good. However, the multi-piece balance showed differences greater than 0.5% full-scale between the two systems [1]. Researchers expected larger differences with the multi-piece as it has larger uncertainties than a single-piece balance, however,

they were not expected to be so large. The sensitivities were replicated across both systems within 0.2% for all components on the single-piece balance [1]. Researchers expected to find better agreement on the primary sensitivities as they are fundamental to the calibration math model. If the systems cannot replicate primary sensitivities, then investigating any other differences seem impractical. The study had planned on moving forward by designing consistent load schedules, adding statistical tools to the analyses and formalizing a comprehensive report. However, the study was stopped when researchers could not find any apparent problems that could reconcile the data for either balance. The current study is moving forward using all of the previous data and with a skeptical mindset toward the multi-piece balance. Development of consistent load schedules is of primary concern to this study.

Another balance calibration study worth noting was performed by The Boeing Company [3]. One of the main points made by the author of this paper, is that the balance itself is much better than the process we use to calibrate it. Meaning that any change in the coefficient matrix from calibration to calibration could likely be attributed to errors made during the calibration process rather than the balance itself. This hypothesis is sensible when a wind tunnel balance is considered as a simple device consisting of beam elements and is governed by its material properties and simple structural mechanics. With the exception of strain gauge damage, debonding, or mechanical overload, there is little to go wrong. In contrast, the calibration process is a complex experiment that may take weeks of work. A calibration on any system mentioned in this paper could be subject to error induced by the calibrator, angle measurement, error in applied loads, drift in the voltmeters, daily temperature changes, and many more factors. Calibration engineers and technicians take great care in eliminating and quantifying many factors that add to the variability in a calibration. Regardless, uncertainties still exist. Additionally, the results from a Boeing study

which replicated calibrations on two single piece balances, the 6176A and the 635M, are outlined [3]. Historical calibrations on either balance were strung together, spanning from 1999 to 2011. The paper focuses on the 635M which was calibrated three times in 1999, once in 2001 and 2003, used seven times over a decade and recalibrated in 2011 [3]. Check loads were performed after each test, and consistent findings were generated with respect to the balance performance. The deviations reported in the balance coefficients between 1999 and 2011 were less than 0.1% for all primary coefficients [3]. If check loads show more than 0.1% deviation, the balance and calibration will be scrutinized for a cause. A new calibration will then be performed and a new matrix will be used for all future check load comparisons [3]. The report claims that most of the variation is due to calibration process error. In 1999, each calibration was performed by a different technician, with different equipment and a semi-randomized load schedule. Each subsequent calibration was performed years apart. The main conclusion of the report is that a single calibration is insufficient to identify the normal variations in a balance matrix. Instead, a running average of all historical calibrations will provide an engineer with a “true” coefficient matrix and provide prediction intervals which allow balance health monitoring.

Most of the ideas and conclusions described in the paper can be further traced back in time to the original study performed by Boeing, which proposes similar conclusions about using multiple calibrations to expose a more true set of coefficients [4]. Data were presented for both pressure transducers and wind-tunnel balance calibrations. For the pressure transducer calibration, the common practice within the facility is to combine the last five calibrations and analyze the data as a single calibration. The hope is to account for uncertainties in the calibration process itself. When a new calibration is performed, the previous five calibrations are compared to the new one and, if no problems are found, the oldest calibration is deleted and replaced [4]. Also presented is

a set of tools that could be used to assess the quality of the new calibration. Figure 1 shows five pressure transducer calibrations with a 95% confidence interval (CI) and prediction interval (PI) also plotted. Unfortunately, the quality of the plot is poor, however, the primary information can still be understood.

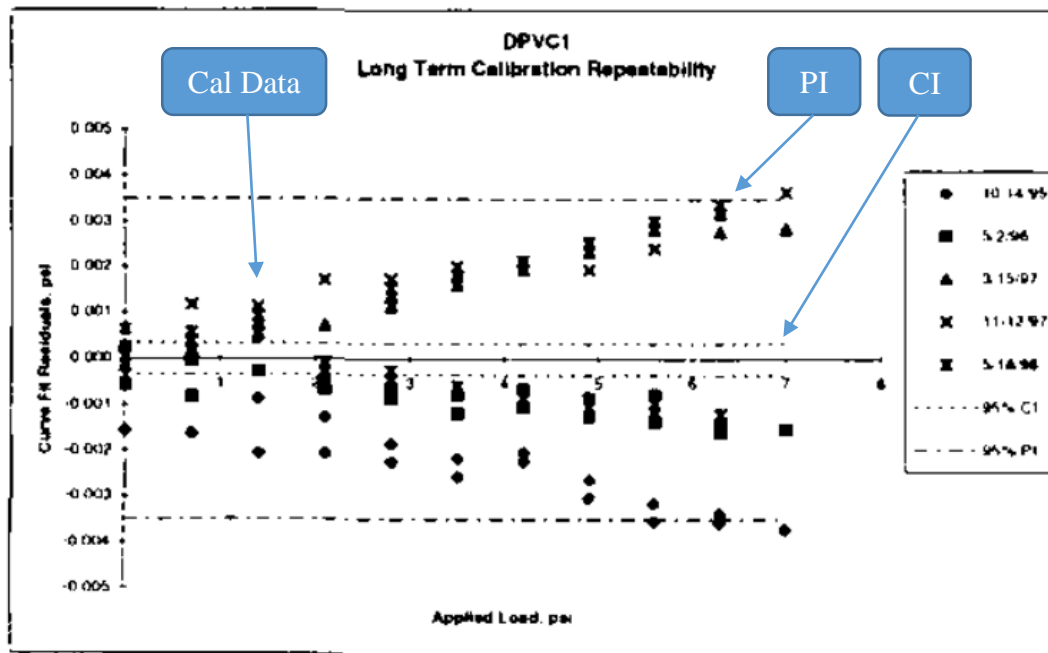


Figure 1: Example of Calibration Quality Monitoring Chart [4]

Figure 1 shows five data sets which are described to fall within the established calibration PI. These five calibrations are combined, used to generate the CI and PI, then fit with a third order polynomial and used in an operational wind tunnel [4]. Figure 2, shown below, is an example of a bad calibration that does not fit within the established PI and would instead be disregarded for future analysis. This method is used to assess the quality of a calibration. The difference in the calibration is clearly visible and a researcher can then begin to look for an assignable cause.

Causes for errors like this are difficult to find as it could be attributed to data acquisition hardware, damaged or nonfunctional pressure transducers, or mistakes during the calibration.

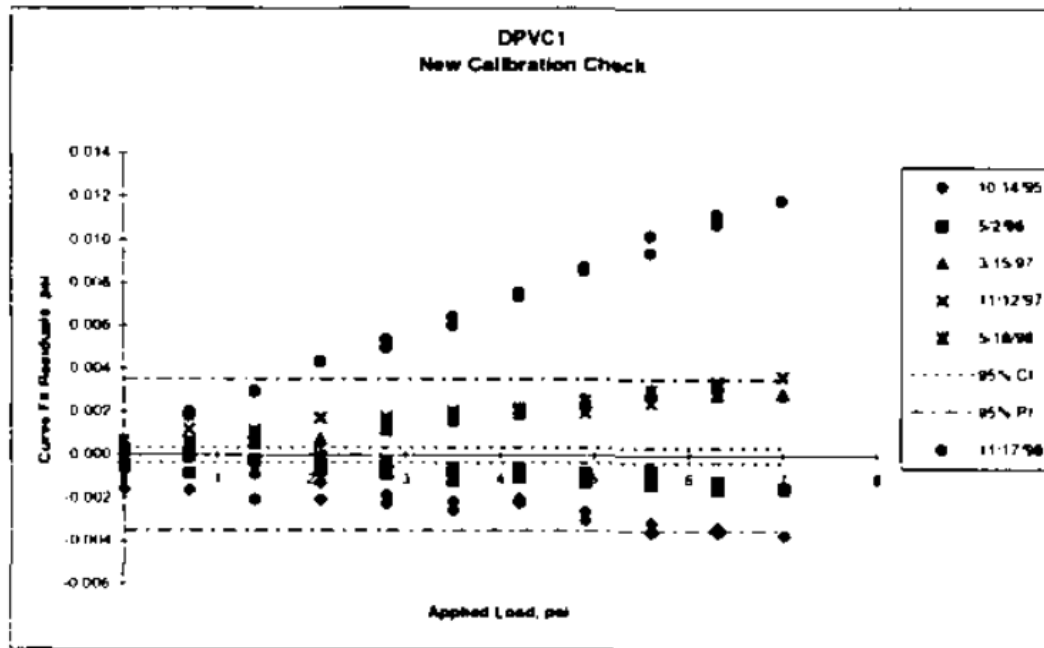


Figure 2: Example of Bad Calibration [4]

A wind-tunnel balance was also included in the paper as an example. The balance is a single-piece balance, designated the 6176A and was mentioned in the previously cited work, however no data were shown [3]. The 6176A was calibrated ten times between 1990 and 1999 [4]. Similarly, the paper proposes that a single calibration does not address the scatter introduced by random errors or unstable biases within the process itself. The paper states, “The calibration process assumes the transfer standard is transported flawlessly, yielding a perfect, repeatable calibration every time a calibration or calibration check is performed” [4]. The method used at this Boeing facility describes combining multiple calibrations to create a matrix of “true coefficients” as was described in Bennett’s paper. This matrix will remain in use until an unexplained excursion from the prediction intervals is observed. If an excursion occurs but no calibration-related cause

can be found, the balance will be inspected for damage and fully recalibrated. The number of calibrations required to derive the “true matrix” is determined by the CI calculated from the combined data. Enough calibrations must be combined to drive the CI to 0.05% full scale (%FS) [4]. Benett indicates that this took three calibrations.

The hope is that a robust method has been developed for balance health monitoring. If successful, a balance user will effectively reduce unnecessary calibrations and test flow time, increase annualized cost savings and provide a baseline for calibration technique improvements.

1.2.2. Statistical Methods Applied to Wind-Tunnel Balance Calibrations

Statistical methods have been employed at Langley for wind-tunnels and wind-tunnel balances for over a decade. Their application to wind-tunnel balance calibration is fully outlined in this work. Parker et al. first proposes the application of statistical practices such as design of experiments (DOE) and response surface methodology (RSM) to efficiently design, execute, and analyze balance calibrations [5]. The paper cited outlined investments in hardware and design schedules, based on statistically rigorous methods that Langley had recently added to its calibration services. This included a small scale SVS which has been used to reduce time and cost of a calibration while simultaneously increasing the quality of the data acquired. The paper concludes that including these statistical tools provides far more rigorous design, calibration, and analysis methods than previously [5]. The insights provided by these methods proved to allow deeper understanding into wind-tunnel balances and how we calibrate them. This lays the groundwork for most of the statistical analysis employed in this thesis. In a later update, Parker further outlines the application of RSM to balances and introduces the application of statistical process control (SPC) for long-term tracking of balance calibrations [6]. Between the 2001 and 2007 publications, Langley had invested in a full-size SVS calibration system capable of calibrating production wind-

tunnel balances. Statistical process control is a method of assessing the variability within a process [7]. Its application is most common in large scale industrial production facilities. In the paper, control charts were proposed as a method to track repeated calibrations of hardware and alert users to any change in the coefficients. Figure 3 shows a set of control charts for a three axis accelerometer.

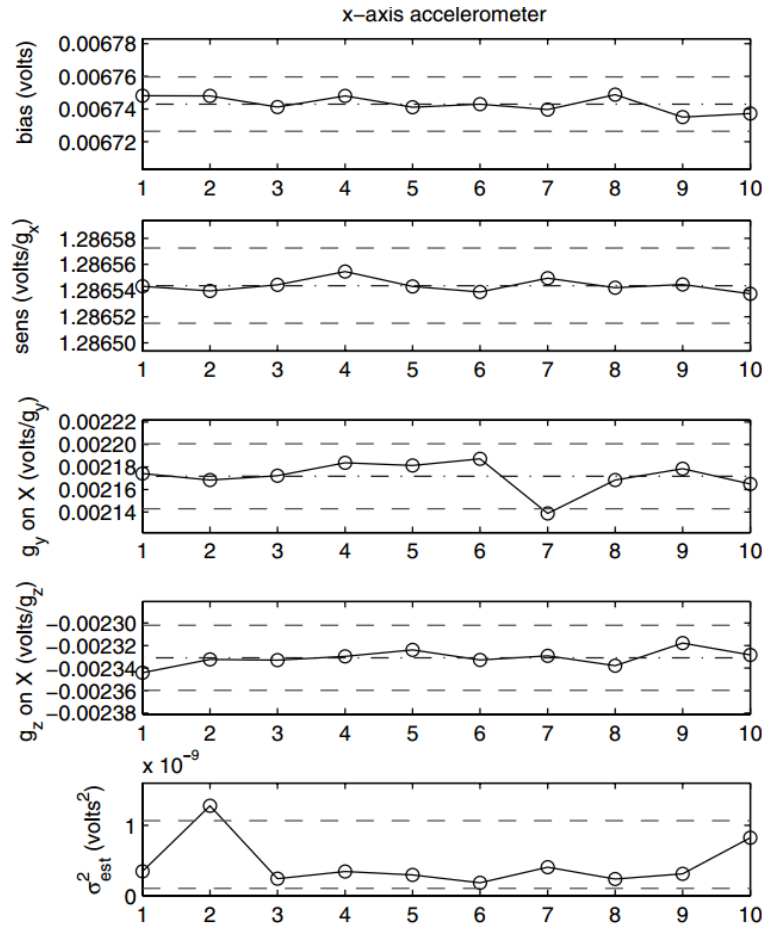


Figure 3: Example of Statistical Process Control Charts [6]

The plots above only show the x-axis model coefficients. There are four coefficients generated per axis when calibrating these instruments. Each component requires a bias coefficient (intercept), a primary sensitivity which represents the predominant behavior and two cross-axis sensitivities (y

and z-axis). The authors use a variation on the standard control charts, where variance is plotted in place of a range (R) or moving range (mR) chart. The limits are calculated from the data and represent three standard deviations from the mean. Excursion from these limits is called a loss in statistical control. A user wishes that all of the points remain in statistical control (within the limits).

Notice that the variance on calibration 2 was found to be out of the limits. At this time, a researcher would begin to search for an assignable cause for the out of control point. Whether or not a cause is found, the researcher then must make a decision to either disregard the data or not. In this case, no cause could be found and the data were not excluded from subsequent analysis.

1.2.3. Statistical Methods Applied to Wind-Tunnel Data Quality

Additionally, statistical methods have been used in Langley wind tunnels for data quality assurance and force balance repeatability. In 1996, an outline for a data quality assurance program was proposed and adopted by the Langley Wind-Tunnel Enterprise (LWTE) which outlined a plan for the addition of SPC, measurement uncertainty predictions, and tunnel-to-tunnel reproducibility to develop a data quality assurance program along with nine other facilities [8]. These methodologies reinforced the need for a repeatable wind-tunnel experiment, estimation of the process standard deviation and continuous checks to ensure the tunnel is operating within standards set forth by NIST.

In 2000, a series of check standard tests were performed and presented in line with the data quality assurance plan outlined four years earlier [9]. The document generated from the tests featured multiple control charts, including a 3-way chart that shows range and moving range. In the document, values such as tunnel temperature, Mach number, tunnel pressure, and wind-tunnel balance data with the model attached was charted and assessed for consistency with previous data

and for statistical control. Statistical control is achieved if all points remain in the control limits. The chart below, shown in Figure 4, is showing values for the Mach number coefficient. This value was tracked in the test section of the Unitary Plan Wind-Tunnel. Three common types of charts are shown, an \bar{X} , R, and mR chart, which track the mean of the process through subgroup averages, the range of each subgroup, and the moving range of the subgroup averages. Again, the limits are calculated from the data and represent three standard deviations.

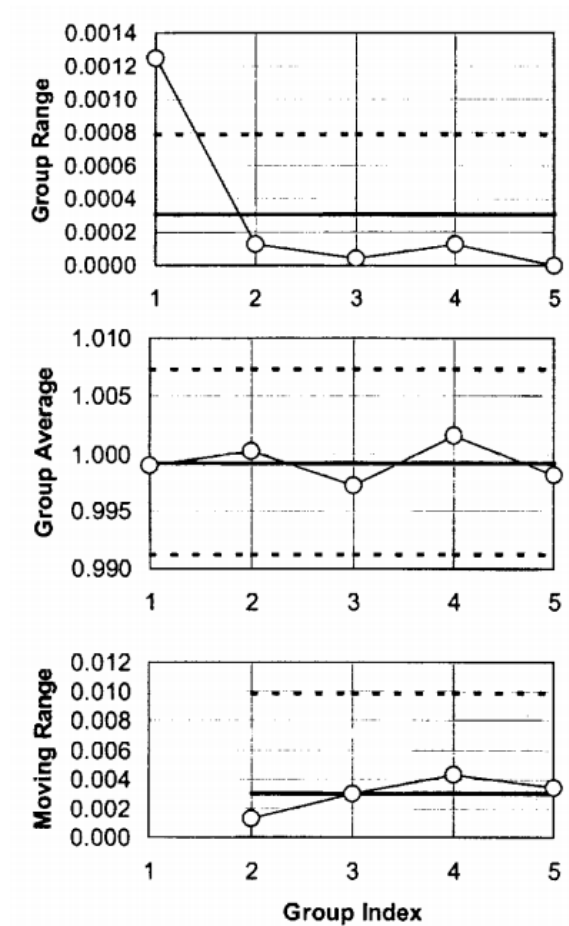


Figure 4: Test Condition Mach Number Coefficient (C_{Mach}) Check Standard Example

This type of charting (using all three charts) is not mentioned much in the traditional literature [7]. Either a range or a moving range chart is used. If replicate observations are available, a range chart is usually the recommended option to accompany a process average chart [7]. From

the literature, the group average chart tracks the process mean between groups of data acquired. The range chart tracks the short-term repeatability/within-group dispersion of the data. The moving range chart tracks between group dispersion. In this case, the moving range tracks the difference between group averages. From the data presented, it seems that not much information can be derived from having both styles of range charts. No out-of-control points were ever seen on the moving range chart though some were detected on the range chart. The document enforces the notion that without a credible estimate of the process standard deviation, no measurement in the wind-tunnel is meaningful. Having adopted some statistical techniques such as control charts, the wind-tunnel can be shown to be a stable system with statistically defensible measurements.

Repeatability is also of large interest to the balance calibration study. Repeatability can be separated into between-group repeatability and within-group repeatability in this context. For example, all of the replicated points within an experiment (calibration) would be within-group repeatability. Being able to match the results from that experiment with an identical experiment would be between-group repeatability.

Further work within the LWTE, as part of the data quality assurance project, has been aimed at determining repeatability statistics for any measurement taken in the wind tunnel [10]. The LWTE data quality assurance plan also involved periodically measuring repeatability in the tunnels by using standard wind tunnel models. For this study, two models are used and two balances are used. The measured repeatability and reproducibility would be characterized in terms of the tunnel itself, the wind-tunnel balance, and the model parameters. This characterization could be used to predict and confirm within-test repeatability and between-test repeatability for any test performed in the wind-tunnel. Figure 5 shows control charts for the within-group repeatability on the left and between-group repeatability on the right. Both wind-tunnel models are included in

these plots. Each of the dots in a colored series represents a series of values obtained for the given angle of attack. The within-group charts show the range of values and the between-group charts show the average of the values.

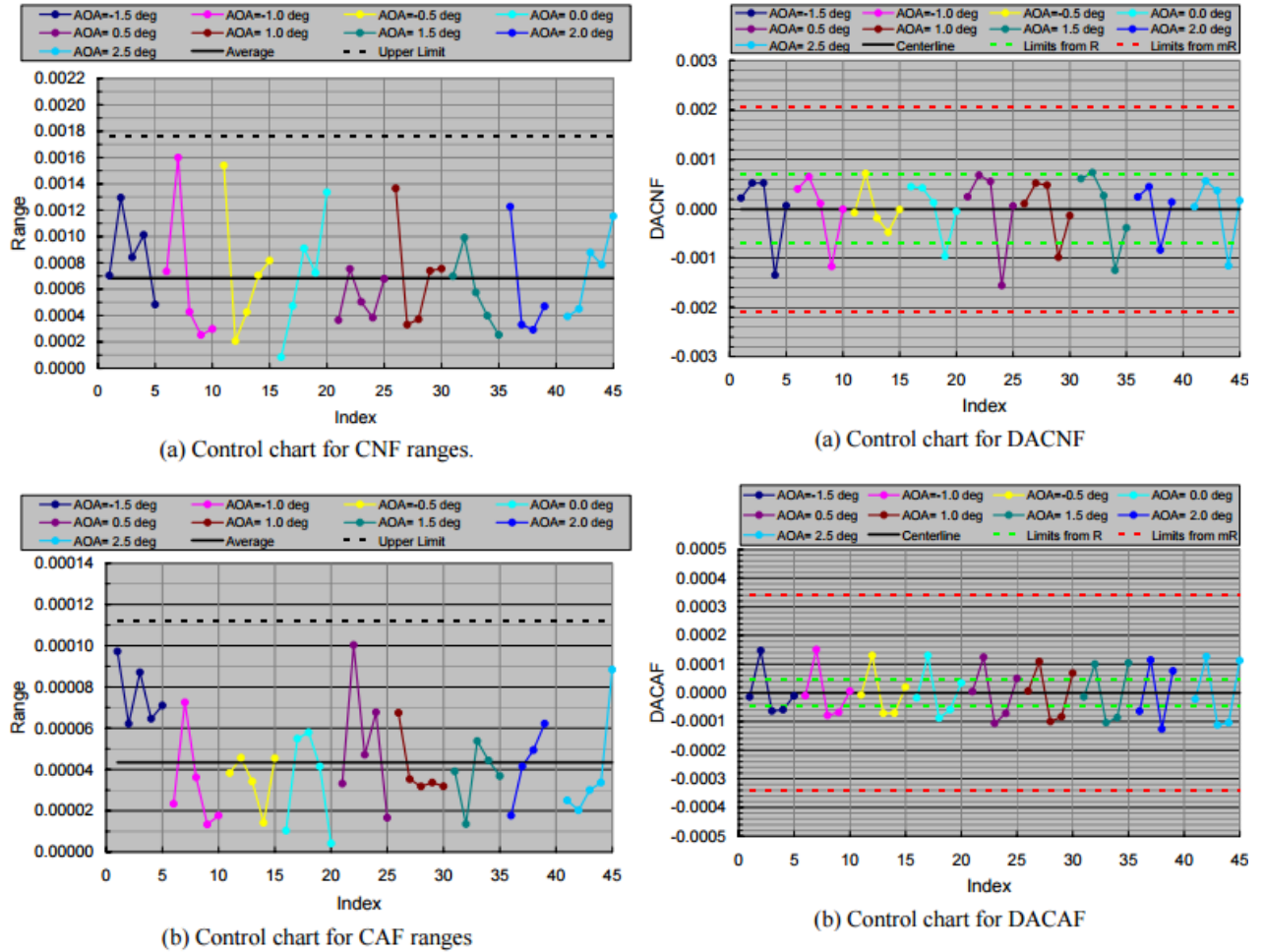


Figure 5: Example of Within-Group Variation/Repeatability (left) and Between-Group Variation/Repeatability (right) for NF and AF [10]

Note that this thesis also uses limits established both with range and moving range. The range limits are generally, the smaller limits. The between-group plots on the right show statistical consistency for the limits given by the moving range and not the range. This implies there is small

within-group variation which drives the range-based limits down. However, significant between-group variation exists which exceeds the range-based limits. This conclusion is corroborated by the left hand plots which show good results for within-group repeatability.

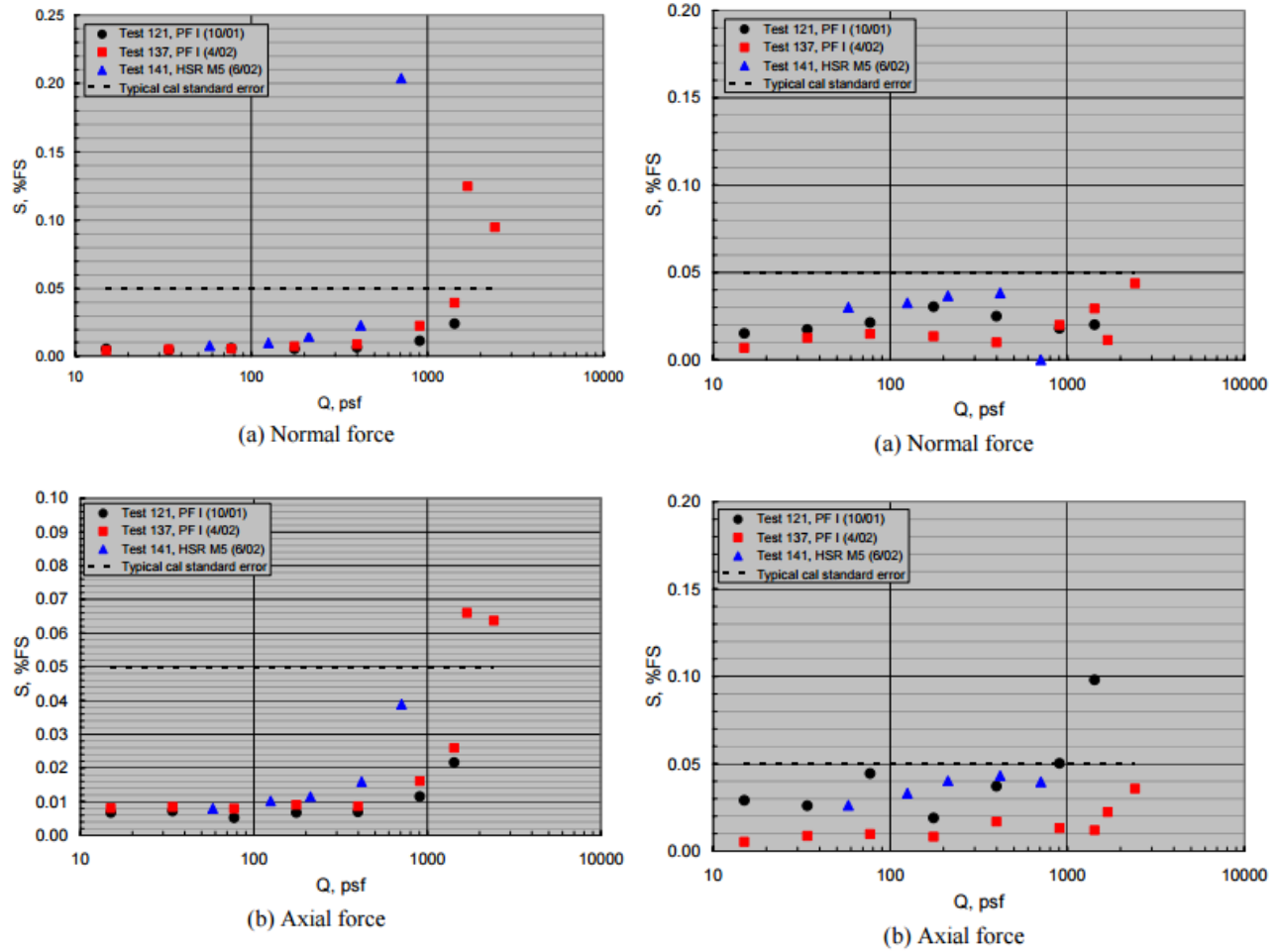


Figure 6: Estimated Standard Deviations for Within-Group Variation/Repeatability (left) and Between Group Variation/Repeatability (right) for NF and AF [10]

Estimates of the within-group and between-group standard deviations are also obtained using the control charts. Figure 6 shows the standard deviation of both the within-group standard deviation on the left and the between-group standard deviation on the right. This statistic would be a very useful value to have for any customer of the tunnel. It could be quoted to lend assurance

to the quality of the test data and will be explored further in this thesis. The standard deviations shown are calculated from each observation and combined either within-group or between-group. The standard deviations are normalized by the full-scale balance output. Also plotted is the typical reported standard uncertainty of a balance calibration (0.05% FS). The agreement looks good on the repeatability side at lower dynamic pressures. The balance output repeatability is nearly equal to the resolution of the balance [10]. However, higher dynamic pressures cause the standard deviations to rise for both balances. Across-test repeatability looks even better and it would appear that across-test repeatability is independent of dynamic pressure and is controllable to a level below the balance calibration uncertainty [10].

A similar study was performed in the Langley 14x22 wind-tunnel using two check standard models as well, one of the models being the same as in the previous study for NTF [11]. This test compares three types of grouping: Within-Group Tests, which are back-to-back measurements of the same parameter Within-Test, which are comparisons of the within-group measurements across the entire wind-tunnel test and Across-Test, which are comparisons of the within-test measurements across multiple experiments. In the previous study, higher dynamic pressure caused higher uncertainty in the results. This experiment shows the opposite trend. The lower dynamic pressure conditions showed more variation than the higher dynamic pressures. Note that the 14x22 is a subsonic tunnel and the dynamic pressure ranges were limited to about 100 psf [11]. It makes sense that both very low dynamic pressure and very high dynamic pressures induce different types of instabilities in the test conditions. Similar methods were used to assess the within and across-test repeatability in this experiment. The author also attempted to model both within and across-test repeatability in this experiment with some success.

Integration of other rigorous statistical methods has continued to be common in the LWTE. NASA LaRC has been experimenting with a DOE approach to wind-tunnel testing since 1997. DOE differs from traditional one-factor-at-a-time (OFAT) methods of testing that many are familiar with. OFAT testing can be understood as holding all but one factor constant. This process would need to be repeated for all factors in the experiment. If interactions were of interest, then each factor would have to be varied in various combinations with all other factors. This method of testing is inefficient and fraught with biases. DOE allows for multiple inputs to be varied simultaneously, thereby reducing the total number of runs needed to characterize a system [12]. DOE leverages statistical knowledge and practices with experimental procedures to increase the quality of research while reducing the cost [12]. For in-depth information on DOE and how it's used in the LWTE, refer to DeLoach [13].

In conclusion, the addition of statistical methods to the calibration of wind-tunnel balances and check standards for wind tunnels has led to significant advancements in the quality, efficiency, and understanding of wind-tunnels and wind-tunnel balances. Long and short-term experimentation on wind tunnel balances has been previously undertaken, however, such an extensive study as proposed here has never been completed. Using methods proposed here, the long-term balance calibration study will derive and further the understanding of the calibration of wind-tunnel balances and the effect calibration systems have on these balances.

CHAPTER 2

BACKGROUND, BALANCES AND CALIBRATION CAPABILITIES

2.1. Internal Strain-Gauge Balances

Balances are classified as either internal or external. Internal balances are mounted inside of the wind-tunnel model, typically at the end of the sting (support), while external balances are external to the model. There are multiple types of internal balances further differentiated by their mechanical and electrical design. Balances can be machined from a single piece or multiple pieces of material. A graphic showing the use of a balance inside a wind tunnel model is shown in Figure 7.

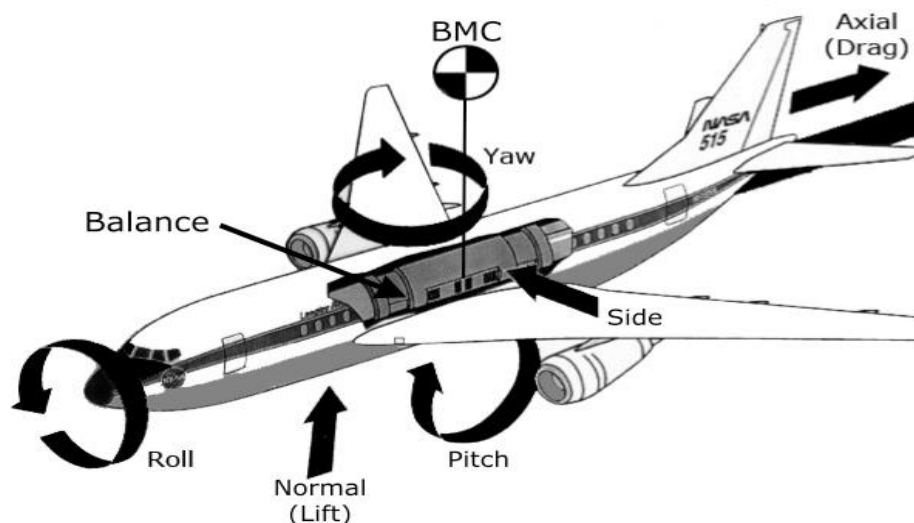


Figure 7: Diagram of Wind Tunnel Model with Balance

Balances are used to measure the six aerodynamic loads encountered by an aircraft: the forces in the X, Y and Z direction, and the moments about them. These forces are more commonly referred to as normal, axial, and side force; the moments being pitching, rolling, and yawing

moments. Balances are designed with a series of structural flexures that deform when a load is applied. Each flexure is instrumented with a strain gauge(s), which vary in resistance proportionally to the deformation of the flexure as load is applied, otherwise commonly called strain. Each strain gauge is an electrical resistance device that makes up one leg of a Wheatstone bridge. By exciting the Wheatstone bridge with a source voltage while a load is applied, the unbalanced bridge will provide a signal voltage proportional to the strain. Each flexure is designed to be sensitive to only a single component of load, meaning that when a normal force is applied, only the normal force bridge returns a response. However, manufacturing restrictions and material and instrumentation imperfections require an understanding of interactions between load components. Therefore, a full characterization is required to understand the balance mechanics. Figure 8 shows a common wind tunnel balance. It is comprised of a metric end, a non-metric end, and three center cage sections. The metric end is directly attached to the wind tunnel model and the non-metric end attaches to the sting support. The outer cage sections hold all of the sensitive flexures and strain gauges for determining all forces and moments except for axial force. The center cage houses a sensitive structure used for measuring forces in axial directions. This section is arguably the most sensitive and difficult to design due to large interactions.

According to common archetypes, balances can further be categorized mechanically as a force balance or a moment balance. The difference is directly related to how the balance flexures are designed. In a force balance, the balance measures five forces and one moment. This means that most of the flexures are designed to be in tension and compression. The rolling moment flexure is usually the only flexure that is subjected to bending. A moment balance is exactly opposite; the balance measures a single force and five moments. Flexures are subjected to bending moments except for the axial force flexure which measures mostly compression. Force balances are often

stiffer than moment balances and, therefore, have a higher load capacity. However, this does reduce the sensitivity of the balance itself. Moment balances rely on bending and are usually more accurate. However, the bending can cause higher order terms to be significant in a model. Balance design is fundamentally a tradeoff between increasing the sensitivity of a balance without sacrificing global stiffness and causing extreme local strain concentrations. By increasing strains within the balance in the flexures, each strain gauge will undergo more strain per pound of load and essentially become more sensitive. The balance must also maintain its structural integrity for use in high load environments. The next sections contain descriptions of the balances and calibration systems used in this study.

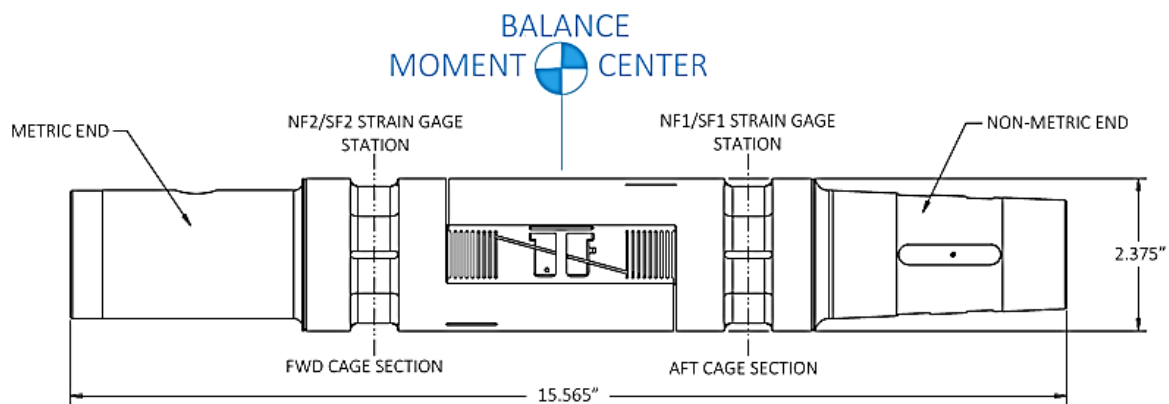


Figure 8: Diagram of Common Wind Tunnel Balance (Single-Piece)

2.1.1. Multi-Piece Balance

As mentioned earlier, several types of balances exist. There are a number of manufacturers with many proprietary designs. However, all of them essentially work the same way. For this study, three common types of balances are being used and will be described in the next sections.

Multi-piece balances are designed exactly as they sound. They are comprised of multiple pieces that are joined together with fasteners and pins and enclosed within an outer shell. Forces

are transferred through the outer shell onto the flexural elements. The shell can make these balances larger than single-piece balances which then requires more space in the model. The shell encapsulates the balance pieces in the event of a catastrophic failure and makes the balance globally stiffer than a single piece. A stiffer balance will deflect less under load which reduces the chance for model fouling and mounting failure. The common Task Multi-Piece Design is fastened together with pins. These pins form the load path between structural pieces inside the balance. Ideally, the balance can accurately measure loads with the added benefit of being able to be simultaneously manufactured and be disassembled when necessary. In practice, the pins have been suspected of loosening over time and with use, which changes the response of the balance. Multi-piece balances often require higher order models to accurately define this behavior. The balance can even develop bidirectionality, which is a term given to balances that exhibit dependence on the sign of the load applied [14, 15]. Figure 9 shows an example of bidirectional behavior.

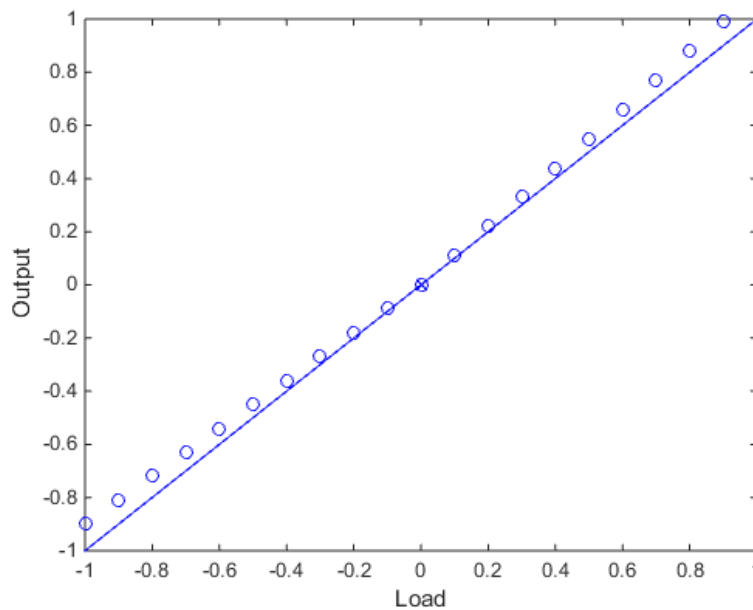


Figure 9: Example of Bidirectional Behavior

Put more simply, a bidirectional balance will respond differently when loaded in the positive direction versus the negative direction. Figure 9 shows a change in slope of the balance response for positive loads or negative loads, where the dots represent the balance data and the line represents a standard linear response. Bidirectional balances are often fit with absolute value terms rather than cubic terms or a separate model for positive and negative loadings. Absolute value terms essentially adjust the base coefficients according to the signs of the component loads [14, 15]. A drawback of higher order modeling is that a larger matrix of coefficients is needed. This, in turn, means more design points are required in the load schedule which costs money and time. Additionally, more coefficients add computational load when reducing data. This is admittedly not an issue for modern computers but can be considered nonetheless. Finally, absolute value terms cause much higher correlation between regression coefficients within the model [14, 15]. Correlation means regression coefficients are linearly dependent on one another and therefore not unique. Multicollinearity, or correlation, is measured by a Variance Inflation Factor (VIF). A rule of thumb from Montgomery [12, 16] states that any VIF below 10 is generally acceptable though, for the purposes of this thesis research, a value of 4 has been chosen. The rationale for using 4 comes from years of historical understanding and expert opinions at NASA LaRC. Having a low VIF means that each coefficient is relatively uncorrelated with any other coefficient and can be uniquely estimated. Adding absolute value terms can cause VIFs to rise on average above 20. This implies that regressors are now correlated with one another. With many large VIF's, it will no longer be clear which regressors are correlated with one another. The correlation structure will become obscured. Note that high VIFs do not imply the model will not predict well. If fit properly and employed on the right balance, absolute value terms will likely improve the model fit. However, the researcher must understand that with high VIFs, no coefficient nominally represents

the primary component or interaction it's named after. Colinearity of the coefficients makes that difficult. It is for this reason that absolute value terms are avoided in this study. The multi-piece balance will be modeled to a quadratic fit so it is comparable with other balances in this study. Unfortunately, the absolute value terms often have significant explanatory power over the response of a multi-piece balance.

The Task MK-29B is the balance selected to represent the multi-piece balances, seen in Figure 10. This balance is owned by NASA Ames Research Center. The balance is a 2-inch diameter balance and 11.5 inches long. Internally, the balance operates using orthogonally placed force elements. The internal makeup can be seen in Figure 11. This balance is considered a force balance as the elements measure applied force and moments are resolved by understanding the distances of each element from the balance moment center (BMC). It is important to note that the normal force and side force elements are at different distances from the BMC and moments must be resolved with that in mind. This is because the pins require room inside the balance. The maximum load limits of the balance can be found below in Table 1. Also found in Table 1 is the accuracy's for each component of the balance. These values are averaged from historical calibrations. They represent two standard deviations of the error in each components fitted model and are in percent full scale. Note also that there are some caveats related to how much force and moment can be applied to the balance. Considering how the balance is designed, having two separate force flexures means that each flexure must be independently considered when applying forces and moments. If a load is centered on any single gauge, the limit is half the total balance quoted load capability. Additionally, the moment created on the opposite element by that force must be considered. The force-moment relationship can be more easily understood using a load rhombus. This graphic was taken directly from the balance schematics. See the rhombus in Figure

12. The rhombus accurately shows that maximum normal force or side force can only be applied at BMC when no moments are generated. Otherwise, a linear reduction of the load must be observed.

There are some advantages of a Task balance. With any other type of balance, there needs to be room allowed in the model for balance deflection. With a Task balance, the deflection happens internally, and therefore, no space is needed inside the model for balance deflection, leaving marginally more room for more instrumentation. Additionally, this balance has a higher load capacity and stiffness than other balances. This is possible because the balance has a separate axial section and a shell. Task balances are also the least costly of all the current balance designs. Disadvantages are the aforementioned inaccuracies from load path disjunction, which is exacerbated by wear and tear. Also, these balances have a minimum diameter of about 0.75 inches. This is due to allowing balance deflection to happen inside of the outer shell.



Figure 10: Multi-Piece Balance – Task Mk-29B

RATED LOADS (MK-29)		2σ (%FS)
NORMAL FORCE (EACH ELEMENT INDIVIDUALLY)	2100 Lbs	0.254%
SIDE FORCE (EACH)	700 Lbs	0.374%
AXIAL FORCE (TOTAL)	350 Lbs	0.073%
ROLLING MOMENT (TOTAL)	3800 In/Lbs	0.347%
PITCHING MOMENT @ ZERO N.F.	15225 In/Lbs	0.503%
YAWING MOMENT @ ZERO S.F.	4200 In/Lbs	0.744%

Table 1: Rated Loads for the Task MK-29 and Associated Accuracy's

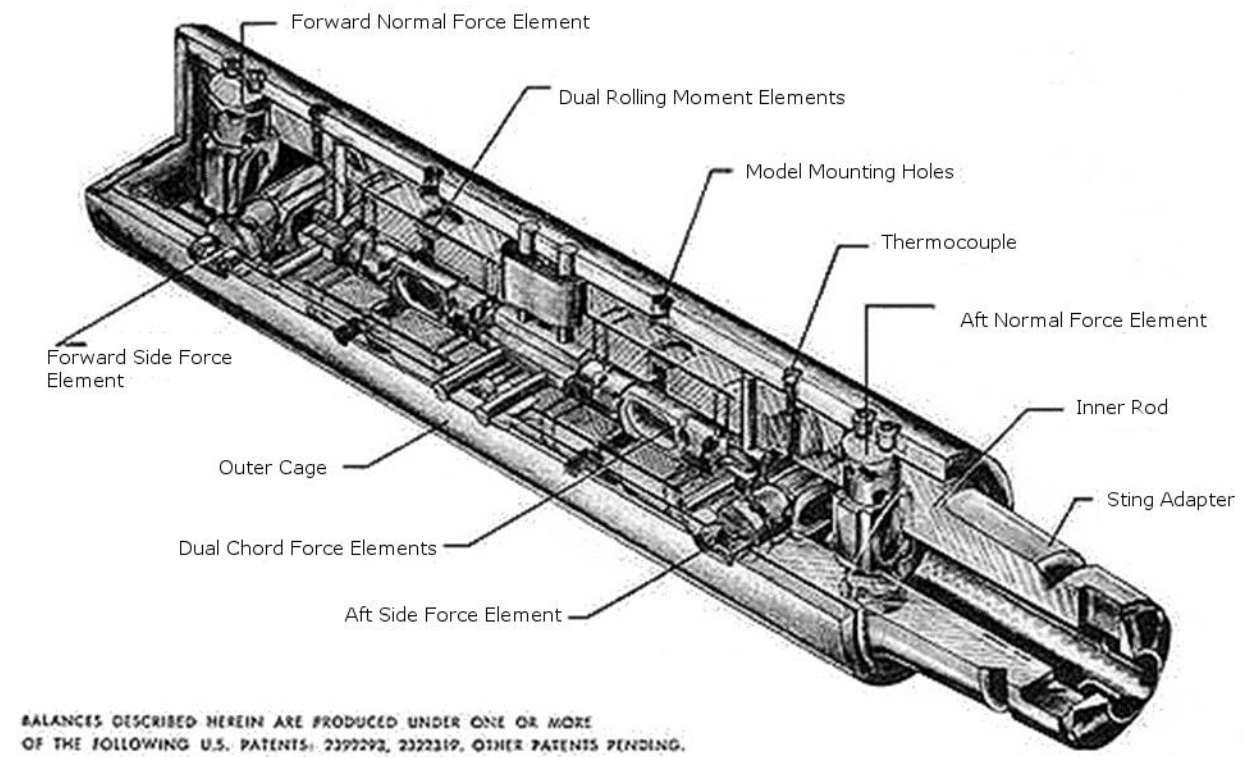


Figure 11: Internal Structure of Task Mk-29

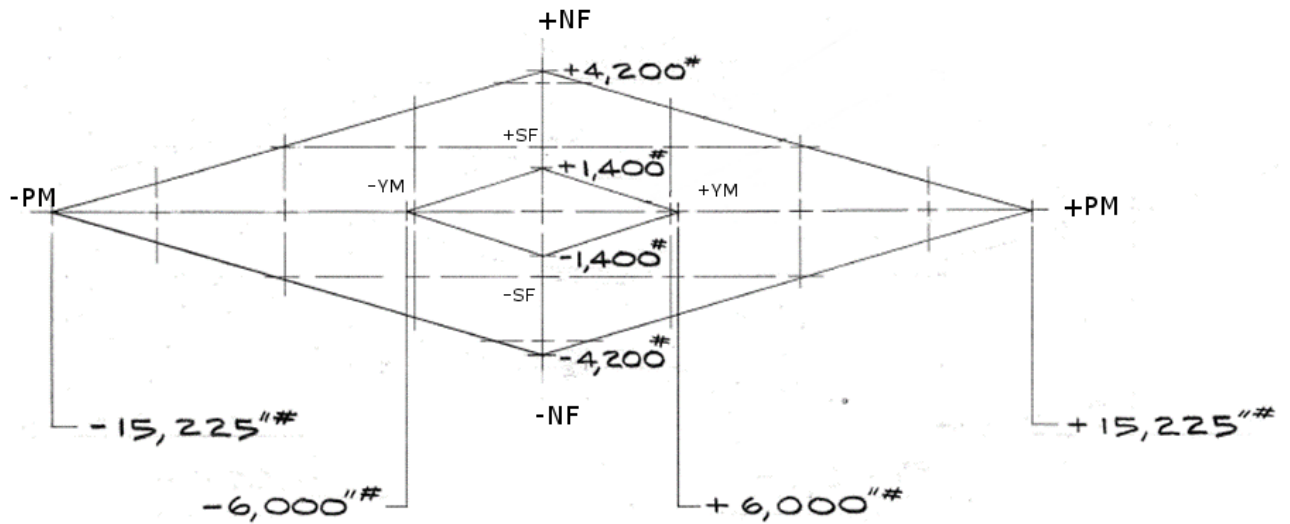


Figure 12: Load Rhombus for the Mk-29

2.1.2. Unitized Multi-Piece Balance

Unitized multi-piece balances are similar to other multi-piece balances as they are made from multiple pieces. They also have an outer shell which protects the internal pieces and acts as a failsafe, much like the Task balance. The loads are transferred through the outer shell into the elements inside of the shell. These balances are also usually force style balances. The similarities between the two multi-piece balances essentially end here. While they may seem similar, it is not necessarily fair to group these balances together as they operate using different methodologies.

The pieces that make up a unitized balance are instead welded together rather than fastened. This essentially unitizes the balance which causes it to act much more like a single-piece balance and can be modeled accurately with a quadratic model. This is commonly known to users and researchers in the industry. Absolute value terms or higher order models are usually not necessary.

The unitized balances are manufactured by Triumph Aerospace – Force Measurement System Division. The balance chosen to represent the unitized design in this study is designated the Triumph MC-60E. This balance is owned by NASA Ames Research Center and can be seen in Figure 13. This balance is 2-inch diameter and 11.12 inches long. Internally, this balance appears more like a single-piece balance with a series of gauged elements; though they measure force, not moment. The elements for normal and side force are placed at either end of the balance. Both normal and side are equidistant from the BMC because they do not need room for pins. Axial and roll are placed in the center of the balance. The outer shell of the MC-60E operates as the metric end of this balance. The model attaches directly to the outer shell and strains are measured at the shell edges in the elements/flexures. The axial and roll sections are more complex. While one can only speculate as the design is proprietary, it is believed that pins connect the outer shell to a symmetric axial section within the inner rod. The dual axial gauges allow for roll moments to be obtained using gauges specifically placed for roll torque. Figure 14 and Figure 15 show the internal workings of the unitized balance in more depth. Much like a Task balance, similar limitations and caveats must be considered when dealing with the load capabilities of a Triumph unitized balance. Again, because the balance measures forces at either end of the balance, moments on the opposite flexure must be considered. The load range of the MC-60E and its historical accuracy can be seen in Table 2. These loads are quoted as shown in the technical drawings for the MC-60E. The load rhombus can be seen in Figure 16. Again, much like the Task, full normal and side force can only be applied directly at the BMC when no pitch or yaw moments are applied respectively.

The advantages of a unitized balance are that the balance is less expensive and easier to manufacture than a single-piece but can be nearly as accurate. The unitization allows the balance to be modeled using a lower order polynomial when compared to a Task multi-piece. These

balances are also often stiffer and offer somewhat higher load capacities. One of the disadvantages are that, once unitized, the balance cannot be disassembled.



Figure 13: Unitized Multi-Piece Balance

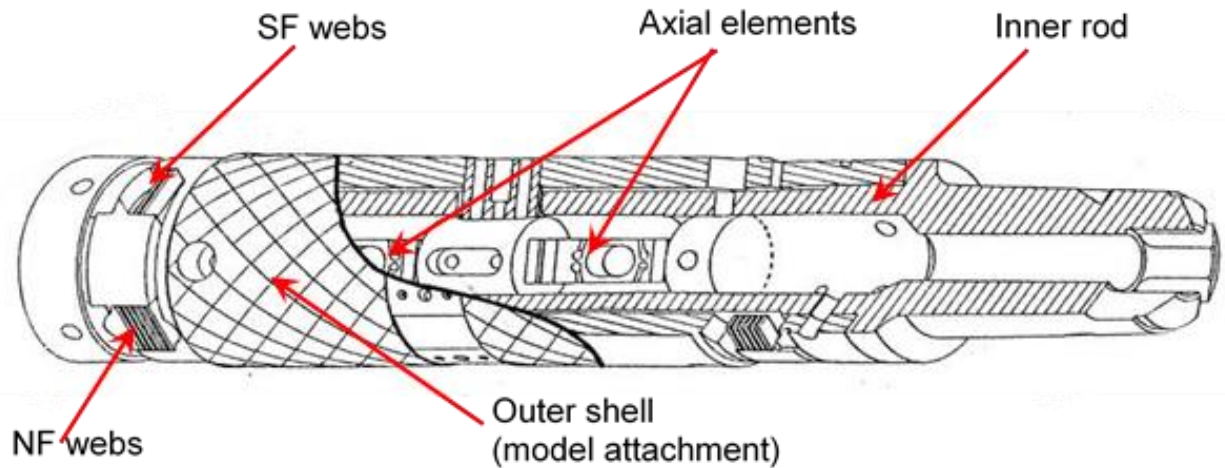


Figure 14: Internal Diagram of Unitized Balance

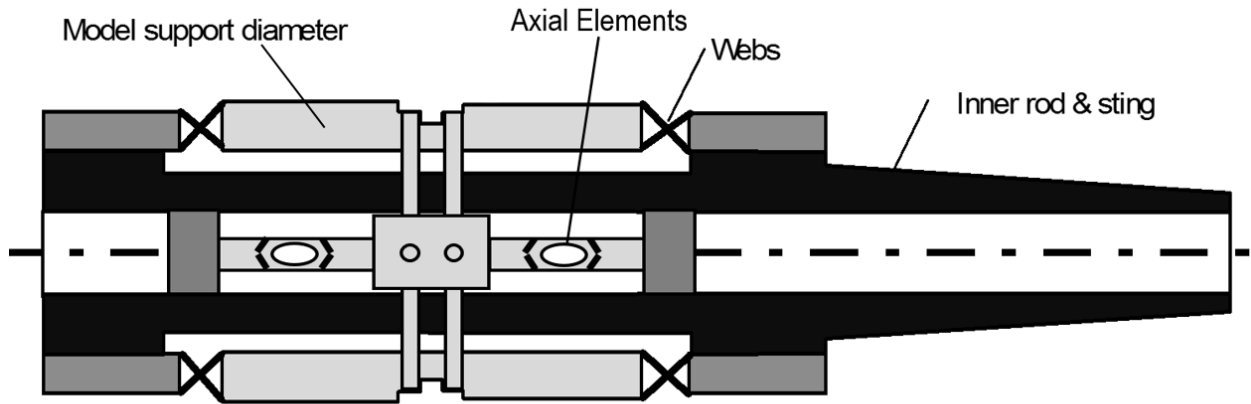


Figure 15: Simple Diagram of Unitized Balance Flexures

RATED LOADS (MC-60E)		2σ (%FS)
NORMAL FORCE (TOTAL)	5000 Lbs	0.077%
PITCH MOMENT @ ZERO N.F.	20000 In/Lbs	0.122%
SIDE FORCE (TOTAL)	2500 Lbs	0.082%
YAW MOMENT @ ZERO S.F.	10000 In/Lbs	0.281%
ROLL MOMENT (TOTAL)	5000 In/lbs	0.134%
AXIAL FORCE (TOTAL)	700 lbs	0.109%

Table 2: Rated Loads for the MC-60E and Associated Accuracy's

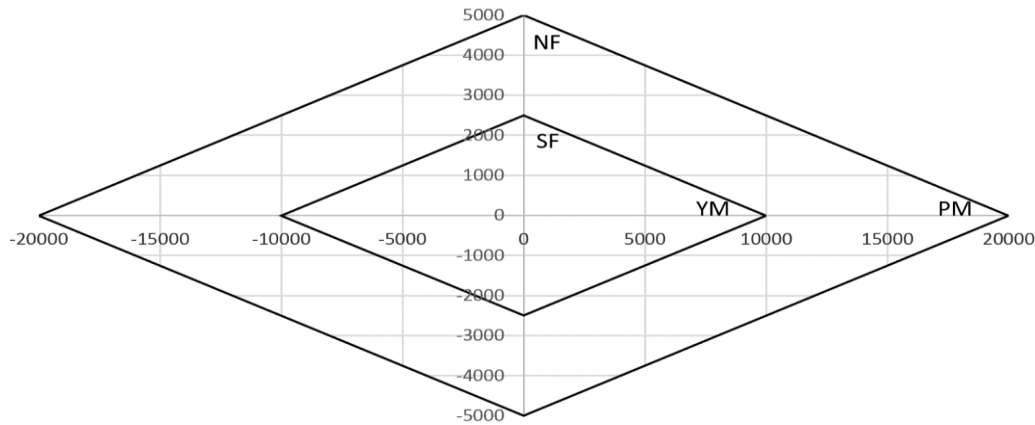


Figure 16: Load Rhombus for MC-60E

2.1.3. Single-Piece Balance

Single-piece balances are machined from a single piece of material which makes them smaller than multi-piece balances. They also don't have a shell which reduces the radius further. Being made of one piece, they are generally more homogeneous and less prone to hysteresis than multi-piece balances. However, the added complexity of machining the single-piece balance makes them much more expensive. Often these balances must be machined using an electrical discharge machine (EDM), which burns away material using rapidly reoccurring electrical discharges between two electrodes. This method is used to achieve the most precise tolerances possible up to $\pm 5\mu\text{m}$. These balances are generally understood to be the most accurate balances available on the market.

Being the most homogeneous balance, single-piece balances respond in a very linear fashion. A quadratic calibration model is used to account for the limited number of true interactions within the balance and to correct for any interactions created from machining limitations, material imperfections, and gauging mistakes. The interactions are very small compared to any other type of balance. The most complex part of this balance is the axial section at the center of the balance. Axial accuracy is highly important to researchers and separating all other components of force

from the axial section is a complicated process that defines the maximum load limits for these balances. In an effort to separate these interactions from the axial section and maintain axial accuracy, the axial section is completely independent from all other cages.

The balances chosen to represent the single-piece balance type in this study are the NTF-113C and the NTF-118B. Two balances were chosen because of damage found on the former balance. Both of these balances are NASA-designed balances, manufactured at Modern Machine and Tool Company. Both balances are 2-inch diameter balances and 15.565 inches long. While most balances at NASA are wired as direct read balances, meaning the gauges combine the loads inside the balance before being read, the NTF balances are wired individually per bridge. This form of electrical design is used for more active temperature compensation. NTF balances can be used in cryogenic conditions and, therefore, can see large temperature gradients within the balance. The normal and side force cage sections are placed at either end of the balance and the axial section is placed in the middle much like the other types of balances. The single piece balance has by far the most complex axial section of all the types of balances. Figure 17 shows the NTF-118B. The axial section is very apparent in the center of the balance. The center beam in the middle of the balance is the axial measuring beam. The smaller beams on either side of the measuring beam are designed to distribute normal force load. The T-shape on the top of the measuring beam isolates any residual normal force and pitching moment from the beam and enforces a single mode of bending. Figure 18 and Figure 19 show different views of the balance and can further the understanding of the complexity of these balances.



Figure 17: Single-Piece Balance

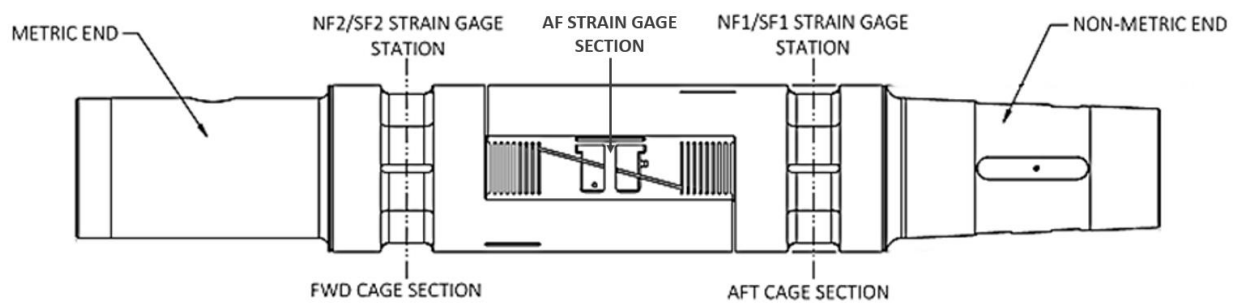


Figure 18: Single-Piece Balance Diagram

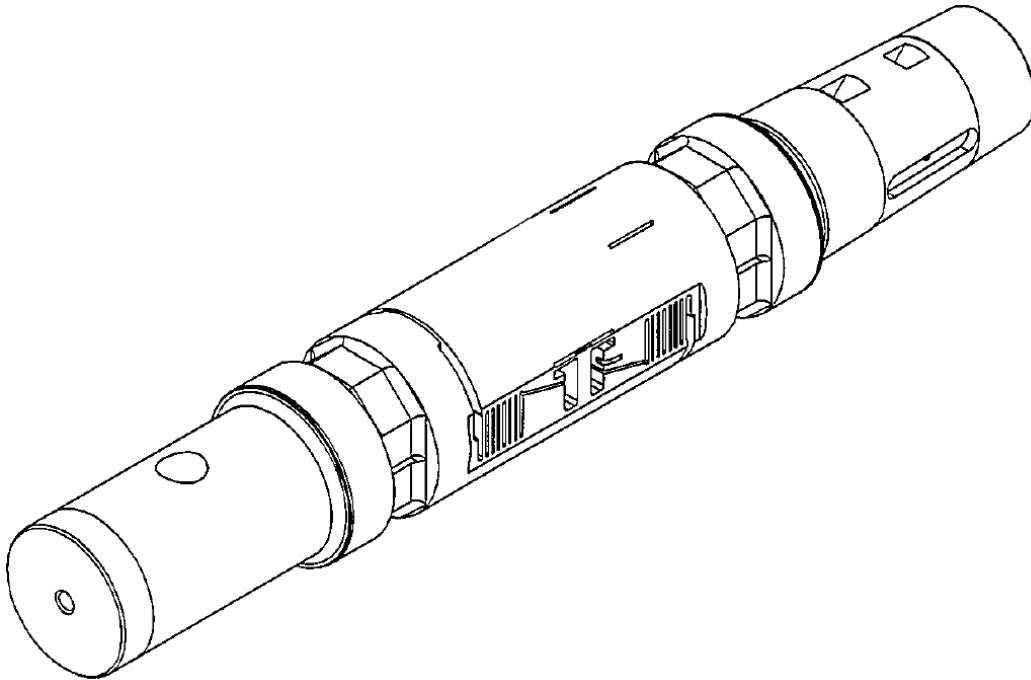


Figure 19: Isometric of a Single-Piece Balance

While generally less stiff than multi-piece balances, the NTF balances can still have respectable load limits. Table 3 shows the load limits and historical accuracy's for both the NTF-113C and the 118B. No load reduction (via a rhombus) is required on the NTF balances. Because the flexures measure bending, moments on the opposite flexure are equal. Therefore, the single piece balance load limits are true limits, unlike the multi-piece balances.

Advantages of the single-piece balances are numerous. These balances are considered by the industry to be the best balances available on the market. Their monolithic design makes them resistant to hysteresis and, therefore, the accuracy tends to be very high. The major disadvantage of the single-piece design is that it is expensive. The use of highly precise machining is required to create the necessary cuts in the balance. This costs time which drives the cost up. Also, single-piece designs are less stiff than other forms of balances. This will allow for more bending under load and requires more space within the model to prevent fouling.

RATED LOADS	NTF-113C	NTF-118B	NTF-113C 2σ (%FS)	NTF-118B 2σ (%FS)
NORMAL FORCE	6500 Lbs	6500 Lbs	.063%	.075%
AXIAL FORCE	400 Lbs	700 Lbs	.240%	.132%
PITCHING MOMENT	13000 in/Lbs	13000 in/Lbs	.107%	.206%
ROLLING MOMENT	9000 in/Lbs	9000 in/Lbs	.186%	.187%
YAWING MOMENT	6500 in/Lbs	6500 in/Lbs	.133%	.102%
SIDE FORCE	4000 Lbs	4000 Lbs	.092%	.043%

Table 3: Rated Loads for the Single-Piece Balances and Associated Accuracy's

2.2. Balance Calibration Methods and Methodologies

There are multiple ways to calibrate a balance using both manual and automatic calibration systems since no standards exist. Calibration methods are divided into gravity loaded systems and force actuator systems. The systems used to calibrate balances in this study are described in this section, and represent the state-of-the-art in balance calibration technique.

2.2.1. Automatic Balance Calibration System (ABCS)

The ABCS is capable of characterizing a six-component balance in a few hours. It employs a series of actuators and load cells to independently apply any combination of forces and moments to the balance, including pure moments. The balance is housed inside a calibration fixture and force is applied to the fixture. The balance is secured by a clamp which simulates a sting support in a wind tunnel. The ABCS can be seen in Figure 20. Optical sensors record the orientation of the balance within the fixture which eliminates the need for repositioning throughout the calibration. Repositioning and re-leveling is required on some systems to ensure that the forces applied to the balance during calibration act in the primary direction only. The automation is an advantage as it minimizes any biases from calibration to calibration and it speeds up the process considerably as

repositioning or re-leveling during calibration is a time-consuming process. The maximum load ranges for the ABCS is shown in Table 4.

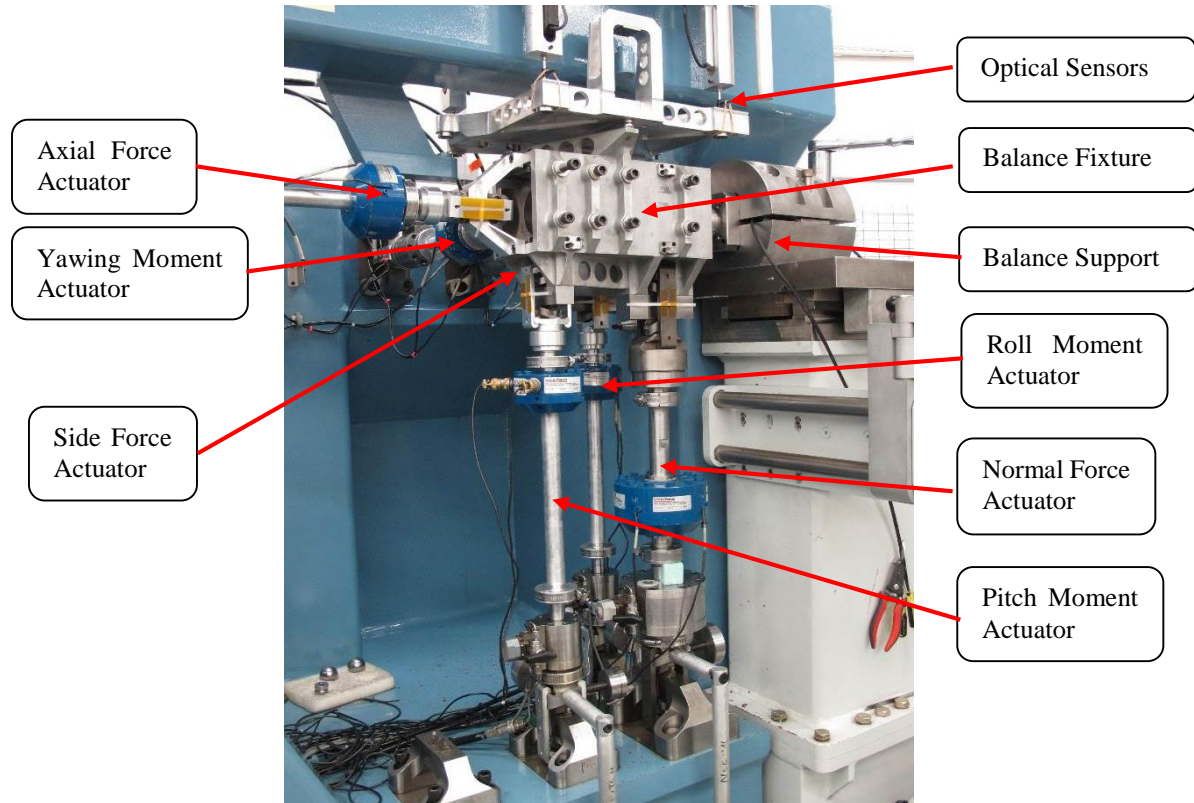


Figure 20: Automatic Balance Calibration System

COMPONENT	MAX ALLOWABLE	MAX USED
NF (lb)	11000	11000
AF (lb)	2500	2200
PM (in-lb)	80000	25000
RM (in-lb)	18000	18000
YM (in-lb)	16000	16000
SF (lb)	5000	5000

Table 4: Automatic Balance Calibration Load Ranges

2.2.2. NASA Langley Single Vector System (SVS)

The SVS combines a unique single-vector load application mechanism with a design of experiments (DOE) approach and has enabled a significant reduction in calibration time and cost while simultaneously increasing the quality of the calibration [5,6]. The SVS generates a six-component load directly on the balance through applied deadweight loads. The weights are attached to the SVS, which consists of both a yoke and a bearing and allows the balance to be reoriented while the applied weight remains in line with gravity. The difference in the balance orientation with respect to gravity creates the six-component loads on the balance. Moments are generated by moving the SVS ring off of the BMC and performing the same procedure. Angle measurement systems (AMS) are used to record the balance's three-dimensional orientation so the forces can be resolved. Moments are resolved as a function of the forces and the location of the force application. Electrolytic sensors are attached to the bottom of the yoke to ensure it is hanging perfectly in line with gravity. The Single Vector System can be seen in Figure 21 and Figure 22.

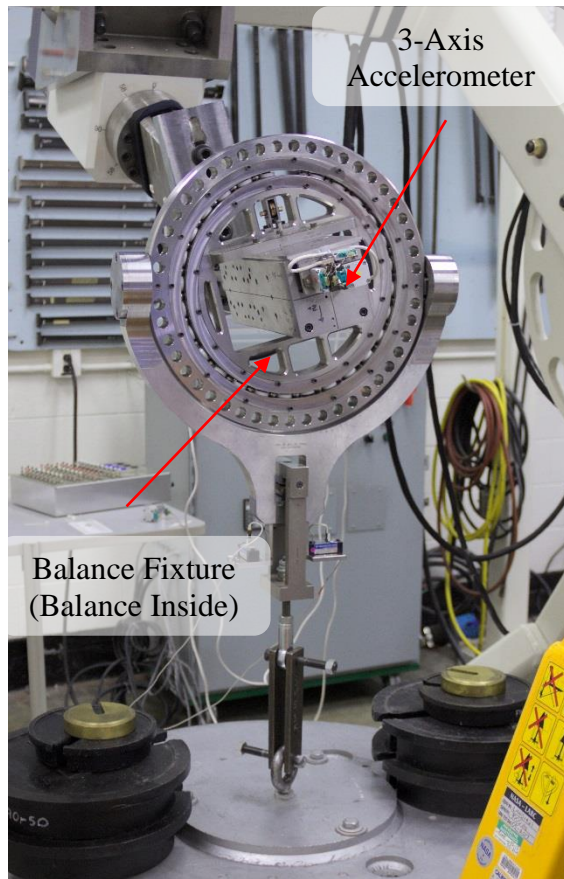


Figure 21: NASA Langley Single Vector System

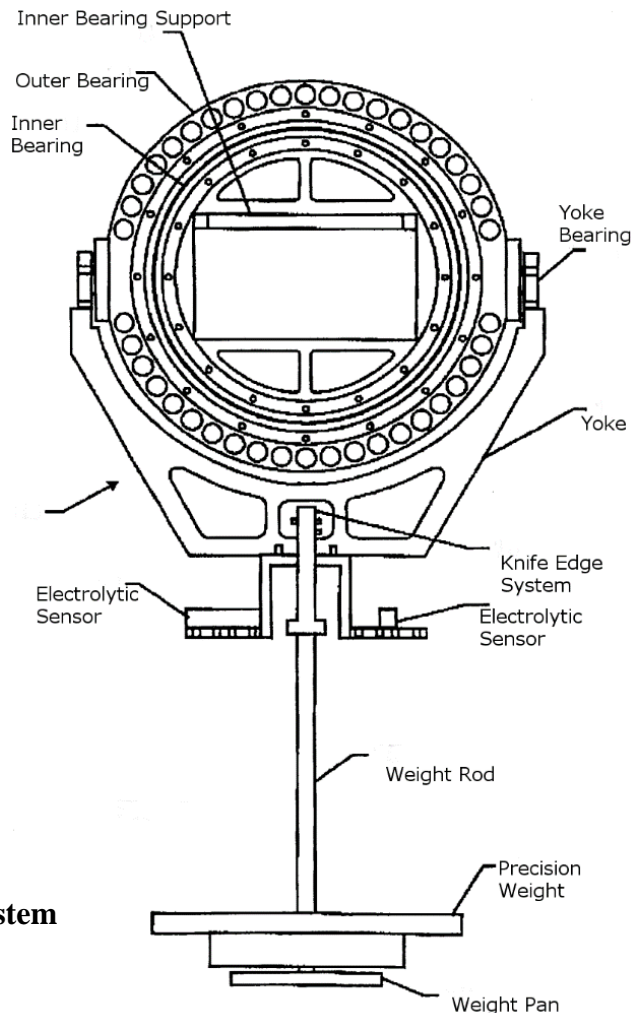


Figure 22: Single Vector System Diagram

2.2.3. Manual Calibration Stand

The Manual Calibration Stand is a traditional approach to calibrating a balance. Deadweight loads are applied to the balance calibration fixture either directly using knife-edges, arms, and hangars or using pulleys, cables, and bell cranks. Arms can generate large moments using small loads, and cable loads can generate side and axial forces. Knife-edges are used as they allow a load to be placed at a specific point with no transfer of moments unless intended. However, the balance

has to remain nominally level to ensure the applied forces remain orthogonal to the balance reference axes. This means that the balance must be re-leveled after each load series.

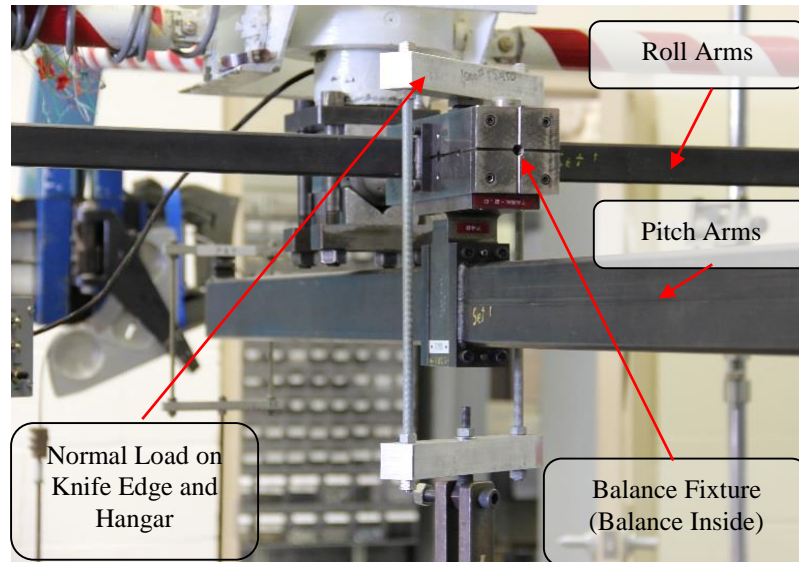


Figure 23: Small Stand Manual Calibration - Close Up

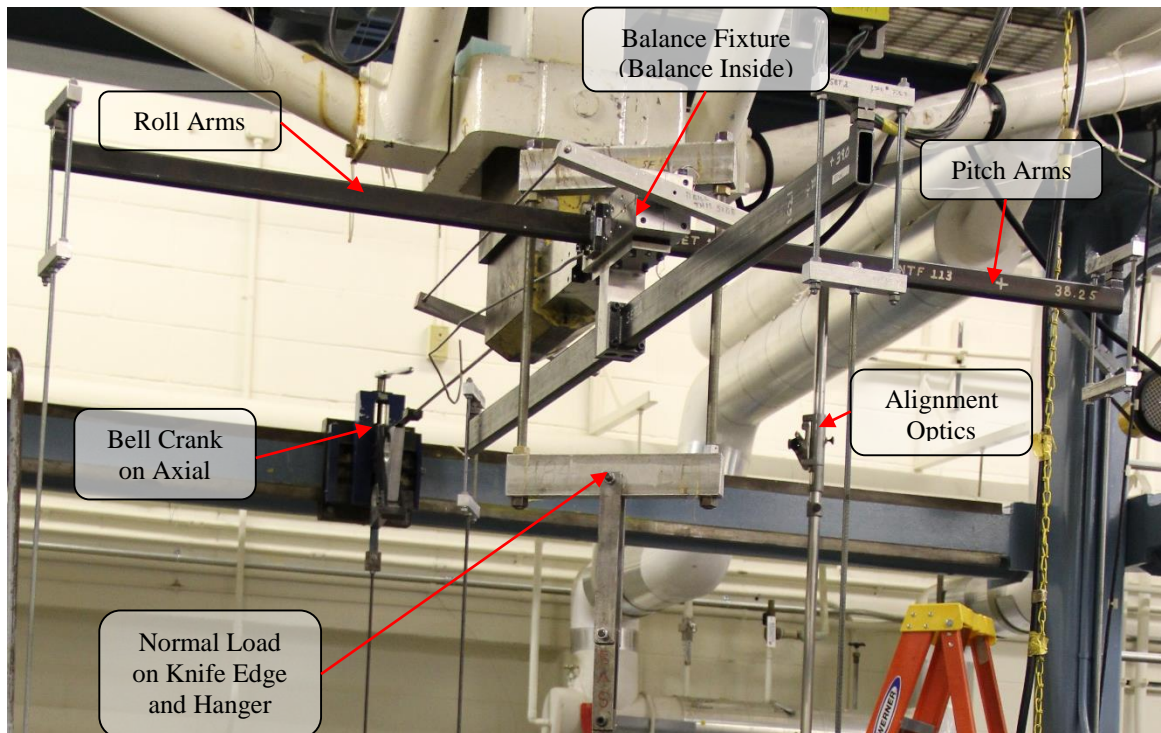


Figure 24: Large Stand Manual Calibration

As seen above in Figure 23 and Figure 24, a manual calibration is a very complex procedure that requires a substantial amount of equipment and specialty knowledge. In Figure 24 alone, a normal load is being applied at the BMC, pitching and rolling moments are being applied over a set of long arms and axial force is being applied over a cable and bell crank. The bell crank eliminates friction while turning the load path 90 degrees and aligning the cable with gravity. Pulleys typically have far too much friction to be used in a high fidelity environment such as required for balance calibration. Once all of the loads are applied to the balance fixture, the balance then needs to be re-leveled. The loads can cause the balance to deflect with respect to its position before it was loaded. Re-leveling ensures all of the forces are in their primary direction. Manual sighted optics are used to ensure that the bell crank is applying a load that is directly in line with the balance axis. A trained user looks through the lens and visually inspects the alignment. The bell crank can be raised, lowered, and jogged left or right to align the load. Additionally, the balance electrical output will be used to ensure alignment with cable loads, such as the one seen above in the axial force direction. Before the experiment, the balance would have been mounted with axial force in line with gravity. From this position, the axial force component will have been loaded and the outputs recorded. This point is called a lineup number and must be done for any cable load used in the calibration (essentially axial loads and side force loads). Then, during the calibration, when the axial component is no longer in line with gravity, the lineup number will be used to get an estimate when the cable is loading in line with the component.

2.3. Rigorous Load Schedule Design

At the center of every characterization initiative is the experimental design. The design of the experiment directly relates to the utility of the instrument being calibrated. The calibration experiment will define how the instrument is modeled, how well the model fits and how accurate

the instrument can be based upon the calibration runs. A load schedule is a term used for the designed experiment which explores the balances' relationship between applied load and output voltage. This relationship is then fully defined in a mathematical model. In many cases, a polynomial model is used and designed to fit responses of any order. This model comes from a Taylor series expansion over a finite design space. One should be cautious of over-fitting a model as interactions above cubic are very rare. It is much more likely that a lurking factor was not accounted for in the characterization experiment.

There are many types of experimental designs, all with different strengths and weaknesses. A design is often optimized for the system of interest using prior knowledge, expertise, and physical observations. Regardless of the system, there are a few desirable criteria that all designed experiments should have. Some examples of these criteria are: the model must result in good model fit, have enough degrees of freedom for lack of fit estimates and defined points for pure error estimates, and the design must be robust to outliers in data and provide a good variance distribution. A list given by Box and Draper outlines these criteria and can be found in standard texts [16]. When designing an experiment, there are numerous factors to consider. It is important for a researcher to choose the correct experimental design to most efficiently answer the question at hand. Montgomery's "*Design of Experiments*" [12] provides an excellent introduction to rigorous experimental design and can be considered a starting point for any newcomer to designed experiments. Further information on experimental design and analysis is given later in Section 3.2.

2.3.1. Automatic Balance Calibration System Load Schedules

The ABCS is the most versatile calibration system. Due to its ability to create any combination of forces and moments, it can execute any load schedule a user wishes. It can also replicate any load schedule performed on any other system. The standard load schedule performed

on the ABCS is a replication of a traditional Manual Stand OFAT experiment. This load schedule can have anywhere from 1000 to 2000 points within and can be completed within a few hours. The automation of the ABCS allows calibrations to have as many points as the customer wishes. A DOE style experiment can also be performed easily on the ABCS. A 62-point excerpt of the common ABCS load schedule can be found below in Table 5 and Figure 25.

As mentioned, the load schedule is a replication of a traditional manual calibration with some minor changes. The balance is loaded in two, three, and four component loads at five or more load increments. In the sample excerpt schedule below, rolling moment, yawing moment and side force are not excited. It is useful to note that currently the standard schedule is biased to normal force excitation. Other components are loaded less frequently and in auxiliary to normal force loading throughout the experiment. Auxiliary loads are loads which are supplementary to the primary load.

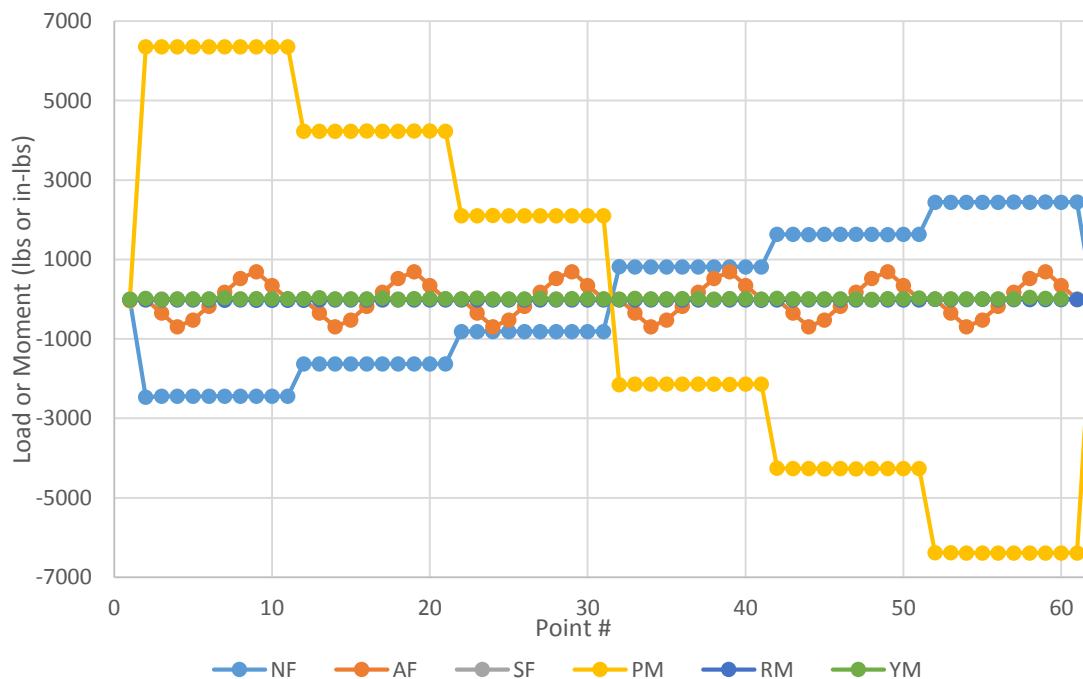


Figure 25: Plot of Load Series Excerpt from ABCS Calibration

POINT #	NF	AF	PM	RM	YM	SF
1	0	0	0	0	0	0
2	-2500	0	6400	0	0	0
3	-2400	-300	6400	0	0	0
4	-2400	-700	6400	0	0	0
5	-2400	-500	6400	0	0	0
6	-2400	-200	6400	0	0	0
7	-2400	200	6400	0	0	0
8	-2400	500	6400	0	0	0
9	-2400	700	6400	0	0	0
10	-2400	300	6400	0	0	0
11	-2400	0	6400	0	0	0
12	-1600	0	4200	0	0	0
13	-1600	-300	4200	0	0	0
14	-1600	-700	4200	0	0	0
15	-1600	-500	4200	0	0	0
16	-1600	-200	4200	0	0	0
17	-1600	200	4200	0	0	0
18	-1600	500	4200	0	0	0
19	-1600	700	4200	0	0	0
20	-1600	300	4200	0	0	0
21	-1600	0	4200	0	0	0
22	-800	0	2100	0	0	0
23	-800	-300	2100	0	0	0
24	-800	-700	2100	0	0	0
25	-800	-500	2100	0	0	0
26	-800	-200	2100	0	0	0
27	-800	200	2100	0	0	0
28	-800	500	2100	0	0	0
29	-800	700	2100	0	0	0
30	-800	300	2100	0	0	0
31	-800	0	2100	0	0	0
32	800	0	-2200	0	0	0
33	800	-300	-2100	0	0	0
34	800	-700	-2100	0	0	0
35	800	-500	-2100	0	0	0
36	800	-200	-2100	0	0	0
37	800	200	-2100	0	0	0
38	800	500	-2100	0	0	0

Table 5: Table of ABCS Load Schedule Excerpt

Table 5 (continued)

POINT #	NF	AF	PM	RM	YM	SF
39	800	700	-2100	0	0	0
40	800	300	-2100	0	0	0
41	800	0	-2100	0	0	0
42	1600	0	-4300	0	0	0
43	1600	-300	-4300	0	0	0
44	1600	-700	-4300	0	0	0
45	1600	-500	-4300	0	0	0
46	1600	-200	-4300	0	0	0
47	1600	200	-4300	0	0	0
48	1600	500	-4300	0	0	0
49	1600	700	-4300	0	0	0
50	1600	300	-4300	0	0	0
51	1600	0	-4300	0	0	0
52	2400	0	-6400	0	0	0
53	2400	-300	-6400	0	0	0
54	2400	-700	-6400	0	0	0
55	2400	-500	-6400	0	0	0
56	2400	-200	-6400	0	0	0
57	2400	200	-6400	0	0	0
58	2400	500	-6400	0	0	0
59	2400	700	-6400	0	0	0
60	2400	300	-6400	0	0	0
61	2400	0	-6400	0	0	0
62	0	0	0	0	0	0

Table 5: Table of ABCS Load Schedule Excerpt

2.3.2. Single-Vector System Load Schedules

The NASA Langley SVS employs a DOE-derived experiment and compliant hardware designed to perform highly efficient, three-force, three-moment load schedules. Currently, the SVS is limited to 3000 pounds of total applied load. This restriction of the load is related primarily to the structural limits of the load bearing. Traditional SVS load schedules are a central composite design (CCD) which efficiently allows quadratic models to be estimated using only 64 points. It is comprised of a half-fraction, fractional factorial experiment and an axial block of pure forces and force/moment couples. The SVS cannot produce pure moments and, therefore, true single

factor axial points are not possible. In general, 20 confirmation points are included to ensure model fit, however, they are not used to estimate the model. A graphic of the CCD can be seen in Figure 26. The cube represents a three-factor space where each vertex is a combination of maximum or minimum values of all three factors. The yellow points on the vertices of the cube are factorial design points and the star points are axial design points. The axial points are single factor points. Choosing this style design gives the experiment a desirable property called rotatability. This means that the variance of the predicted value is a function only of radius from the center of the design. In other words, the variance is constant on a given radius. This is a desirable property because it allows a researcher to tailor their experiment to a specific prediction variance.

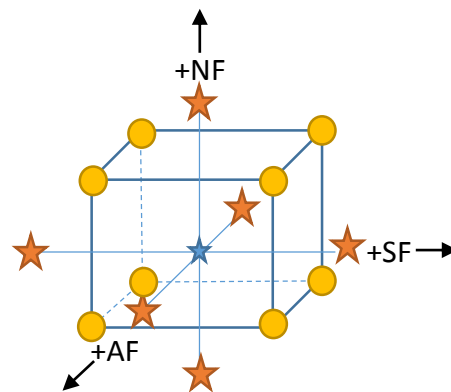


Figure 26: Graphic showing CCD Design

The hardware and experiment are designed to run a CCD experiment and to fit a quadratic model. The quadratic model works very well for single piece balances as it models physical attributes and interactions directly related to the balance's structural mechanics. However, if a balance is suspected of higher order interactions, the SVS can still be used to estimate them. A nested factorial experiment or an optimal design has been used in the past to allow limited cubic terms to be estimated. An optimal load schedule is optimized for each balance using an optimal

algorithm such as I-Optimal or D-Optimal. I-Optimal design algorithms will seek to reduce the total average prediction variance over a design space. D-Optimal designs seek to minimize the variance of the factor effect estimates. These designs are usually generated for about 75 points which provide acceptable statistical power and VIF's. A sample of the CCD can be found below in Figure 27 and Table 6. The first 10 points are a sample of the factorial block and the remaining points are axial points. The plot below appears much more random than the ABCS. An SVS experiment uses statistical insurance policies such as blocking and randomization to minimize the effect of any persistent variables that are unaccounted for.

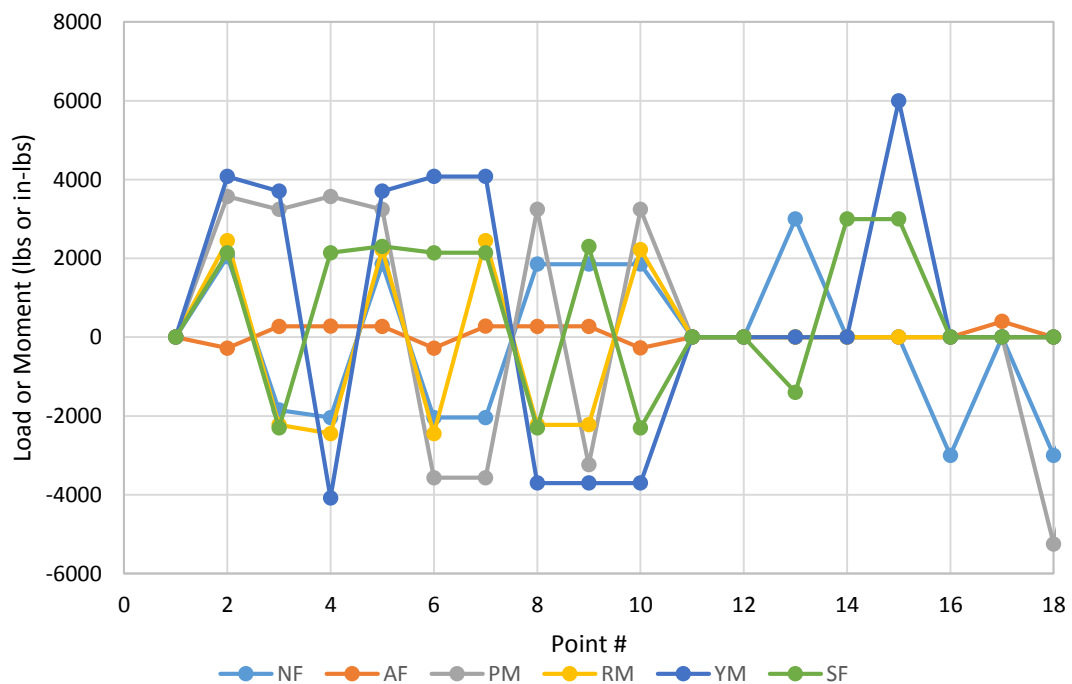


Figure 27: Excerpt from Standard Load Schedule for Single Vector System

POINT #	NF	AF	PM	RM	YM	SF
1	0	0	0	0	0	0
2	2039.955	-278.255	3569.922	2447.946	4079.91	2140.574
3	-1852.12	275.1532	3241.218	-2222.55	3704.249	-2305.39
4	-2039.96	278.2548	3569.922	-2447.95	-4079.91	2140.574
5	1852.125	275.1532	3241.218	2222.55	3704.249	2305.39
6	-2039.96	-278.255	-3569.92	-2447.95	4079.91	2140.574
7	-2039.96	278.2548	-3569.92	2447.946	4079.91	2140.574
8	1852.125	275.1532	3241.218	-2222.55	-3704.25	-2305.39
9	1852.125	275.1532	-3241.22	-2222.55	-3704.25	2305.39
10	1852.125	-275.153	3241.218	2222.55	-3704.25	-2305.39
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	3000	0	0	0	0	-1398.61
14	0	0	0	0	0	3000
15	0	0	0	0	6000	3000
16	-3000	0	0	0	0	0
17	0	400	0	0	0	0
18	-3000	0	-5250	0	0	0
19	0	0	0	0	0	0
20	-3000	0	0	0	0	1398.606
21	3000	0	5250	0	0	0
22	3000	0	0	-3600	0	0
23	-3000	0	0	3600	0	0
24	0	0	0	0	0	3000
25	0	0	0	0	0	0
26	-3000	0	0	0	0	-1398.61
27	0	0	0	0	0	-3000
28	3000	0	0	0	0	1398.606
29	3000	0	0	0	0	0
30	3000	0	-5250	0	0	0

Table 6: Excerpt from Standard Load Schedule for Single Vector System

2.3.3. Manual Stand Load Schedules

Manual Stand calibrations use an OFAT design. The design limits multi-component loads down to two components simultaneously. This is due to the complexity of setting up a six component load using cables. All six primary loads are excited individually and all two component combinations of load are excited. Cables are used to generate auxiliary loads, not in line with

gravity. Often multiple load levels are included, sometimes up to nine. Shown below in Figure 28 and Table 7 are a sample of 56 points from a traditional Manual Stand calibration. Secondary axes are used to allow the forces to be seen through the moment components. There are 82 loading sequences of between five and nine load steps [17]. For each sequence, loads are applied in equal increments up to full scale and then decremented to zero. An experiment with five load steps would have two positive increments, two equivalent negative increments and the zero point. Logically, a nine-step experiment will have four positive and four negative increments with a zero point. This calibration type was designed to fully determine both first and second order interactions. The load increments serve also to expose any hysteresis in the balance.

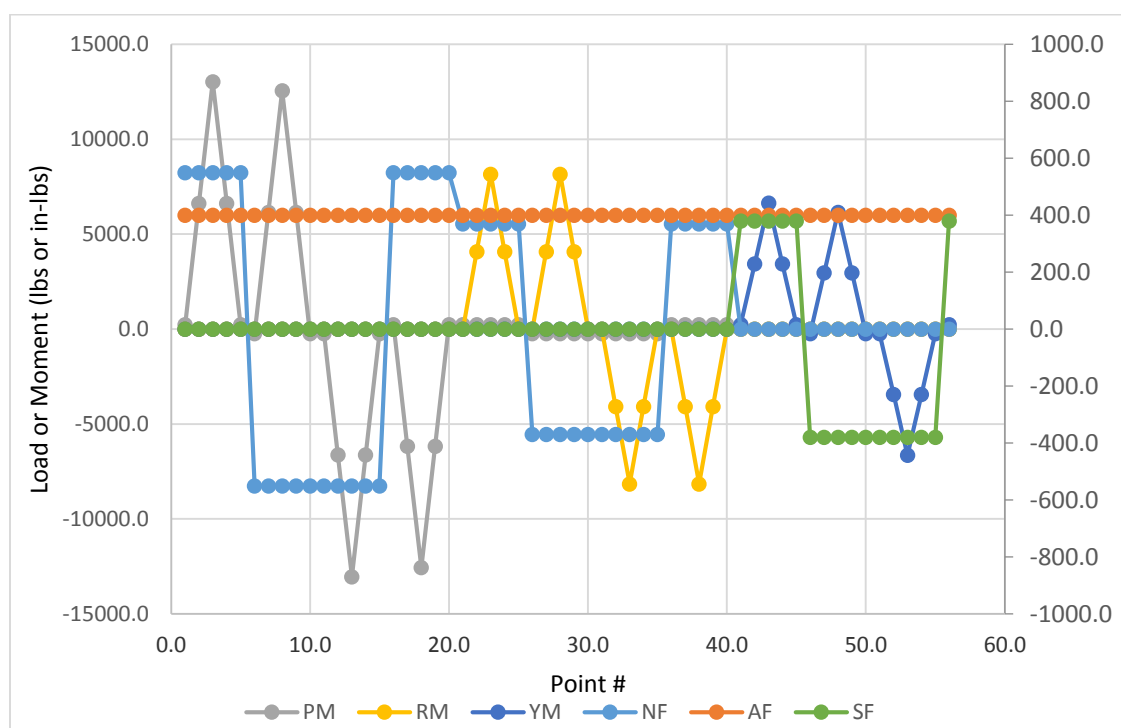


Figure 28: Excerpt from Standard Load Schedule for Manual Stand with 5 Load Levels

POINT #	NF	AF	PM	RM	YM	SF
1	550	400	230	0	0	0
2	550	400	6630	0	0	0
3	550	400	13030	0	0	0
4	550	400	6630	0	0	0
5	550	400	230	0	0	0
6	-550	400	-230	0	0	0
7	-550	400	6160	0	0	0
8	-550	400	12560	0	0	0
9	-550	400	6160	0	0	0
10	-550	400	-240	0	0	0
11	-550	400	-240	0	0	0
12	-550	400	-6630	0	0	0
13	-550	400	-13040	0	0	0
14	-550	400	-6630	0	0	0
15	-550	400	-240	0	0	0
16	550	400	240	0	0	0
17	550	400	-6160	0	0	0
18	550	400	-12560	0	0	0
19	550	400	-6160	0	0	0
20	550	400	230	0	0	0
21	370	400	230	0	0	0
22	370	400	230	4080	0	0
23	370	400	230	8150	0	0
24	370	400	230	4080	0	0
25	370	400	230	0	0	0
26	-370	400	-230	0	0	0
27	-370	400	-230	4080	0	0
28	-370	400	-230	8150	0	0
29	-370	400	-230	4080	0	0
30	-370	400	-230	0	0	0
31	-370	400	-230	0	0	0
32	-370	400	-230	-4080	0	0
33	-370	400	-230	-8150	0	0
34	-370	400	-230	-4080	0	0
35	-370	400	-230	0	0	0
36	370	400	230	0	0	0
37	370	400	230	-4080	0	0
38	370	400	230	-8150	0	0
39	370	400	230	-4080	0	0

Table 7: Excerpt from Standard Load Schedule for Manual Stand with 5 Load Levels

Table 7 (Continued)

POINT #	NF	AF	PM	RM	YM	SF
40	370	400	230	0	0	0
41	0	400	0	0	240	380
42	0	400	0	0	3440	380
43	0	400	0	0	6640	380
44	0	400	0	0	3440	380
45	0	400	0	0	240	380
46	0	400	0	0	-240	-380
47	0	400	0	0	2960	-380
48	0	400	0	0	6160	-380
49	0	400	0	0	2960	-380
50	0	400	0	0	-240	-380
51	0	400	0	0	-240	-380
52	0	400	0	0	-3440	-380
53	0	400	0	0	-6640	-380
54	0	400	0	0	-3440	-380
55	0	400	0	0	-240	-380
56	0	400	0	0	240	380

Table 7: Excerpt from Standard Load Schedule for Manual Stand with 5 Load Levels

For this study, a new more efficient manual calibration load schedule based on a modified Box-Behnken Design (M-BBD) was used. The design is capable of supporting a quadratic model and does not contain any fraction of a factorial experiment. Design points are edge centers rather than vertices and is conceptually illustrated in Figure 29. A modified version of the BBD design for wind-tunnel balance calibration was first presented in a 2005 conference paper titled, “*Calibrating Large Capacity Aerodynamic Force Balance Instrumentation Using Response Surface Methods*” [18], which proposed that the design was a more rigorous alternative to the traditional Manual Stand load schedule [18]. The design featured a full load schedule which consisted on only two-component loads. A small balance was calibrated using an M-BBD design in NASA Langley’s calibration labs [18]. This study uses a further modified version of the M-BBD detailed in Simpson et al. [18]. The primary concern during the development of this design

was to reduce run time and workload over the entire calibration process. NASA Langley already used an efficient design on the SVS but is limited to 3000 lbs of applied load. NASA was in need of a more efficient design for use on the Manual Stand which is capable of calibrating most balances to full scale. As proposed, the design effectively reduces the number of loads, set-up time, the total uncertainty in the calibration, turnaround time for researchers, and the total cost of the calibration. The design used has around 65 total points including centers and is modified for limitations of the Manual Stand and limiting multi-component loads to two factors. These modifications are designed to break correlations and even further reduce calibration time and workload over the calibration process. Multi-component loads on a Manual Stand are extremely complex and by using a BBD and cutting out multi-component loads, the calibration will run quickly and feature reduced set point errors that propagate from the added complexity. The design is also near-rotatable and offers some choices for orthogonal blocking. It is comparable to a CCD in that it has a similar number of runs, uses three-factor levels, and provides estimates of lack of fit and pure error. Where a CCD changes all the factors at once, the M-BBD only changes two. Table 8 shows assorted statistics for comparison of the traditional Manual Stand versus the M-BBD versus the CCD used with the SVS.

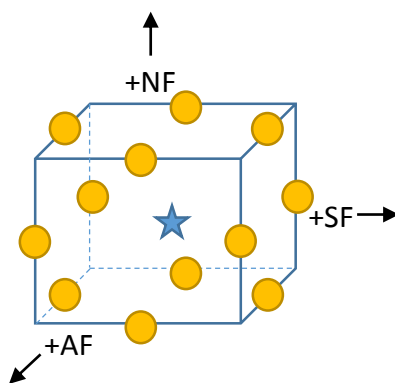


Figure 29: Box-Behnken Design

D-Optimality is a measure of the variance across the entire design. It seeks to minimize the determinant of the variance-covariance matrix, which is obtained from $(X_M^T X_M)^{-1}$. It can be seen in Table 8 that the result from the M-BBD is the lowest of all three designs. G-Efficiency is related to the maximum value of the standard error within the design and, in this case, is equivalent to the CCD. If you consider all of the factors in the table below, it out performs the traditional Manual Stand in every way. It is estimated that this design can be completed in 30 to 50 % less time versus the traditional manual design.

Criterion	Modified BBD	CCD-SVS	OFAT-729
Number of points	65	64	729
D-optimality	7.5	40	2 E+17
G-efficiency	70%	70%	16%

Table 8: Assorted Statistics for Design Comparisons

CHAPTER 3

MODELING AND UNDERSTANDING WIND-TUNNEL BALANCES

3.1. Wind Tunnel Balance Characterization

There are six responses from the balance, force in the normal (Z) direction, axial (X) direction, and side (Y) direction and moments about each of those axes; yawing, rolling, and pitching moment, respectively. A model is estimated for each of these responses using ordinary least squares regression with significance testing through analysis of variance (ANOVA). Only statistically significant terms are retained in the model. Statistical significance is determined by estimating the full-scale effect and comparing it to an estimate of the system noise at a given significance level. At NASA LaRC, the standard 95% confidence level is used. Typically if a regression model coefficient's significance is found to be less than 5%, it is statistically not differentiable from noise and removed from the model. Model estimation begins using the AIAA recommended math model seen in Equation (1). This model includes all components through the pure cubic terms as well as absolute value terms. At NASA LaRC, a second order model has been found to be sufficient to capture most, if not all, of the balances' primary and interaction behaviors. This model can be seen in Equation (2). Note that R_t is the electrical output of the balance, a_i is the intercept coefficient, $b(n)_{i,j,k}$ are the coefficients and $F_{j,k}$ are the applied forces.

$$\begin{aligned}
 R_t = & a_i + \sum_{j=1}^n b1_{i,j} F_j + \sum_{j=1}^n b2_{i,j} |F_j| + \sum_{j=1}^n b3_{i,j} F_j^2 + \sum_{j=1}^n b4_{i,j} F_j |F_j| + \sum_{j=1}^n \sum_{k=j+1}^n b5_{i,j,k} F_j F_k + \\
 & \sum_{j=1}^n \sum_{k=j+1}^n b6_{i,j,k} |F_j F_k| + \sum_{j=1}^n \sum_{k=j+1}^n b7_{i,j,k} F_j |F_k| + \sum_{j=1}^n \sum_{k=j+1}^n b8_{i,j,k} |F_j| F_k + \sum_{j=1}^n b9_{i,j} F_j^3 + \sum_{j=1}^n b10_{i,j} |F_j^3| \quad (1)
 \end{aligned}$$

$$R_t = a_i + \sum_{j=1}^n b1_{i,j} F_j + \sum_{j=1}^n b2_{i,j} F_j^2 + \sum_{j=1}^n \sum_{k=j+1}^n b3_{i,j,k} F_j F_k \quad (2)$$

3.2. Statistical Characterization of Data

The wind-tunnel balance is a critical part of any wind-tunnel model system. The interval between calibrations varies among users and is often driven by financial and/or program requirements. The variance associated with repeated calibrations has seen few studies and is not well understood. Applying rigorous statistical methods to a balance undergoing repeated calibrations will allow a more in-depth look into exactly how repeatable a balance can be across calibrations. To date, the long-term studies that have been undertaken for repeated wind-tunnel balance calibrations have suffered from small sample size. The expectation of this study is to conduct enough repeated calibrations to establish a baseline.

3.2.1. Design of Experiments and Response Surface Methodology

Design of experiments has origins in the agricultural field. R.A. Fisher first developed factorial experiments and ANOVA to perform experiments on large amounts of crop data [12]. Response Surface Methodology (RSM) has roots in industrial and chemical applications. Box and Wilson founded what is known as RSM with the goal of optimizing the response of a process [16]. If resources were unlimited, designing experiments would not be of large concern, however, this is not the case for any modern organization. DOE and RSM allow users to learn as much as possible about a process including characterization, optimization and error quantification. These tools are specifically designed to reduce time in designing or developing new products or processes, improve the performance, reliability and robustness of an existing process, evaluate design alternatives and define system tolerances [12, 16]. Overall, these methods provide a rigorous method to characterize any system and fine tune the response to a user's specific needs. Design of experiments and response surface methodology are used extensively in this study to facilitate rigorous experimental design, execution, and analysis of the calibration data.

3.2.1.1. *Application of Design of Experiments*

Following the outline from Montgomery [12], the recommended steps for designing, executing and analyzing an experiment are as follows.

1. Problem Statement
2. Selection of factors and response variables
3. Selection of factor levels
4. Selection of experimental design
5. Performing the experiment
6. Statistical analysis of the data
7. Conclusion

While this procedure is very useful, experience has shown that proper planning requires, at least, an iterative approach and at most a simultaneous approach to experiment planning. The first step in any experiment is to formulate a problem statement. This step may seem somewhat intuitive, however depending on the problem, this is not always straightforward. Ultimately, it is up to the experimenter to think ahead and determine what the expected outcome of a given experiment might be, what resources are available, and what is required in terms of analysis to achieve the goal defined in the problem statement. Selection of the factor levels comes directly from the operational space that the system is used in. For existing systems, these levels are often the maximum and minimum possible values for that factor. The selection of the experimental design is possibly the most important aspect of the characterization initiative. Without the proper experiment, the responses may be impossible to understand. The experimenter must first make a selection as to what type of model is to be fit. This will often narrow the options for what type of experiment is chosen. Common candidates are pure factorials, central composite designs and Box-

Behnken designs. Design tools such as Design Expert® can be used to help the process along. Performing the experiment comes next. This step is also very important and must be completed carefully. Any mistakes may cost time and money. A researcher should consider including statistical insurance policies such as randomization and blocking to increase the robustness of the design. Replicates and center points should be added to allow estimates of pure error. The analysis of a multi-factor experiment will usually require ANOVA which will be discussed at length in the next section. Using the statistics calculated in the ANOVA table, a conclusion as to each factor's statistical importance can be made.

3.2.1.2. Overview of Analysis of Variance (ANOVA) and Regression Modeling

Analysis of variance or ANOVA is an analysis technique most often used in calibration for testing the significance of individual model terms and allows for model reduction. A sample ANOVA table can be seen in Table 9. Each of the columns will be discussed in length throughout the next few sections. Criteria by which to evaluate this table will also be described.

In general, multiple linear regression is used to fit a model to a set of data, such as the one seen in Equation (2). The regression model is typically a polynomial of order equal to the number of factor levels minus one. Hence in this study, the minimum number of factor levels is three to create a regression model of order two or less. A model of order two is otherwise called a quadratic model.

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	P-VALUE PROB > F
MODEL	2.515E+007	11	2.287E+006	1.991E+007	< 0.0001
<i>A-NF</i>	2.482E+007	1	2.482E+007	2.162E+008	< 0.0001
<i>B-AF</i>	83.66	1	83.66	728.56	< 0.0001
<i>C-PM</i>	10.91	1	10.91	94.97	< 0.0001
<i>D-RM</i>	4.92	1	4.92	42.82	< 0.0001
<i>E-YM</i>	12.05	1	12.05	104.95	< 0.0001
<i>F-SF</i>	785.25	1	785.25	6838.64	< 0.0001
<i>BC</i>	0.99	1	0.99	8.66	0.0048
<i>DE</i>	1.32	1	1.32	11.50	0.0013
<i>DF</i>	69.94	1	69.94	609.12	< 0.0001
<i>EF</i>	3.49	1	3.49	30.37	< 0.0001
<i>F^2</i>	0.53	1	0.53	4.64	0.0358
RESIDUAL	6.09	53	0.11		
COR	2.515E+007	64			
TOTAL					

Table 9: Sample ANOVA Table

The regression model is developed by computing the least squares estimator (b) of regression coefficients (β) given the model matrix (X_M) and response matrix (y) as shown in Equation (3) [12].

$$b = (X_M^T X_M)^{-1} X_M^T y \quad (3)$$

For the second order response surface for six factors, the fitted model becomes:

$$\begin{aligned}
 \hat{y} = & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + b_6 x_6 \\
 & + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{14} x_1 x_4 + b_{15} x_1 x_5 + b_{16} x_1 x_6 \\
 & + b_{23} x_2 x_3 + b_{24} x_2 x_4 + b_{25} x_2 x_5 + b_{26} x_2 x_6 \\
 & + b_{34} x_3 x_4 + b_{35} x_3 x_5 + b_{36} x_3 x_6 \\
 & + b_{45} x_4 x_5 + b_{46} x_4 x_6 \\
 & + b_{56} x_5 x_6 \\
 & + b_1^2 x_1^2 + b_2^2 x_2^2 + b_3^2 x_3^2 + b_4^2 x_4^2 + b_5^2 x_5^2 + b_6^2 x_6^2
 \end{aligned} \quad (4)$$

Note that there are 27 terms in total, 6 first order coefficients, 15 interaction terms and 6 pure quadratic terms.

The significance of a regression is found by determining if there is a linear relationship between the response variable and any subset of the regressor variables. An ANOVA is generally used to determine if this linear relationship exists through statistical hypothesis testing. If such a relationship exists, then the fitted model can be considered significant. The hypothesis used to determine significance can be seen in Equation (5). If the null hypothesis (H_0) can be rejected, then at least one regressor contributes significantly to the model.

$$\begin{aligned} H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1: \beta_j \neq 0 \text{ for at least one } j \end{aligned} \tag{5}$$

The ANOVA partitions the total variability in all observations, which is computed in terms of a sum of squares (SS_T), into the variability associated with the regression model (SS_R) and the variability associated with the residual error (SS_E) as shown in Equation (6).

$$SS_T = SS_R + SS_E \tag{6}$$

Following the procedure detailed in Montgomery [12], the total corrected sum of squares (SS_T) is found by computing the overall variability in all observations as shown in Equation (7).

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 = y^T y - \frac{(\sum_{i=1}^n y_i)^2}{n} \tag{7}$$

Similarly, the error sum of squares (SS_E) measures the variability not accounted for by the model. This variability is can be further separated and understood as a combination of pure (experimental) error and regression modeling lack-of-fit. This quantity is estimated by computing total

disagreement between each observed response and the value estimated by the model as shown in Equation (8).

$$SS_E = y^T y - b^T X_M^T y \quad (8)$$

The variability associated with the individual factors or treatments can then be determined by subtracting the error sum of squares from the total sum of squares. It is desired that the variability associated with the individual factors represents the largest portion of the total variability within the observations. This would imply the model is representative of the system of interest and the error is small.

$$SS_R = SS_T - SS_E \quad (9)$$

The overall variance observed within each factor's level and the variance between factors can be estimated by computing mean square quantities. The mean square quantities represent a weighted average of the variances and can be used to employ a statistical significance test on each variance contribution. Equations (10) and (11) show the mean square estimates for the variance within factor levels and the variance between factors respectively.

$$MS_E = \frac{SS_E}{n - k - 1} \quad (10)$$

$$MS_R = \frac{SS_R}{k} \quad (11)$$

In the above equations, n is the total number of observations and k is the total number of regression factors included in the model. It should be noted that the error variance (σ^2) can be estimated from the mean square for error as follows.

$$\hat{\sigma}^2 = MS_E \quad (12)$$

The F_0 test statistic is used to accept or reject the null hypothesis and is calculated from the mean squares as follows.

$$F_0 = \frac{MS_R}{MS_E} \quad (13)$$

To reject H_0 , implying the model is significant, F_0 must be greater than a critical statistic called $F_{critical}$. The critical F value is based on a specified confidence level (i.e. for 95% confidence, $\alpha = 0.05$) and the degrees of freedom associated with the regression (k) and the error ($n - k - 1$). Thus, if the inequality expressed in Equation (14) is true, the model is considered significant.

$$F_0 > F_{\alpha, k, n-k-1} \quad (14)$$

The F test determines significance in a strict sense, however, no indication is given as to how significant any given model or term might be. It is possible to have terms that are only marginally significant and minimally influence the model while others are very significant and drastically influence the model. It is therefore useful to consider P-values in regression analysis. The P-value reports the value of the smallest level of significance for which the model or term becomes significant. Therefore, for 95% confidence, any value P-value smaller than $\alpha = 0.05$ would be considered significant within the model. For example, $P = 0.0001$ would be considered highly significant while $P = 0.049$ would be considered only marginally significant.

3.2.1.3. Lack of Fit and Pure Error

The total error term SS_E can be further decomposed provided that the experiment has some replicated design points. With replicated points, a formal test can be conducted for the lack of fit of the regression model and the pure experimental error can be quantified. Specifically, the total error SS_E is a combination of error due to lack of fit (SS_{LOF}) and pure error (SS_{PE}). Both lack-of-fit and pure error can be similarly quantified using a sum of squares. Lack of fit refers to how well the regression model fits the experimental observations and is calculated by determining the variability of all terms not included in the regression model. Thus, lack of fit is considered a model

dependent measure. Pure error refers to the errors present in repeated experimental measurements. Unlike lack of fit, pure error simply quantifies random error and is a model independent measure. As detailed in Myers et al. [16], lack of fit (SS_{LOF}) is computed by subtracting the sum of squares for pure error (SS_{PE}) from the sum of squares for total error (SS_E) as shown in Equation (15).

$$SS_{LOF} = SS_E - SS_{PE} \quad (15)$$

In general, a design with replication will have several observations (n_{rep}) at the same factor level (m). The sum of squares for pure error is calculated from the replicated observations as shown in Equation (16).

$$SS_{PE} = \sum_{i=1}^m \sum_{j=1}^{n_{rep}} (y_{ij} - \bar{y}_i)^2 \quad (16)$$

Similar to the model significance test, mean squares can be calculated for both lack of fit and pure error. Mean square calculations for lack of fit and pure error are shown in Equations (17) and (18).

$$MS_{LOF} = \frac{SS_{LOF}}{(m - p)} \quad (17)$$

$$MS_{PE} = \frac{SS_{PE}}{(n - m)} \quad (18)$$

Note that there are $m - p$ degrees of freedom associated with SS_{LOF} where p is the number of model parameters including the mean. Similarly, there are $n - m$ degrees of freedom for SS_{PE} . A test statistic for lack of fit can then be calculated as follows.

$$F_0 = \frac{MS_{LOF}}{MS_{PE}} \quad (19)$$

This test is a ratio of the variance due to model fit versus the variance due to random error. Lack of fit is significant if the test statistic is larger than the critical F value as shown in Equation (20). A significant lack of fit is not desired, however, in a low noise environment, like that found in this

study, pure error may become very small and lack of fit can become significant while the fitted model is acceptable. This effect can also be seen when a large number of replicated observations have been made. Thus, other fit statistics such as R^2 must be consulted before any definite conclusions can be made regarding goodness of fit.

$$F_0 > F_{\alpha, m-p, n-m} \quad (20)$$

3.2.1.4. *Model Adequacy*

It is always necessary to ensure that a fitted model adequately approximates the system of interest. Several statistics are available to quickly judge the model adequacy and should be considered for all regressions. Regression model fits are often judged using a family of summary statistics called the coefficient of multiple determination (R^2).

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad (21)$$

The R^2 family of statistics are all measurements of goodness of fit or how well the data fit the regression model. A value of one is a perfect fit in which all of the variability in the response is explained by the model. Zero indicates that the model explains none of the variability in the response. In general, R^2 can be increased by adding additional model terms, regardless of significance to the regression model. Thus, an additional statistic is available that considers model size and penalizes over fitting. The adjusted R^2 or R_{adj}^2 is shown in Equation (22). The value of R_{adj}^2 will often decrease when extra terms are included in a model. The hope is R_{adj}^2 will yield a better estimate of the model's ability to explain the variability using the fewest terms possible.

$$R_{adj}^2 = 1 - \frac{n-1}{n-p} (1 - R^2) \quad (22)$$

While the previous statistics help to quantify how well a given model fits the data, additional understanding can be gained by examining the model's ability to predict future

observations. It should be noted that large R^2 does not imply that the given model is capable of making good predictions of the response, but rather that it fits the data used to generate the model. A prediction error sum of squares (PRESS) can be used to examine each observation's influence on the developed model and incorporate this information into a quantifiable model adequacy statistic. The end result of the PRESS residual provides an estimate of how well the model predicts future observations. Computation of the PRESS statistic requires considering all possible subsets of $n - 1$ observations. Each reduced data set is then used to fit a model which results in n model's total with 1 less degree of freedom compared to the original model. Next, each new model is used to predict the withheld observation and the associated PRESS residual error (e). Finally, the sum of squares of the n PRESS residual errors is computed, yielding the PRESS statistic. This statistic can be efficiently calculated by first mapping the vector of observed values into a vector of fitted values using the so-called hat matrix H as shown in Equation (23). Next, the residual vector from the original fitted model is calculated in Equation (24).

$$H = X_M(X_M^T X_M)^{-1} X_M^T \quad (23)$$

$$e = y - X_M b \quad (24)$$

Finally, the PRESS residual is computed by weighting the ordinary residual (e) according to the diagonal elements of the hat matrix (h_{ii}) as shown in Equation (25).

$$PRESS = \sum_{i=1}^n \left(\frac{e_i}{1 - h_{ii}} \right)^2 \quad (25)$$

The PRESS statistic can be used to compute an approximate R^2 that gives some indication of the predictive capability of the regression model. This value is predicted R^2 or R_{pred}^2 .

$$R_{pred}^2 = 1 - \frac{PRESS}{SS_T} \quad (26)$$

Expected values for R_{pred}^2 will range between zero and one with one being the ideal case. An $R_{pred}^2 = 1$ would mean that the model is expected to explain 100% of the variability in predicting new observations. All three parts of the R^2 family, R^2 , R_{adj}^2 and R_{pred}^2 , should be considered when assessing model fit [16].

3.2.1.5. *Model Reduction*

In the previous section, the use of ANOVA was described for the purpose of determining if a regression model was statistically significant. When modeling a system, it is advantageous to reduce the model to include only statistically significant model terms. Model reduction reduces the degrees of freedom required to estimate the model, thereby adding to the degrees of freedom available for the estimation of error. Additionally, reduced models are far less cumbersome to employ. To reduce a model, it is necessary to determine the significance of each model term estimated. The significance of any given model term can be determined through the extra sum of squares method using a partial F-test. The extra sum of squares measure the change in the error sum of squares from the addition or exclusion of a regressor. The hypothesis to test in this case is as follows, where β_i refers to any single model term.

$$\begin{aligned} H_0: \beta_i &= 0 \\ H_1: \beta_i &\neq 0 \end{aligned} \tag{27}$$

Following the extra sum of squares method detailed in Myers et al. [16], the partial sum of squares for each model term is calculated by first computing the sum of squares for the original regression model (SS_R). Next, a reduced model is created which excludes the model term in question. Finally, the sum of squares for the reduced regression model ($SS_{Reduced}$) is calculated

and subtracted from the original regression sum of squares. For a model with two factors, this would be expressed mathematically as shown below, where β_1 represents the term of interest.

$$SS_R(\beta_1|\beta_0, \beta_2) = SS_R(\beta_1, \beta_2|\beta_0) - SS_R(\beta_2|\beta_0) \quad (28)$$

In the above relation, $SS_R(\beta_1, \beta_2|\beta_0)$ is the regression sum of squares for the original model and $SS_R(\beta_2|\beta_0)$ is the regression sum of square for the reduced model. The partial sum of squares for the β_1 term would then be given by $SS_R(\beta_1|\beta_0, \beta_2)$. Noting that the single term of interest has one degree of freedom, the mean square from the partial sum of squares for the model term of interest can be obtained as follows:

$$MS_{partial} = \frac{SS_{Reduced}}{1} \quad (29)$$

For the case of the first order example, this would be expressed mathematically as shown below.

$$MS_{partial} = \frac{SS_R(\beta_1|\beta_0, \beta_2)}{1} \quad (30)$$

The test statistic F_0 is then computed based upon the partial mean squares for the term of interest ($MS_{partial}$) and the mean square for error (MS_E) of the original regression model as shown in Equation (31).

$$F_0 = \frac{MS_{partial}}{MS_E} \quad (31)$$

F_0 is then compared to the critical F-value ($F_{\alpha, n_{terms}, n-p}$) where $\alpha = 1 - confidence$, n_{terms} is the number of terms being tested (there could be more than one), n is the total number of observations, and p is the number of model parameters including the mean. If F_0 is greater than the critical F-value, the null hypothesis H_0 can be rejected which implies that the model term in question (β_i) is significant and should be included in the regression model. As before, the P-value can be used to determine level of significance at which the factor becomes significant. The P-value is then compared to the level of significance as stated before.

Since most models will have more than one term, a backward elimination procedure is used to reduce the model. In this procedure, terms are tested for significance and eliminated one at a time starting with the highest order term and ending with the lowest order term if they are found insignificant. Once complete, the model is reduced to include only significant terms.

3.2.1.6. *Confidence Intervals and Prediction Intervals*

From the understanding that there is uncertainty associated with any estimated parameter, it is often useful to quantify this uncertainty with bounds referred to as a confidence interval (CI). In the case of a regression model, uncertainty exists within the individual regression coefficients, the mean response at any particular point and the prediction of future responses. As detailed in Myers et al. [16]., the $100(1 - \alpha)\%$ CI for an individual regression coefficient is shown below.

$$b_i - t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 C_{ii}} \leq \beta_i \leq b_i + t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 C_{ii}} \quad (32)$$

where β_i is the true regression coefficient, b_i is the estimated regression coefficient, C_{ii} is the diagonal element of the matrix $(X_M^T X_M)^{-1}$, α is the significance level, n is the number of observations, $t_{\frac{\alpha}{2}, n-p}$ is the t-statistic, $\hat{\sigma}^2$ is the error variance, and p is the number of model parameters including the mean.

Similarly, as defined by Myers et al. [16]., the $100(1 - \alpha)\%$ CI for the mean response is given as shown in Equation (33).

$$\begin{aligned} \hat{y}(x_0) - t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 x_0^T (X_M^T X_M)^{-1} x_0} &\leq \mu_{\hat{y}(x_0)} \\ &\leq \hat{y}(x_0) + t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 x_0^T (X_M^T X_M)^{-1} x_0} \end{aligned} \quad (33)$$

where $\mu_{\hat{y}(x_0)}$ is the true mean response, $\hat{y}(x_0)$ is the estimated mean response at the design point x_0 , X_M is the model matrix, α is the confidence level, n is the number of observations, $t_{\frac{\alpha}{2}, n-p}$ is the t-statistic, $\hat{\sigma}^2$ is the error variance estimated by the MS_E , and p is the number of model parameters including the mean.

Finally, from Myers et al. [16]., the $100(1 - \alpha)\%$ CI for the prediction of future observations is given as follows, commonly called a prediction interval.

$$\begin{aligned} \hat{y}(x_0) - t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 (1 + x_0^T (X_M^T X_M)^{-1} x_0)} &\leq y \\ &\leq \hat{y}(x_0) + t_{\frac{\alpha}{2}, n-p} \sqrt{\hat{\sigma}^2 (1 + x_0^T (X_M^T X_M)^{-1} x_0)} \end{aligned} \quad (34)$$

where y is the actual future response, $\hat{y}(x_0)$ is the predicted future response at the design point x_0 , X_M is the model matrix, α is the significance level, n is the number of observations, $t_{\frac{\alpha}{2}, n-p}$ is the t-statistic, $\hat{\sigma}^2$ is the error variance, and p is the number of model parameters including the mean.

3.2.1.7. Residual Analysis

Residual analysis is a process that can be done post experimentation. Residual diagnostics are used to ensure that the normality, independence, and constant variance assumptions used in the development of the ANOVA are valid for a given set of observations. The residuals (e) are computed by comparing the observed response (y) with the predicted response (\hat{y}).

$$e = y - \hat{y} \quad (35)$$

In general, each and every data point from an experiment would have a small residual that hopefully resulted from random sampling error. If all of the residuals are truly attributable to random errors, an experimenter would expect these residuals to be normally distributed (conform

to a Gaussian distribution). If some factor was unaccounted for, the residuals would be affected. They may appear not normally distributed, biased positive or negative, or abnormally large. In many instances, it is helpful to scale the residuals as studentized residual (r_i) as shown in Equation (36). This accounts for the variance of the residuals associated with the location in the design space, and scales the residual accordingly.

$$r_i = \frac{e_i}{\sqrt{\sigma^2(1 - h_{ii})}} \quad (36)$$

Figure 30 through Figure 32 show some examples of common forms of residual analysis taken from Design Expert®.

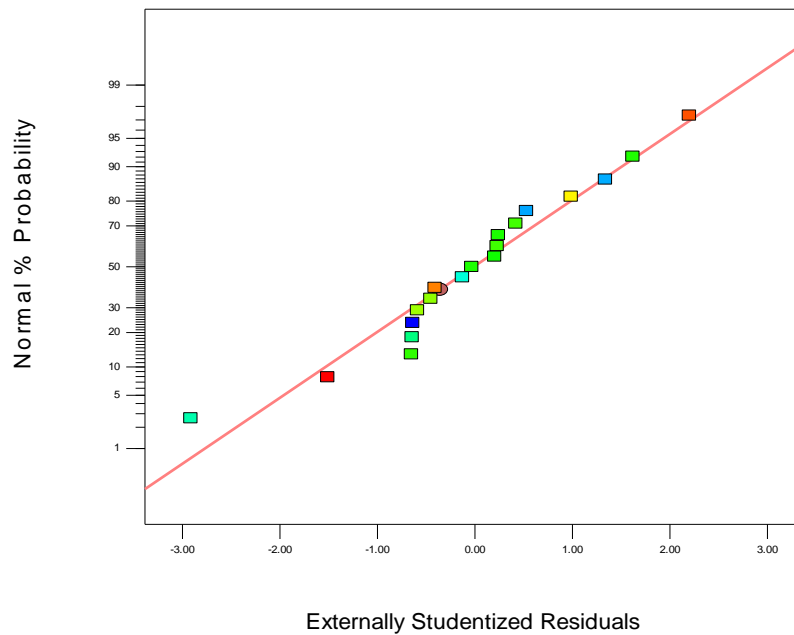


Figure 30: Normal Plot of Residuals

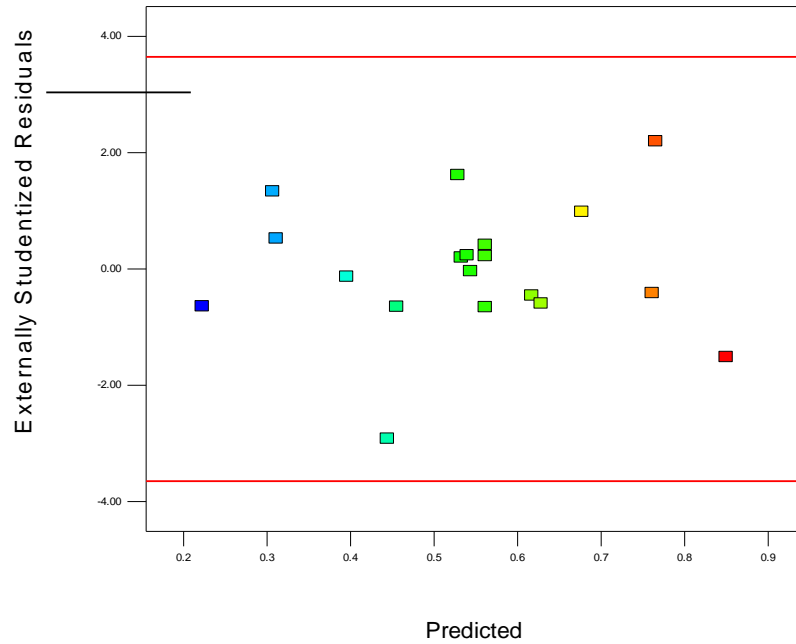


Figure 31: Residuals vs. Predicted

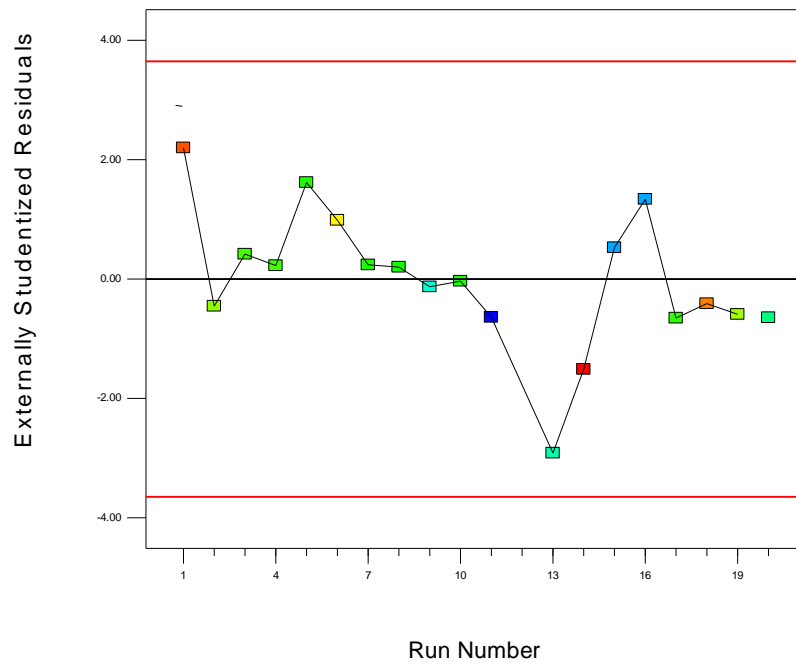


Figure 32: Residuals vs. Run

In these three plots, a user would be looking for normality, independence, and constant variance. Figure 30 plots the residuals over a transformed normal probability curve; if all of the points lie roughly on the line across the plot, the residuals are normally distributed. From the

example plot, it can be seen that the sample data is normally distributed. Figure 31 is a plot of residuals versus predicted. In this plot, a user is looking for constant variance. The purpose is to determine if the model predictions vary equally throughout the design space. In this case, the prediction variance is fairly constant. Features such as a cone shape or diamond shape would imply that the variance is not constant and the model's predictive power is better in some parts of the design space. Finally, Figure 32 plots residuals versus run. A user should look for independence in this plot. Are the residuals independent of run number (time), or does the variance increase or decrease over time? There are some trends that can be seen in the example residuals versus run plot. The data trend positive or negative for a few points before reversing and crossing the center axis. This trend may imply residual dependence on a factor that switches every couple of runs. However, the trends are not persistent and the data cross zero multiple times, so this observation would not raise concern for such a test. Note that the interpretation of these plots is subjective and a knowledgeable practitioner should make the final decision. If an experiment passes the test for normality, independence, and constant variance (homoscedasticity), then a user should be confident that the experiment sufficiently eliminated any biases and lurking variables. The errors seen are a product of sampling error.

3.2.2. Statistical Process Control

Statistical Process Control (SPC) is a method of quality control implemented on long-term sources of data. Typically used in the manufacturing industry, this method allows one to monitor, control and ultimately improve a process through the use of statistically rigorous methods. Fundamentally, data are placed on a chart and used to calculate control limits based on a measure of variance about a nominal value [7]. These limits, along with other criteria, are intended to alert a user to a process which is out of statistical control. In other words, the process is not operating

within expectations. One would look for an assignable cause for the lack of control and rectify it. The origin of control charts can be traced to Dr. Walter A. Shewhart, who developed these methods with the understanding that every process suffers from some variation. However, there is a difference between controlled variation and uncontrolled variation [7]. Controlled variation allows predictions of the future to be made while uncontrolled variation makes that impossible.

Control charts are usually intended to be used with large and continuous sources of data, where each point plotted on the control chart is the average value of multiple data points. These averages are taken over constant time intervals. This method of charting is called an \bar{X} chart. The \bar{X} chart is usually shown in conjunction with a range chart, or R chart. The R chart shows the difference between the maximum and minimum values in each grouped set of data plotted in the \bar{X} chart [7]. These two charts together give one an idea of both the trend in the mean as well as the variation within each set of data. For this study, most of the data taken from each balance calibration is a single data point and cannot be plotted using the standard \bar{X} -R method. Instead, individual points are plotted on a control chart called an X chart. Variation is tracked using a moving range or mR chart [7]. A moving range is simply the difference between a given point and the one preceding it in time. Figure 33 provides a sample of each type of chart with labels of each component. The \bar{X} -R and the X-mR charts are combined and labeled together as the chart style is essentially the same.

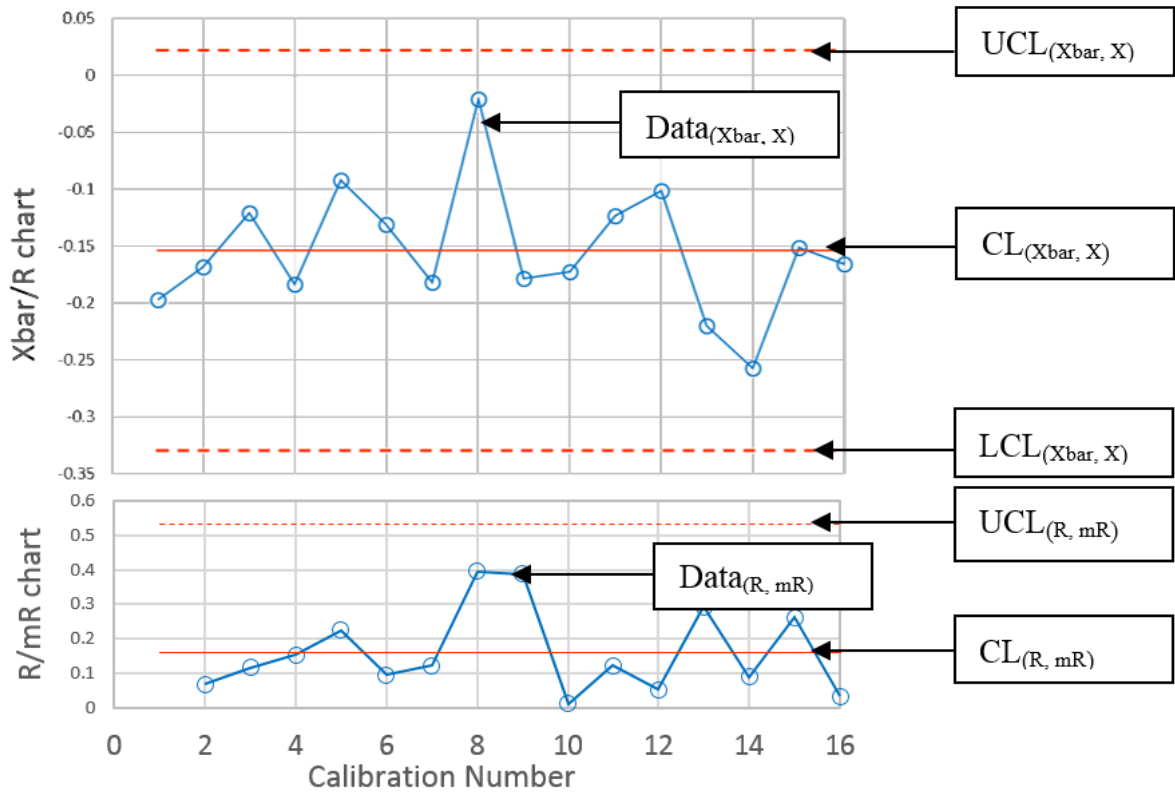


Figure 33: Example of Control Charts

Control limits for each of these chart types are calculated using equations derived by Shewhart [7]. These equations are shown below, numbered 37-40. The values for A_2 and D_4 are based on the number of observations in each subgroup and their values are pre-calculated from bias correction factors [7].

$$\begin{aligned}
 CL_{\bar{X}} &= \bar{\bar{X}} \\
 UCL_{\bar{X}} &= \bar{\bar{X}} + A_2 \bar{R} \\
 LCL_{\bar{X}} &= \bar{\bar{X}} - A_2 \bar{R}
 \end{aligned}
 \tag{37}$$

Equations for the \bar{X} Chart

$$\begin{aligned}
 CL_R &= \bar{R} \\
 UCL_R &= D_4 \bar{R}
 \end{aligned}
 \tag{38}$$

Equations for the R Chart

$$\begin{aligned}
 CL_X &= \bar{X} \\
 UCL_X &= \bar{X} + 2.660\overline{mR} \\
 LCL_X &= \bar{X} - 2.660\overline{mR}
 \end{aligned}
 \tag{39}$$

Equations for the R Chart

$$\begin{aligned}
 CL_{mR} &= \overline{mR} \\
 UCL_{mR} &= 3.268\overline{mR}
 \end{aligned}
 \tag{40}$$

Equations for the mR Chart

3.2.3. Developed Metrics for Calibration Data Comparison

Several parameters of interest have been chosen, each providing valuable insight into the long-term behavior of the balance as well as the calibration methods. Parameters of interest at this time are the aforementioned primary sensitivity coefficients from each calibration, a set of root mean square error (RMSE) estimates for both the model points and confirmation points and replicated normal force loads.

The primary sensitivities are a set of 6 values that represent the balance's primary response for each component. They are derived from the main effect first order coefficients of the fitted model. From the fitted coefficients, determining the primary sensitivity is shown below in Equation (41).

$$Sens_{comp} = \frac{1000}{b_{ME_{comp}}} \tag{41}$$

In the equation above, the main effect coefficient ($b_{ME_{comp}}$) is converted to a primary sensitivity ($Sens_{comp}$). Each primary sensitivity is in units of pounds per millivolt per volt (lbs/mV/V). As presented, the primary sensitivities are shown in change in percent full scale. This transformation is achieved using Equation (42).

$$Sens_{i,comp}(\Delta\%FS) = \frac{Sens_{i,comp}}{Sens_{comp}} * 100 - 100 \tag{42}$$

Each sensitivity for each of the six components in each of the calibrations ($Sens_{i,comp}$) is normalized by the average sensitivity for that component over all of the calibrations ($\overline{Sens_{comp}}$). By dividing by the average, the sensitivity is changed to a percent change. Multiplying by 100 and subtracting 100 converts the units to a true percentage and normalizes the percent change to zero. The change in sensitivities can be assessed between calibrations easily after this conversion.

A root mean square error (RMSE) can be calculated from the design points using ANOVA, and a similar RMSE can be estimated from the confirmation points, hereinafter referred to as $RMSE_M$ and $RMSE_C$, respectively. The RMSE is a frequently used measure of the average difference between an observed value and a predicted value. The metric is computed from the root sum square of every residual. In an ideal world, all of the residuals would be zero. This would imply that the model predicts the actual response perfectly every time. In the real world, random noise can be found in any system, producing observations that are oscillate above and below the actual measurand. This would then send the average of all of the residuals to zero and prove that the residuals are not biased positively or negatively. Seen another way, RMSE is looking at the collective magnitude of all of the residuals. In addition to randomly distributed residuals, small residuals are also desired. By looking at the magnitude of the residuals, the total error of the system can be assessed. All residuals are in mV/V. The equation can be seen below in Equation (43).

$$RMSE = \sqrt{\frac{1}{v} \sum (Y_i - \hat{Y}_i)^2} \quad (43)$$

Similarly, the RMSE values are normalized by the full-scale voltage output of each of the balance's bridges. Each residual is in units of percent of full-scale voltage output. The maximum voltages of each bridge are found by substituting the maximum loads into the fitted model.

$$RMSE (\%FS) = \frac{RMSE}{V_{max}} \quad (44)$$

Replicated normal force loads within each calibration are being tracked and evaluated for both within-calibration repeatability and across-calibration repeatability. Each calibration has a set of approximately four identical points. The balance is oriented in the same position every time and the same load is applied. The balance output is recorded and then compared to other replicates. This analysis combines the error from both the balance itself and the calibration process. The calibration process involves the accuracy of the hardware involved with the calibration process and the calibrating technician. For example, these replicates will lend insight into how well the calibration technician can return the balance to the same position every time. In this case, the average electrical zero for each calibration is subtracted from each normal force replicate. This will remove zero shift biases. The values are reported in raw voltage which is model independent and therefore, model error is left out of the normal force replicates. Without model error, the replicates will allow an estimate of the pure error for the balance. Pure error sets the lower bound for the accuracy of the balance.

Confidence intervals (CI) fit to the data for cross-system comparisons. They will form the statistical metric of comparison between two systems. The CI is a band about the estimated value in which the true value will lie. If any CI overlaps with another CI, then the two points cannot be statistically differentiated. This interval is defined by defining the standard deviation between two of the same measurements [16].

CHAPTER 4

ANALYSIS OF BALANCE CALIBRATION DATA

4.1. Introduction of Results and Initial Remarks

There are two major systems of interest in this research: the wind-tunnel balances and the calibration systems. When calibrating a balance on a system, it is difficult to separate the variability introduced by the balance versus the variability introduced by the calibration system. This is because the errors from each are confounded with one another. Uncertainty analyses have been performed on calibration systems and balances individually, which are of interest to this research. However, in general, the balance and the calibration system must be considered as a single system together. The results and analysis of this research will be presented in two broad categories: a comparison of calibration methods across a common balance and long term repeatability of a balance on a single calibration system. This will not fully separate the effects of a balance on a calibration system or vice versa, however, more in-depth inferences can be made.

From this analysis, the qualitative differences between balances and systems will be discussed at length for the user's benefit. Finally, metrics will be proposed that could give a user tools to judge the quality of a calibration. To date, at least some data exist for all of the balances on multiple systems. This analysis leverages both data taken during the balance calibration study and historical data from NASA's and Triumph Force Measurement System's archives.

Table 10 has tabulated values from a sample calibration. This was a single-piece balance which was calibrated on the SVS. Note the highly significant factor effects, insignificant lack-of-fit, high R^2 statistics and low VIF's. These results are typical of an ultra-low noise environment such as the balance calibration laboratory at NASA LaRC.

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	Factor	Coefficient Estimate	Standard Error	VIF
Model	25198857	12	2099905	12413956	2E-160	Intercept	0.348646	0.09128	
A-NF	25147949	1	25147949	1.49E+08	2E-166	A-NF	947.4401	0.07770	1.002
B-AF	67.95342	1	67.95342	401.7186	8E-26	B-AF	-1.97519	0.09855	1.067
C-PM	12.78901	1	12.78901	75.60449	1E-11	C-PM	0.86276	0.09922	1.002
D-RM	9.752537	1	9.752537	57.65383	6E-10	D-RM	-0.75711	0.09971	1.010
E-YM	12.62989	1	12.62989	74.66379	1E-11	E-YM	-0.84877	0.09823	1.013
F-SF	856.8156	1	856.8156	5065.216	1E-52	F-SF	5.813212	0.08168	1.011
CE	1.350774	1	1.350774	7.98534	7E-03	CE	-0.48961	0.17326	1.017
DE	2.409604	1	2.409604	14.24479	4E-04	DE	-0.6504	0.17233	1.002
DF	43.80161	1	43.80161	258.9409	1E-21	DF	-2.40792	0.14964	1.001
EF	3.6763	1	3.6763	21.7331	2E-05	EF	0.556828	0.11944	1.016
C^2	0.954089	1	0.954089	5.640266	2E-02	C^2	0.444012	0.18696	1.016
F^2	0.894462	1	0.894462	5.287769	3E-02	F^2	0.357062	0.15528	1.017
Residual	8.626996	51	0.169157			Std. Dev.	0.411287	R-Sq	1.000
Lack of Fit	7.51492	43	0.174766	1.257221	4E-01	Mean	-6.28643	Adj R-Sq	1.000
Pure Error	1.112075	8	0.139009			C.V. %	6.542455	Pred R-Sq	1.000
Cor Total	25198866	63				PRESS	13.47262	Adeq Prec	10407.4

Table 10: Tabulated Statistics from a Common Balance Calibration (Single-Piece Balance – SVS)

4.2. Comparison of Calibration Methods across a Common Balance

Multiple types of balance calibration systems exist, each using a different methodology from the last. The question remains, do the systems produce the same results? In actuality, each of these systems is estimating the behavior of the balances with some degree of inaccuracy. Each system employs a methodology to reduce the errors in the estimation process and understand the unexplained error that remains after the model is estimated. In general, it is understood that a result from any one of these calibration systems is wrong in different ways. The true answer is unknown to the user as regression analysis is at its very core, an estimation process using the least squares method. The only way one has to judge each system is to compare it to the other systems and attempt to understand the differences. If the differences are small, more confidence in the answers will be justified.

Balance accuracy is quoted as being at best within one-twentieth of one percent of full scale [2]. The value represents an informed estimate of the uncertainty in the balance, calibration system, and calibration process together. Specifically, the value represents the compounded error from the calibrated weights, load actuators, measurements of moment arms, the resolution of the data acquisition system, temperature and humidity variations in the calibration lab, and any other uncertainties involved. The primary sensitivities are thought to be fundamental to the balance electrical and structural mechanics. These values are first order main effects which define the response of the balance. While they are estimated by a model, it is thought that these values should not shift from calibration to calibration. This means that, a change in primary sensitivity much larger than 0.05% would be perceived as a significant difference and would generally warrant investigation. For these systems to be considered similar, each of the systems should be capable of determining the sensitivities.

Figure 34 through Figure 57 will show the primary sensitivities of each balance as calculated from a calibration on each system. Not all balances have been calibrated on each system currently, however, all comparisons that can be made are included. Much of these data are historical and were performed prior to the balance calibration study. The data for some of the historical calibrations are not readily available for analysis and therefore only calculated coefficients are shown in this section. The primary sensitivities are shown in change in percent of full scale units. The average coefficient was taken for all the calibrations over all the systems. This value is divided out of every individual calibration which puts the values in percent full scale of the estimated average, see Equation (42). Figure 34 through Figure 39 show the cross-system comparisons for the NTF-113C. Figure 40 through Figure 45 show the cross-system comparisons

for the MC-60E. Figure 46 through Figure 51 show the cross-system comparisons for the MK-29B and Figure 52 through Figure 57 show the cross-system comparisons on the NTF-118B.

4.2.1. NTF-113C Cross-System Comparisons

The NTF-113C has been calibrated a total of 20 times on the SVS, five times on the Manual Stand and a single time on the ABCS. Confidence intervals (CI) are fit to each coefficient estimate based on a significance level of 5%. If the CIs overlap, the values are statistically not differentiable from one another. The ABCS value does not have a CI because there is only one calibration. An estimate of standard deviation is therefore, not possible. Each coefficient has been normalized into change in percent full scale units. This normalization will allow for easy comparison of each coefficient. Informal limits at 0.05% are included which are based on the average coefficient estimated from all three systems. Note that, the average normalized coefficient (zero) is biased toward the SVS because of the large number of data points. This places the SVS coefficients within the 0.05% limits more often. Be aware that the average and limits are not intended to be an estimate of the “true coefficient”, but instead a way to compare the systems on a percent scale. The plots do not imply the SVS is correct because the coefficients sit within the limits. It does imply that the calibration systems produce results that are more or less than 0.05% different based on location of a data point.

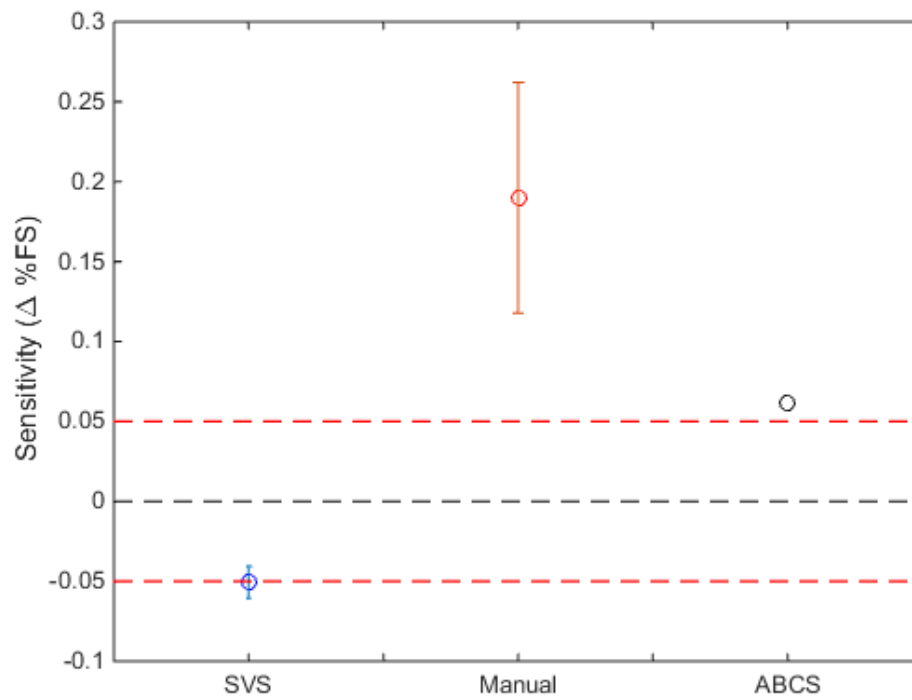


Figure 34: Cross-System Comparisons, NF Primary Sensitivity (NTF-113C)

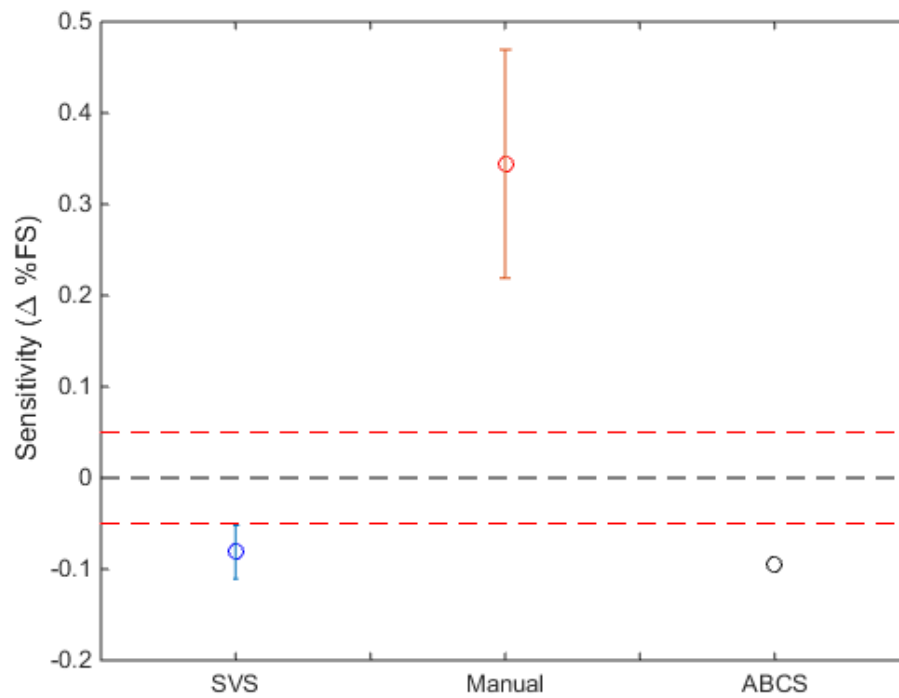


Figure 35: Cross-System Comparisons, AF Primary Sensitivity (NTF-113C)

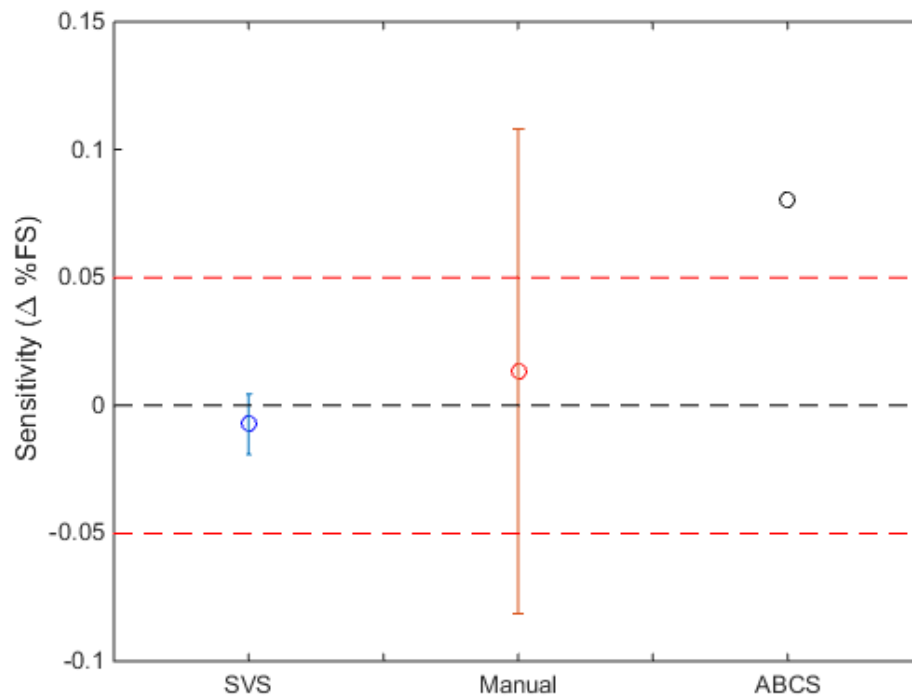


Figure 36: Cross-System Comparisons, PM Primary Sensitivity (NTF-113C)

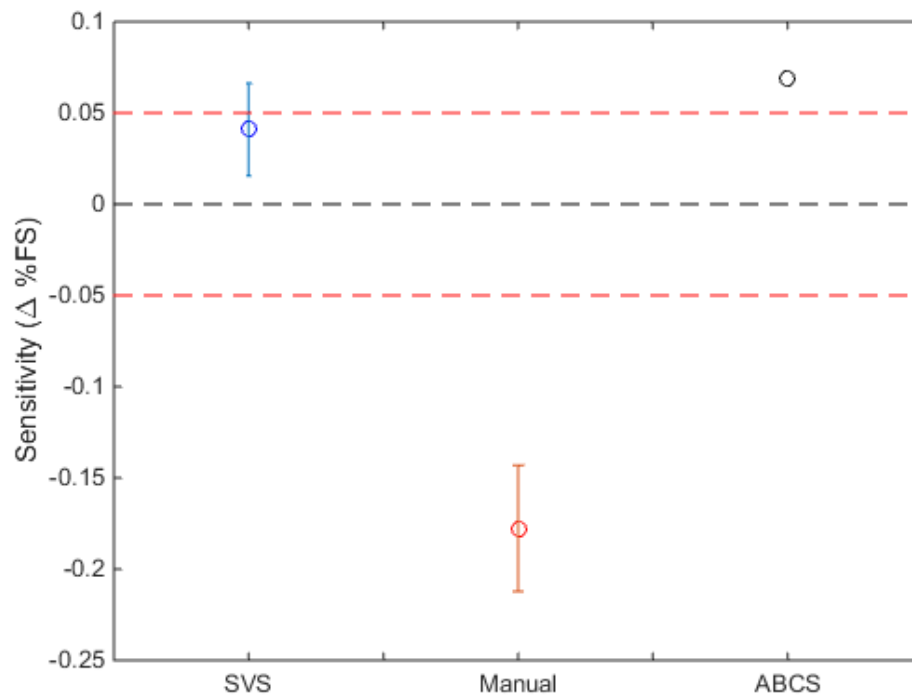


Figure 37: Cross-System Comparisons, RM Primary Sensitivity (NTF-113C)

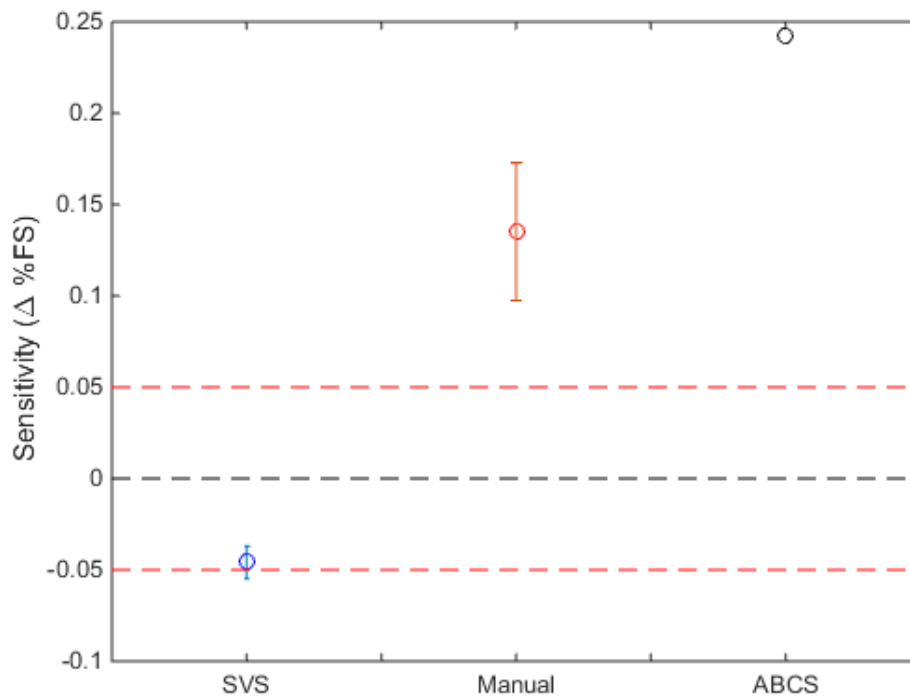


Figure 38: Cross-System Comparisons, YM Primary Sensitivity (NTF-113C)

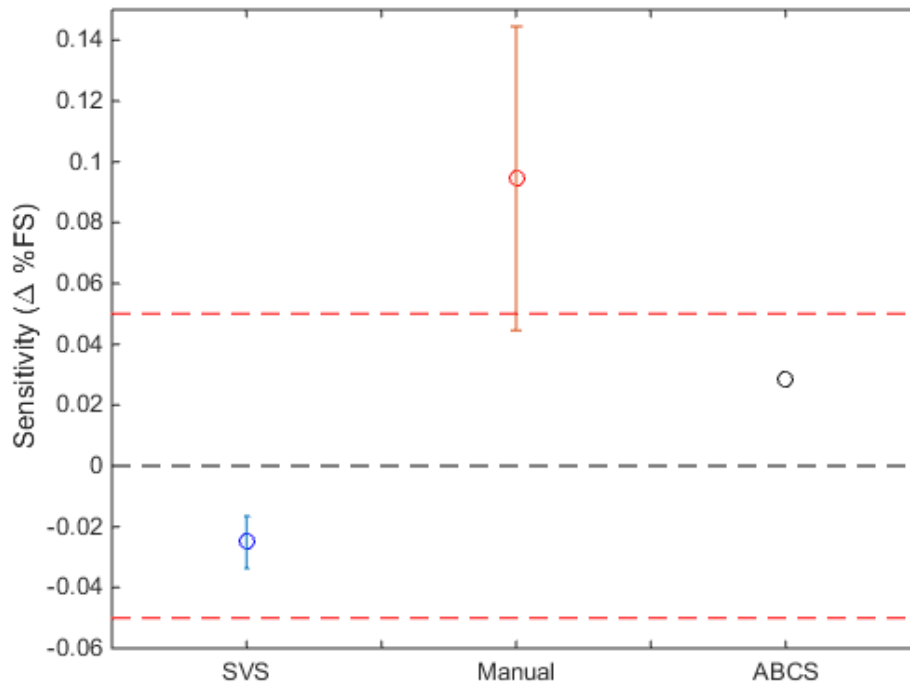


Figure 39: Cross-System Comparisons, SF Primary Sensitivity (NTF-113C)

From closer inspection of the plots, none of the CIs overlap, except pitching moment. This further implies that each of these estimates of coefficients are statistically different than one another. Had they overlapped, the systems would be producing the same results with 95% confidence. In some cases, a CI overlaps the limits but not another coefficient. This implies that, that coefficient may have been replicated within the expected value of 0.05%, but is still different from the other estimates of the coefficient. Looking closer at the normal force plot, the SVS value lies within the 0.05% limits and any reasonable CI would place the ABCS within the limits and possibly overlapping the SVS CI. The axial force plot shows that the SVS and ABCS have CIs that overlap. The pitching moment plot shows that the SVS and the Manual Stand agree and that the Manual Stand and the ABCS agree. Additionally, a reasonable CI on the ABCS value may place it within the 0.05% limits. Rolling moment shows agreement between the SVS and ABCS again. The yawing moment plot shows no agreement between any of the coefficients. Finally, the side force plot shows the SVS and ABCS are within the error limits and that the Manual Stand CI overlaps the 0.05% limits. A reasonable CI on the ABCS may overlap the SVS. It would seem that in many cases, the ABCS and the SVS have similar estimates. The Manual Stand appears to be different from the other estimates by at least 0.1%.

4.2.2. MC-60E Cross-System Comparisons

The comparisons across calibration systems for the MC-60E can be seen in Figure 40 through Figure 45. This balance has been calibrated twelve times on the SVS. The SVS calibrations have been separated into the ten more recent calibrations and two historical calibrations. CIs have been plotted for these estimates as well, except for the ABCS, which again only has one calibration. The CIs are somewhat hard to see when compared to the scatter of the data.

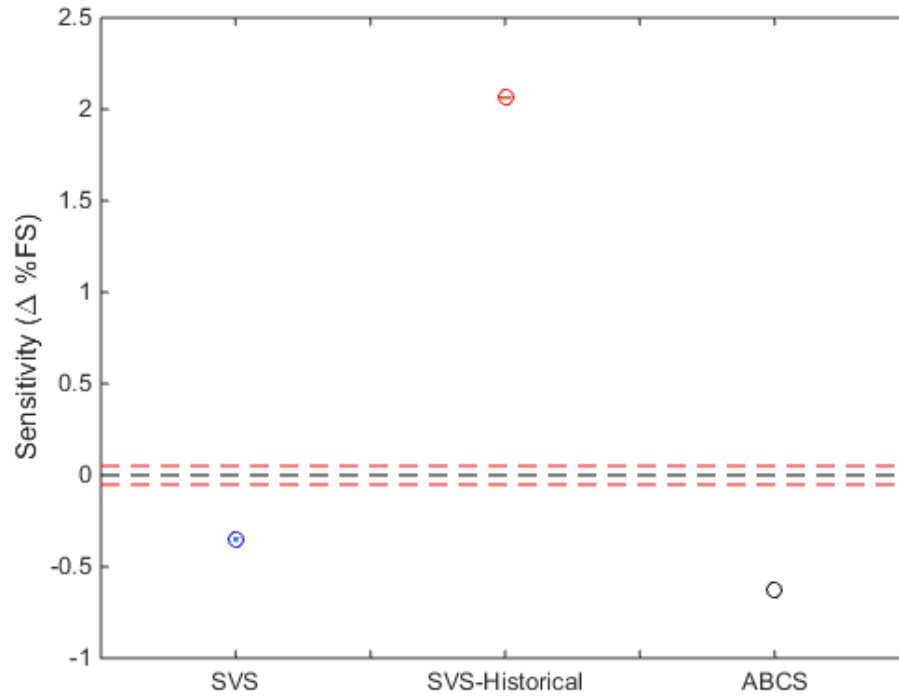


Figure 40: Cross-System Comparisons, NF Primary Sensitivity (MC-60E)

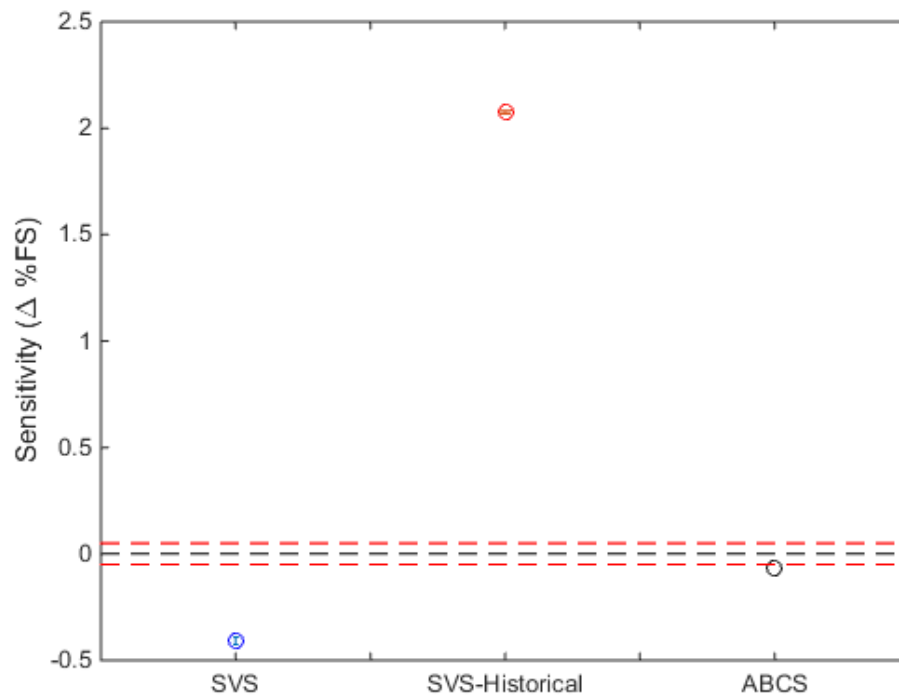


Figure 41: Cross-System Comparisons, AF Primary Sensitivity (MC-60E)

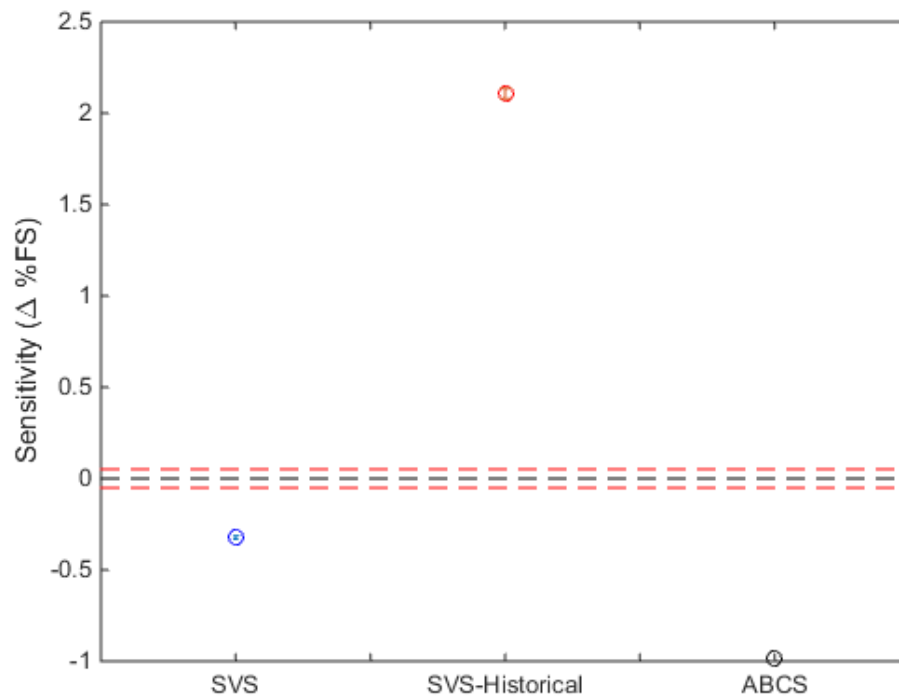


Figure 42: Cross-System Comparisons, PM Primary Sensitivity (MC-60E)

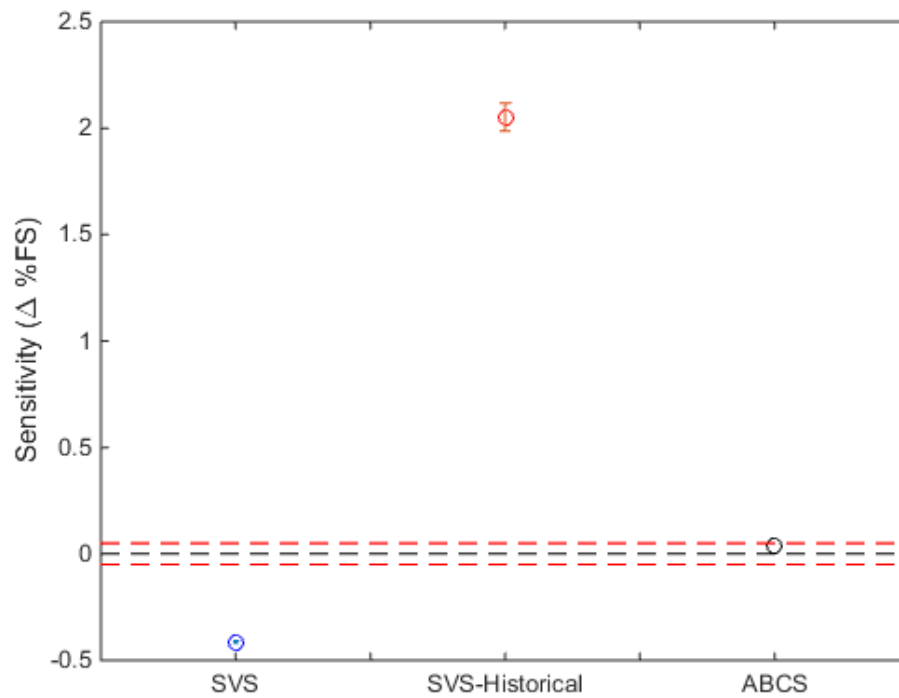


Figure 43: Cross-System Comparisons, RM Primary Sensitivity (MC-60E)

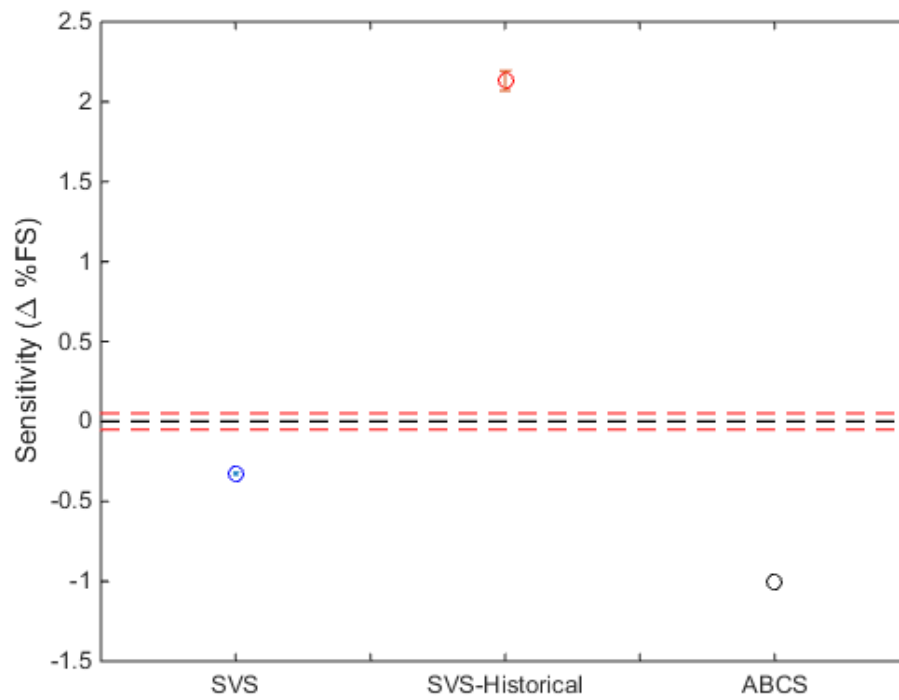


Figure 44: Cross-System Comparisons, YM Primary Sensitivity (MC-60E)

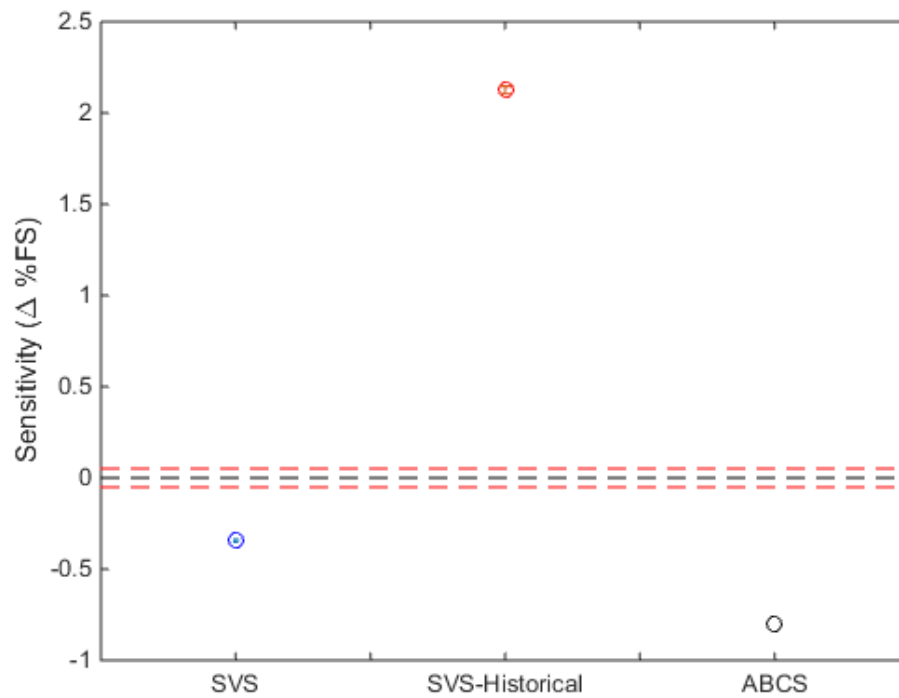


Figure 45: Cross-System Comparisons, SF Primary Sensitivity (MC-60E)

It is very apparent that the calibration systems produced very different results for this balance. What is most disconcerting is the large difference between the historical SVS calibration and the more current SVS calibrations. The historical calibrations were performed by the previous calibration study which was undertaken at NASA LaRC in 2006. A standard SVS CCD was performed on the MC-60E for the historical calibrations. The ten current calibrations, performed by the current calibration study, were performed using an I-optimal calibration design which was instead devised because of the unfamiliarity with the unitized balances. It was decided that an optimal design would allow for higher order terms to be estimated if necessary, while keeping the design run number efficient. The difference in load schedule design could be a reason for the differences seen between the calibrations. However, ideally a sensitivity should be estimable with any adequately designed and performed experiment. Additionally, the more recent calibrations on the SVS agree more readily with the ABCS calibration. The ABCS calibration was performed by the company that designed the balance, and that lends confidence in the data from the ABCS. Triumph should know if the sensitivities didn't agree with the way the balance was designed. In summary, the comparisons for the MC-60E are not perfect, however they are favorable between the current SVS calibrations and the ABCS calibration. The historical SVS calibration is significantly different from any other estimates of sensitivities.

4.2.3. MK-29B Cross-System Comparisons

The MK-29B has long been suspected of having mechanical issues. The drawbacks of a multi-piece design have been discussed in previous sections. This consideration should be kept in mind for the MK-29B. The MK-29B has undergone three SVS calibrations, two traditional Manual Stand calibrations and one M-BBD manual calibration. Figure 46 through Figure 51 show the cross-system comparisons for the MK-29B.

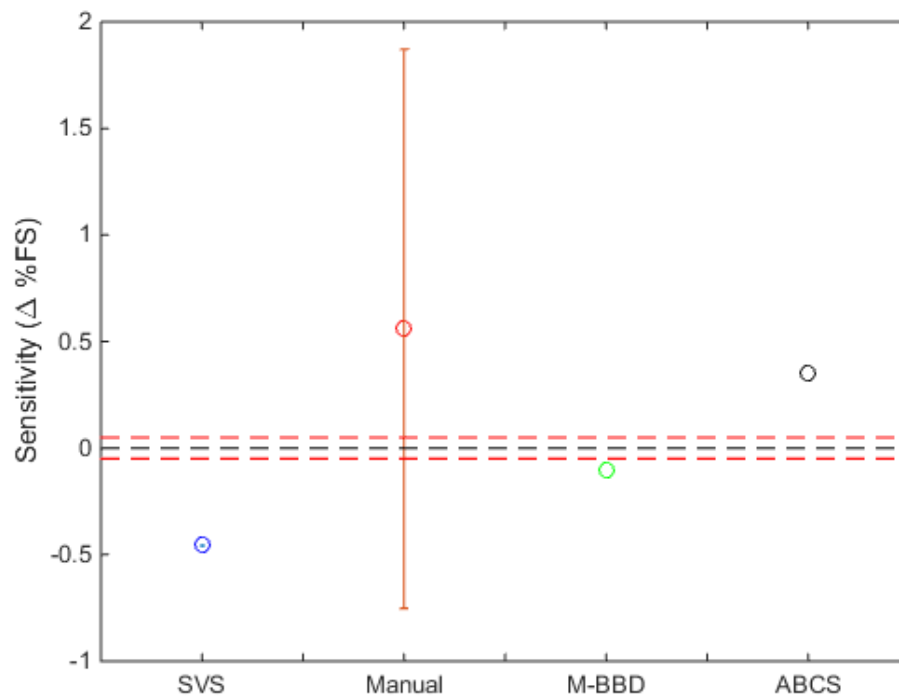


Figure 46: Cross-System Comparisons, NF Primary Sensitivity (MK-29B)

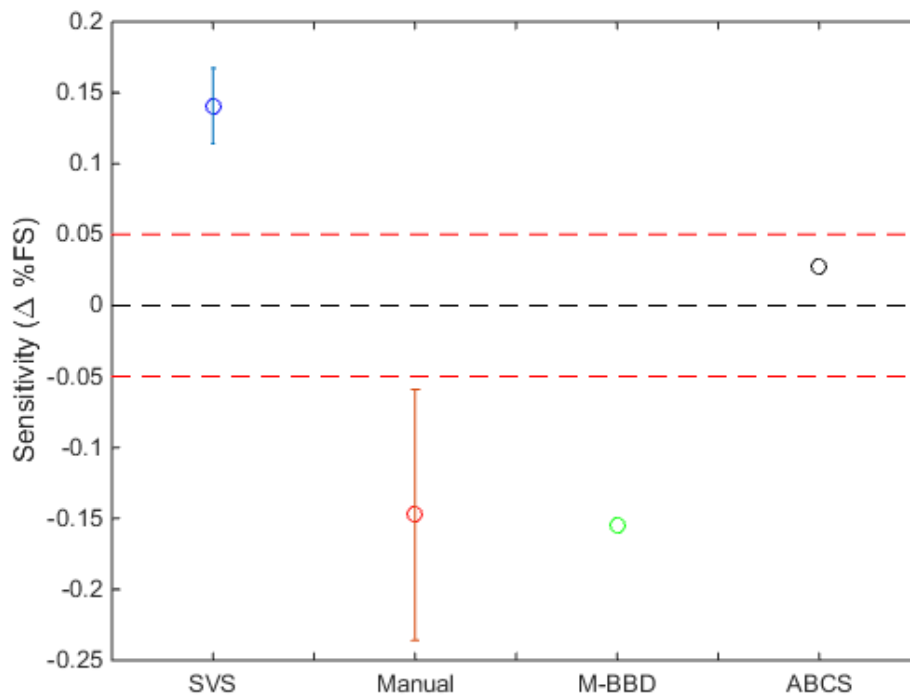


Figure 47: Cross-System Comparisons, AF Primary Sensitivity (MK-29B)

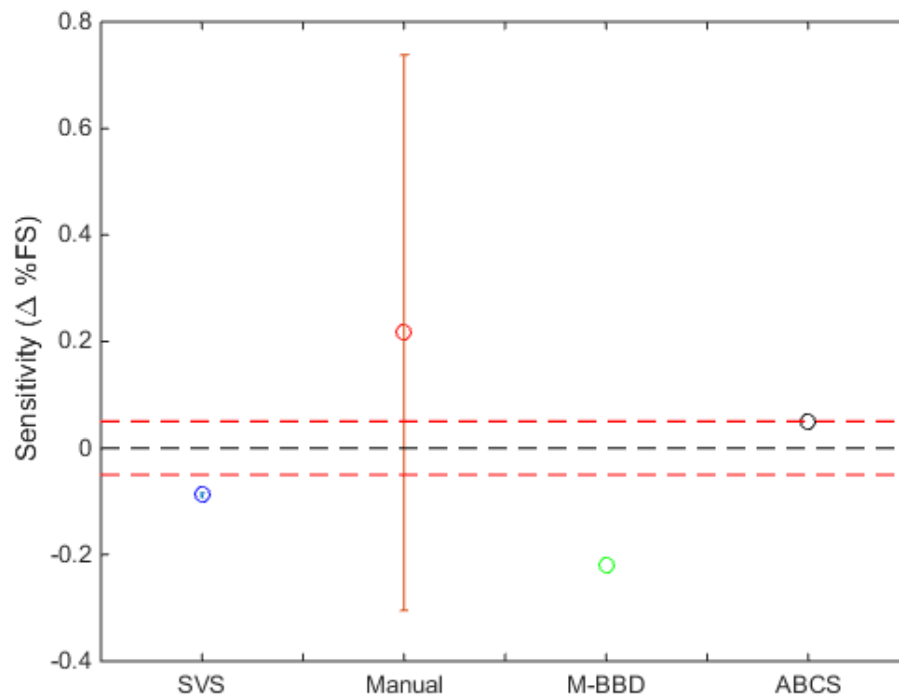


Figure 48: Cross-System Comparisons, PM Primary Sensitivity (MK-29B)

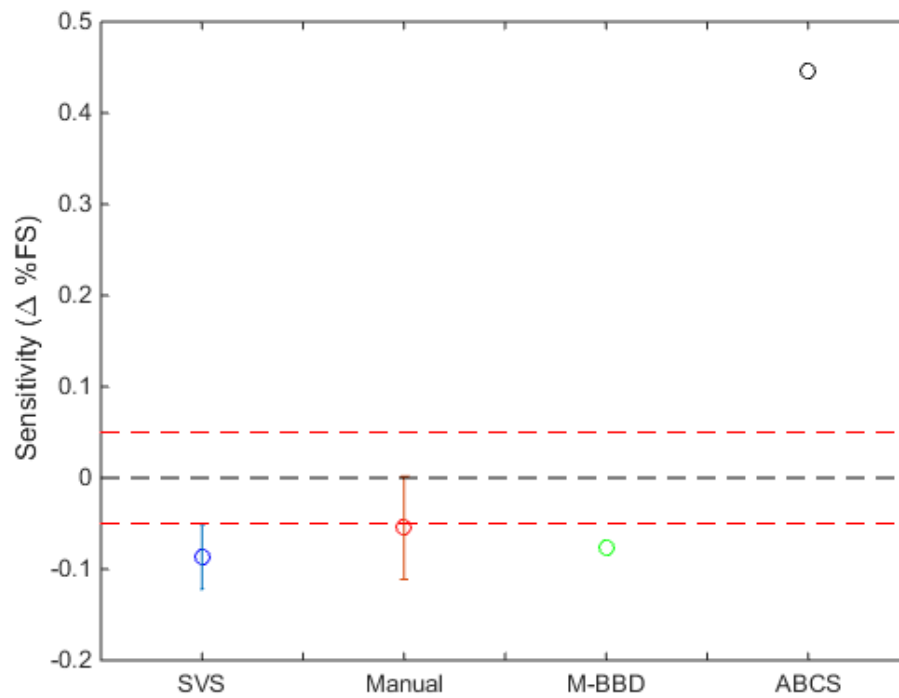


Figure 49: Cross-System Comparisons, RM Primary Sensitivity (MK-29B)

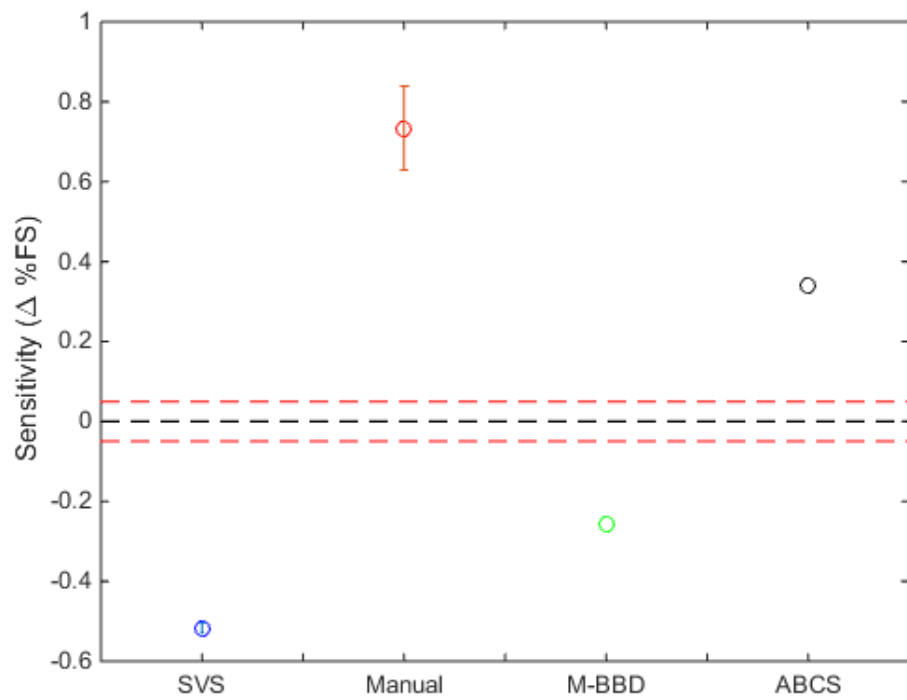


Figure 50: Cross-System Comparisons, YF Primary Sensitivity (MK-29B)

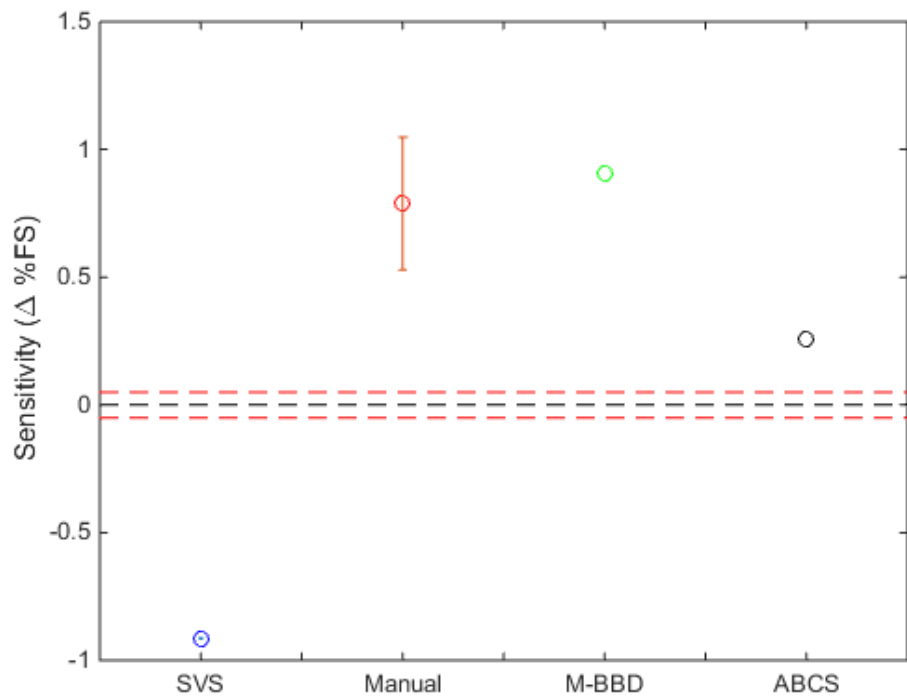


Figure 51: Cross-System Comparisons, SF Primary Sensitivity (MK-29B)

From interrogation of the plots, the balance replicated very well on the SVS. The CI for that data is extremely small. Three calibrations were performed which all agree very well with one another. Two traditional manual calibrations were performed as well. However, the CIs are much larger on that system than compared to the SVS. Additionally, a single manual M-BBD calibration and a single ABCS calibration was performed on the MK-29B. These points do not have CIs for lack of replicate calibrations. The MK-29B and the NTF-118B are the only balances to have undergone the M-BBD calibration. The M-BBD is performed on the Manual Stand, however the design is more comparable to the CCD for the SVS. Comparisons for a full-scale calibration between a traditional Manual Stand, SVS and M-BBD calibration have never been shown. From looking at the normal force plot, the Manual Stand has a large confidence interval. The variability between both the calibrations is much larger than on the SVS. It also means that the Manual Stand potentially agrees with all of the other calibration's coefficients shown in the plot. Variation between the systems can be seen at levels up to $\pm 0.5\%$. From the axial force plot, variability can be seen which remains within $\pm 0.2\%$. Additionally, the Manual Stand calibration agrees well with the M-BBD for axial. The pitching moment coefficients show variation around 0.2% as well, and much like the normal force plot, the Manual Stand coefficient has a large CI which covers all of the other coefficients. Rolling moment shows the best agreement of all. The SVS, Manual Stand and M-BBD design all agree with one another almost exactly. However, the ABCS produced a primary sensitivity about 0.5% larger. YM shows the worst agreement as none of the coefficients with CIs agree. Total variability on the plot exceeds 1.0% . Side force shows agreement between the Manual and M-BBD, however differences between systems range up to $\pm 1.0\%$.

To conclude with the MK-29B, the occasional agreement between these systems on this balance is not very convincing. Variation greater than $\pm 0.2\%$ is common between the systems.

Even comparable load schedules do not generate agreement on a wide scale. Unfortunately, it is difficult to determine if this is because of the balance or the systems, however, cross-system agreement for the NTF-113C and MC-60E proved much better (without the historical SVS calibration) than this balance has shown. Additionally, the comparisons between systems on the NTF-118B also appear much better.

4.2.4. NTF-118B Cross-System Comparisons

The NTF-118B has the fewest calibrations of all other balances in this study. It has been calibrated twice on the SVS, twice with a traditional Manual Stand and once using a M-BBD design. Figure 52 through Figure 57 show the cross-system comparisons for all calibrations on the NTF-118B.

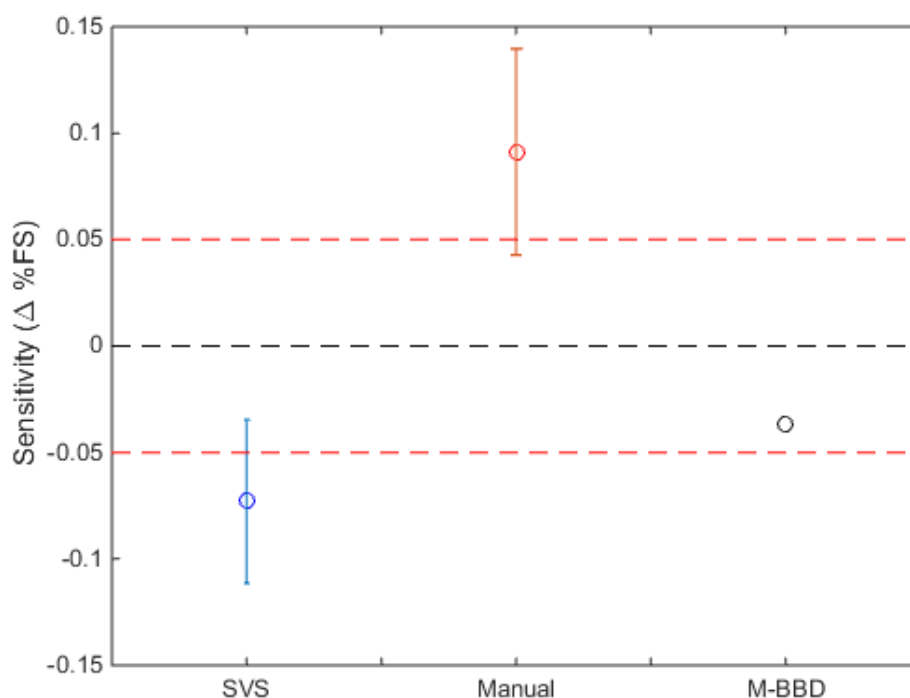


Figure 52: Cross-System Comparisons, NF Primary Sensitivity (NTF-118B)

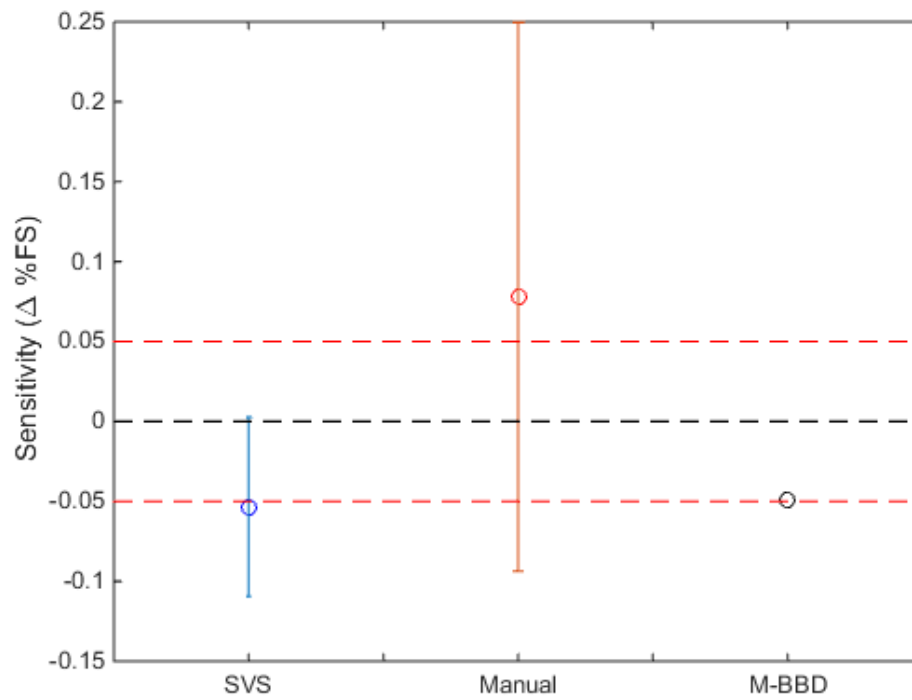


Figure 53: Cross-System Comparisons, AF Primary Sensitivity (NTF-118B)

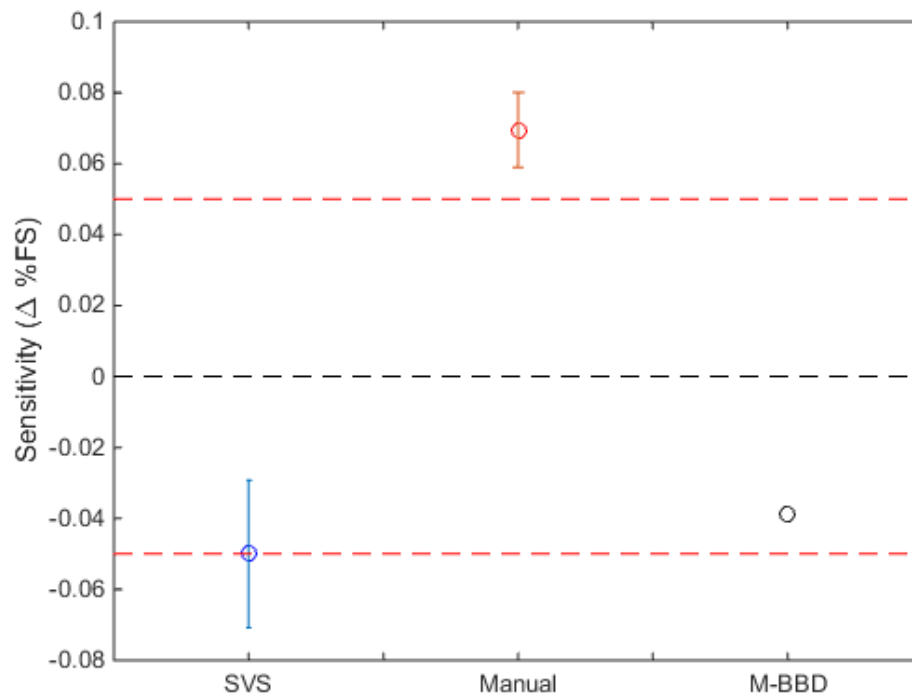


Figure 54: Cross-System Comparisons, PM Primary Sensitivity (NTF-118B)

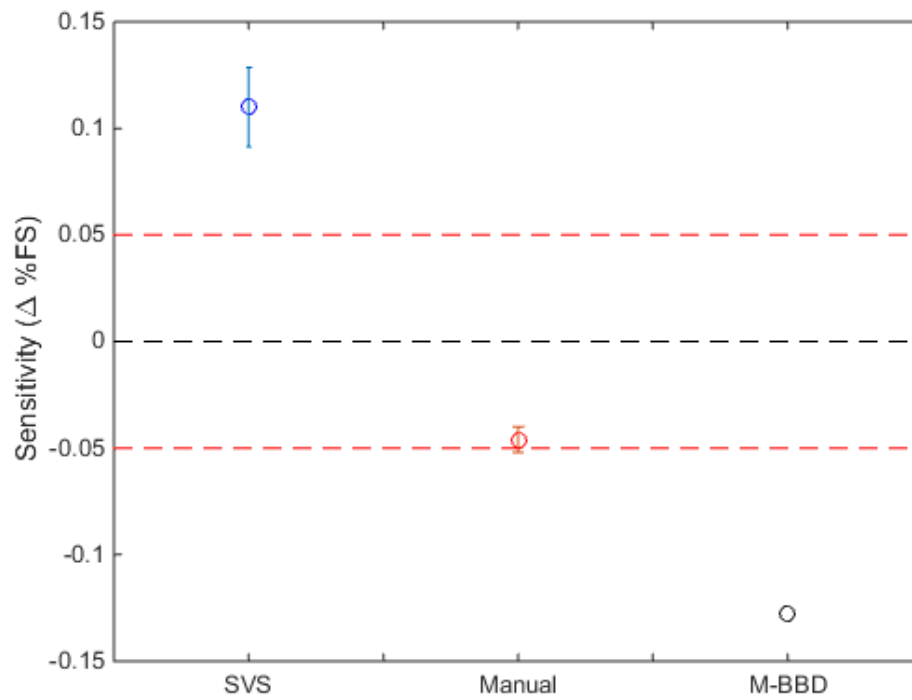


Figure 55: Cross-System Comparisons, RM Primary Sensitivity (NTF-118B)

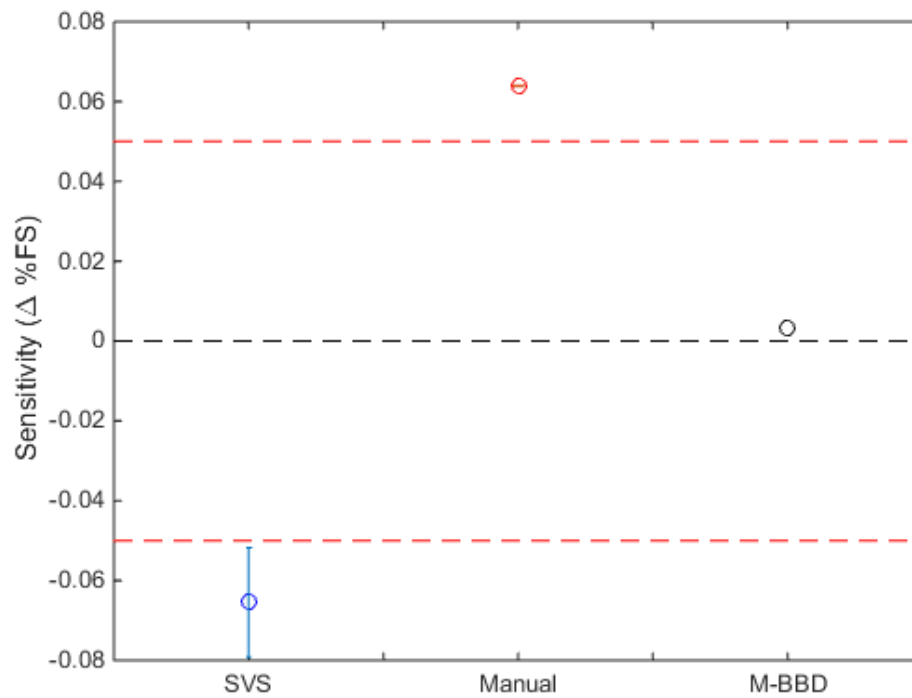


Figure 56: Cross-System Comparisons, YM Primary Sensitivity (NTF-118B)

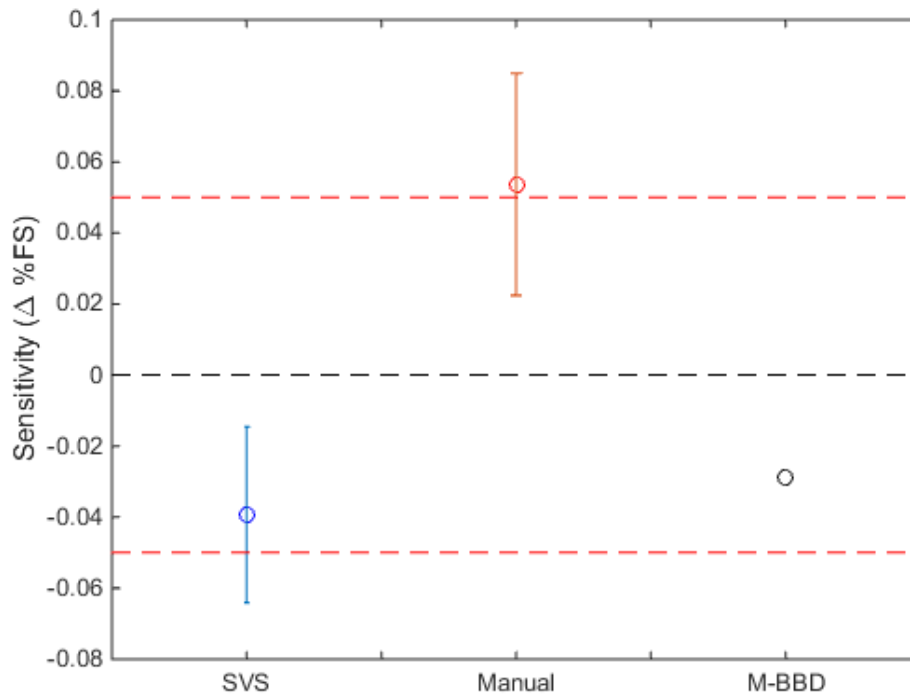


Figure 57: Cross-System Comparisons, SF Primary Sensitivity (NTF-118B)

From examination of the cross-system comparisons, it can immediately be seen that all but one of the confidence intervals do not overlap. The Manual Stand CI does overlap the SVS CI for axial force. However, it can be said that most all of the confidence intervals overlap the $\pm 0.05\%$ limits. This implies that, while most of the coefficients are statistically different from one another, in many cases, the systems are producing similar results within $\pm 0.05\%$ of one another. This balance has shown the best cross-system reproducibility of all the balances thus far, with the NTF-113C close behind. The NTF-118B has also undergone a M-BBD calibration. It can be seen that the M-BBD agrees very well with the other systems. Most of the plots, show the M-BBD produced a coefficient in-between the traditional Manual Stand and the SVS. Rolling moment shows the worst agreement of all the plots, however, the difference between the largest coefficient estimate and smallest coefficient estimate is about 0.25%.

To summarize, for all the balances and calibration systems, the overall results leave the reader with few conclusions. The NTF-113C, MC-60E and NTF-118B have shown decent agreement. The systems often don't have overlapping CIs nor do they always fall within the 0.05% error bands. This leads to the conclusion that the systems are fundamentally different and that unknown factors affect the estimates from each system. It would seem in the best cases, the calibration systems can replicate another system's results around $\pm 0.15\%$. Note that this result was established by comparing the differences between each system's coefficients and includes the entire CI width if applicable. This result however is limited to the single-piece balances shown. There is no effort planned in the foreseeable future to address this issue. Finding and fixing any issues with any of the systems or procedures is a process that is not currently feasible. An additional issue to note is which system would be considered the standard to which all other systems are compared against? An obvious choice might be the Manual Stand as it has seen the most use historically. However, this document has shown the largest variation in coefficient estimates belongs to the Manual Stand. Additionally, the SVS uses the least hardware, all forces are resolved in one vector and the calibration can be completed in a week. This drives the total uncertainty in the calibration process down, which makes the SVS a possible choice. Alternatively, the ABCS is fully automated and any calibration can be completed in a single day. This procedure should reduce errors and biases from the calibration process itself as each calibration should be the same every time. Unfortunately, this thesis has no replicate ABCS calibrations to judge the repeatability of the ABCS. Each of the systems are unique and use a different methodology to reduce errors and accurately estimate balance performance. Understanding these differences is no easy task and currently the calibration community accepts these unknowns.

4.3. Long-Term Repeatability of a Balance on a Single System

In this section, SPC will be used to analyze data from multiple replicated calibrations. All data are presented in chronological order and normalized to percent full-scale values if possible. All of the calibrations shown were completed on the NASA Langley SVS. The same calibration technician was used for every calibration. The balance was removed from the stand between every calibration and reinstalled but was not removed from the SVS load template. This procedure introduces variability that actually exists between production calibrations. The balance was not removed from the template between every calibration to prevent wear and tear on the dowel pins which hold the balance in place inside of the fixture.

4.3.1. NTF-113C Statistical Process Control

The control charts for the NTF-113C can be seen in Figure 58 through Figure 84. Figure 58 through Figure 63, shown in the next section, show the primary sensitivities of the NTF-113C. These plots are generated by normalizing each calculated coefficient by average coefficient from all calibrations over the long-term process. Each calibration point is therefore shown in percent deviation from the centerline, which has been set to zero. Equation (42) defines how this conversion was done. A goal of 20 calibrations was set to establish a representative baseline for the SPC process. The NTF-113C was calibrated 21 times over the course of about 27 months.

4.3.1.1. NTF-113C Primary Regression Coefficients

It is important to note that Figure 58 has an extra point included that all other plots omit. This point, designated as calibration 18, is a calibration that was left out purposefully from the data due to hardware problems. The point is included in Figure 58 simply to demonstrate the power of SPC applied to a balance calibration system. It was easily differentiated from the rest of the data using the charts. Calibration 18 clearly exceeded control limits on all of the primary coefficients, RMSE of both the model points and confirmation points and the normal force replicates. The magnitude of change of Calibration 18 as compared to all the others warranted an investigation for a cause. A faulty voltmeter was discovered and replaced for subsequent calibrations. Due to the obvious issues with Calibration 18, it was omitted from the remaining charts as well as from calculation of the control limits on Figure 58. Calibration 21 was performed to complete the 20 calibration goal without Calibration 18.

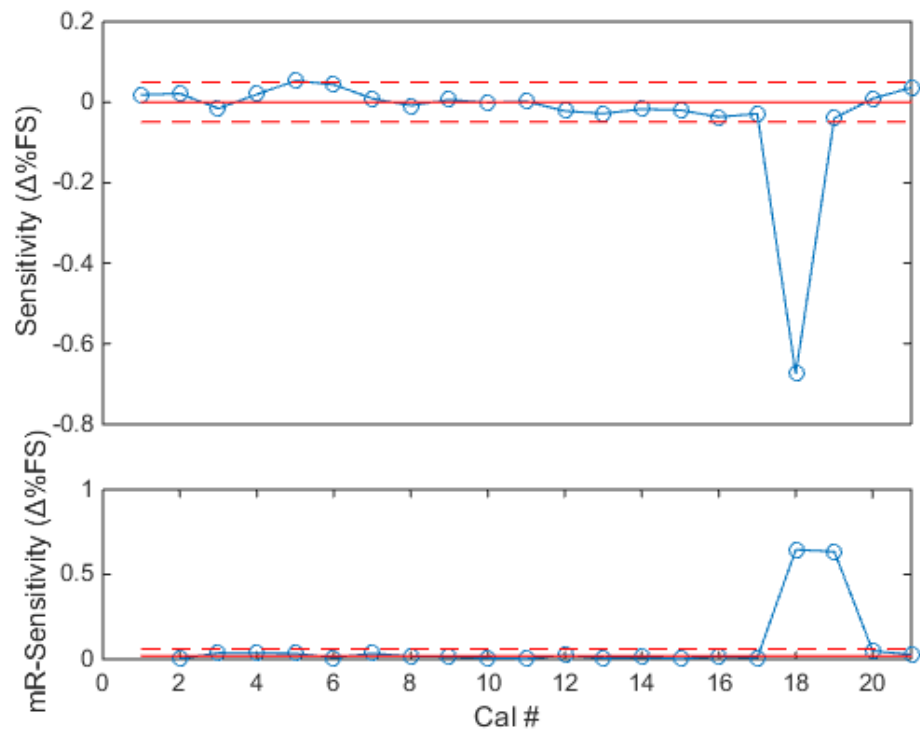


Figure 58: X-mR Control Chart for NF Primary Sensitivity (NTF-113C)

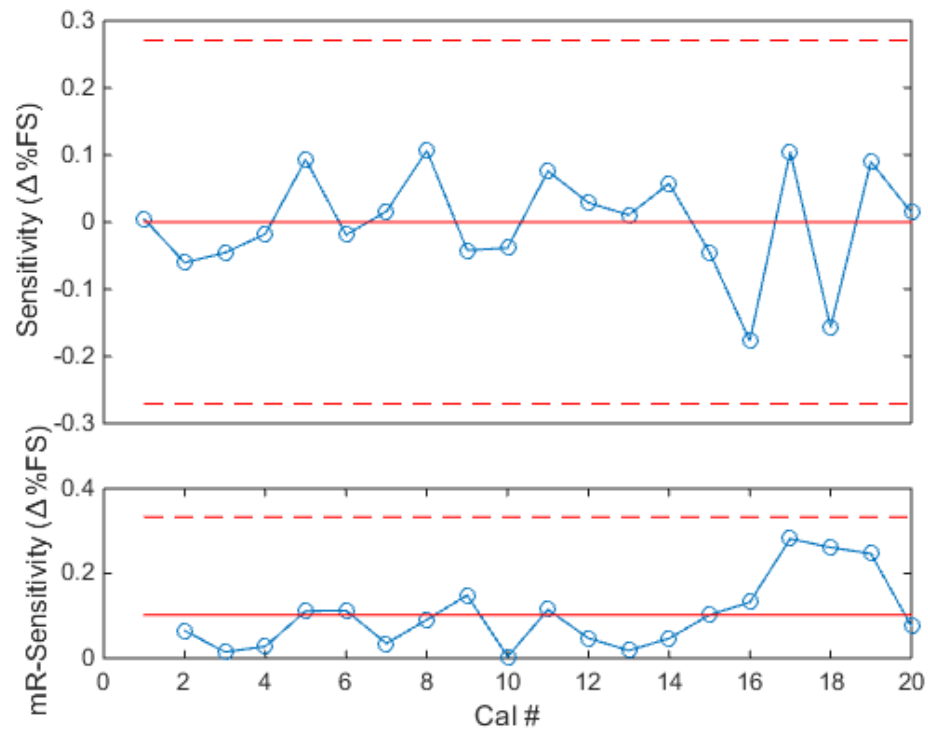


Figure 59: X-mR Control Chart for AF Primary Sensitivity (NTF-113C)

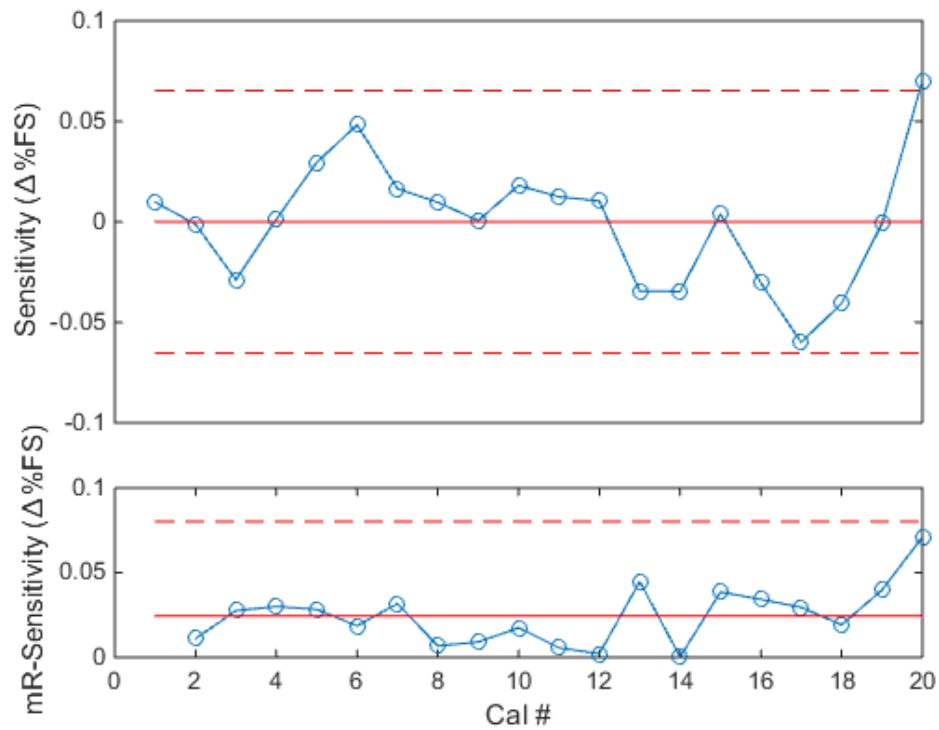


Figure 60: X-mR Control Chart for PM Primary Sensitivity (NTF-113C)

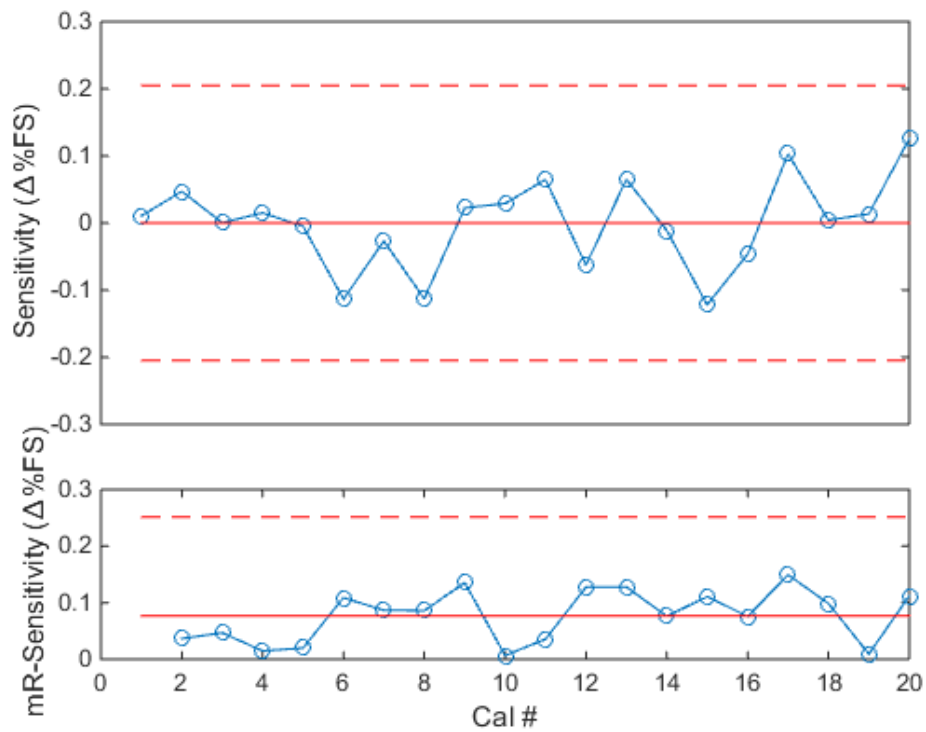


Figure 61: X-mR Control Chart for RM Primary Sensitivity (NTF-113C)

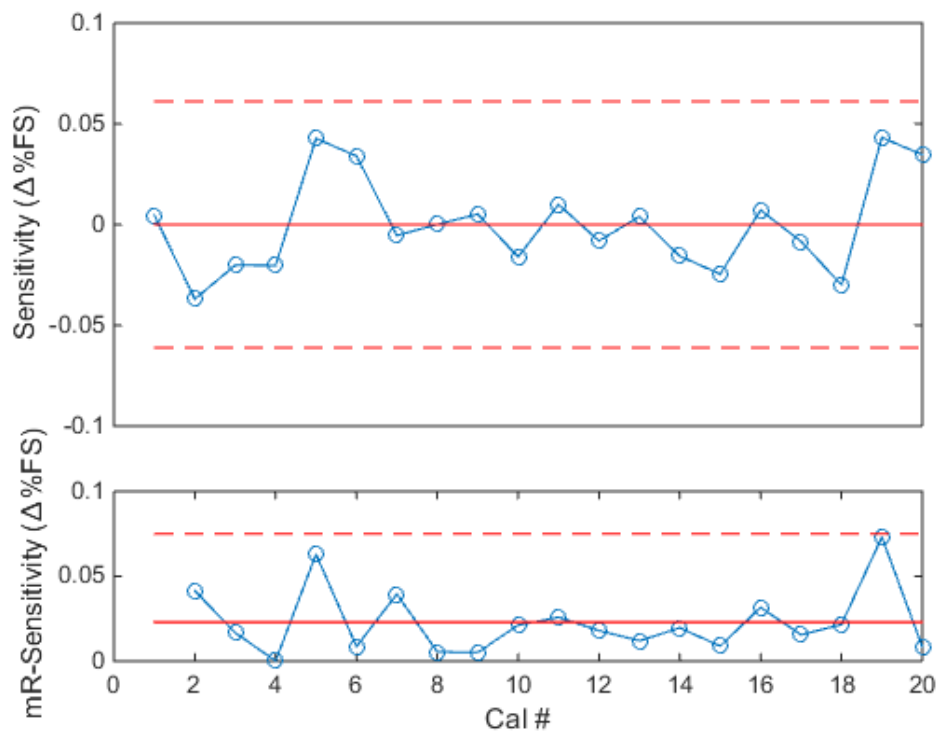


Figure 62: X-mR Control Chart for YM Primary Sensitivity (NTF-113C)

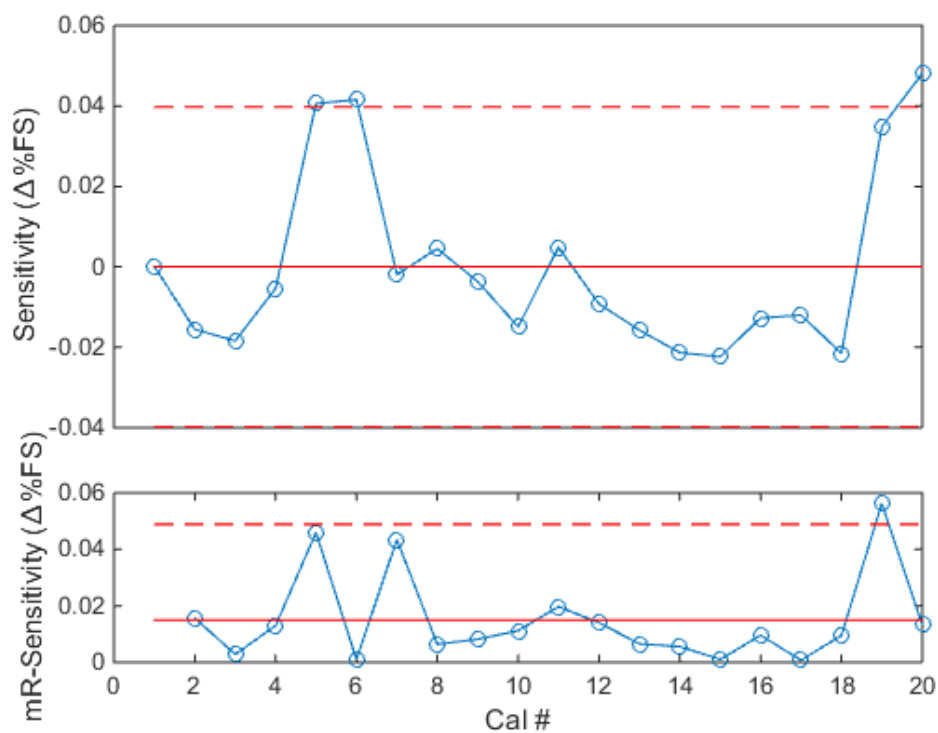


Figure 63: X-mR Control Chart for SF Primary Sensitivity (NTF-113C)

The plots shown above are representative of all coefficients not shown and valuable conclusions can be drawn from the plots provided. These plots show the change in the primary regression coefficients between calibrations in percent full scale. First order coefficients can be found in Appendix I, however all two factor interactions and second order coefficients have been omitted from this document for brevity.

From the data acquired for this calibration system and balance, it can be said that the balance and calibration technique are in relative statistical control (omitting Calibration 18). Most of the regression coefficients over the calibrations have stayed within the control limits, however the normal force, pitching moment and side force responses have shown points outside of the control limits. As these first twenty calibrations are considered a baseline (phase I), there is not much that should be done to reconcile the points that are near or slightly out of bounds. There is simply not enough information on the repeatability of this balance to truly know if this is abnormal or not.

As per the SPC method, successive calibrations provided an adjustment to the bounds. A second issue to note is how much variability is seen in the axial force and rolling moment components. Balance accuracy is quoted as being at best within one-twentieth of one percent of full scale. The primary sensitivities are considered to be fundamental to the balance mechanics and primary sensitivity changes much larger than 0.05% could be perceived as an unexplained change. The value represents an informed estimate of the uncertainty in the balance and calibration system together [2]. Specifically, the value represents the compounded error from the calibrated weights, measurements of moment arms, the resolution of the data acquisition system, temperature and humidity variations in the calibration lab and any other uncertainties in the process. The accuracy of the balance can never be any better than the uncertainty in the calibration process. Prediction

accuracy of the balance is truly set by a combination of the pure error and modeling error. However, primary sensitivities should be very repeatable due to how fundamental they are to the balance. Most of those plots show variation in the primary sensitivity within 0.05%, however, the axial force and rolling moment components show variation that exceeds 0.1%. Axial force is the primary component that measures drag when the aircraft model is at low angles of attack. This component is extremely important to aircraft cruise studies, often performed with this balance. These plots seen above led to the decision to have the NTF-113C inspected for damage at the end of the 20 calibration campaign. A small crack was found at the root of the axial force flex beams. It is unclear as to whether this occurred during the study or whether it was an existing problem. It is clear that during the study, degradation in the axial bridge was observed from the model point residuals (shown later in Figure 71). As for the rolling moment response, it is believed that the larger errors are due to the large torsional stiffness of the NTF balances. The NTF balances usually will not deflect much due to rolling moment loads. The roll estimates can therefore suffer from low resolution as it is often difficult to find a good location to locate the roll strain-gauges. Additionally, estimating roll effects on the SVS is also an issue as the moment arms are short and the rolling motion is accounted for using the large outer bearing. This can cause small errors from bearing friction. These combined issues can make rolling moment become less accurate and repeatable between calibrations. The average half-width for the control limits, not including the axial force and rolling moment bridges is 0.0538%. Note that this is a limit based on three standard deviations, which means a more traditional confidence interval, based on two standard deviations, would be even smaller. The half width of the axial and rolling moment coefficient are over four times larger than the other four components. While balance accuracy is minimally 0.05%, modeling error is also involved in the actual accuracy of the balance when used for prediction.

Table 3 shows the estimated accuracy of the balance per component. These values are historical averages. The replication of the primary sensitivities seen in the plots above is consistently better than the actual uncertainty of the balance, including axial force and rolling moment. Recall, that the limits shown here are three standard deviations and the ones in Table 3 are two standard deviations.

One further observation with regard to the load schedule shows a relatively lower statistical power in the axial force model. This could be another source of the larger variation in the axial force primary response and may warrant changes to future load schedules to attempt to raise the power.

4.3.1.2. *NTF-113C Confirmation and Model Point Error*

Root-mean-square errors (RMSE) can be derived for both the model points and the confirmation points included in the SVS load schedule. Figure 64 through Figure 69 show the confirmation point RMSE ($RMSE_C$).

Figure 70 through Figure 75 show the model point RMSE ($RMSE_M$). The residuals are reported in units of millivolts per volt (mV/V). The RMSE for either the model or confirmation points is computed by calculating the absolute value of the average residual per calibration. This value is then normalized by the component full-scale voltage output, expressing the residual in percent full-scale units. To achieve the full-scale voltage, the maximum loads are run through the fitted model from each calibration to get full-scale voltage output. The RMSE is then divided by that voltage. Because the desired residual is always zero, the following plots will allow an in-depth look at both the magnitude of the errors with respect to full-scale output and how they are changing over time.

The $RMSE_C$ gives an idea of the calibration robustness and can be seen in Figure 64 through Figure 69. The confirmation points are points not used to fit the model, but they are simply used to verify that the balance predicts the load correctly. Each confirmation point is unique and different from any model point (design point). The $RMSE_C$ values are plotted in X-mR control charts and tracked over the course of the calibration study.

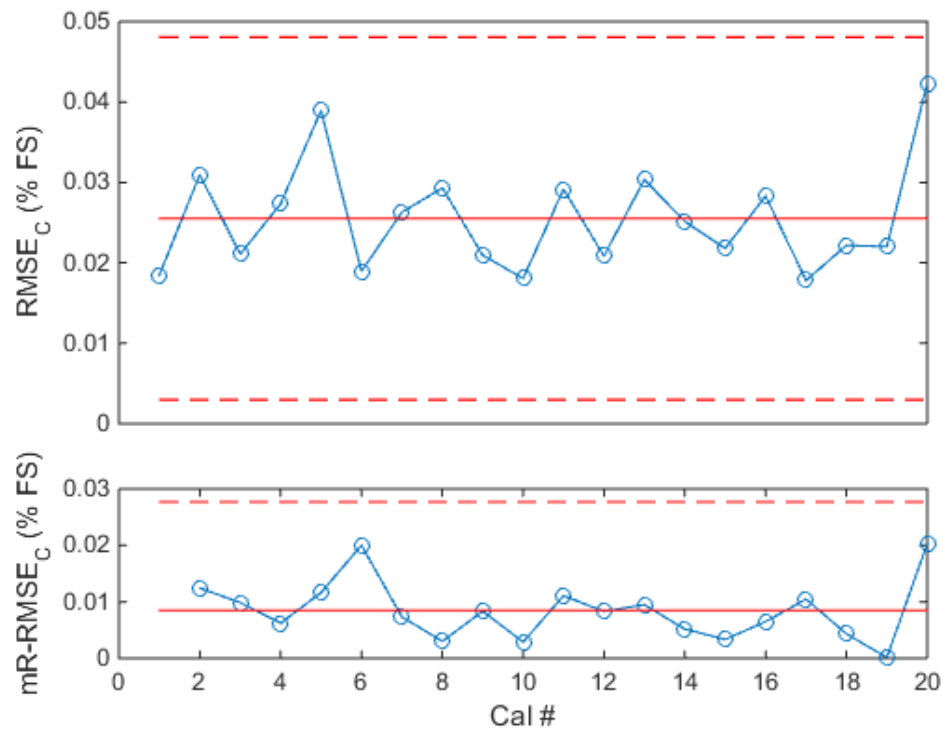


Figure 64: X-mR Control Chart of RMSE on NF Confirmation Point Residuals (NTF-113C)

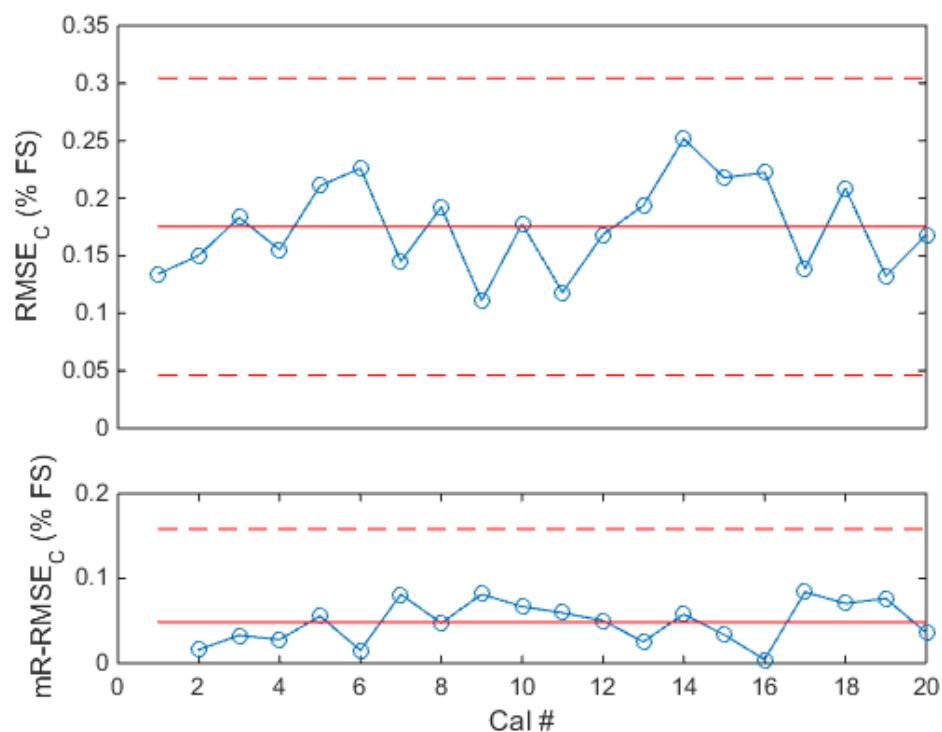


Figure 65: X-mR Control Chart of RMSE on AF Confirmation Point Residuals (NTF-113C)

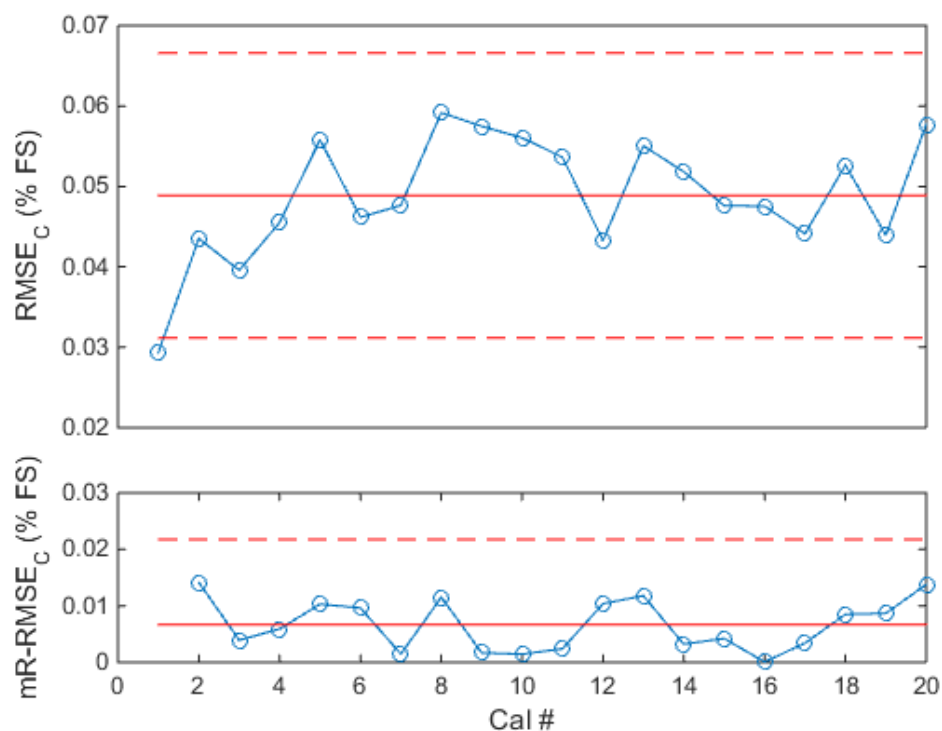


Figure 66: X-mR Control Chart of RMSE on PM Confirmation Point Residuals (NTF-113C)

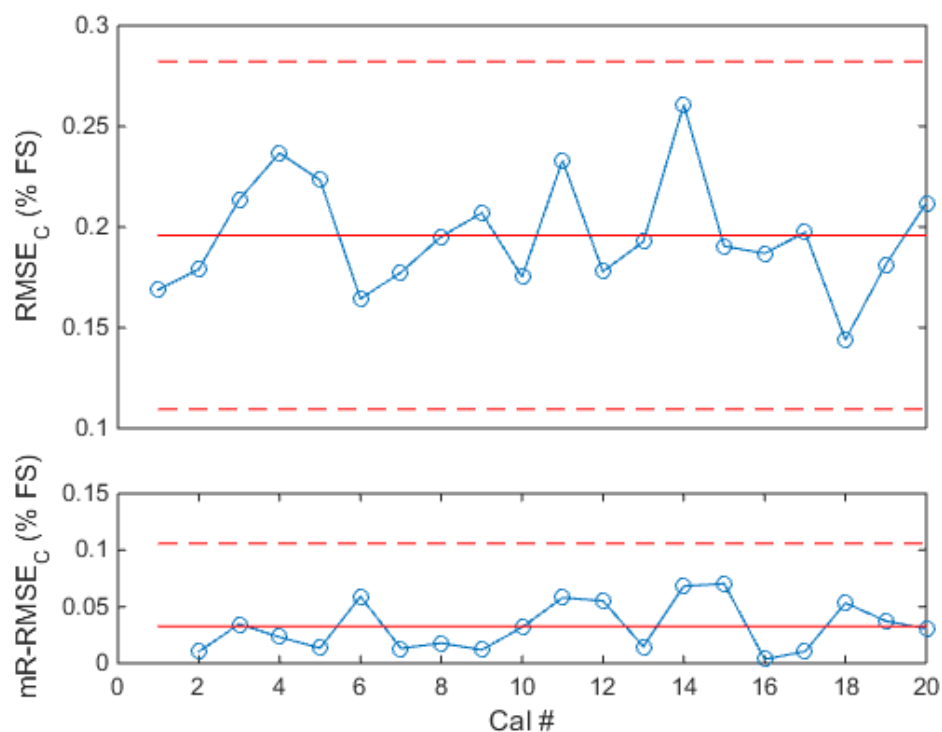


Figure 67: X-mR Control Chart of RMSE on RM Confirmation Point Residuals (NTF-113C)

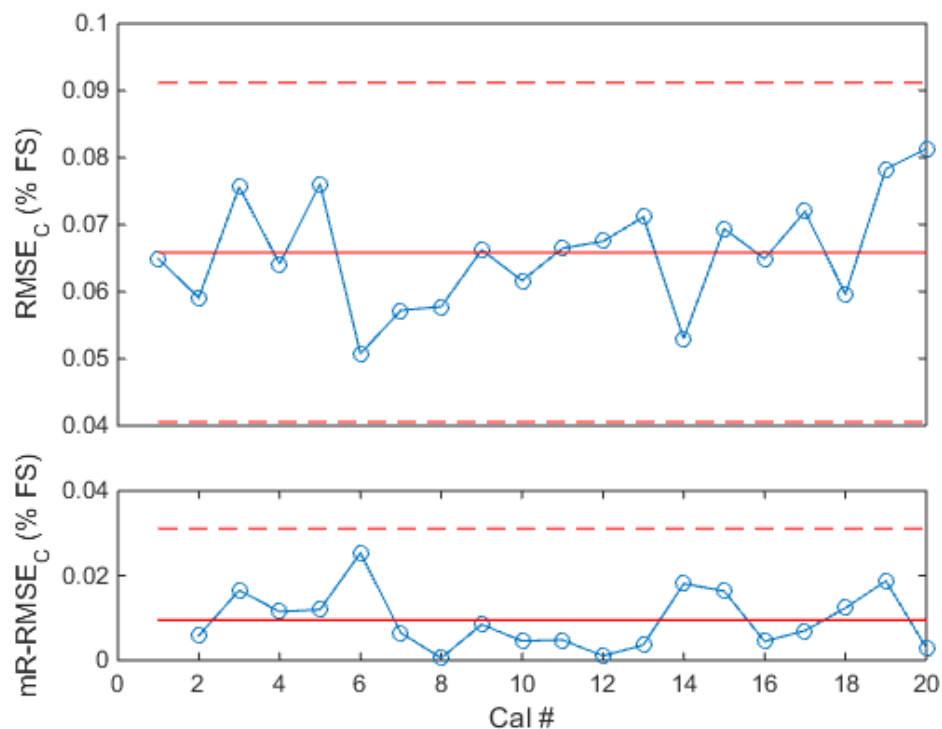


Figure 68: X-mR Control Chart of RMSE on YM Confirmation Point Residuals (NTF-113C)

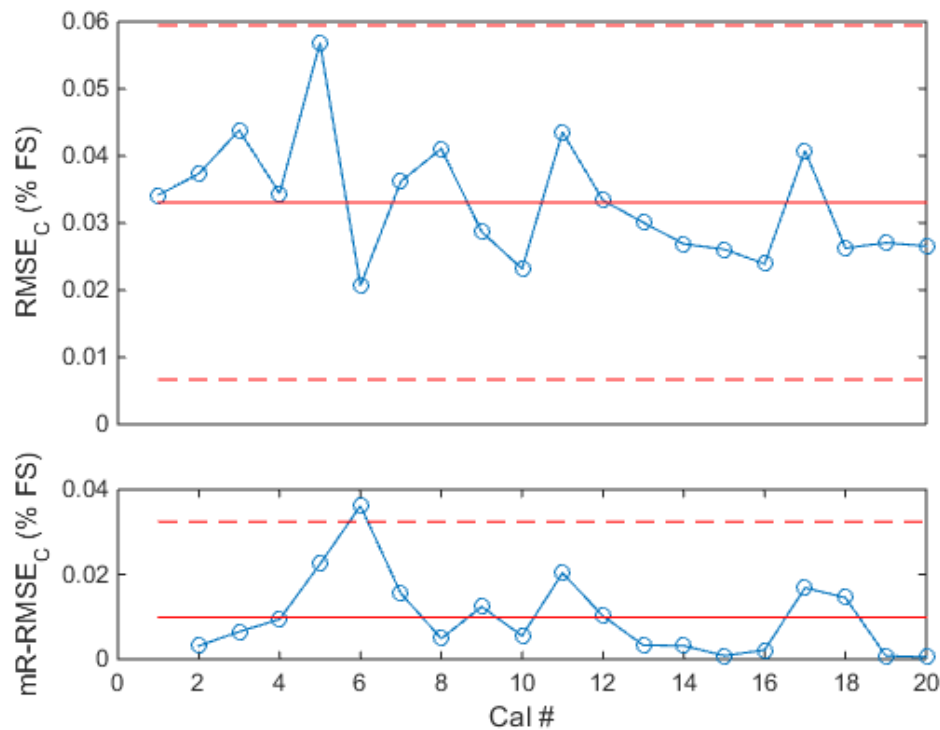


Figure 69: X-mR Control Chart of RMSE on SF Confirmation Point Residuals (NTF-113C)

It can be seen that all of the points remain within the control limits. Axial force and rolling moment show much poorer performance than the other four components. Both average about 0.2% error for all calibrations. Again, axial force is the most sensitive component of the balance, and arguably the most important component for researchers. The errors seen here are likely due to the damage found on the axial force flexures and the issues with roll as mentioned earlier. In general, errors nearing 0.2% is not ideal, and are large in magnitude when compared to the total accuracy of the balance, shown in Table 3. Note, that RMSE is similar to one standard deviation of observed errors. The total accuracy in the axial and roll sections are quoted at 0.240% and 0.186% respectively. These values should be divided by 2 for comparison to the RMSE values, which equates to 0.12% for axial and 0.09% for roll. The RMSE_C for axial and

roll averages about 0.2% full scale for both, which exceeds the historical values by a tenth of a percent. All other components show better performance than the historical values.

The $RMSE_M$, seen in Figure 70 through Figure 75, come from the actual model points and is a combination of pure error and model lack of fit. Both are errors associated with how well the model fits the data. Recall, pure error is model independent and derived from true replicated points, and lack-of-fit is a model dependent estimate of the variability accounted for by variance from terms not used in the model. They are normalized and plotted in the same way as the confirmation point residuals.

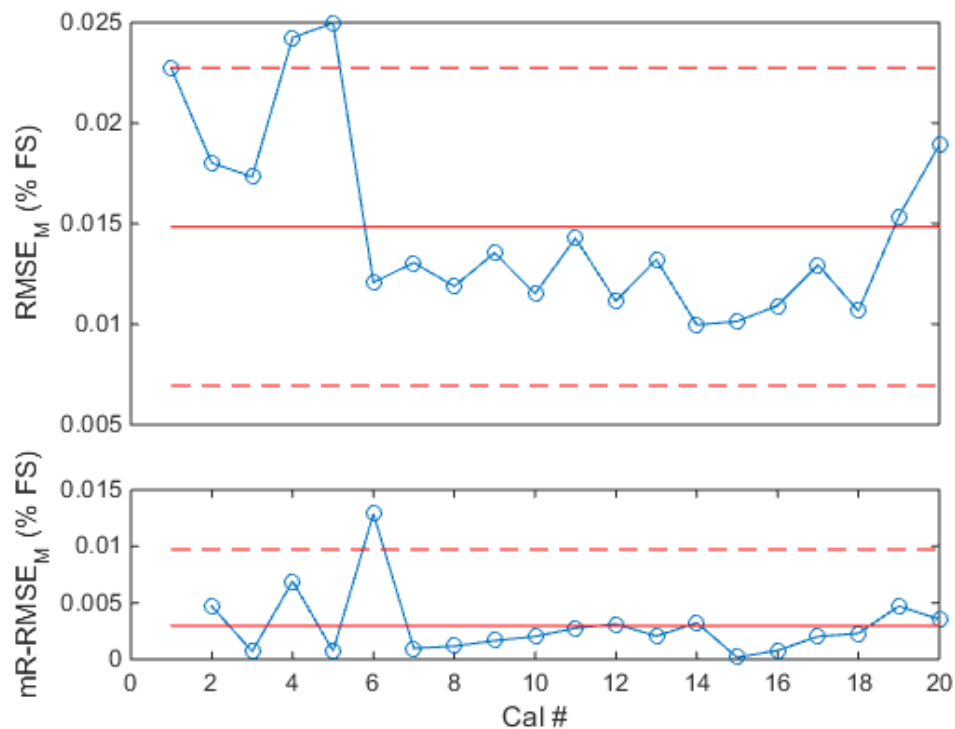


Figure 70: X-mR Control Chart of RMSE on NF Model Point Residuals (NTF-113C)

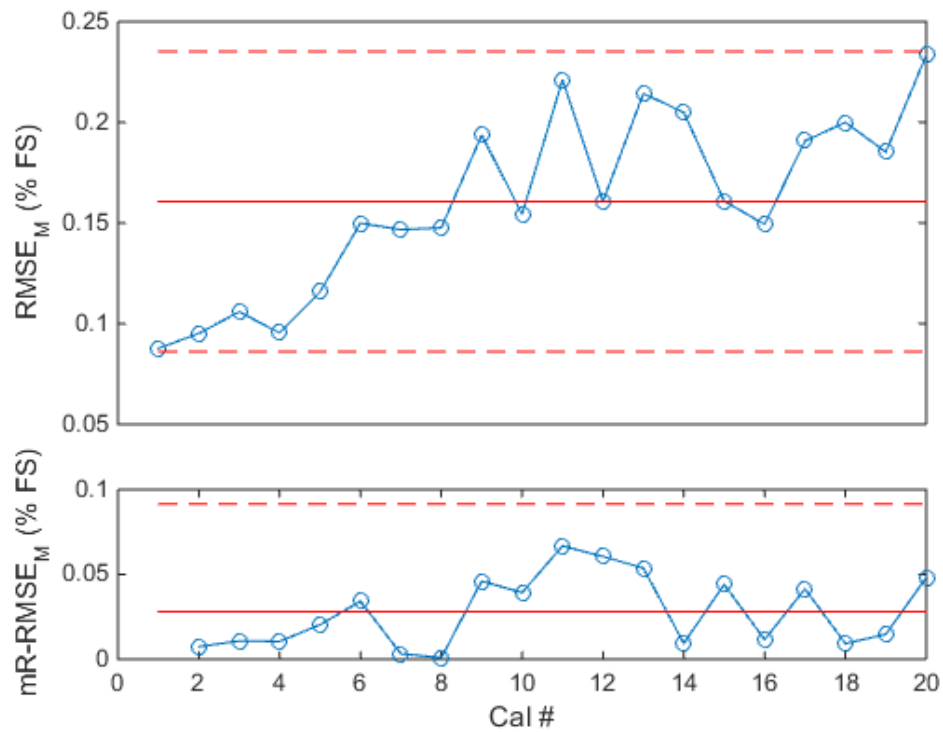


Figure 71: X-mR Control Chart of RMSE on AF Model Point Residuals (NTF-113C)

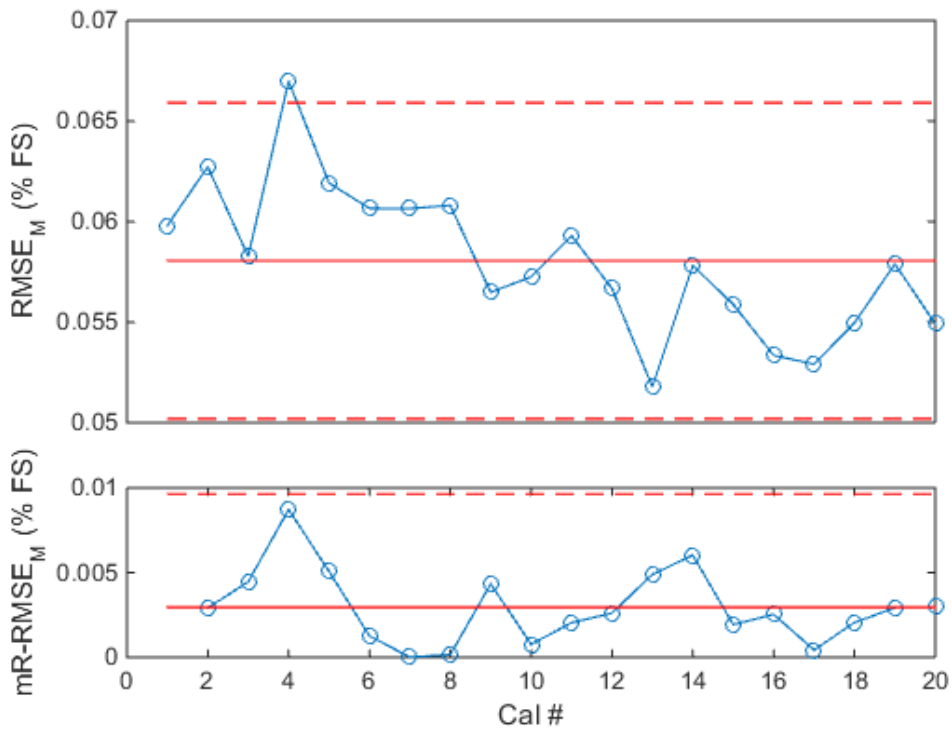


Figure 72: X-mR Chart of RMSE on PM Model Point Residuals (NTF-113C)

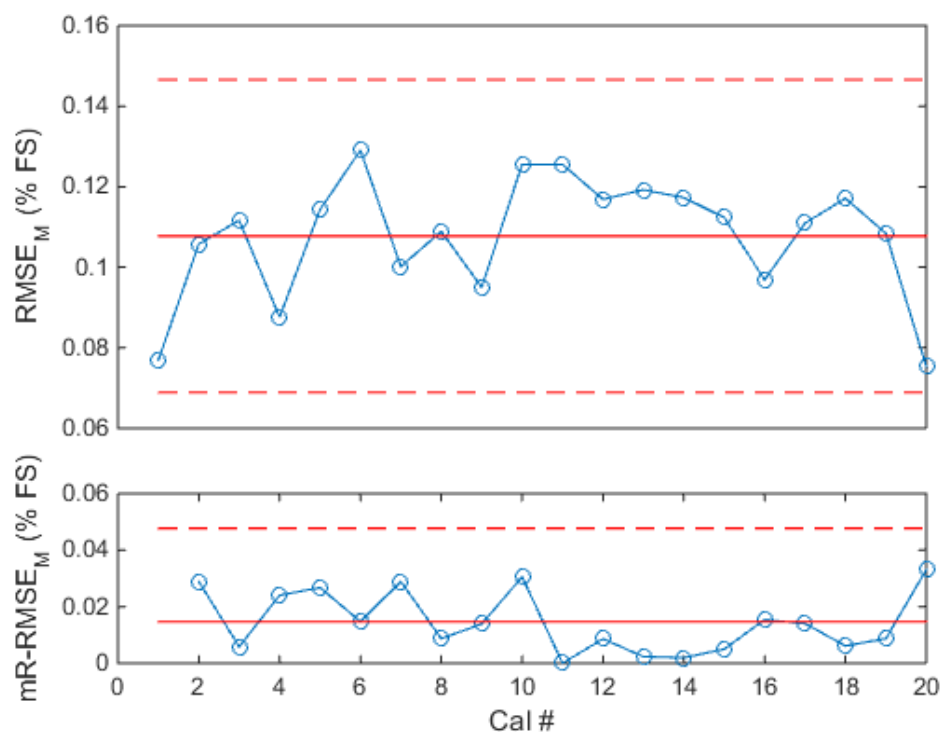


Figure 73: X-mR Control Chart of RMSE on RM Model Point Residuals (NTF-113C)

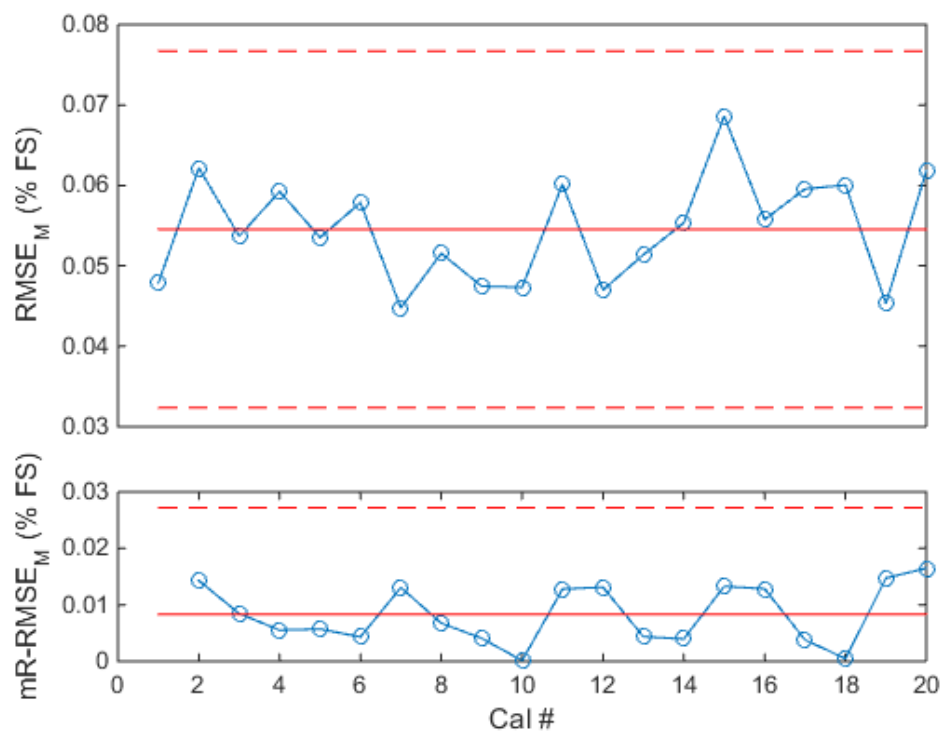


Figure 74: X-mR Control Chart of RMSE on YM Model Point Residuals (NTF-113C)

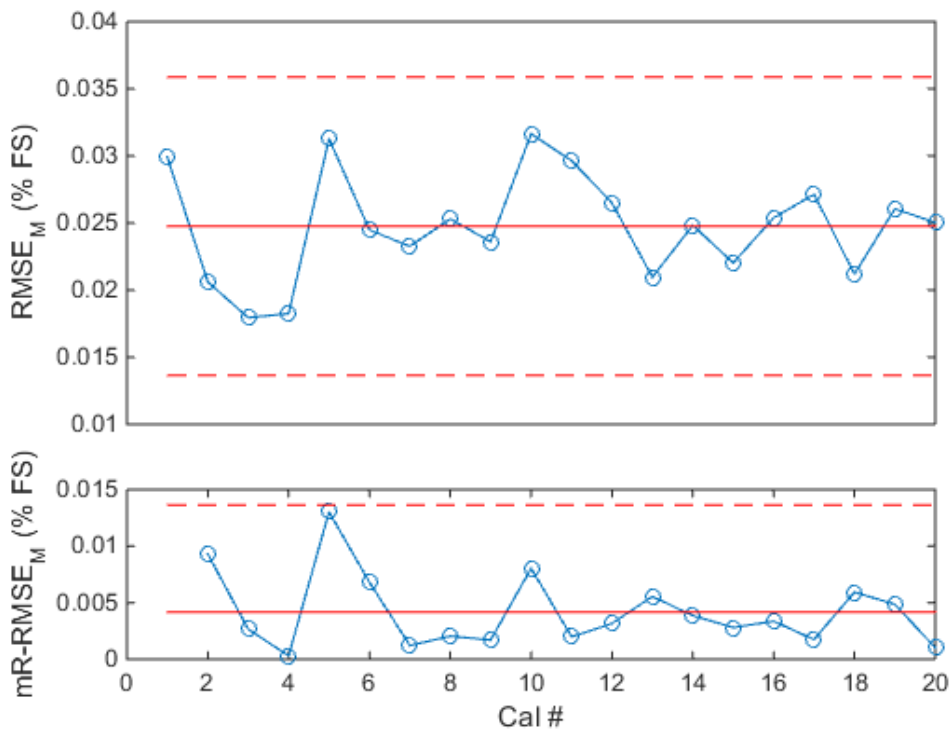


Figure 75: X-mR Control Chart of RMSE on SF Model Point Residuals (NTF-113C)

From inspection of the $RMSE_M$ plots, it can be seen that up to now, nearly all the points are within the limits, and the average RMSE is fairly small for all components other than axial and roll components. Pitching moment is slightly elevated but not to the extent that axial and roll are. One specific trend to notice is the generally increasing errors in the axial force model point residuals. This is an alarming trend which is believed to exemplify the damage on the axial section. The trend implies that repeated calibrations caused further deterioration over time to the axial section. Likely, the crack was being elongated from repeated stress cycles which caused the errors to continually increase for every calibration after Calibration 4. Calibration 4 appears to be the point where axial force began to deteriorate. It's not clear if the balance was damaged prior to the calibration study or if some event in-between Calibration 3 and 4 caused the damage itself. In most cases, it appears that the $RMSE_M$ is smaller or about equal to the $RMSE_C$ for all components. This

is expected as the error associated with points used to fit the model should be smaller when compared to points not used to fit the model. While this is not a desirable result, the tracking of axial force residuals led to a discovery that may not have been identified without careful SPC tracking. The errors seen in all plots average near or below 0.05% full-scale, omitting axial force and rolling moment, which are near 0.15% and 0.11% respectively. However, rolling moment has remained lower than the historical accuracy limit. All other $RMSE_M$ values remain near half or lower than the historical values for accuracy found in Table 3. This is a promising result as it can be seen that all of the undamaged components are operating as expected from historical data. While the average axial RMSE is not very high, the trend is alarming and the maximum RMSE seen in Calibration 20 is showing errors that are double the historical values.

4.3.1.3. *NTF-113C Normal Force Replicates*

Replicates were added to the calibration procedure in Calibration 8. There are no loaded replicates prior to Calibration 8 as can be seen in the figures below. With replicates, an \bar{X} -R chart can be used. Recall that these charts track the process mean and the range for each subset of data. Figure 76 through Figure 78 show the normal force replicated points. Note that normal force replicates are reported in raw voltage form and have units of millivolts per volt (mV/V). This choice of units keeps the normal force replicates model independent. The average zero load output in mV/V for each calibration has also been subtracted from each of the replicated normal force loads. Subtracting the zero load output is a common practice for most measurement devices. This removes any bias from shifting zeroes that change with daily conditions, but keeps the input/output relationship of the device the same. The replicated points will set the baseline for the repeatability of the balance, calibration system and calibration process together. Each replicate is an identical point where the balance was oriented with normal force in line with gravity and the same weight

was placed on the balance. There are two electrical bridges that make up the full normal force response in an NTF balance. They are designated N1 and N2 respectively and are placed at opposite ends of the balance. These bridge outputs will be identified separately for this analysis rather than combined. The other electrical bridges, AF, RM, S1 and S2, will be excluded from the normal force replicate analysis.

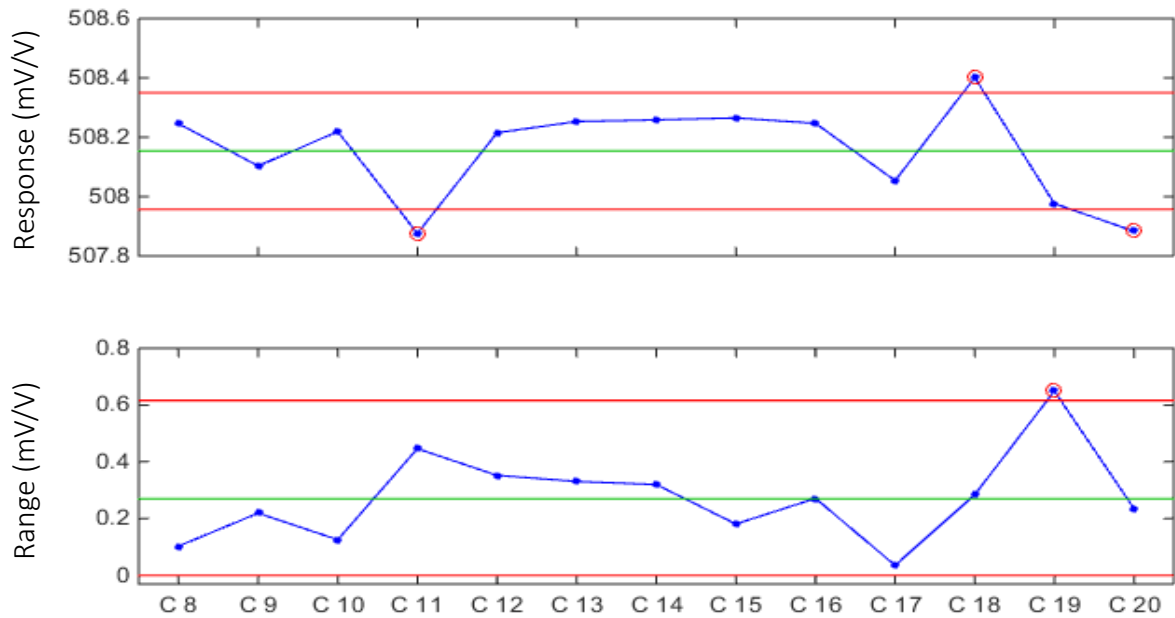


Figure 76: \bar{X} -R Control Chart for N1 Bridge, NF Replicates (NTF-113C)

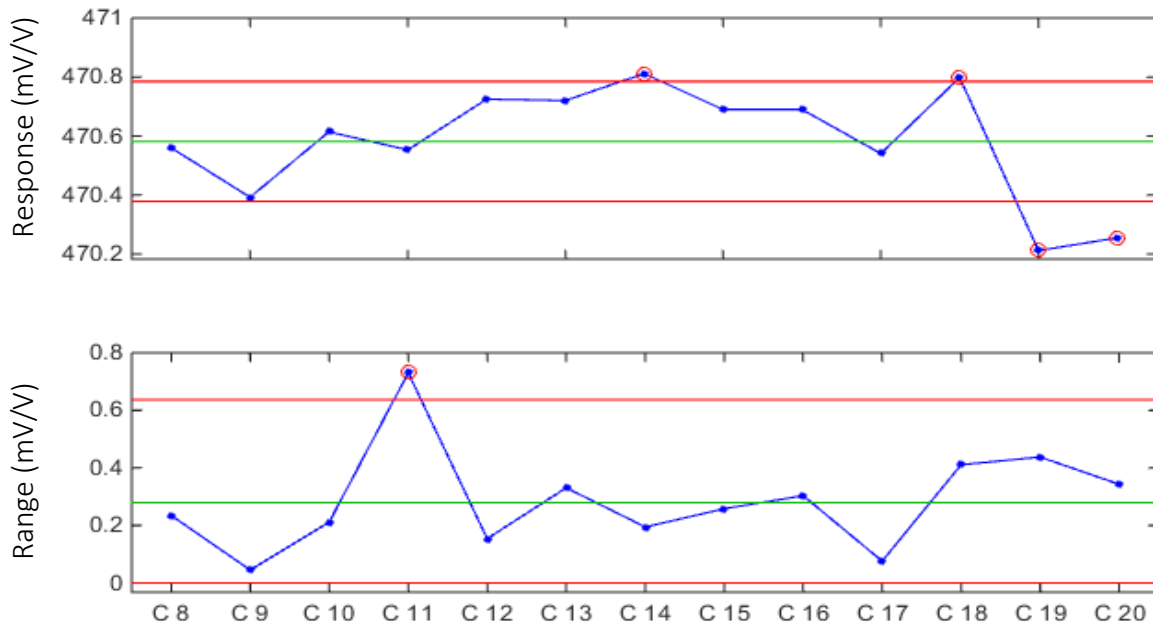


Figure 77: \bar{X} -R Control Chart for N2 Bridge, NF Replicates (NTF-113C)

The plots above show the variability, or repeatability, of the balance under a pure normal force loading. These plots can be used to understand both the within-calibration repeatability and across-calibration repeatability. The \bar{X} chart shows the time history of the mean output for each calibration over all 13 calibrations (Calibrations 8 to 20). The R chart displays the variation of N1 and N2 within each calibration. Looking at the plots, it can be seen that the data are scattered mostly within the bounds. It would seem that the balance shows similar repeatability within a calibration as when compared to across calibrations. The average variation seen in either of the \bar{X} plots is approximately $\pm 0.2 \text{ mV/V}$. This equates to approximately 0.02% change in full scale voltage per bridge, for either bridge. Outliers leave the control chart limits if they vary any more than about 0.2 mV/V. The associated range charts show within-calibration variation to be about the same. There are a few more outliers on the \bar{X} chart than the R charts. It would seem that the across-calibration repeatability is just slightly less stable than the within-calibration repeatability.

From further examination of the \bar{X} charts, the outliers can usually be found on both the N1 and N2 plots. This leads to the conclusion that there was some external factor or factors that caused both N1 and N2 gauges to read high or low. The R charts on the other hand, show only a single point out of bounds for each bridge. Oddly enough, each of these points are not the same point for both bridges. Due to the fact that the control charts did not flag the same calibration on both bridges, it is believed that there is no serious extraneous cause for this excursion. It can be concluded that within-calibration repeatability is acceptable and slightly better than the across-calibration repeatability. The limits for the \bar{X} chart are determined by the average range of the data. If the within-calibration repeatability is generally good, the limits of the \bar{X} chart become smaller. If there is a lot of unexplained variation between calibrations, then the points on the \bar{X} chart will leave the limits. It makes sense that there are some extra unexplained factors that can become important between calibrations as compared to within calibration. These factors could be related to atmospheric conditions such as temperature, humidity and pressure.

To assess the consequences of these changing replicates, each average replicate for each calibration can be converted back into loads. This is achieved by combining the N1 and N2 bridges (N1+N2) and using an independent calibration matrix (not the matrix related to the chosen zeroes). The effect of the zero shifts can therefore be independently assessed. Using the fitted model for each calibration will force a good correlation between the electrical output and the loads, making the effect of the changing voltages obfuscated. Using a separate matrix will make the effect easier to see and quantify. The matrix used to convert the voltages into loads is from Calibration 7. To make this a worst case comparison, the tare model was also used from Calibration 7. Table 11 shows the maximum percent difference of the normal force load prediction for each set of averaged

replicates. The predicted loads are also shown as normalized by the average predicted load and plotted in Figure 78.

<i>CALIBRATION</i>	<i>NF PREDICTION (LBS)</i>
<i>8</i>	3049.03
<i>9</i>	3048.25
<i>10</i>	3049.15
<i>11</i>	3048.56
<i>12</i>	3050.05
<i>13</i>	3050.8
<i>14</i>	3050.76
<i>15</i>	3051.19
<i>16</i>	3050.86
<i>17</i>	3050.26
<i>18</i>	3051.63
<i>19</i>	3047.98
<i>20</i>	3048.55
<i>MAX % DIFFERENCE</i>	0.119
<i>Standard Deviation (lbs)</i>	1.239

Table 11: NF Load Prediction from NF Replicates (NTF-113C)

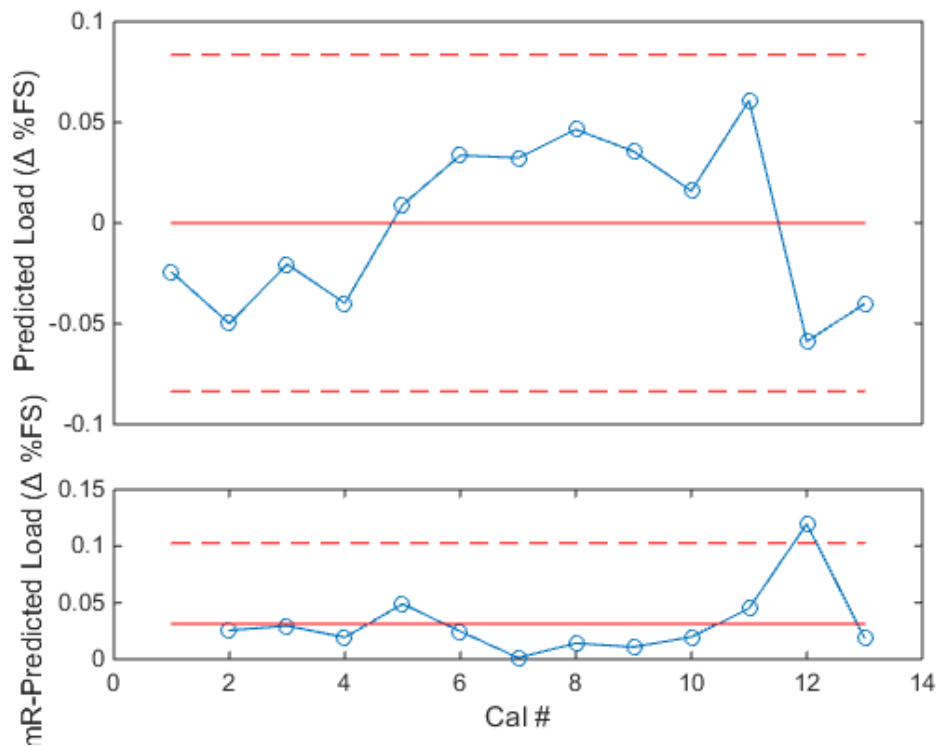


Figure 78: Change in Predicted Load from Average NF Replicated Points (NTF-113C)

Note the percent difference is about 0.12% between the maximum predicted load and the minimum predicted load, the standard deviation is 0.02% FS. This can be compared to the historical value for balance accuracy on the normal force component. The value from Table 3 is 0.063% which is a two standard deviation estimate of accuracy of the NTF-113C. It can be seen that the base repeatability of the NTF-113C is at worst 0.02%. This result is very important because repeatability must always be better than the accuracy of the balance. The repeatability of the calibration process both-within calibration and across-calibration are good.

4.3.1.4. NTF-113C Phase 2 Analysis

The purpose of SPC is to establish a baseline from a representative population of data by creating control charts and control limits which will represent the acceptable range of operation

for that system. These first runs are known as Phase I. All of the benefits from SPC come from a period of time that is called Phase II. Figure 79 through Figure 84, shown below, show a phase II SPC analysis on the primary coefficients collected for the NTF-113C. Phase II describes the point from which the control limits have been established and new data is compared to the limits (calculated in Phase I), but not added into the control limit calculations. For demonstration purposes, the first five calibrations are used to set the limits (Phase I) and the last 15 were plotted over those limits. Five points were chosen because it has been proven that the first five calibrations are different than the last fifteen on the NTF-113C. This effort was intended to both examine the procedure and to determine if the damage on the axial section was affecting coefficient estimates.

From the literature, it is clear that the more points used in the baseline, the more representative and robust the SPC process will be. It is not recommended to construct limits from a small sample size and base any conclusions on such a comparison. However, due to the small sample size in this study, the primary sensitivities will be used to demonstrate a Phase II analysis procedure. The first five calibrations were used as the baseline to generate the control limits, and the last fifteen calibrations were plotted within the established control limits. Figure 79 through Figure 84 show the primary sensitivity phase II charts for the NTF-113C.

To clarify the differences between these charts and the charts seen previously, note that the light blue data points are the data used to generate the limits and the solid red centerline is the average of that data. The dark blue data points are the data not used to adjust the control limits and the dotted red line in the center of that data is its average. The dotted red lines near the top and bottom of the charts are the limits established by the first five points.

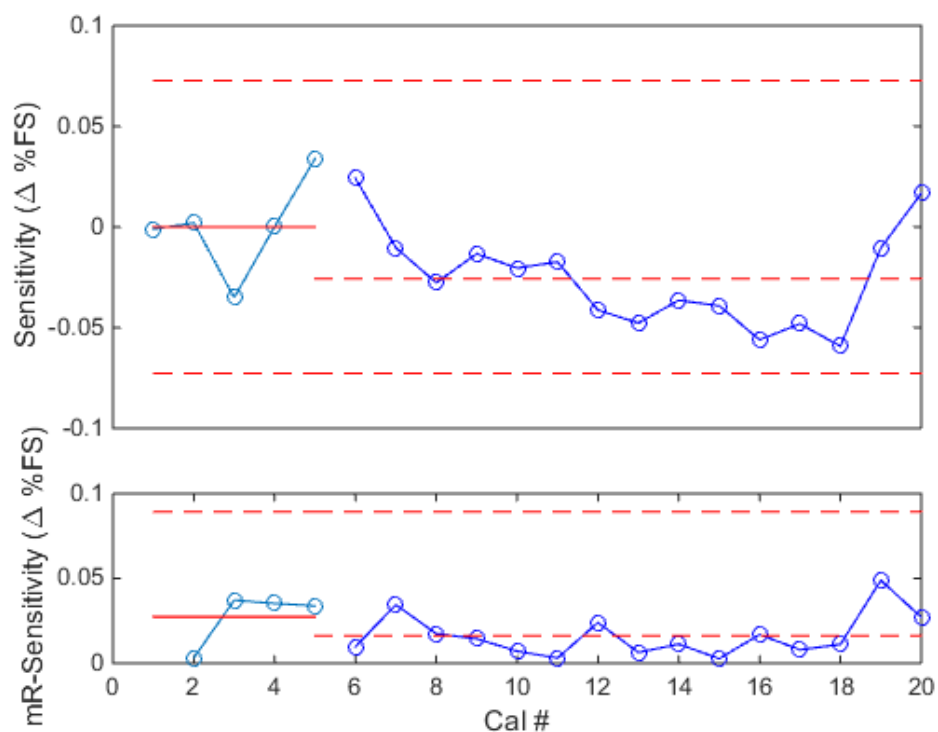


Figure 79: Phase II Control Chart for NF Primary Sensitivity (NTF-113C)

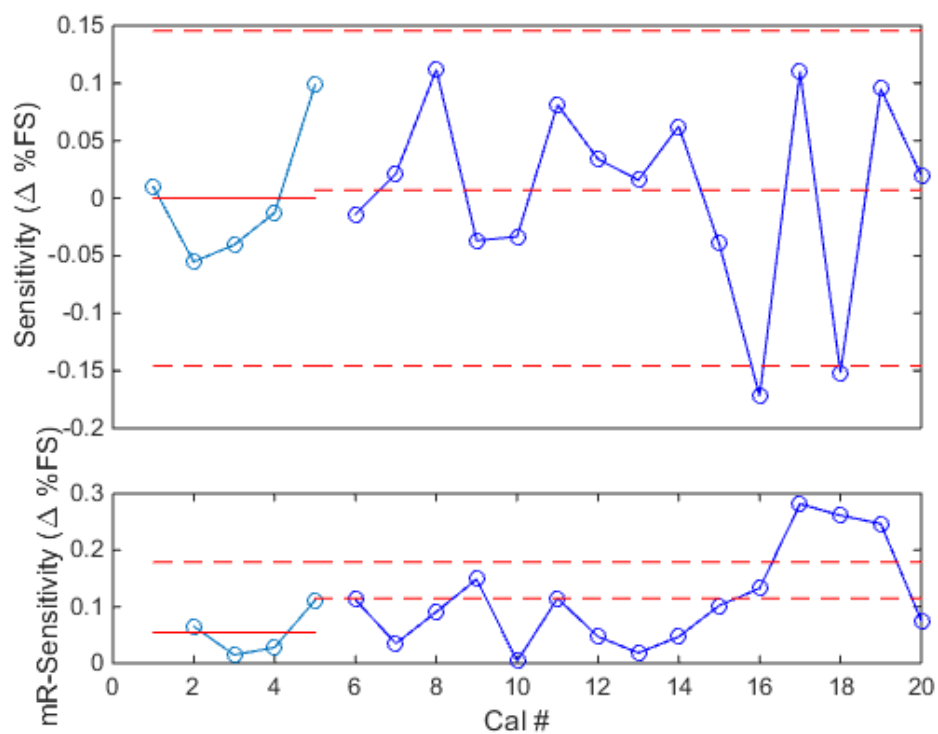


Figure 80: Phase II Control Chart for AF Primary Sensitivity (NTF-113C)

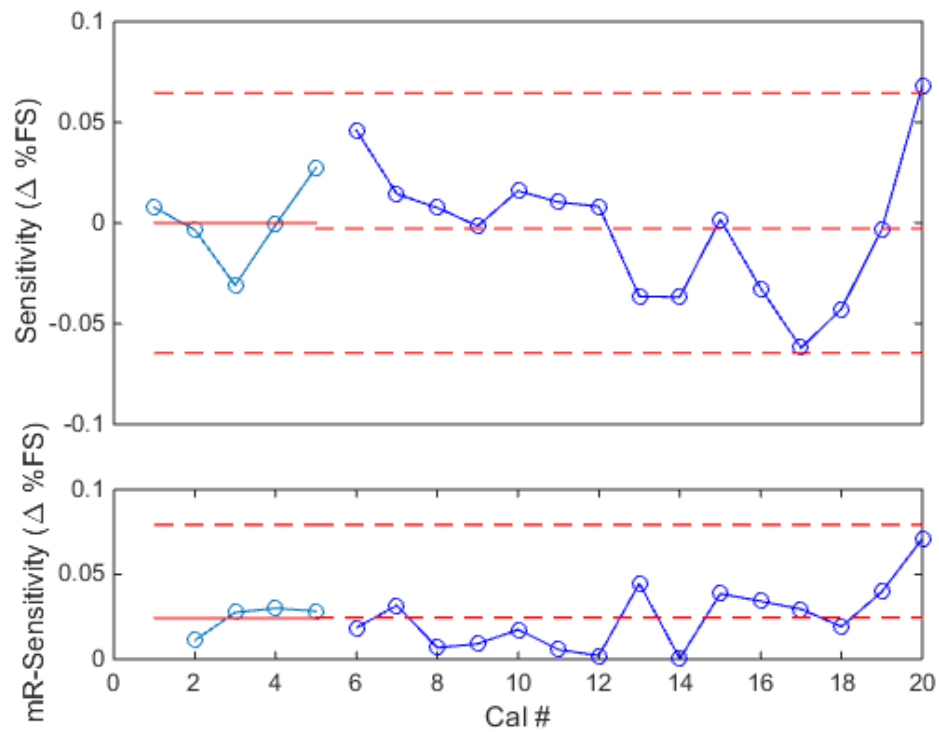


Figure 81: Phase II Control Chart for PM Primary Sensitivity (NTF-113C)

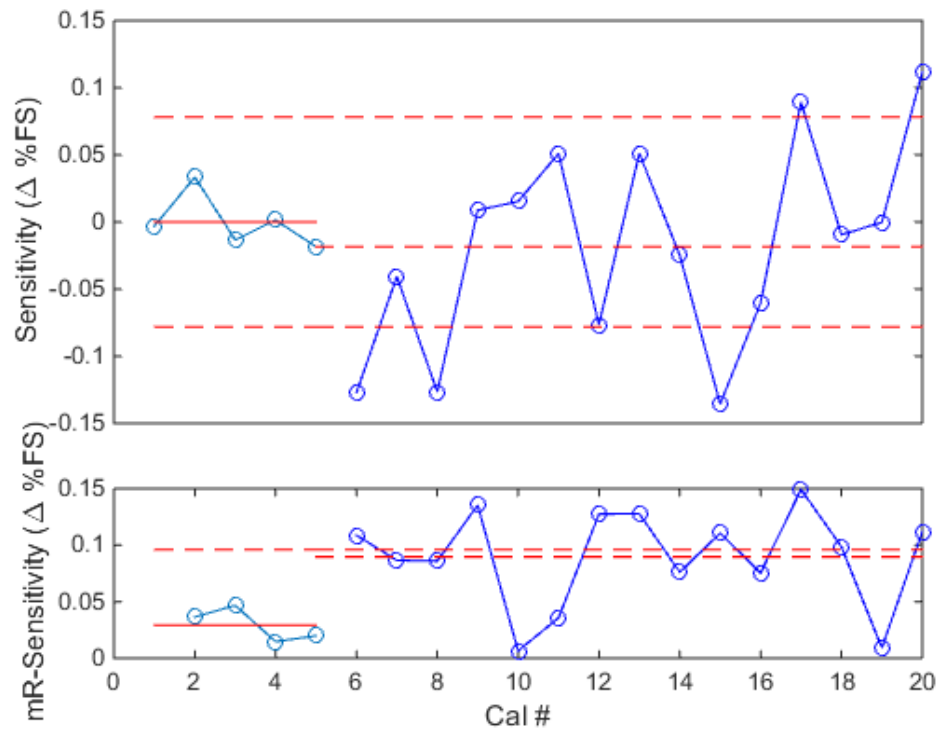


Figure 82: Phase II Control Chart for RM Primary Sensitivity (NTF-113C)

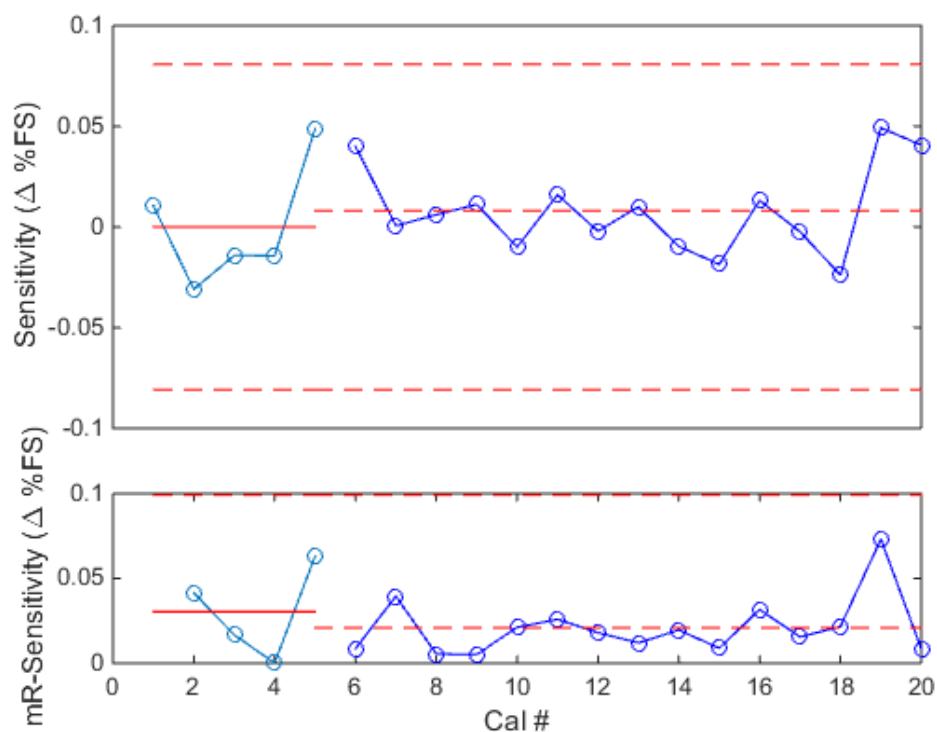


Figure 83: Phase II Control Chart for YM Primary Sensitivity (NTF-113C)

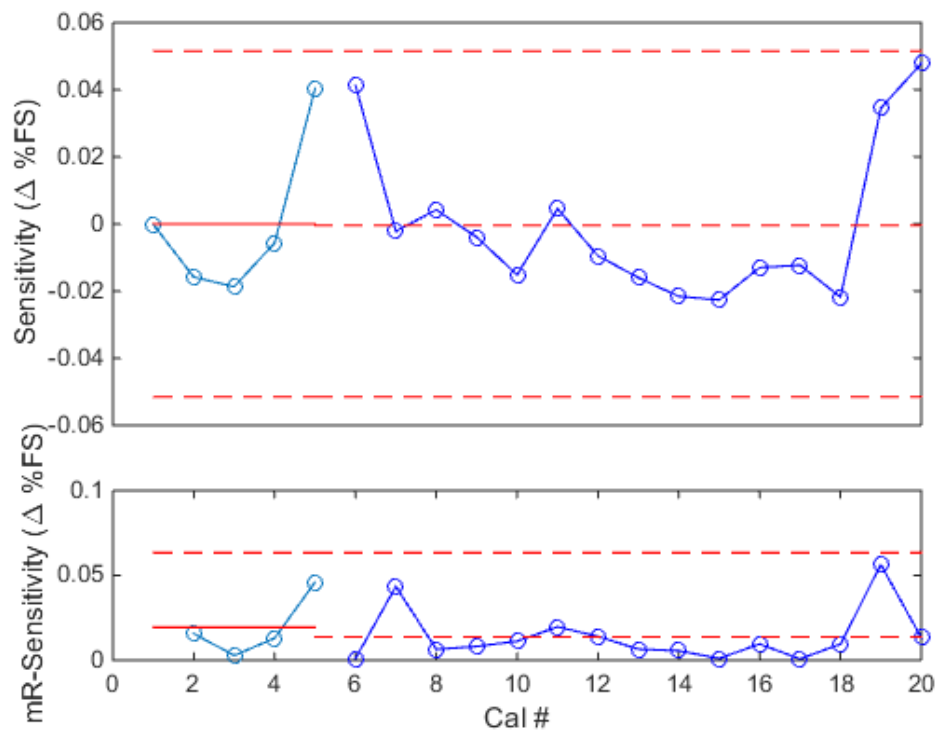


Figure 84: Phase II Control Chart for SF Primary Sensitivity (NTF-113C)

From the plots above, trends show that the limits seem to capture most of the data. Axial force and rolling moment appear to show the lowest capture rate of all the plots. However, with the information presented previously, this is not surprising. It seems feasible that with only five calibrations, a control chart can be constructed for practical monitoring of balance calibrations over long term. It is recommended to use as many points as possible when establishing the limits. As more data is added, the limits become smaller and more sensitive to outliers. Using the charts, it has been shown that for most components (not axial), the axial force damage has not affected the estimated primary sensitivities dramatically. However, for axial and roll, the balance appears to be behaving more erratically. Where resources are limited, such charts could be used to flag potential problems such as faulty gauges or damaged flexures.

4.3.2. Triumph MC-60E Statistical Process Control

The following plots and associated analyses show data from the Triumph MC-60E. SPC was applied to the MC-60E much like the NTF-113C and is shown in Figure 85 through Figure 111. Each calibration was performed on the NASA Langley SVS over the course of about two months. There are only ten calibrations for this system instead of the preferred minimum of twenty for a baseline (Phase I) study. Valuable information can still be derived for this balance. Figure 85 through Figure 90 show the primary sensitivities. Figure 91 through Figure 96 show the confirmation point RMSE. Figure 97 through Figure 102 show the model point RMSE. Figure 103 through Figure 105 show the normal force replicated points. Finally, Figure 106 through Figure 111 show the Phase II analysis for the MC-60E.

4.3.2.1. MC-60E Primary Regression Coefficients

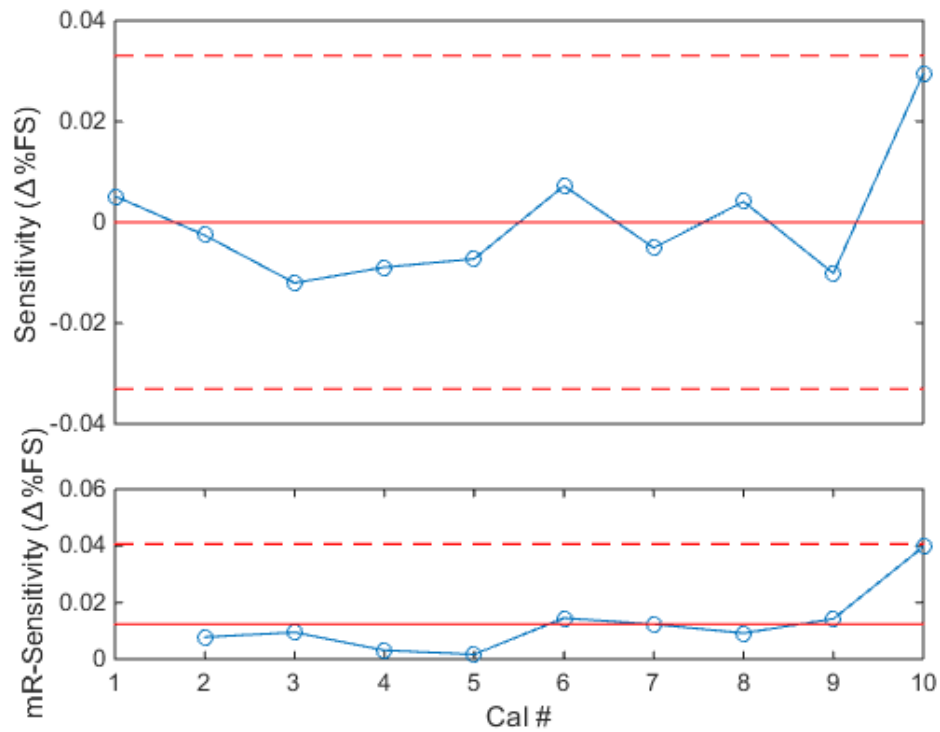


Figure 85: X-mR Control Chart for Normal Force Primary Sensitivity (MC-60)

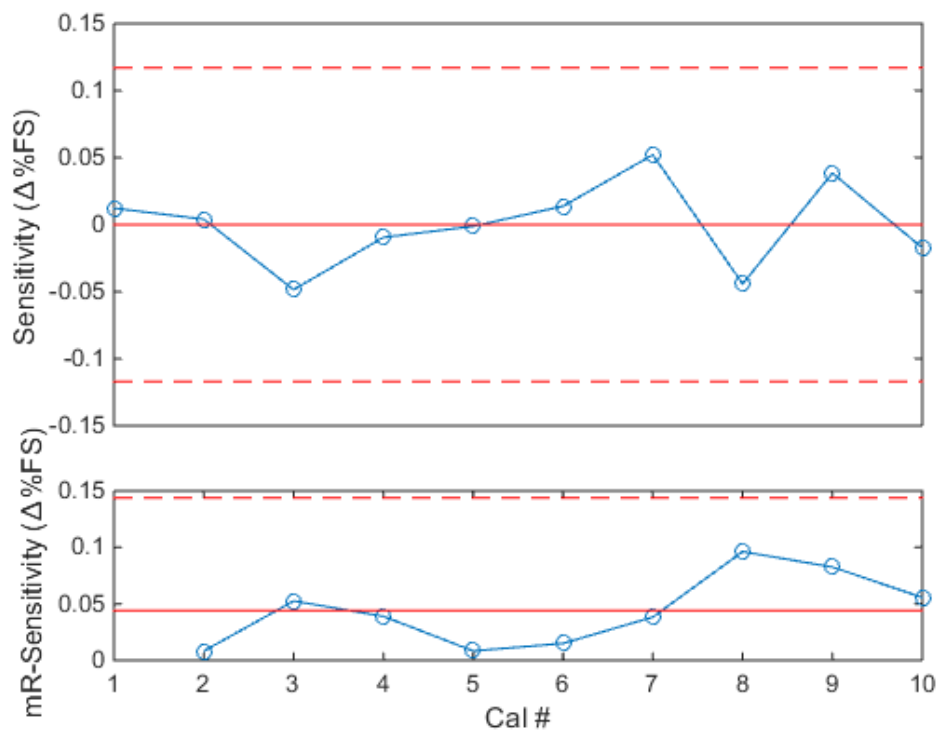


Figure 86: X-mR Control Chart for Axial Force Primary Sensitivity (MC-60)

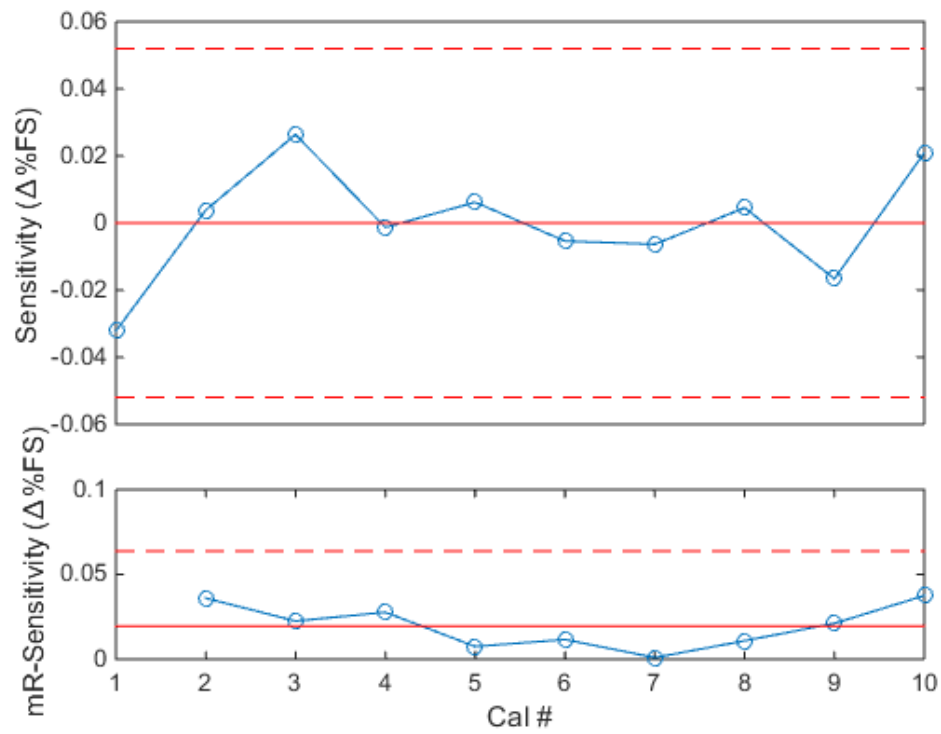


Figure 87: X-mR Control Chart for Pitching Moment Primary Sensitivity (MC-60)

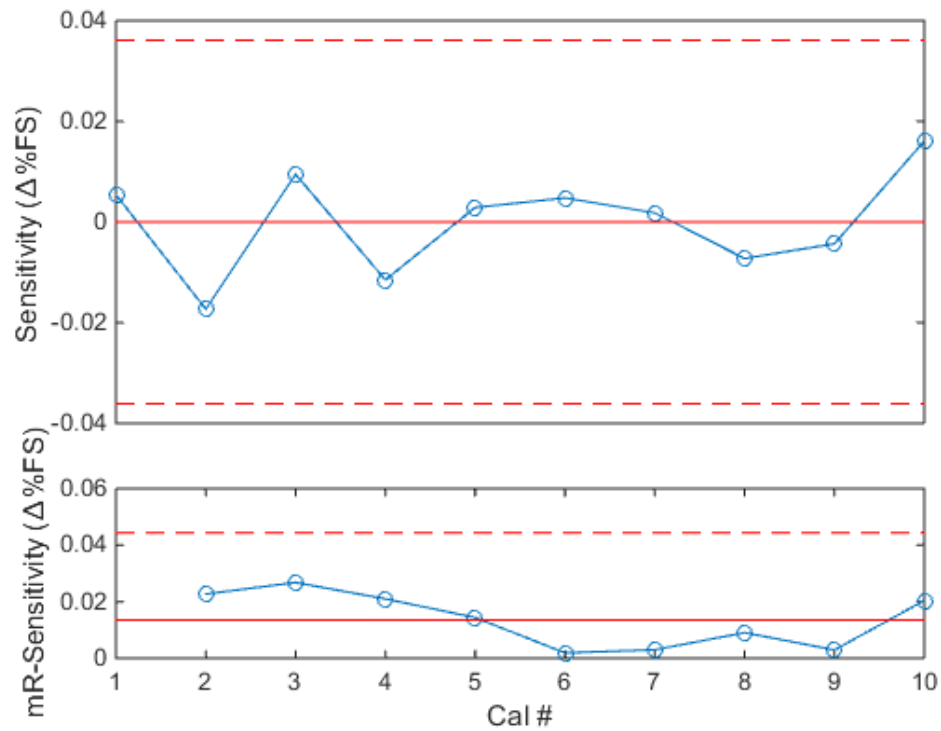


Figure 88: X-mR Control Chart for Rolling Moment Primary Sensitivity (MC-60)

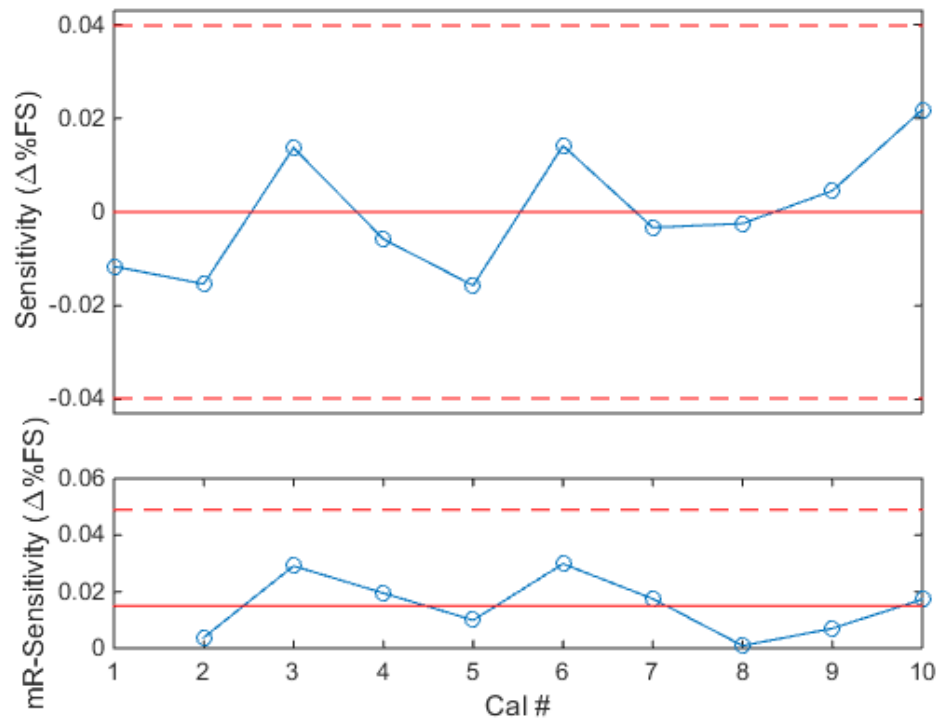


Figure 89: X-mR Control Chart for Yawing Moment Primary Sensitivity (MC-60)

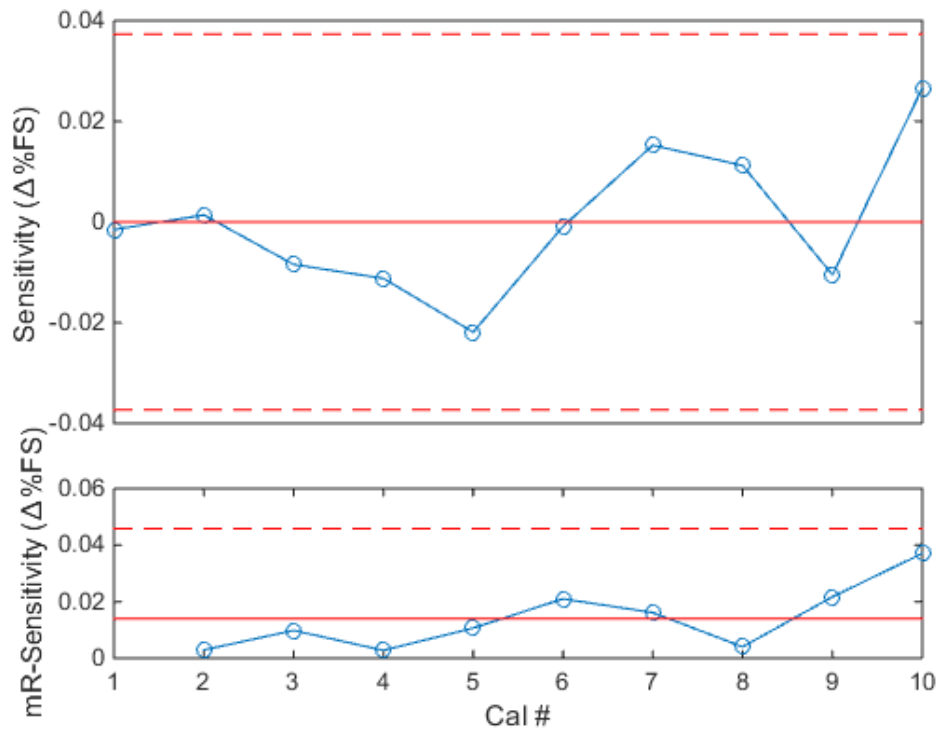


Figure 90: X-mR Control Chart for Side Force Primary Sensitivity (MC-60E)

Figure 85 through Figure 90 show the primary regression coefficients in delta percent full scale. Again, this unit represents the change in primary coefficients from the average of the group. Each coefficient is plotted in an SPC chart where the limits represent three standard deviations of the data within. From the charts, across-calibration variation of the primary coefficients differs less on average than 0.04% FS for all coefficients except for axial force. Axial force shows just slightly more variation than 0.05%. The 3 sigma limits are enlarged to about 0.12% though no data ever reach the limits. The extra variation might be due to the increased sensitivity that the axial section has over all other components. Regardless, the across-calibration repeatability of this balance seems to be on the same order of magnitude as the NTF-113C. Note that with only ten calibrations, the control chart limits have not fully stabilized yet. If extra calibrations were added to the charts, provided that the new calibrations are similar to the points already collected, the limits would likely begin to narrow. This would make the SPC more rigorous and make the limits more representative of the variation in the system. Currently the limits appear to be large compared to the data within the chart. The primary sensitivities appear to vary within acceptable limits when compared to the historical values for balance accuracy found in Table 4.

4.3.2.2. *MC-60E Model and Confirmation Point Error*

The next set of figures show both the model point and confirmation point RMSE. Again, note that the confirmation point error is an indicator of calibration robustness and the model point error is a combination of the lack of fit of the model and pure error. The RMSE values have been normalized into percent full scale so both the overall magnitude of the errors and the change from calibration to calibration can be easily observed.

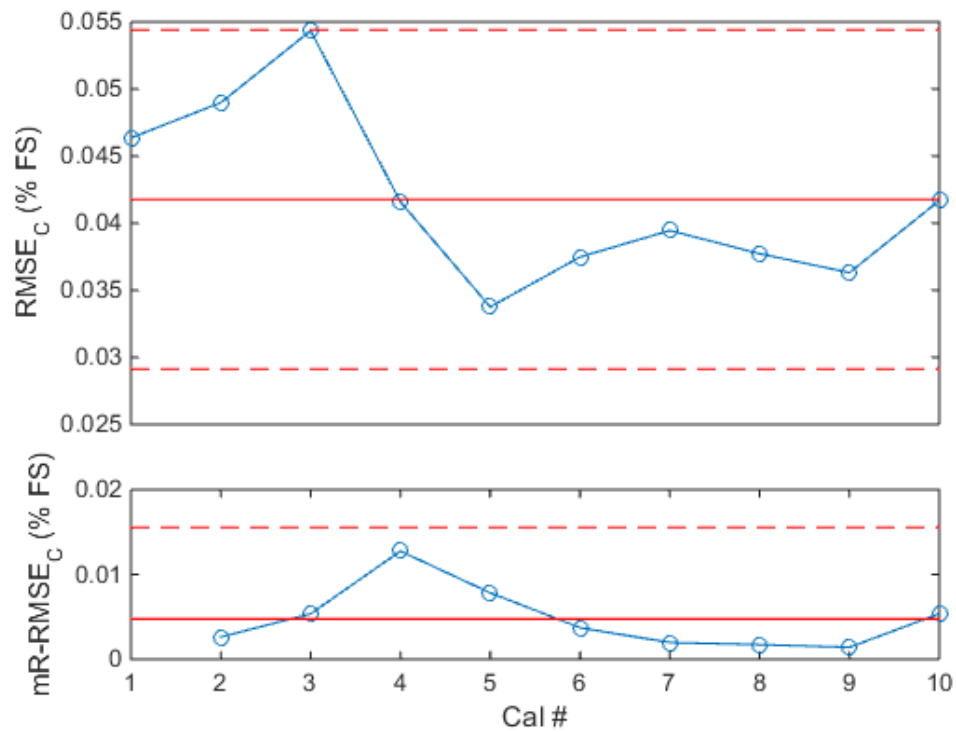


Figure 91: X-mR Control Chart of RMSE on NF Confirmation Point Residuals (MC-60E)

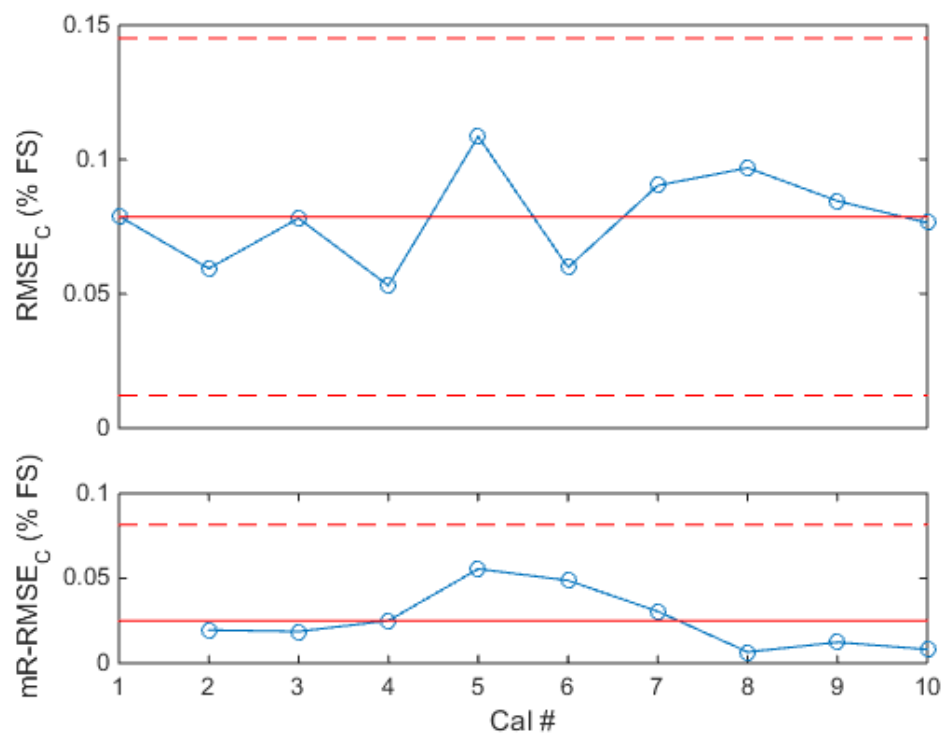


Figure 92: X-mR Control Chart of RMSE on AF Confirmation Point Residuals (MC-60E)

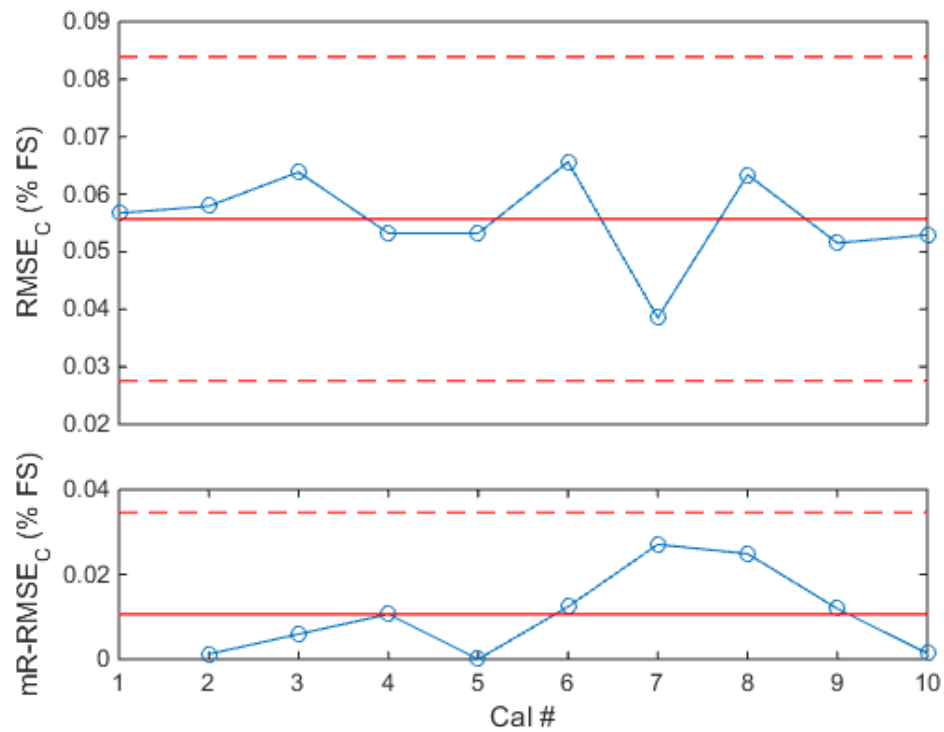


Figure 93: X-mR Control Chart of RMSE on PM Confirmation Point Residuals (MC-60E)

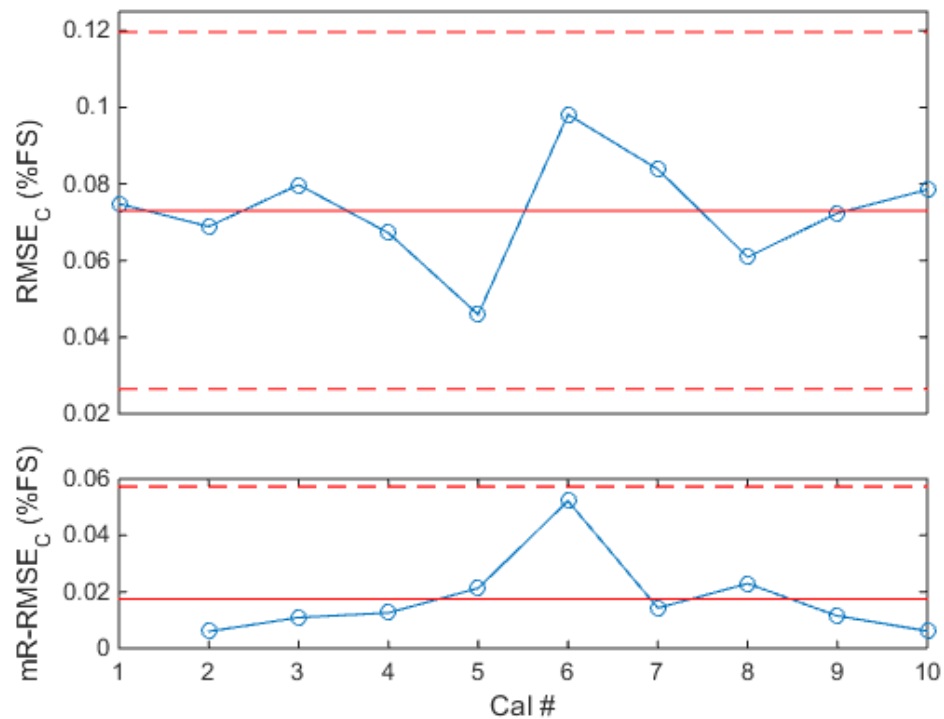


Figure 94: X-mR Control Chart of RMSE on RM Confirmation Point Residuals (MC-60E)

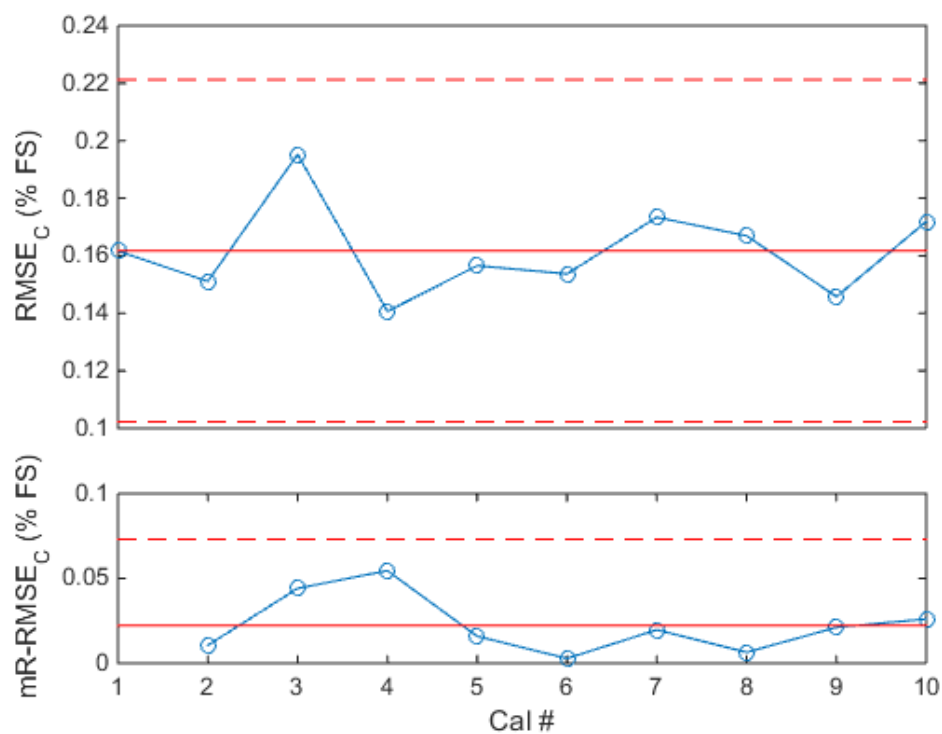


Figure 95: X-mR Control Chart of RMSE on YM Confirmation Point Residuals (MC-60E)

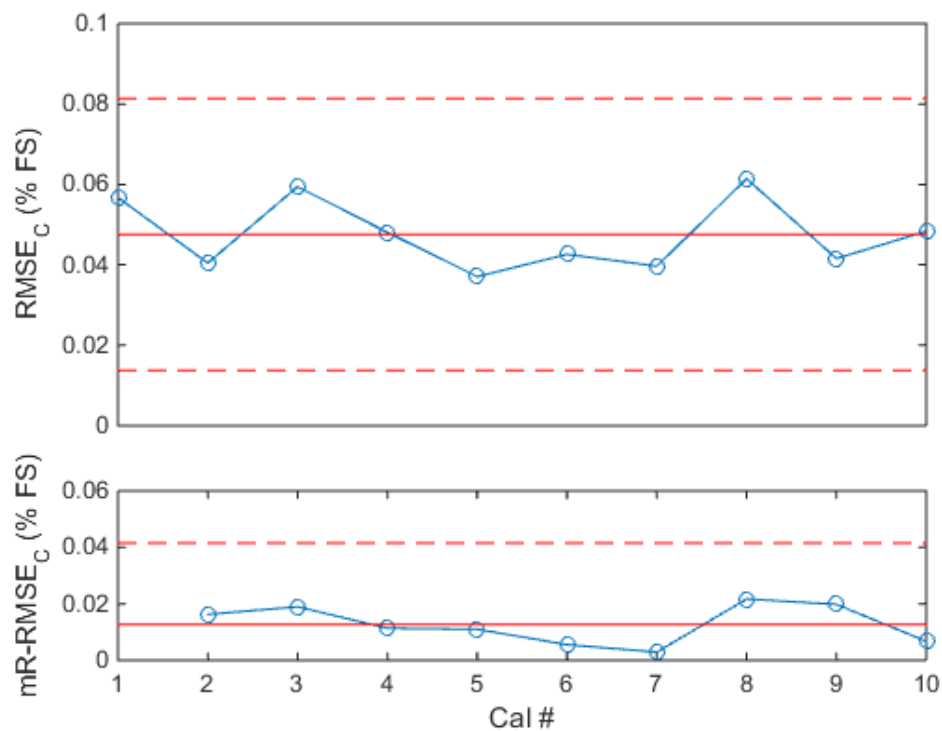


Figure 96: X-mR Control Chart of RMSE on SF Confirmation Point Residuals (MC-60E)

The values for all of the calibrations seem to hover around or just below 0.05% full-scale error. In this case, the axial force and yawing moment appear to have errors that are slightly larger than expected. Yawing moment specifically has errors that extend up to 0.2% full scale. Without more insight into the construction of this balance, it is not entirely clear why yawing moment error is elevated. Further, all of the points remain within the limits. While there are not much data here, initial results are promising. With further calibrations, a more rigorous SPC analysis can be generated. It is interesting to note that almost all $RMSE_C$ values have just slightly exceeded the historical accuracy of the balance. These small excursions above quoted accuracy are not unusual as the confirmation points are not used to fit the model and often have more error associated. The true test is the $RMSE_M$. The $RMSE_M$, is shown for the model points below in Figure 97 through Figure 102. The plots below give a good understanding of how well the model is predicting.

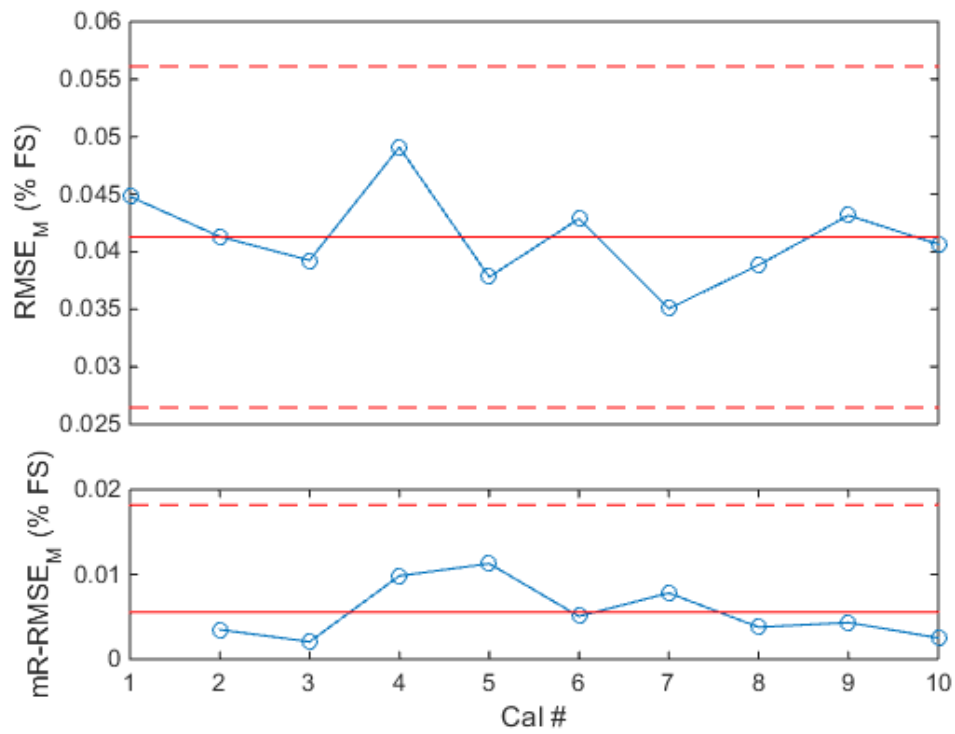


Figure 97: X-mR Control Chart of RMSE on Normal Force Model Point Residuals (MC-60E)

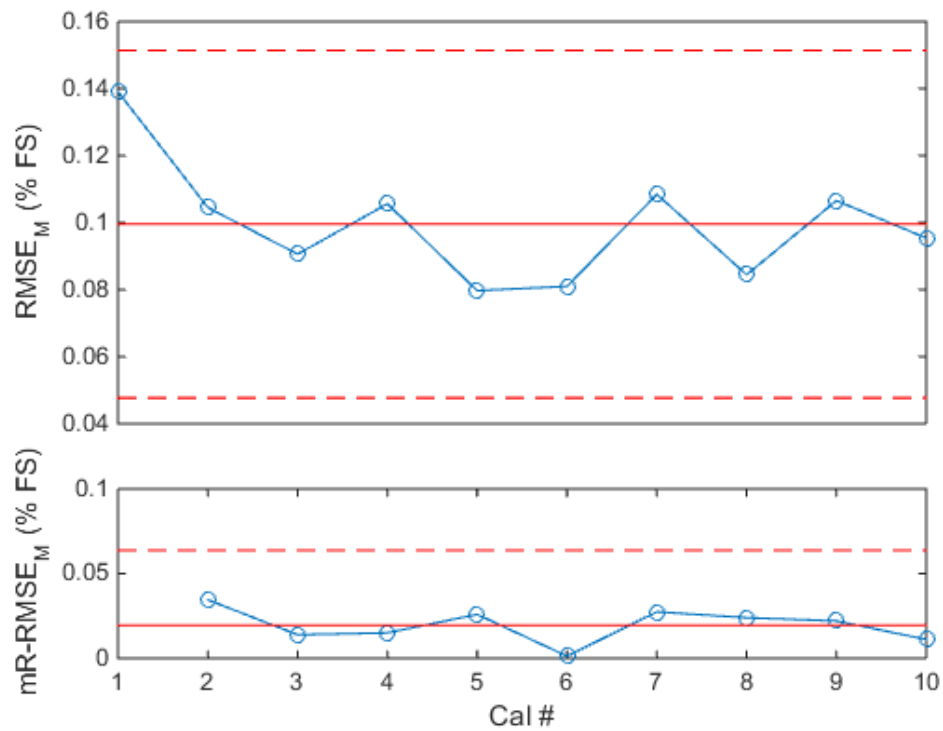


Figure 98: X-mR Control Chart of RMSE on Axial Force Model Point Residuals (MC-60E)

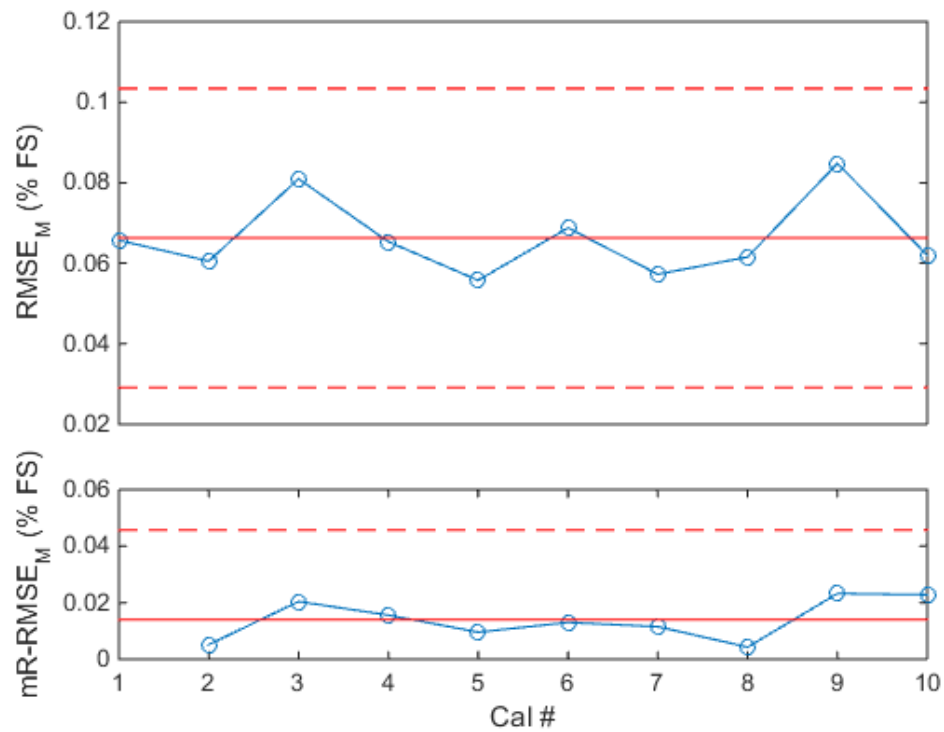


Figure 99: X-mR Control Chart of RMSE on Pitching Moment Model Point Residuals (MC-60E)

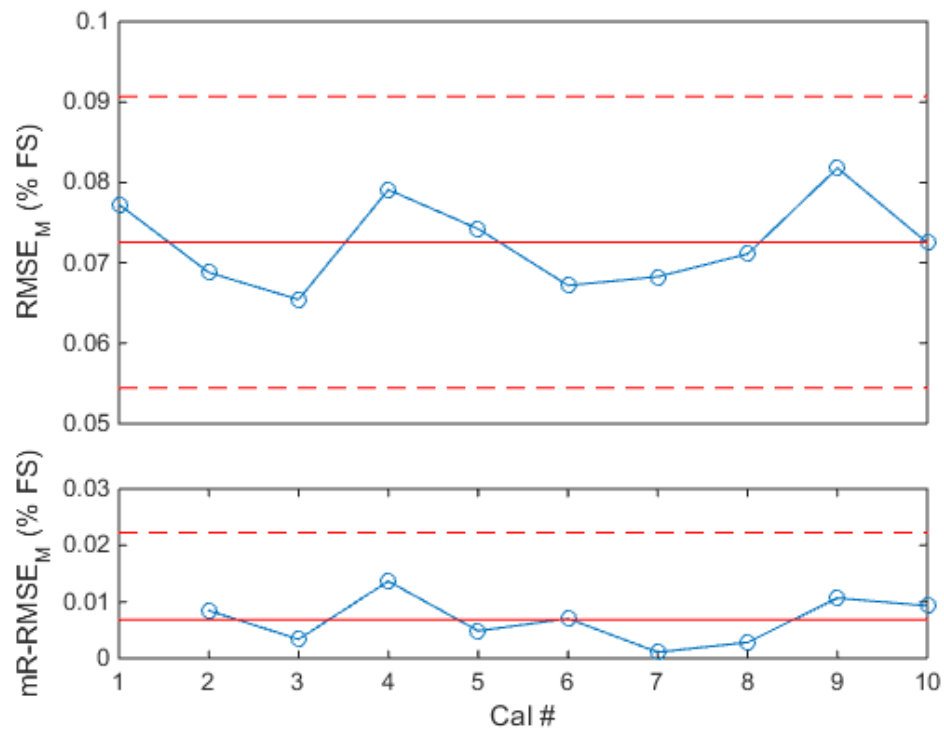


Figure 100: X-mR Control Chart of RMSE on Rolling Moment Model Point Residuals (MC-60E)

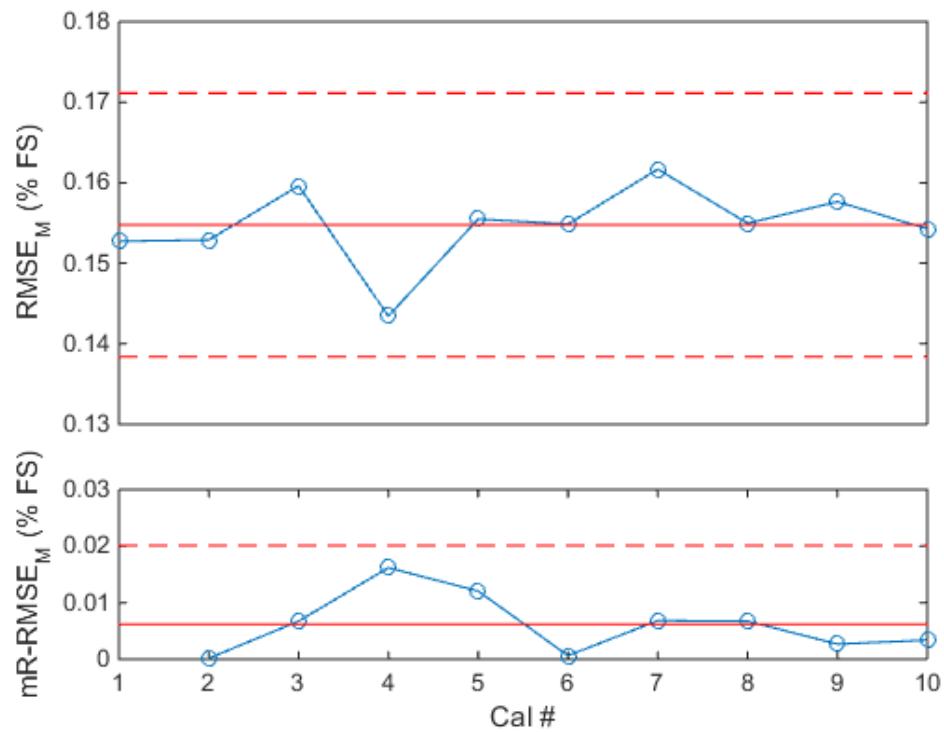


Figure 101: X-mR Control Chart of RMSE on Yawing Moment Model Point Residuals (MC-60E)

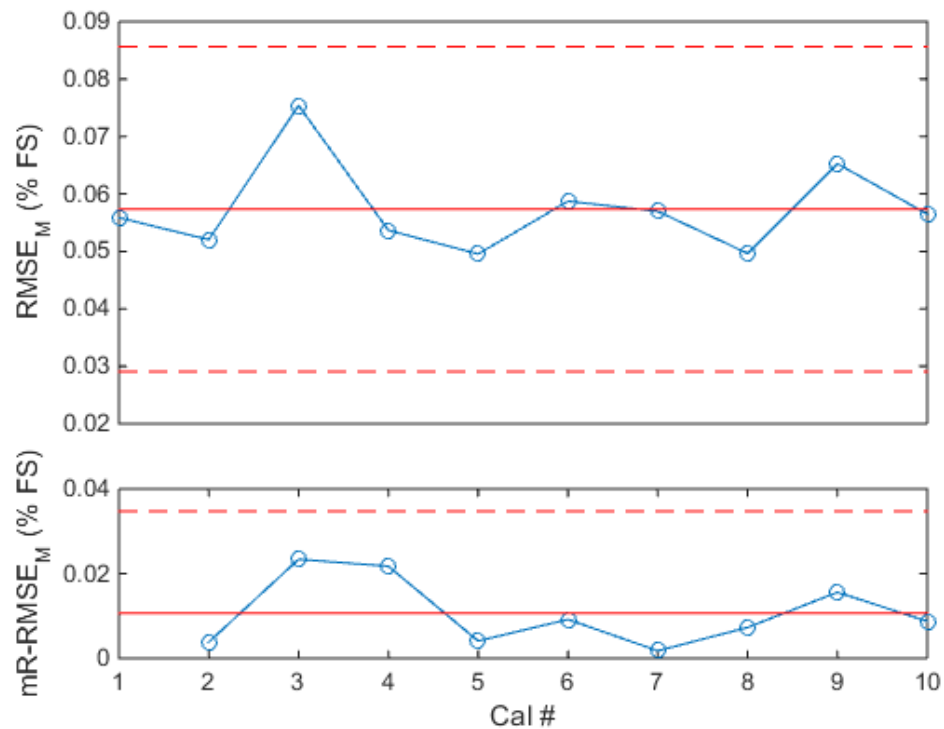


Figure 102: X-mR Control Chart of RMSE on Side Force Model Point Residuals (MC-60E)

From examination of the $RMSE_M$ plots, most of the errors seem to be between 0.05% and 0.08% FS. However, the axial force and yawing moment residuals again look somewhat high. Both component errors average above 0.1%. The larger values can be seen both in the model and confirmation RMSE across ten calibrations. The fact that the RMSE is fairly constant for all of the calibrations means that a systematic error exists that affects the axial and yaw models predictive capability. It could be a balance design problem because yawing moment residuals were not very high for the NTF-113C and the axial residuals can be explained. Without more information on the balance, it is impossible to tell. Additionally, examination of the ANOVA for all MC-60 E calibrations shows that there is no significant interaction between axial and yawing moment. The errors are therefore not correlated with each other in the model. Furthermore all of the points have remained in statistical control on both the $RMSE_C$ and $RMSE_M$ plots. As stated earlier, the limits appear somewhat large at this time. Regardless, all of the $RMSE_M$ values agree with the $RMSE_C$ and are very close to the historically quoted accuracy of the balance. If the values in Table 4 are divided in half, the comparisons can be made.

4.3.2.3. *MC-60E Normal Force Replicates*

The next two plots show the electrical output of the balance for the normal force replicates taken in all calibrations. The I-optimal routine that generated the load schedule was specifically modified to include the replicate points from the beginning. The plots show both the trend in the mean and the range of observations within each calibration. Note that the zeroes have been subtracted from the normal force replicates. The bridges are again separated into N1 and N2 and presented in raw voltages (mV/V). This means there is no model dependency for any of the data presented.

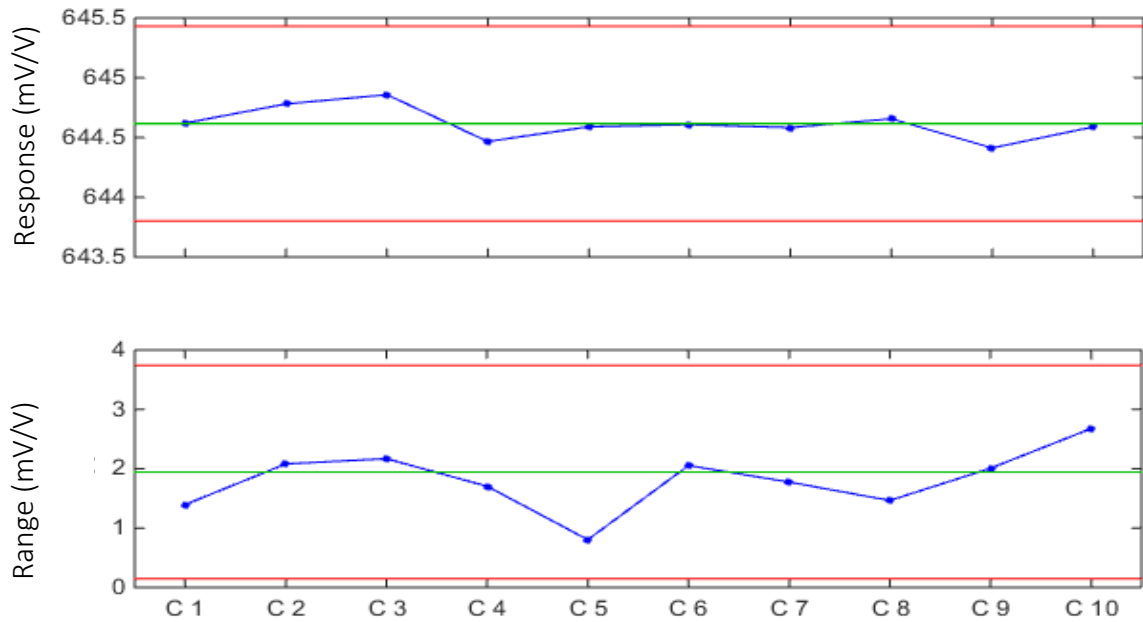


Figure 103: \bar{X} -R Control Chart for N1 Bridge, NF Replicates (MC-60E)

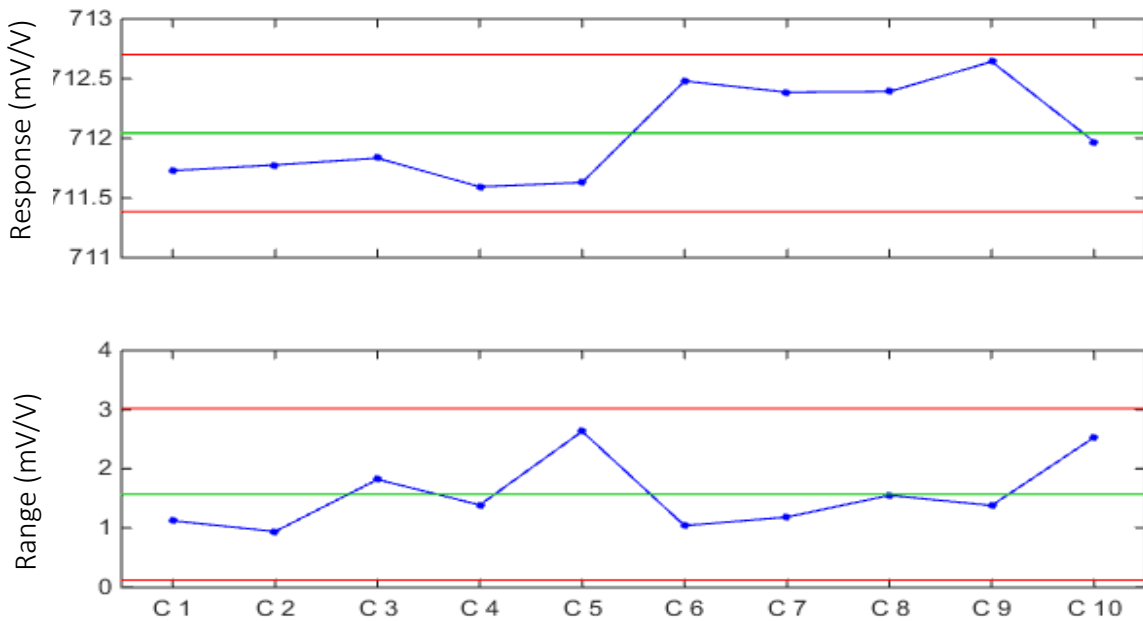


Figure 104: \bar{X} -R Control Chart for N2 Bridge, NF Replicates (MC-60E)

From simple inspection of the SPC limits on the control charts, it appears that the balance is fairly repeatable. However, the size of the SPC limits raises further questions about the stability of the balance. The range of the data within calibration appears to be much larger than the range

between the average replicates across calibrations. The range of data across calibrations also appears to be much larger than seen on the NTF-113C. It would seem that the balance exhibits significant variation in voltage within calibration. However, the averages across calibration appear to be fairly stable in comparison. The large variation in the range would drive the \bar{X} limits to become much larger and reduce the sensitivity of the limits. The root cause for these large variations is unknown, however one possibility may be set point errors. However, it was proven for the NTF-113C that the calibration technician is very good at replicating load conditions. It is unlikely that the quality of calibration has suffered for the MC-60E. It is again useful to understand how much a load prediction would be affected by the shifting normal force replicate voltages.

The replicates can be used again in conjunction with a single calibration matrix to evaluate the effect they have on load prediction. In this case, Calibration 1 was used because no other calibration exists. This data are presented in Table 12. Figure 105 shows these points in percent deviation from the average.

CALIBRATION	NF PREDICTION (LBS)
1	3034.873
2	3036.061
3	3037.551
4	3036.112
5	3036.874
6	3036.117
7	3035.337
8	3034.071
9	3032.737
10	3027.772
Max % Difference	0.322
STANDARD DEVIATION	2.81592

Table 12: NF Load Prediction from Normal Force Replicates (MC-60E)

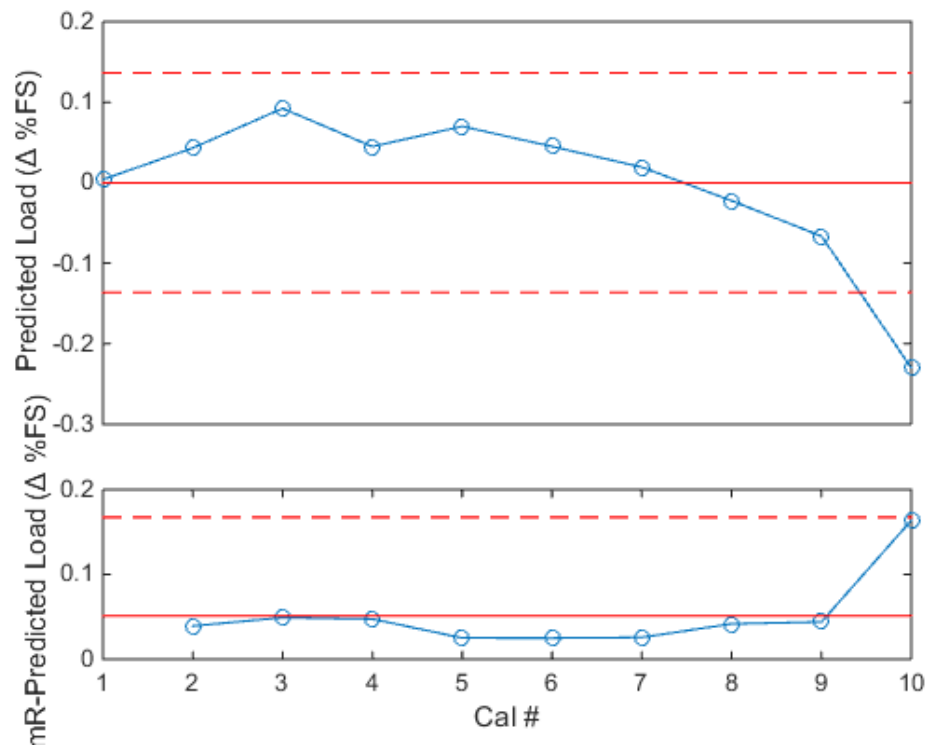


Figure 105: Change in Predicted Load from Average NF Replicated Points (MC-60E)

From the replicates, it can be seen that the MC-60E shows much more variability in the across calibration load prediction. Once again, only the average voltages are taken from each calibration. The zero load voltages and tare loads are taken from the single calibration (Calibration 1). The average voltages appeared to drift about 1 mV/V between all of the calibrations. Once converted to load using calibration 1, the predicted load standard deviation is 0.056% or almost 3 lbs. In comparison to the NTF-113C, the MC-60E has greater variability in repeated measurements. Replication is directly related to pure error which sets the lower limit for the uncertainty of any instrument. This result does not match well with the quoted accuracy of the balance found in Table 4. However, once again this analysis represents a worst case scenario for the repeatability. A single calibration was used to tare the data and reduce the data for comparison. While this method proved useful on the NTF-113C, it appears to have not worked as well here.

Additionally, with only ten calibrations, the repeatability of this balance cannot truly be assessed. It does appear that the across-calibration repeatability is worse than the within-calibration repeatability.

4.2.3.4. MC-60E Phase II Analysis

Similarly, a Phase II SPC analysis can be developed for the MC-60E even with the limited number of replicate calibrations. Again, the first five calibrations will serve as the Phase I portion and the last five calibrations will serve as the phase II portion. Figure 106 through Figure 111 show the primary sensitivities for the MC-60E, split into Phase I and Phase II analyses. Again, note that the Phase II data are not used to calculate the control limits.

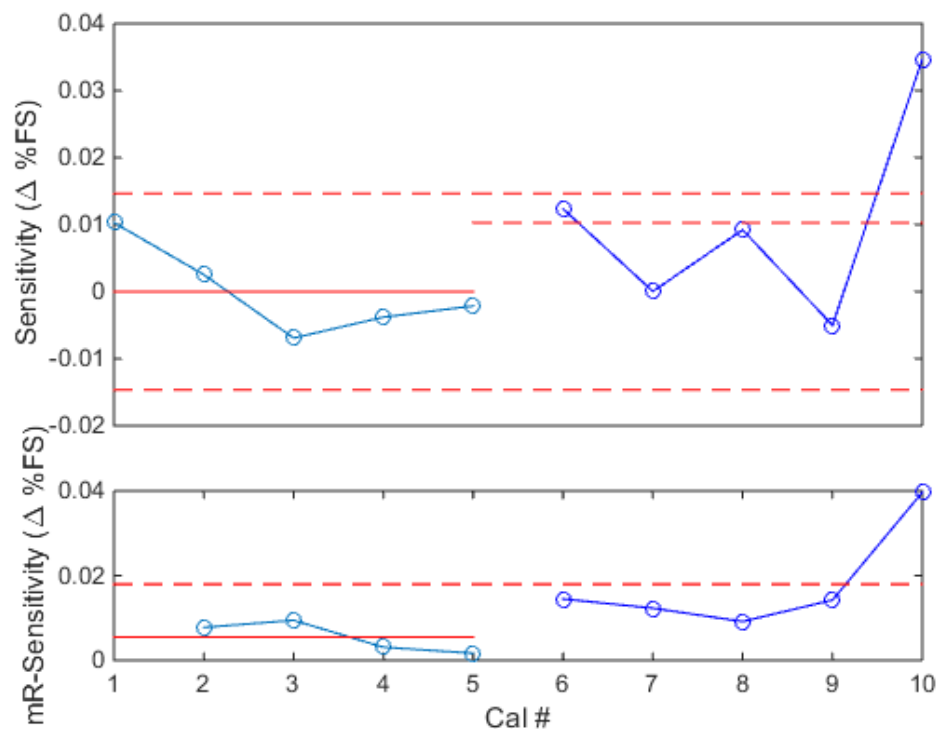


Figure 106: Phase II Control Chart for NF Primary Sensitivity (MC-60E)

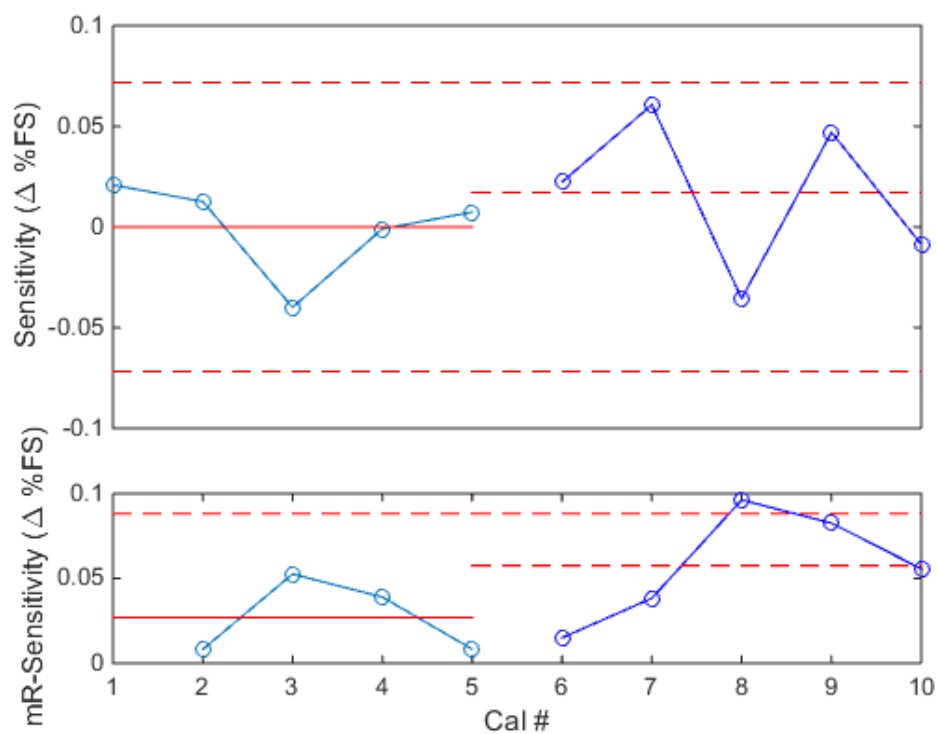


Figure 107: Phase II Control Chart for AF Primary Sensitivity (MC-60E)

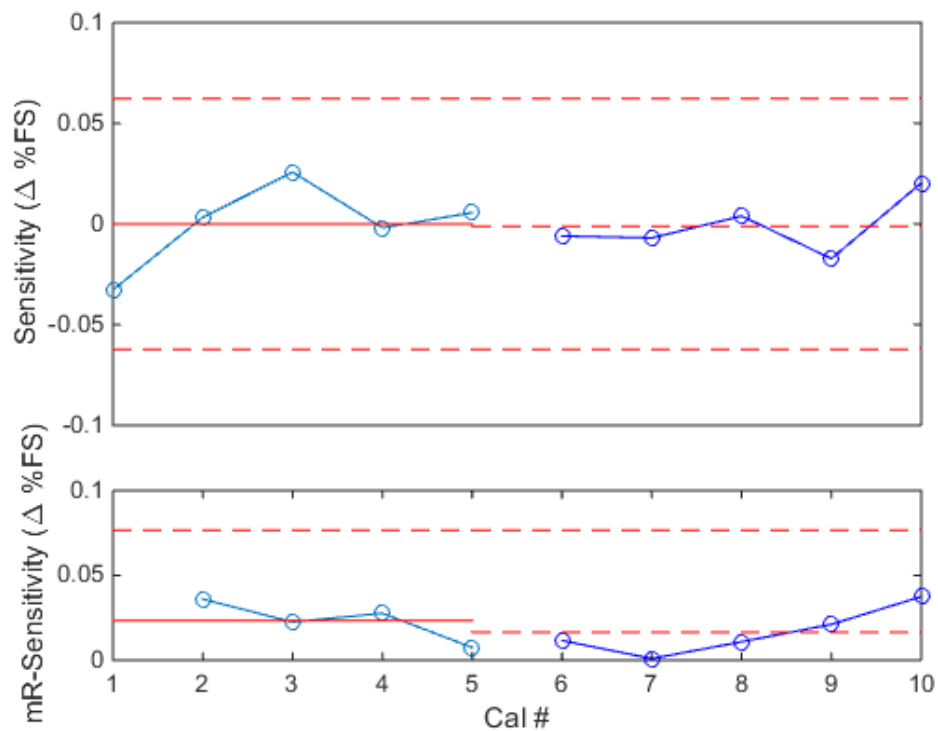


Figure 108: Phase II Control Chart for PM Primary Sensitivity (MC-60E)

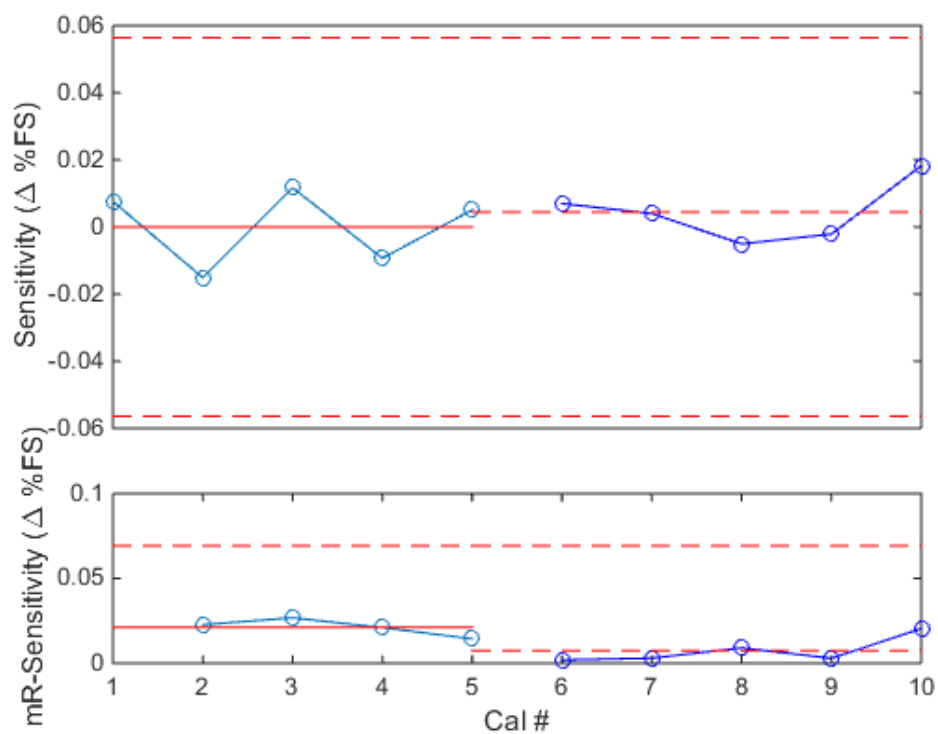


Figure 109: Phase II Control Chart for RM Primary Sensitivity (MC-60E)

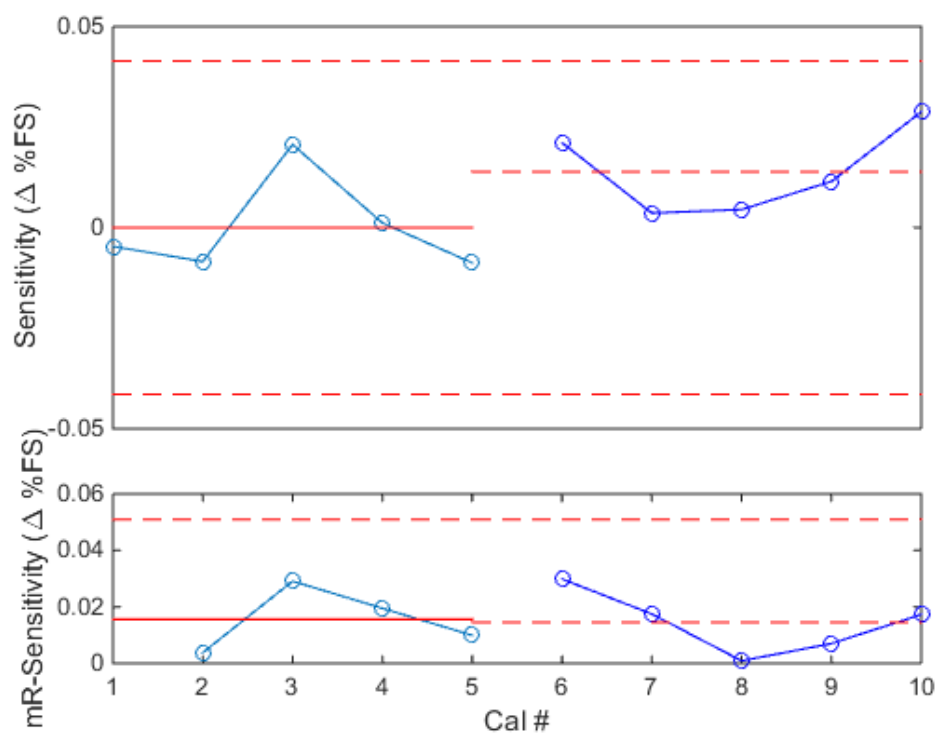


Figure 110: Phase II Control Chart for YM Primary Sensitivity (MC-60E)

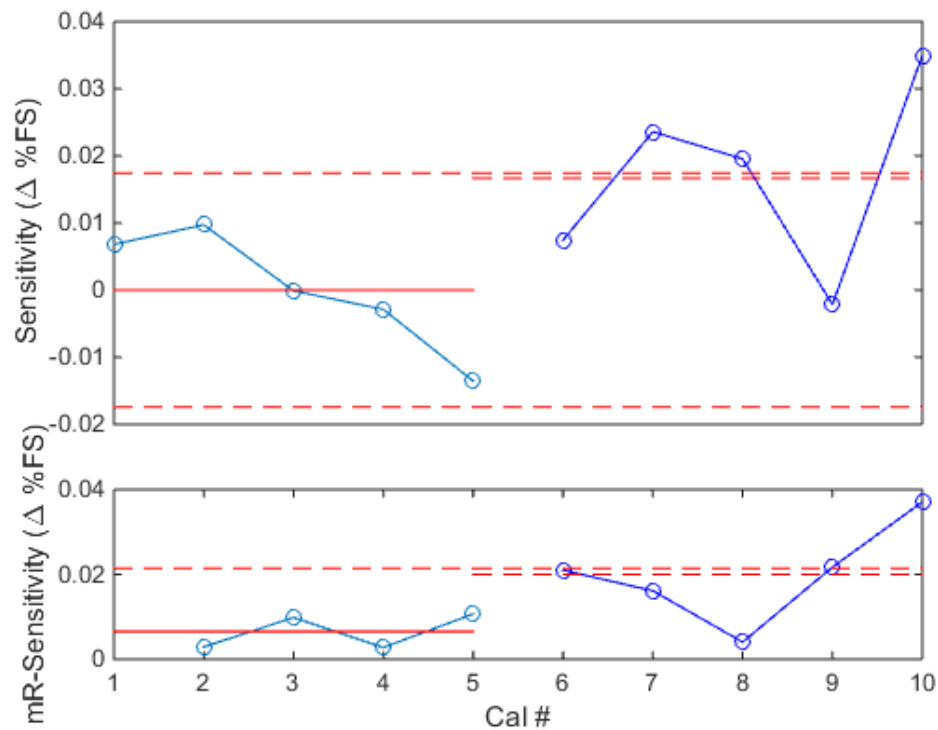


Figure 111: Phase II Control Chart for SF Primary Sensitivity (MC-60E)

The Phase II analysis for the MC-60E looks favorable for most components of the balance. However, NF, AF and SF show points which exceed limits on either the X or mR chart. It is difficult to make definitive conclusions with these charts because of the small sample size. However, it would seem that the MC-60E is showing less agreement between calibrations using Phase II analysis when compared to the NTF-113C (excluding axial and roll issues). Lack of data for this analysis makes conclusions difficult.

4.4. Historical Summary of Balance Calibrations

Table 13, below, shows a listing of the data that were analyzed in this thesis. It details how many calibrations are available for each balance on each system. Table 14 shows a more in-depth deconstruction of each of the calibrations performed on each of the balances. In this table, extra

information such as the type of load schedule used, the load limits used for that load schedule, the number of design points and the number of confirmation points per calibration can be found.

BALANCE	SVS	MANUAL STAND	ABCS
NTF-113C	21	5	1
NTF-118B	2	3	0
MC-60E	12	0	1
MK-29B	3	3	1

Table 13: Summary Table of Calibrations for Each Balance on Each System

Balance	Type	Max Load limits								
		NF	AF	PM	RM	YM	SF			
NTF-113C	Single-piece	6500	400	13000	9000	6500	4000			
Date	Cal Type	Cal Load Limits						Design	P#	C#
13.01.30	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
13.03.05	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
13.04.09	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
14.01.09	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
14.08.25	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
14.09.10	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
14.10.08	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
14.10.20	SVS	3000	400	5250	3600	6000	3000	CCD	64	20
14.11.04	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
14.11.11	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
14.11.19	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
14.12.02	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
14.12.10	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
14.12.17	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
15.01.05	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
15.01.13	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
15.03.02	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
15.03.26	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
15.04.09	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
15.04.28	SVS	3000	400	5250	3600	6000	3000	CCD	67	22
98.10.28	Manual	6520	400	12800	8050	6400	4000	OFAT	728	
04.07.22	Manual	6520	400	12800	8050	6400	4000	OFAT	729	
02.01.03	Manual	6520	400	12800	8050	6400	4000	OFAT	729	
02.10.08	Manual	6520	400	12800	8050	6400	4000	OFAT	729	
08.05.09	Manual	6520	400	12800	8050	6400	4000	OFAT	729	
08.07.01	ABCS	6500	400	13000	9000	6500	4000	OFAT	1070	
Balance	Type	Max Load limits								
		NF	AF	PM	RM	YM	SF			
MC-60E	Unitized	5000	700	20000	5000	10000	2500			
Date	Cal Type	Cal Load Limits						Design	P#	C#
15.06.05	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30
15.06.10	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30
15.06.14	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30
15.06.19	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30
15.06.24	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30

Table 14: In-depth Table of Balance History Included in this Thesis

Table 14 (Continued)

15.07.16	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30			
15.07.22	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30			
15.07.28	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30			
15.08.01	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30			
15.08.07	SVS	3000	520	6560	5000	4100	2120	I-optimal	75	30			
13.11.06	SVS	2600	500	7800	4342	6560	1600	CCD	64	20			
13.11.26	SVS	2600	500	7800	4342	6560	1600	CCD	64	20			
09.01.30	ABCS	5000	700	20000	5000	10000	2500	OFAT	1923				
Balance	Type	Max Load limits											
		NF	AF	PM	RM	YM	SF						
Mk-29B	Multi-piece	4200	350	15225	3800	4200	1400						
Date	Cal Type	Cal Load Limits									Design	P#	C#
09.04.16	SVS	2100	350	7700	3500	2100	700				CCD	64	20
09.05.04	SVS	2100	350	7700	3500	2100	700	CCD	64	20			
09.05.07	SVS	2100	350	7700	3500	2100	700	CCD	64	20			
09.07.22	Manual	1600	350	5000	2500	1800	600	OFAT	729				
10.03.02	Manual	1600	350	5000	2500	1800	600	OFAT	729				
15.10.01	Manual	1600	350	5000	2500	1800	600	MBBD	64				
08.07.23	ABCS	3600	350	13050	3800	3975	1325	OFAT	954				
Balance	Type	Max Load limits											
		NF	AF	PM	RM	YM	SF						
NTF-118B	Single-Piece	6500	700	13000	9000	6500	4000						
Date	Cal Type	Cal Load Limits									Design	P#	C#
15.09.08	SVS	2850	400	5000	3420	5700	3850				CCD	64	20
15.09.15	SVS	2850	400	5000	3420	5700	3850	CCD	64	20			
15.12.00	Manual	5500	600	11000	8050	5000	3400	MBBD	64	64			
08.11.04	Manual	6520	700	12800	8150	6400	4000	OFAT	729				
09.12.17	Manual	6520	700	12800	8150	6400	4000	OFAT	729				

Table 14: In-depth Table of Balance History Included in this Thesis

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1. Conclusions

Many examples of wind-tunnel balance calibrations have been shown and comparisons have been made both across systems and for long-term repeated balance calibrations. The conclusions drawn from this study are founded in rigorous statistical analysis and provide valuable insight into the current state of balance calibrations in the ground test community.

From examining the estimated primary sensitivities from each balance on each calibration system, it has been shown that the differences between the systems is measurable. Each calibration system operates on different methodologies and despite efforts to remove bias, there are still measurable differences. While producing something as simple as a linear first order coefficient should be easy, the unfortunate truth is each system's individual complexities make each experiment biased in different ways. The previous study had shown the same problems, and this study sought to make comparisons using more consistent load schedules. In this case, the M-BBD load schedule for the Manual Stand was designed to be comparable to the SVS CCD and SVS I-Optimal design in terms of run number, design metrics, and power. Cross-system comparisons between these load schedules showed more favorable agreement on the NTF-118B. The MK-29B also was calibrated using the M-BBD load schedule and fared no worse in agreement when compared to the other two systems or load schedule design. This lends confidence in the ability of the M-BBD design to efficiently characterize a balance. It was previously untested on a large capacity balance. The best overall comparisons between calibration systems and load schedules are found for the single-piece balances. In most cases, damage aside, primary sensitivity replication within $\pm 0.2\%$ was possible. The MC-60E showed favorable comparisons between a

current SVS calibration and the ABCS calibration. However, two historical calibrations showed fairly poor agreement with the aforementioned data. The MK-29B showed the worst agreement of the balances tested. It is still unclear if this balance has mechanical issues which may explain these poor results. A long-term replicate study on the MK-29B might expose the mechanical issues if they exist.

From the data collected, it would appear that the system which is least repeatable is the Manual Stand. Often the CI's for calibration coefficients which were generated using the Manual Stand are the largest. There is a large amount of hardware and time involved with a Manual Stand calibration, and while it is the most widely used and understood system, it would seem that a comparatively larger amount of variability exists for experiments performed on the Manual Stand. To assess differences between systems, a single system must agree with itself. A better understanding of the Manual Stand calibration and the variability is required for a more comprehensive comparison. Gathering a few extra calibrations will be necessary as the enlarged CI's may partially be due to a lack of sample size. Gathering more calibrations will lend more insight into the repeatability of the system in general. With the new M-BBD design, this data should be more readily achievable as the design is much more efficient in time and cost. Further work must also be aimed at assessing the repeatability of the ABCS. With only one ABCS calibration per balance, the variability cannot be assessed. Once an understanding of the repeatability of each system has been established, the true differences might be more easily understood. The primary source of weakness in cross-system comparisons has been the lack of consistent sample size for each system. The limited data obtained to date has allowed tentative comparisons between these systems to be made. The comparisons thus far have shown a best case agreement around $\pm 0.15\%$ between primary coefficients. However, to make any more definitive

comparisons and begin to understand the true differences between these systems with the goal of making recommendations to improve agreement between these systems, determining the source of variability in the Manual Stand and assessing the variability of the ABCS is imperative. This will require a full understanding of both system's entire process including setup, calibration and data analysis. At NASA LaRC, the Manual Stand is well understood and all procedures are well documented, however reviewing them might bring forward new insight. Admittedly, the ABCS is the least understood system of all. Future work should be aimed at further understanding these two systems.

This long-term calibration study features more repeated calibrations on a single balance and a single system than previously published. The goal of the process was to use statistical process control techniques to establish a baseline for the repeatability of a balance and calibration system. With this further understanding of the long-term capabilities of a balance, balance engineers and users potentially have more tools to assess the quality of the calibrations over a long term. This SPC baseline can also help assess the health of a balance over long term. Potentially, with the data shown, a balance user could be more confident in the long term stability of the regression model. After establishing a baseline, the frequency of calibrations required per balance is then reduced which will save researchers money.

From the replicate calibrations on the NTF-113C and the MC-60E, it can be concluded that both the balance and calibration process are in relative statistical control. The calibration process used with the SVS sufficiently eliminates or accounts for most extraneous factors that can affect the balance during calibration. The model derived can consistently replicate sensitivities around 0.05%. Further analysis showed that RMSE on most components appeared near or within acceptable limits for all replicate calibrations on either system. The rising RMSE in the model

point residuals of the NTF-113C was detected prior to completing 20 calibrations. A choice was made to continue with the repeated calibrations to see if the problem persisted. This many calibrations on a single balance had never been witnessed and there was little reason to expect damage to the balance. The plots of RMSE did directly lead to the choice to have the balance inspected upon completion of the 20 calibrations. It was shown for both balances that the SVS process is very good at replicating a load condition, at least in normal force. Both charts for both electrical bridges showed acceptable statistical control for the normal force replicates. The MC-60E appeared to vary much more than expected, both within calibration and across calibration. This led to replication error of above .05% FS. Finally, the full SPC process was exercised by comparing the first five calibrations (phase I) against the rest of the calibrations (phase II) for both balances. It is not recommended to compute limits using such a small sample size, but nevertheless good agreement was seen for most components, even with only five calibrations.

SPC has been shown to be a valuable tool in assessing the repeatability of balances. The balances which were repeatedly calibrated appear to be fairly stable. The calibration system and process also appear to be fairly stable and capable of reproducing similar results every time. Real issues with a calibration were found in the case of Calibration 18 on the NTF-113C. Calibration 18 was affected by a multimeter failure. The control charts made the issue simple to detect. The magnitude of the departure and the pattern of residuals also quickly indicated that the issue was related to voltage. The problem was quickly found and remedied. Additionally, the long term behavior of the $RMSE_M$ plot for axial force, led to the decision to have the NTF-113C inspected for damage. The discovery of a crack at the base of the axial flex beams answered many questions that arose over the course of the SPC process. The axial force primary sensitivities were not greatly

affected by this crack and only through careful examination of the RMSE was the crack found. A future wind-tunnel test may have suffered a catastrophic failure if this balance was used.

Statistical tools have been shown to work to assess the long-term repeatability of a wind-tunnel balance. Statistical tools have also been used in the LWTE and the balance calibration laboratory for over a decade to increase the quality of the data acquired. Further research continues to prove the usefulness of these methods and further usage of these methods continues to yield substantial advancements in the ground-test capabilities at NASA Langley Research Center.

5.2. Future Work

Future work with the balance calibration study will focus on strategically performed calibrations that will lend the most insight into balance repeatability. Calibrations will be gathered that will fill in the gaps where there are no calibrations of a balance on a system. There is the suspicion that the MK-29E is not a stable balance from the data taken from the original balance calibration study. However, multiple calibrations on the order that was presented in this paper have never been completed on a task balance. Replicating 10 or 20 calibrations would lend further insight into the true nature of the Task balance. Insight may be drawn from calibrations of a second Task balance as well to rule out possible mechanical failures of the first. This would also allow further comparisons to the NTF and Triumph balance.

Additionally, further research will be directed at the M-BBD that was developed for this study. Both the NTF-118B and the Task MK-29E were calibrated using the M-BBD. The design appeared to work very well for the NTF balance, however questions remain regarding the MK-29E. A better comparison of the design could be made with a calibration of the MC-60E. Further information as to the health of the MK-29 can also be derived from this comparison. Furthermore using the M-BBD may allow further insight into use of the Manual Stand. This document has

shown that the Manual Stand shows more variability in regression coefficient estimates than the SVS. This may partially be a function of having fewer Manual Stand calibrations available. More data is required to assess the Manual Stand in general.

Increasing the sample size of calibrations on the ABCS is also imperative for assessing its actual behavior as compared to the other systems. The SVS has undergone rigorous scrutiny to compare its results with the accepted standard calibration method (Manual Stand). Only limited analyses and comparisons have ever been made with the ABCS and Manual Stand. There is some belief that the ABCS may be biased when compared to the gravity loaded systems. By gathering additional data, the differences could be much more easily quantified. Additionally, the ABCS can be made to replicate any load schedule. While value is added from understanding the ABCS using its own load schedule, by replicating a true manual schedule or SVS schedule, a truly consistent comparison can be made between the systems.

Finally, it is useful to note that while this study has been more comprehensive than any previous endeavor, assessing the true state of balance calibration services today will require more input from wind tunnel users. A good next step may be to bring more users into the study and gather user input on a continuous basis. With more users, a more substantial discussion can be had about the results thus far and how to proceed into the future.

REFERENCES

- [1] Rhew, Ray D. "Wind Tunnel Force Balance Calibration Study - Interim Results." *8th International Symposium on Strain-Gauge Balances*.
Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120008941.pdf>>.
- [2] Conversation with Dr. Peter Parker, NASA LaRC. Expert opinion in balance design.
- [3] Bennett, Henry. "Assessing Long-Term Wind Tunnel Balance Performance and Uncertainty from Multiple Calibrations - Revised." *51st AIAA Aerospace Sciences Meeting* (2013). AIAA. doi: 10.2514/6.2013-546
- [4] Belter, Dale, and David Holler. "Assessing Long-term Instrumentation Performance and Uncertainty from Multiple Calibrations." *21st Aerodynamic Measurement Technology and Ground Testing Conference* (2000). AIAA. doi: 10.2514/6.2000-2202
- [5] Parker, P., M. Morton, N. Draper, and W. Line. "A Single-vector Force Calibration Method Featuring the Modern Design of Experiments." *39th Aerospace Sciences Meeting and Exhibit* (2001). AIAA. doi: 10.2514/6.2001-170
Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010011040.pdf>>
- [6] Parker, Peter A., and Tom D. Finley. "Advancements in Aircraft Model Force and Attitude Instrumentation by Integrating Statistical Methods." *Journal of Aircraft* 44.2 (2007): 436-43. AIAA. doi: 10.2514/1.23060
- [7] Wheeler, Donald J., and Chambers, David S. *Understanding Statistical Process Control*. 3rd ed. SPC, 2010.
- [8] Hemsch, Michael. "Development and Status of Data Quality Assurance Program at NASA Langley Research Center - Toward National Standards." *Advanced Measurement and Ground Testing Conference* (1996). AIAA. doi: 10.2514/6.1996-2214
Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19960028785.pdf>>
- [9] M. Hemsch, J. Grubb, W. Krieger, and D. Cler. "Langley Wind Tunnel Data Quality Assurance - Check Standards Results." *21st AIAA Advanced Measurement Technology and Ground Testing Conference* (2000). AIAA. doi: 10.2514/6.2000-2201
Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000099747.pdf>>
- [10] Hemsch, Michael, David Tuttle, Heather Houlden, and A. Graham. "Measurement of Force Balance Repeatability and Reproducibility in the NTF." *42nd AIAA Aerospace Sciences Meeting and Exhibit* (2004). AIAA. doi: 10.2514/6.2004-771

- [11] Hemsch, Michael J., and Heather P. Houlden. "Repeatability Modeling for Wind-Tunnel Measurements: Results for Three Langley Facilities." *52nd Aerospace Sciences Meeting* (2014). AIAA. doi: 10.2514/6.2014-0096
- [12] Montgomery, Douglas C. *Design and Analysis of Experiments*. 6th ed. John Wiley & Sons, 2005.
- [13] Deloach, R. "Improved Quality in Aerospace Testing through the Modern Design of Experiments." *38th Aerospace Sciences Meeting and Exhibit* (2000). AIAA. doi: 10.2514/6.2000-825
Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000025557.pdf>>
- [14] Deloach, Richard, and Norbert M. Ulbrich. "A Statistical Theory of Bidirectionality." *AIAA Ground Testing Conference* (2013). AIAA. doi: 10.2514/6.2013-2995
Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20140000503.pdf>>
- [15] Kammeyer, Mark, and Norbert M. Ulbrich. "The Use of Absolute-Value Terms in Regression Modeling of Multi-Piece Force Balances (Invited)." *31st AIAA Aerodynamic Measurement Technology and Ground Testing Conference* (2015). AIAA. doi: 10.2514/6.2015-3382
- [16] Myers, Raymond H., and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: Wiley, 1995.
- [17] Ferris, A.T. Judy. "Strain Gauge Balance Calibration And Data Reduction at NASA Langley Research Center." *First International Symposium on Strain-Gauge Balances* (1996). Ntrs.NASA.gov. NASA
<<http://ntrs.nasa.gov/search.jsp?R=19990072775>>
- [18] Simpson, James, Drew Landman, Rupert Giroux, Michelle Zeisset, Brian Hall, and Ray Rhew. "Calibrating Large Capacity Aerodynamic Force Balance Instrumentation Using Response Surface Methods." *2005 U.S. Air Force T&E Days* (2005). AIAA. doi: 10.2514/6.2005-7601

Additional References:

- Rhew, Ray D. "NASA LaRC Strain Gauge Balance Design Concepts." *First International Symposium on Strain Gauge Balances* (1999): 525-41.
Ntrs.NASA.gov. NASA.
<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19990072764.pdf>>.

Parker, Peter, and Tianshu Liu. "Uncertainty Analysis of the Single Vector Force Balance Calibration System." *AIAA Aerodynamic Measurement Technology and Ground Testing Conference* (2002). AIAA. doi: 10.2514/6.2002-2792

Ntrs.NASA.gov. NASA

<<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20030000847.pdf>>

APPENDICES

A. First Order Interaction Coefficients NTF-113C

The following plots show the first order coefficients for all six components of force for the NTF-113C. All values are in percent full scale, where the primary coefficient would be 100%. The primary coefficients are not shown here as they were previously shown in the NTF-113C primary sensitivity section.

a. Normal Force Model

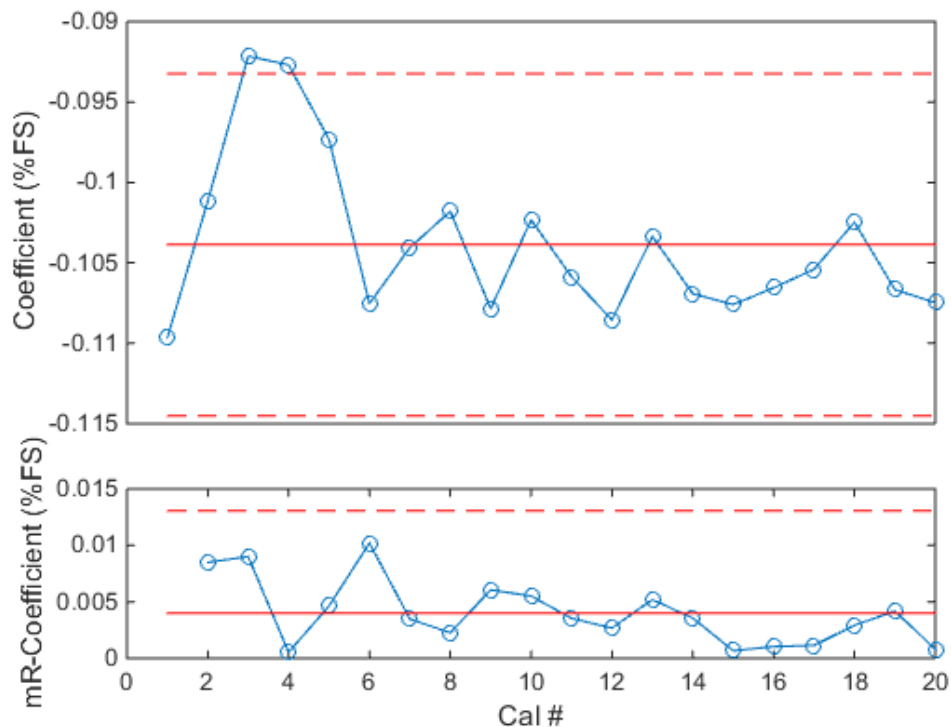


Figure 112: Normal Force Model - AF First Order Coefficient (NTF-113C)

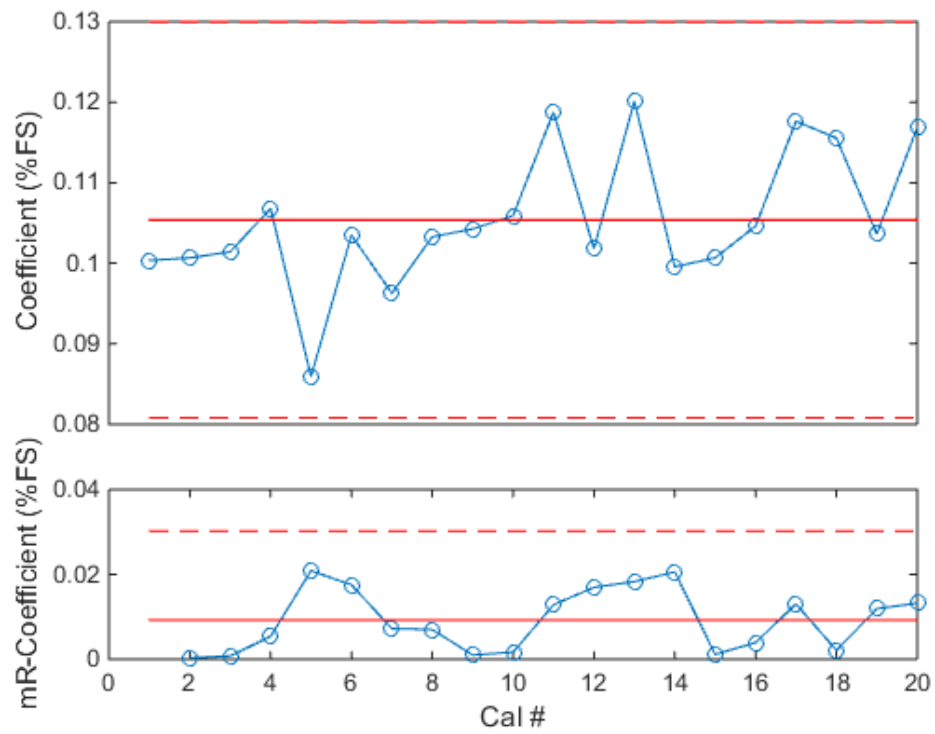


Figure 113: Normal Force Model - PM First Order Coefficient (NTF-113C)

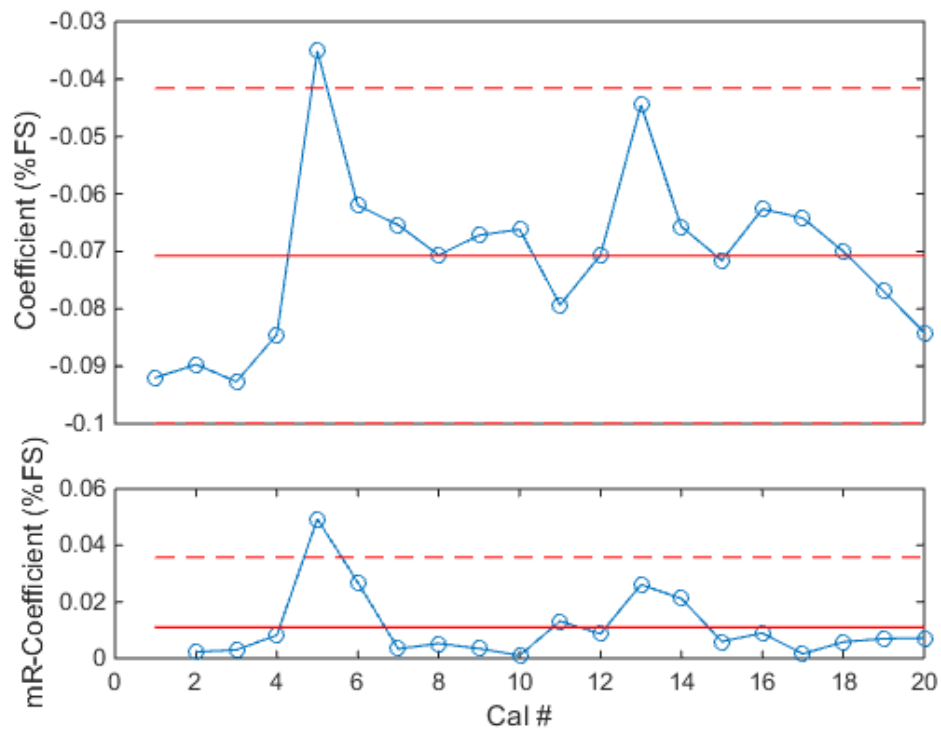


Figure 114: Normal Force Model - RM First Order Coefficient (NTF-113C)

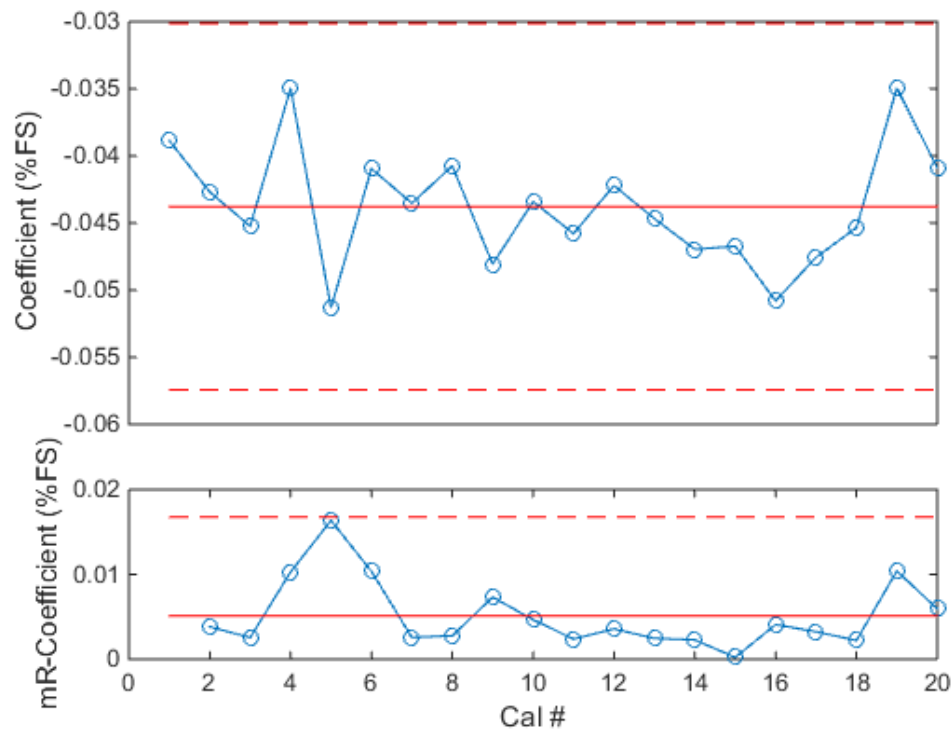


Figure 115: Normal Force Model - YM First Order Coefficient (NTF-113C)

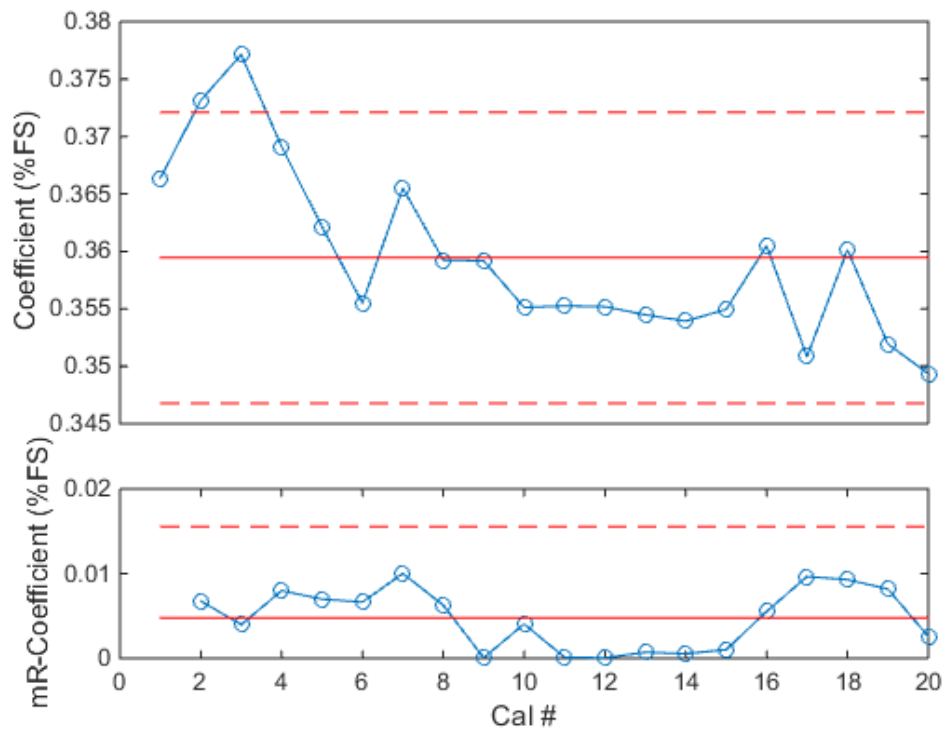
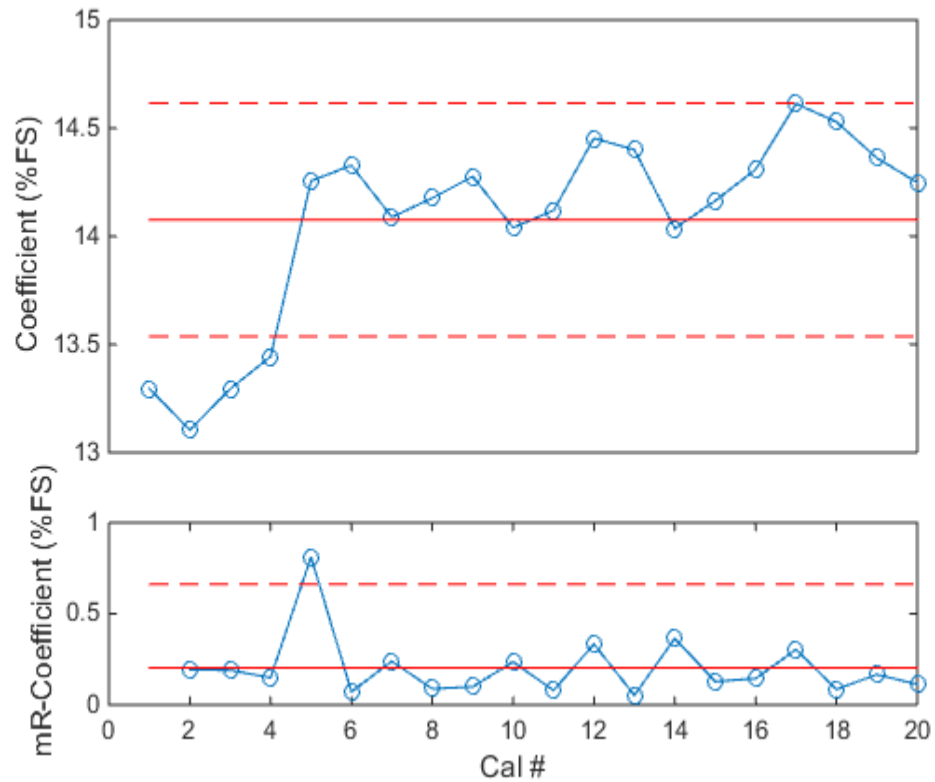
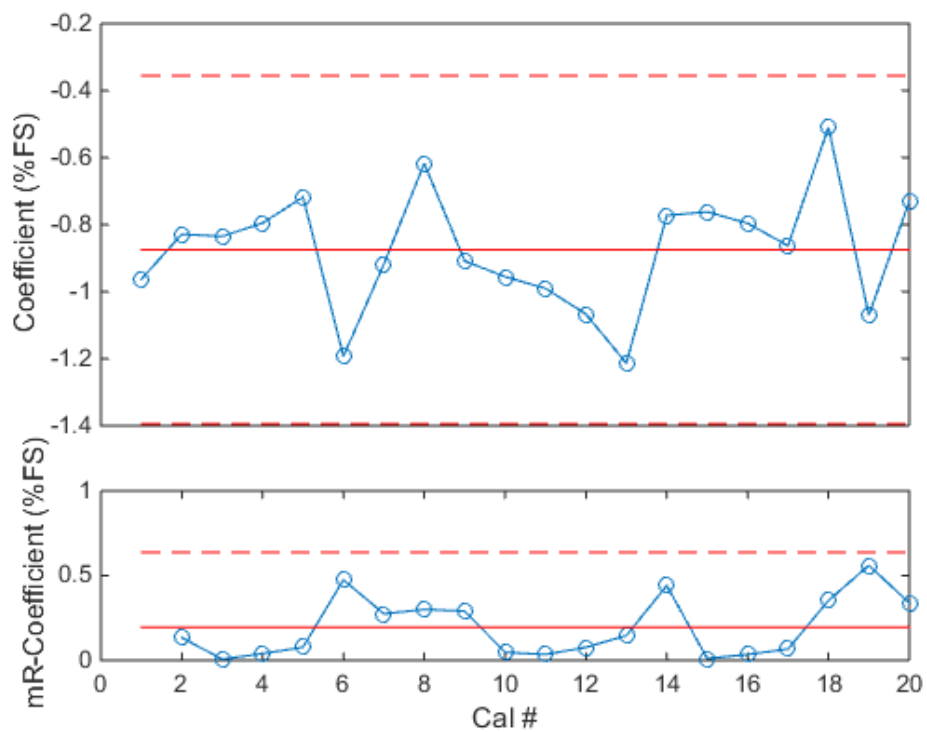


Figure 116: Normal Force Model - SF First Order Coefficient (NTF-113C)

b. Axial Force Model**Figure 117: Axial Force Model - NF First Order Coefficient (NTF-113C)****Figure 118: Axial Force Model - PM First Order Coefficient (NTF-113C)**

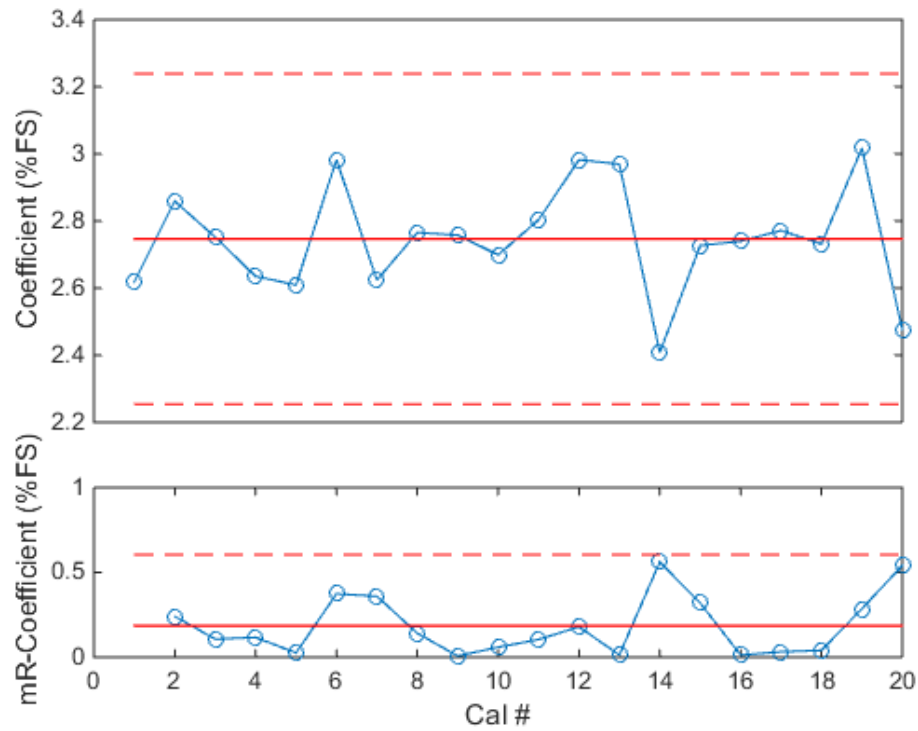


Figure 119: Axial Force Model - RM First Order Coefficient (NTF-113C)

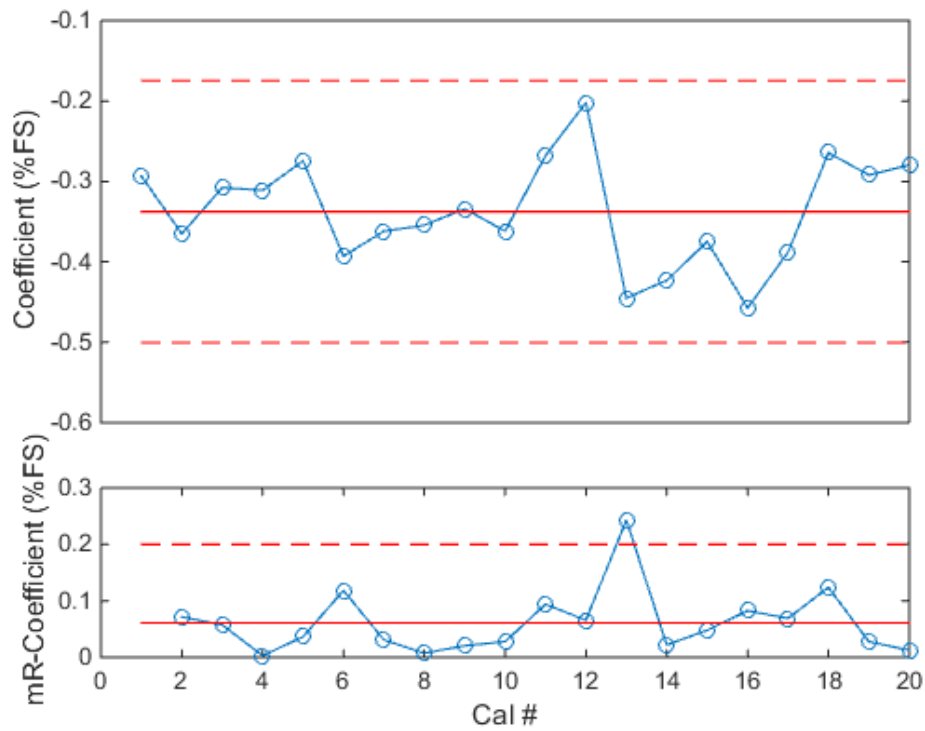


Figure 120: Axial Force Model - YM First Order Coefficient (NTF-113C)

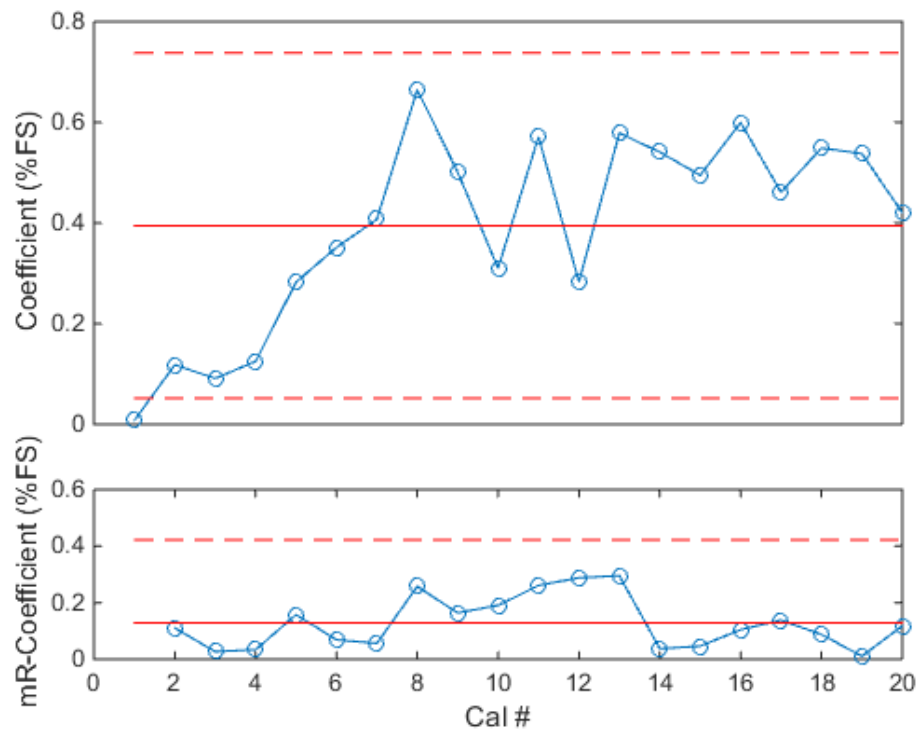


Figure 121: Axial Force Model - SF First Order Coefficient (NTF-113C)

c. Pitching Moment Model

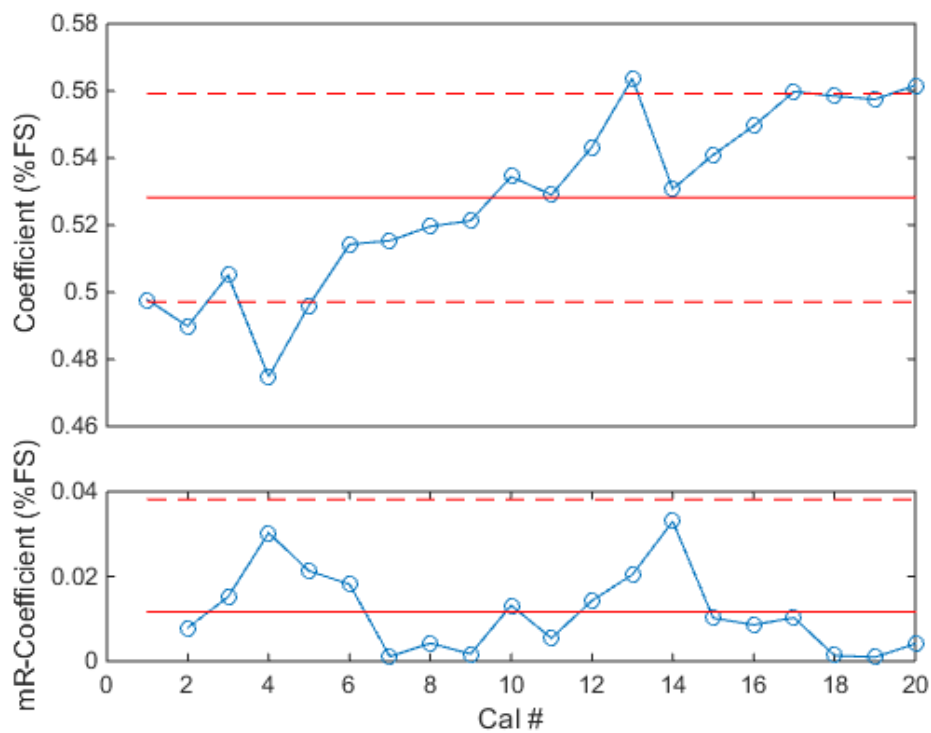


Figure 122: Pitching Moment Model - NF First Order Coefficient (NTF-113C)

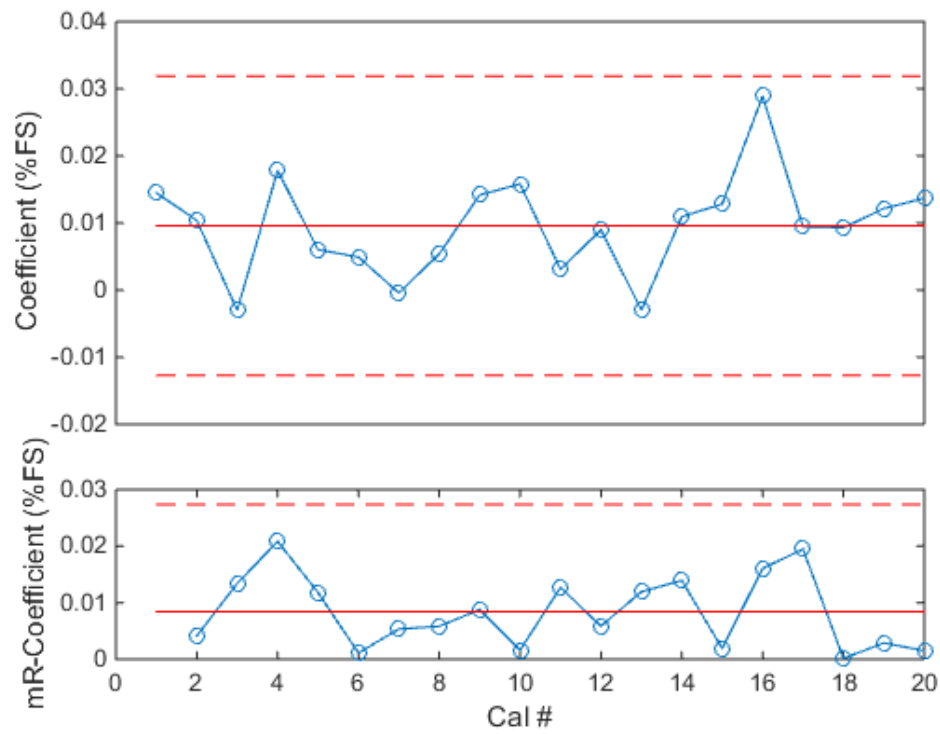


Figure 123: Pitching Moment Model - AF First Order Coefficient (NTF-113C)

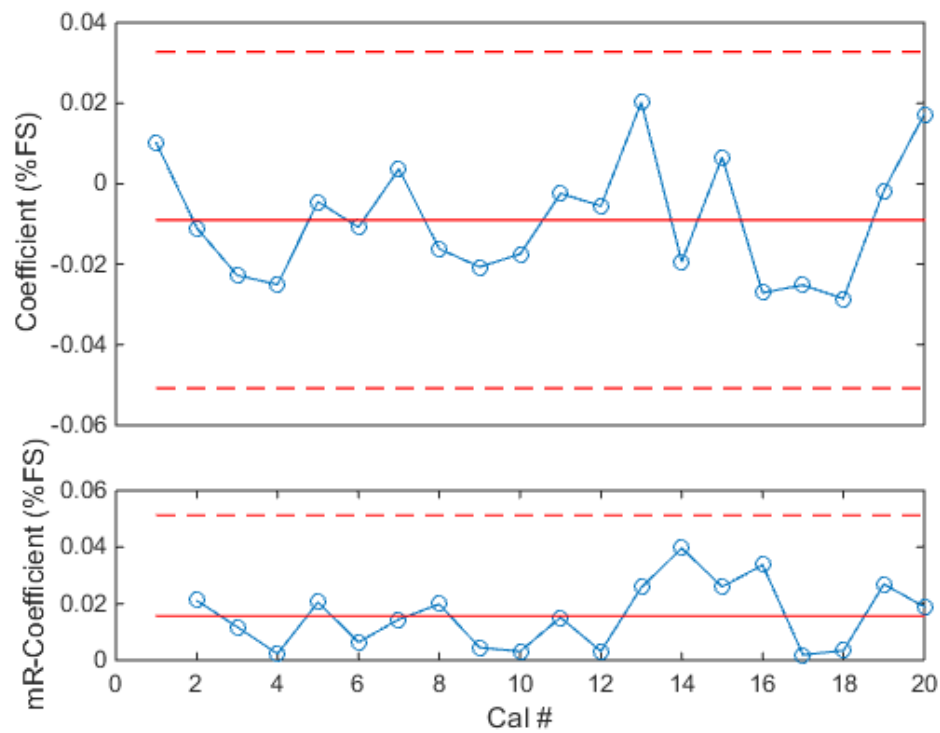


Figure 124: Pitching Moment Model - RM First Order Coefficient (NTF-113C)

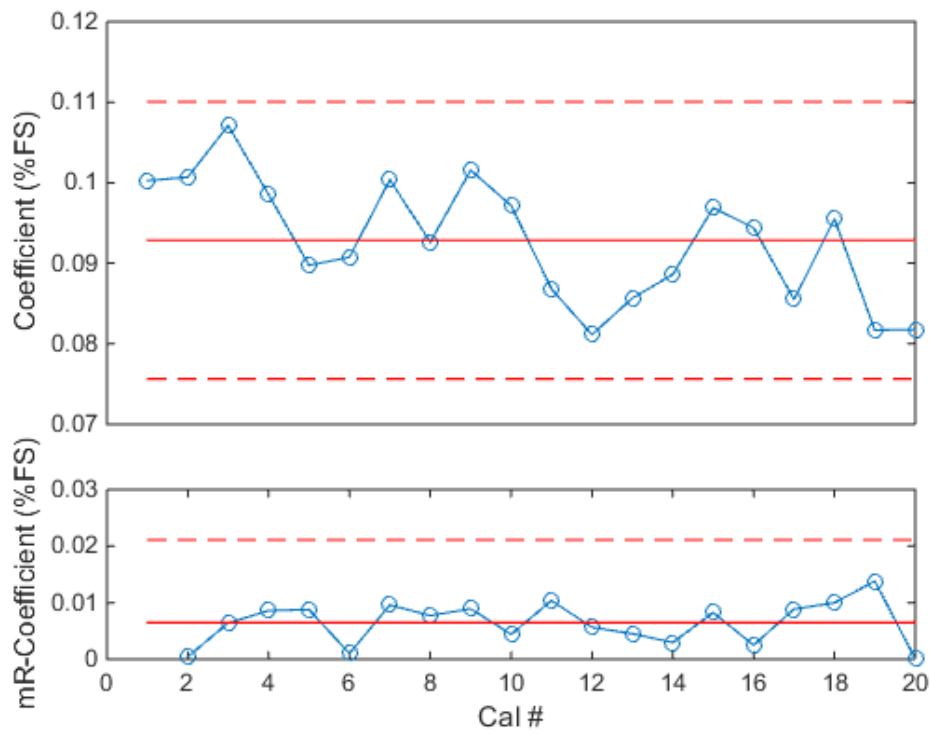


Figure 125: Pitching Moment Model - YM First Order Coefficient (NTF-113C)

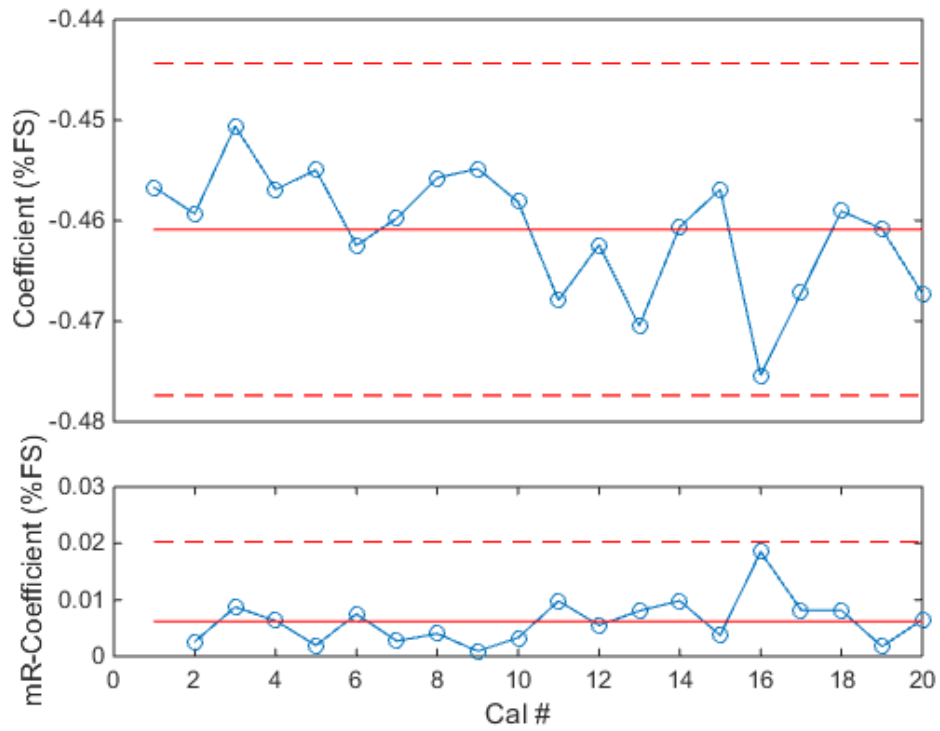


Figure 126: Pitching Moment Model - SF First Order Coefficient (NTF-113C)

d. Rolling Moment Model

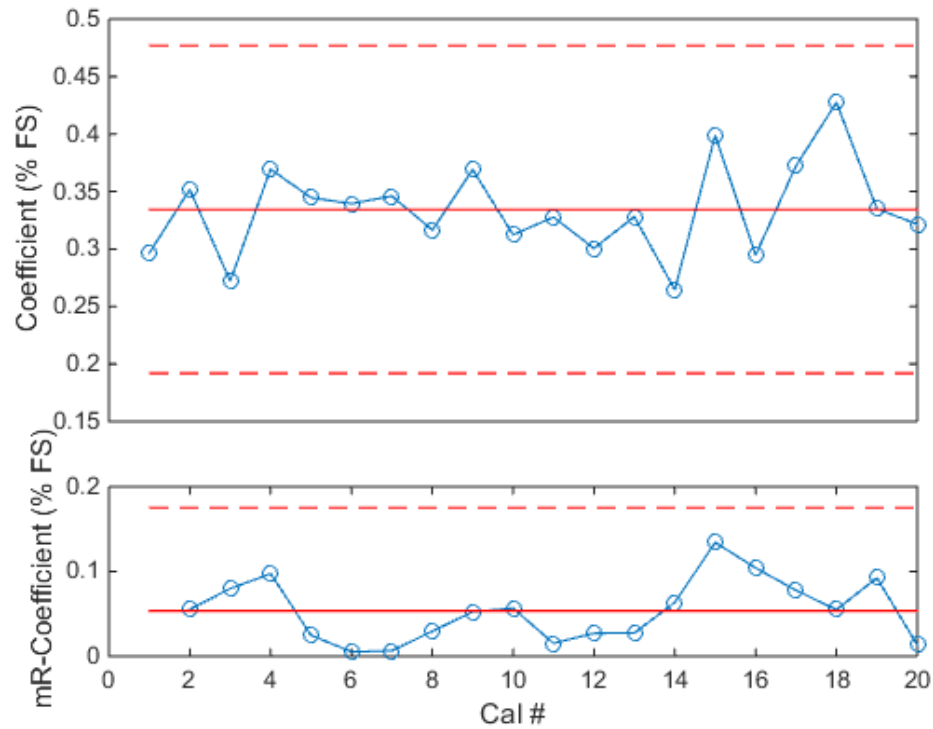


Figure 127: Rolling Moment Model - NF First Order Coefficient (NTF-113C)

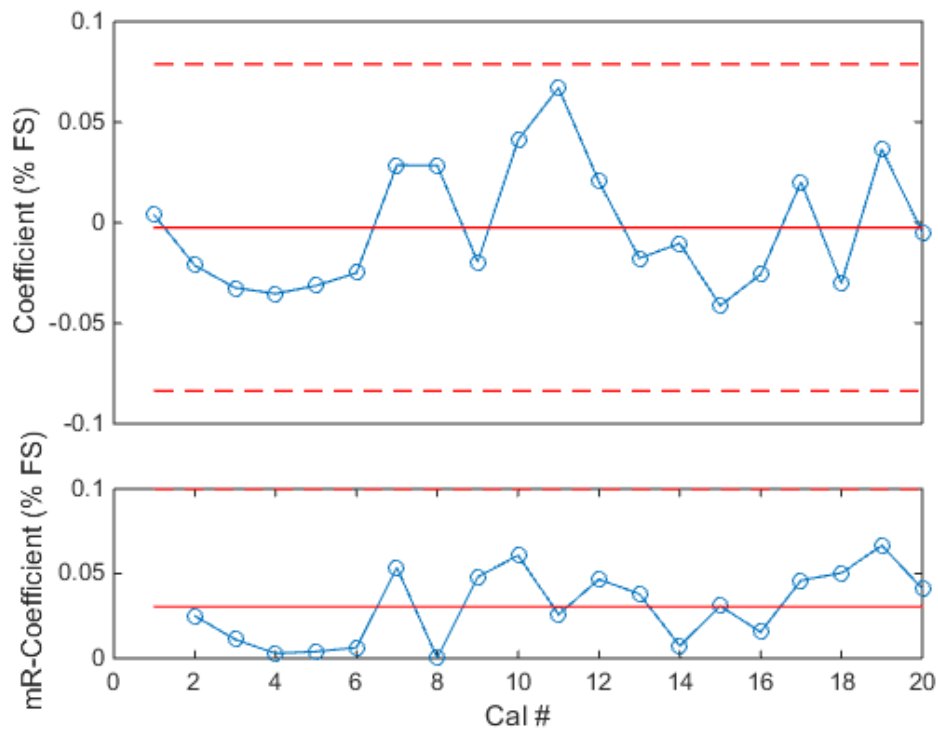


Figure 128: Rolling Moment Model - AF First Order Coefficient (NTF-113C)

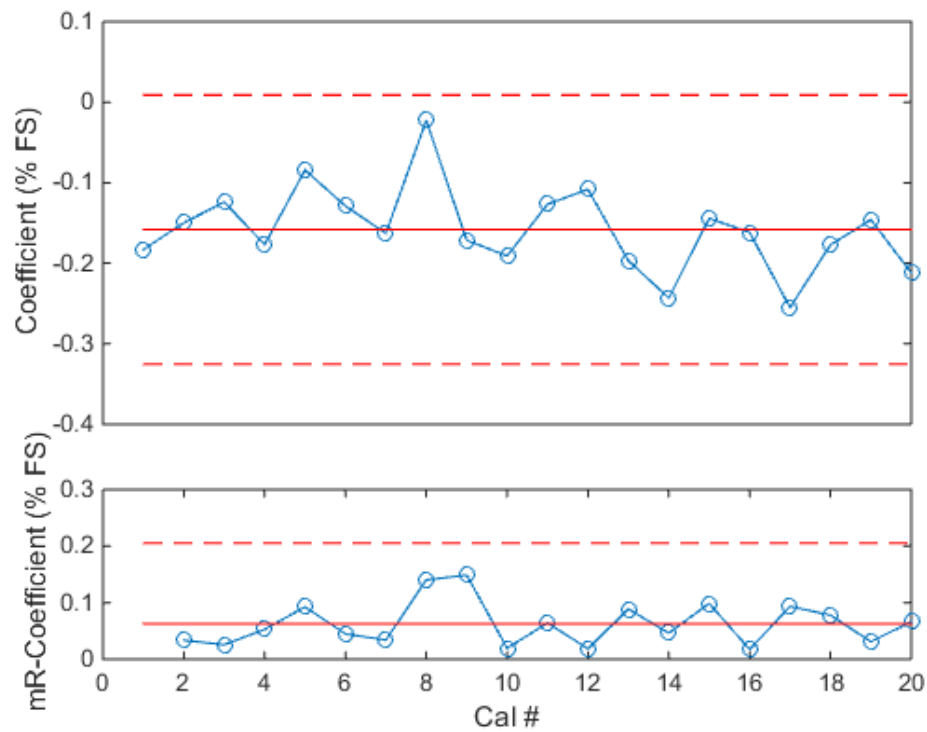


Figure 129: Rolling Moment Model - PM First Order Coefficient (NTF-113C)

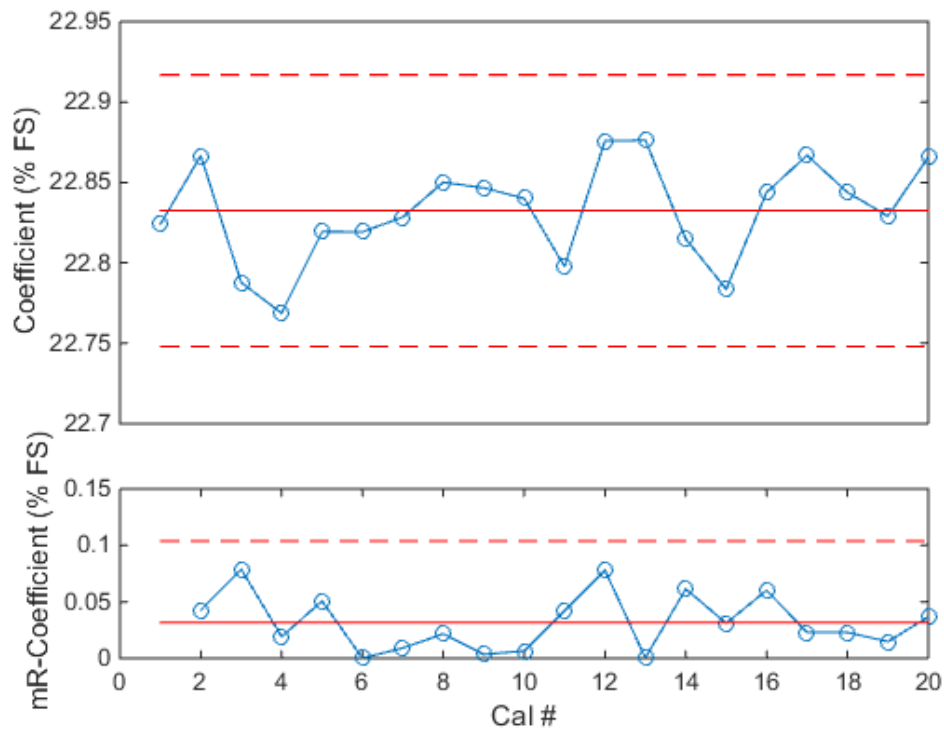


Figure 130: Rolling Moment Model - YM First Order Coefficient (NTF-113C)

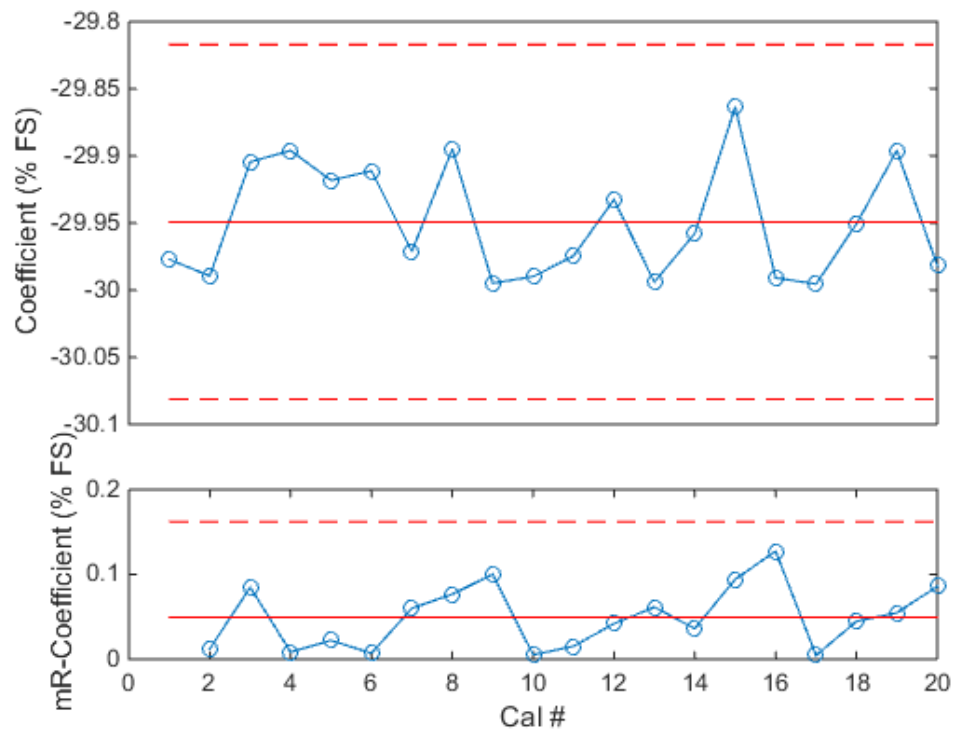


Figure 131: Rolling Moment Model - SF First Order Coefficient (NTF-113C)

e. Yawing Moment Model

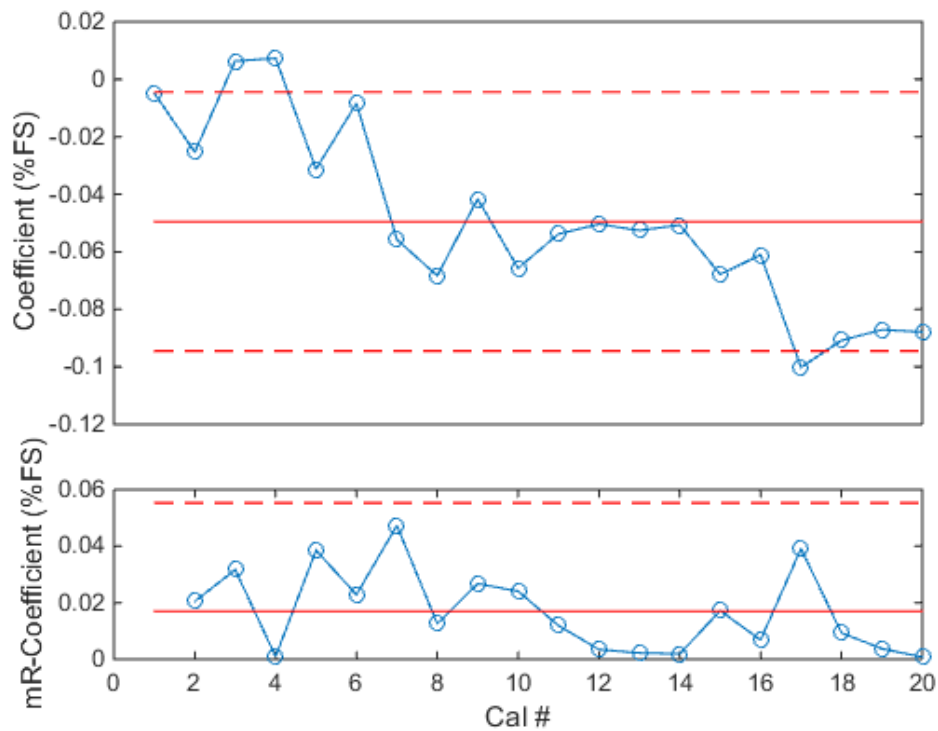


Figure 132: Yawing Moment Model - NF First Order Coefficient (NTF-113C)

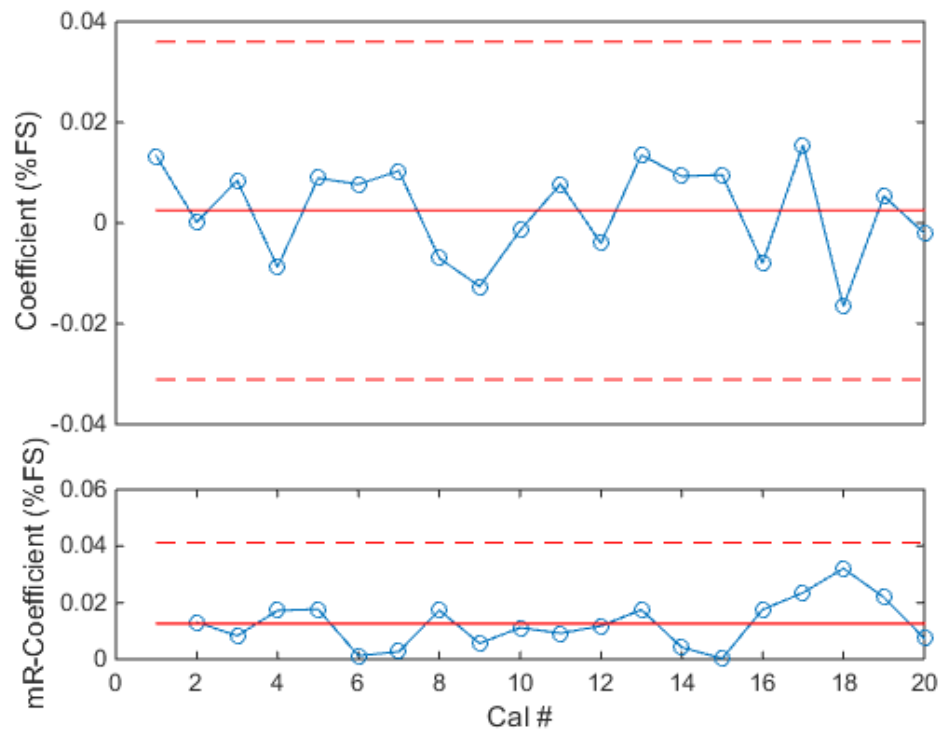


Figure 133: Yawing Moment Model - AF First Order Coefficient (NTF-113C)

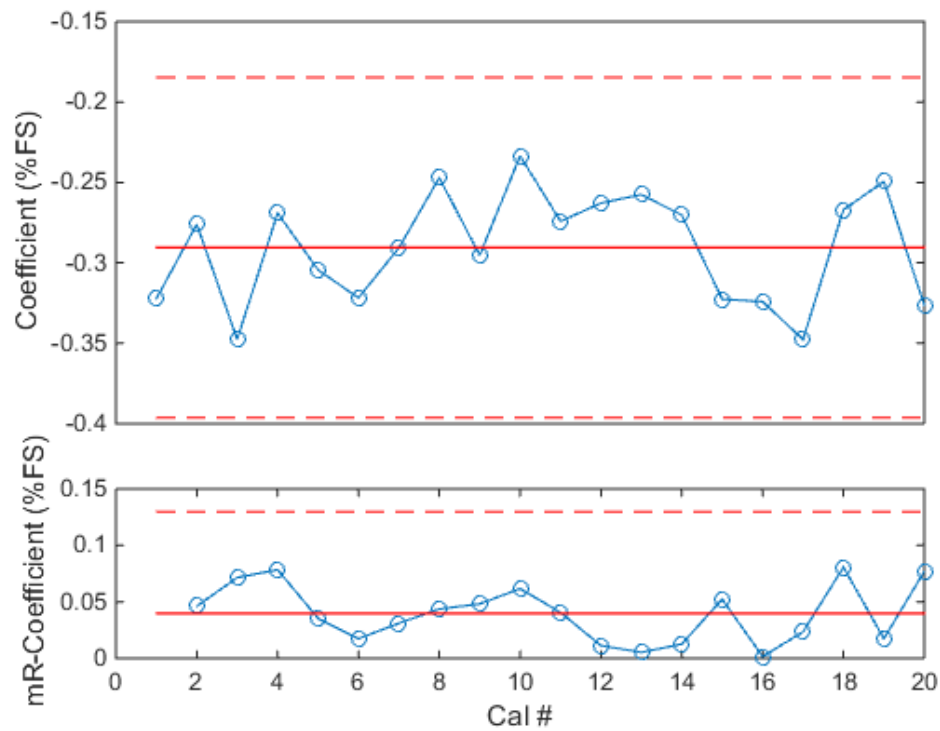


Figure 134: Yawing Moment Model - PM First Order Coefficient (NTF-113C)

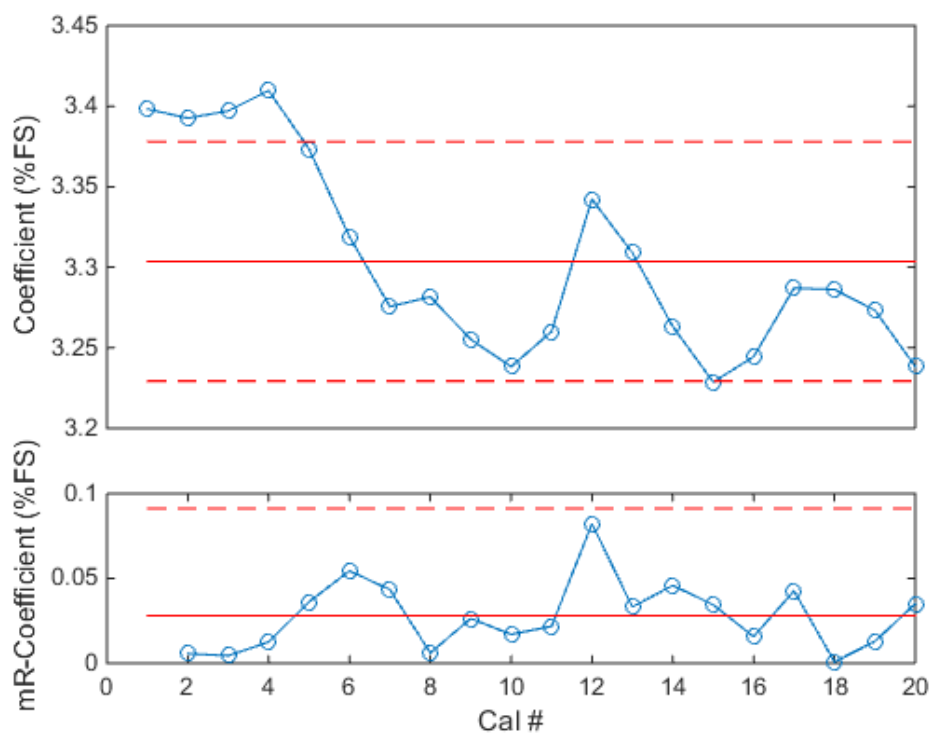


Figure 135: Yawing Moment Model - RM First Order Coefficient (NTF-113C)

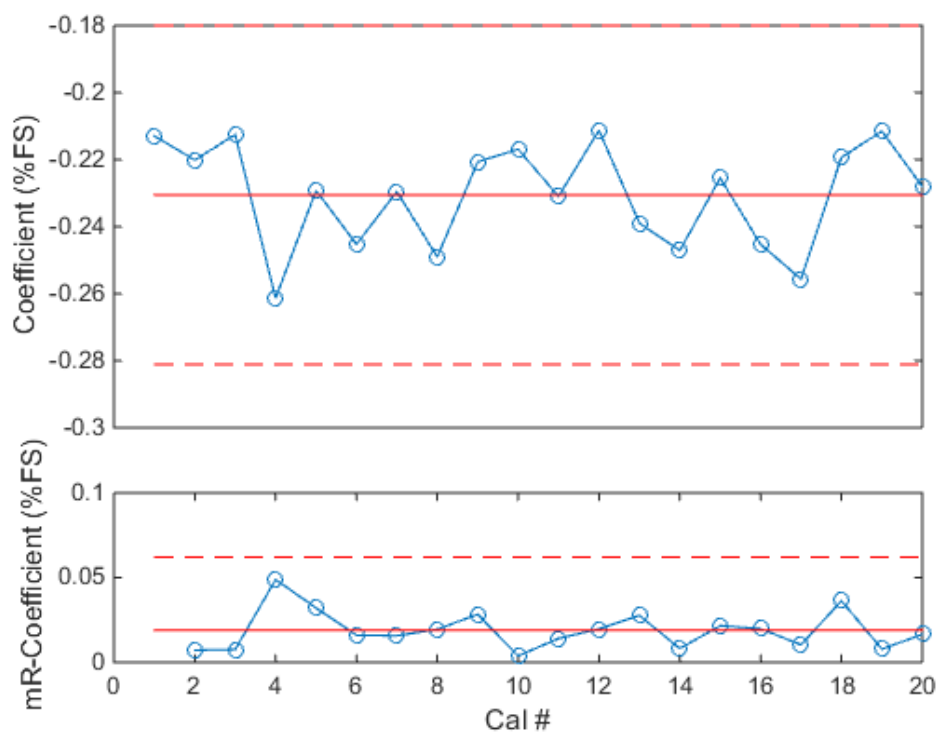


Figure 136: Yawing Moment Model - SF First Order Coefficient (NTF-113C)

f. Side Force Model

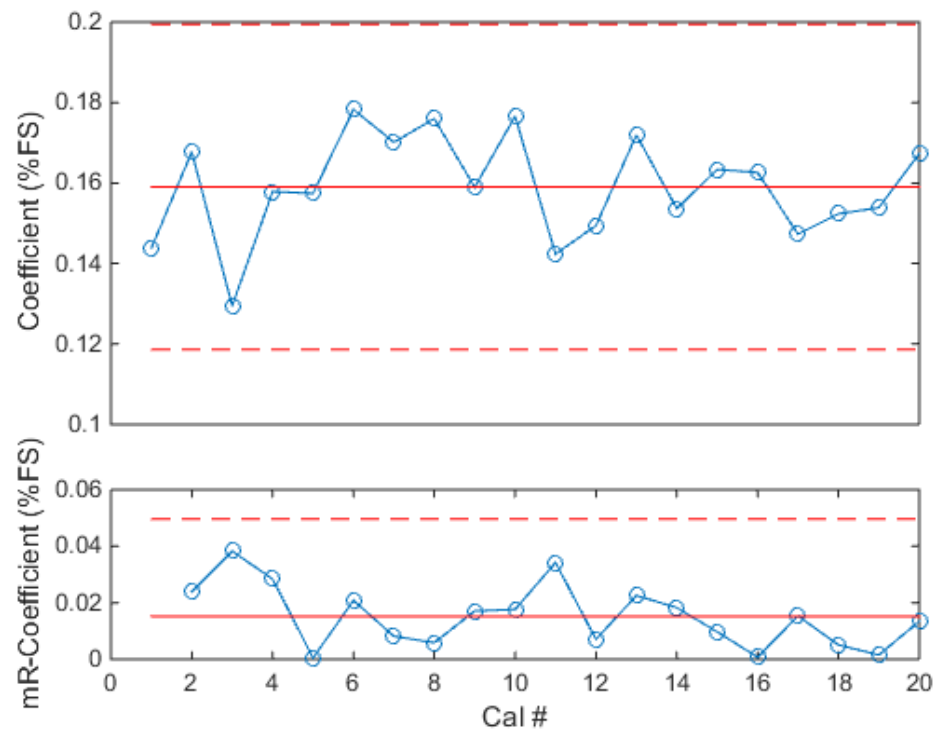


Figure 137: Side Force Model - NF First Order Coefficient (NTF-113C)

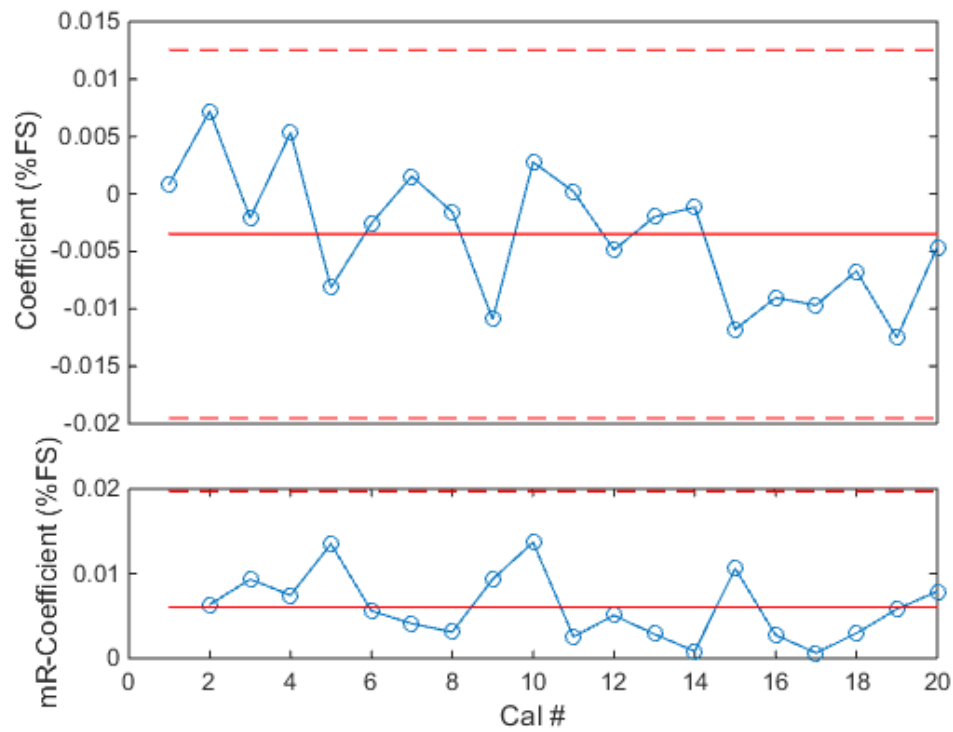


Figure 138: Side Force Model - AF First Order Coefficient (NTF-113C)

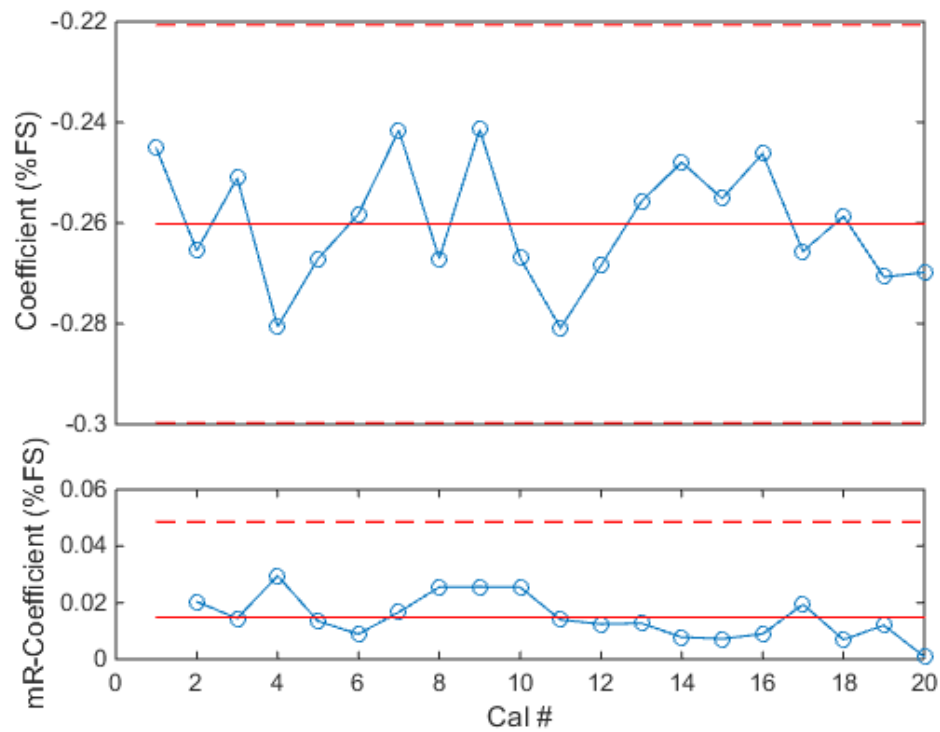


Figure 139: Side Force Model - PM First Order Coefficient (NTF-113C)

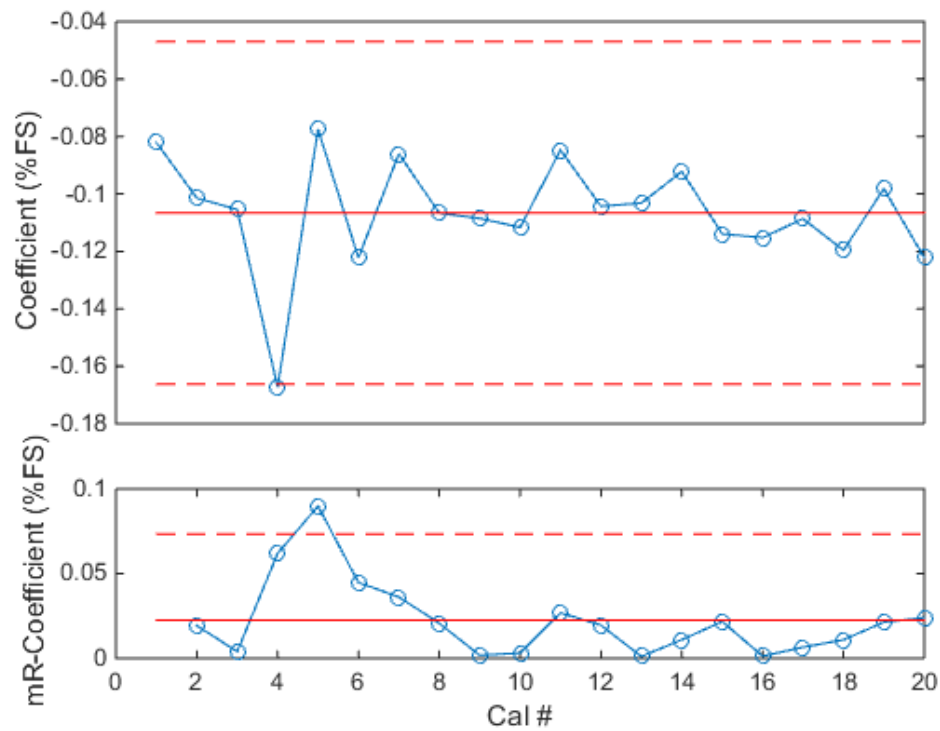


Figure 140: Side Force Model - RM First Order Coefficient (NTF-113C)

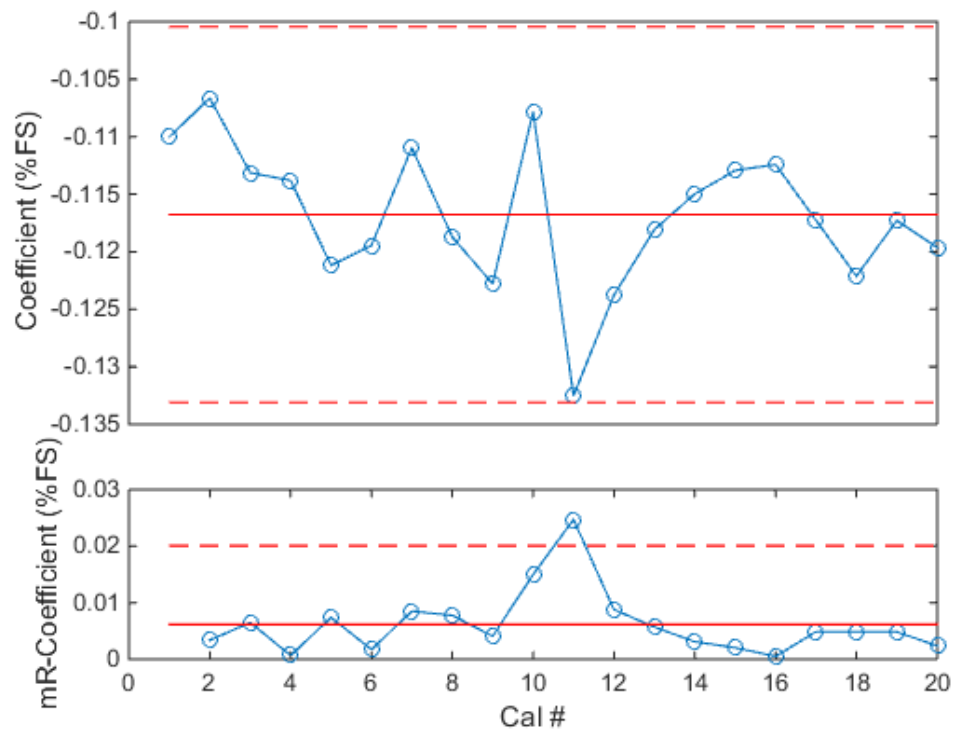


Figure 141: Side Force Model - YM First Order Coefficient (NTF-113C)

B. First Order Interaction Coefficients MC-60E

The following plots show the first order coefficients for all six components of force for the Triumph MC-60E. All values are in percent full scale, where the primary coefficient would be 100%. The primary coefficients are not shown here as they were previously shown in the MC-60E primary sensitivity section.

a. Normal Force Model

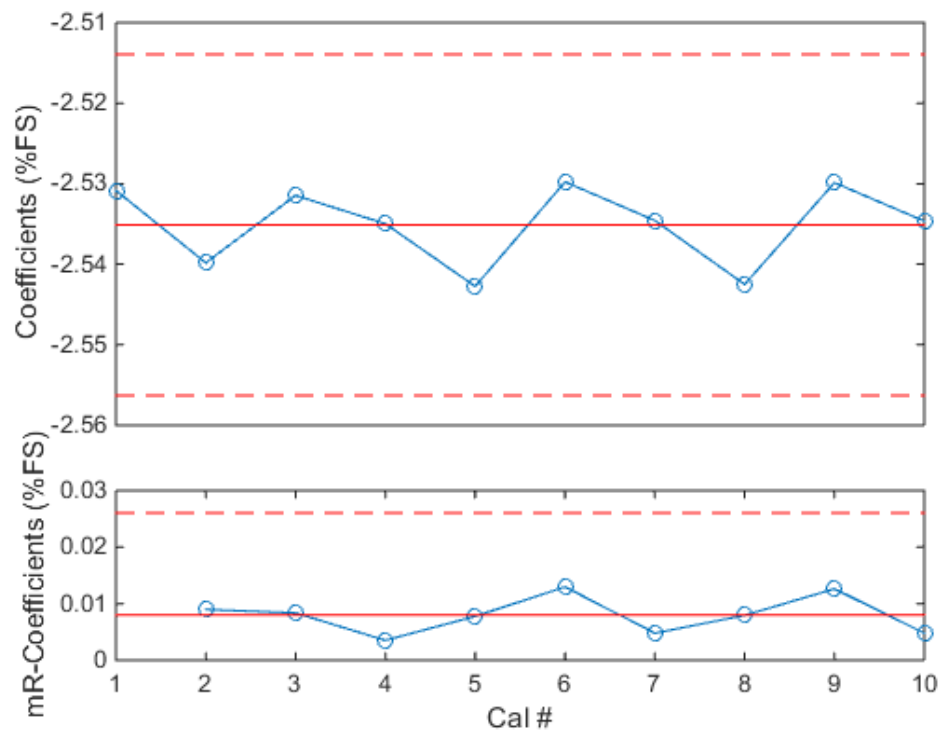


Figure 142: Normal Force Model - AF First Order Coefficient (MC-60E)

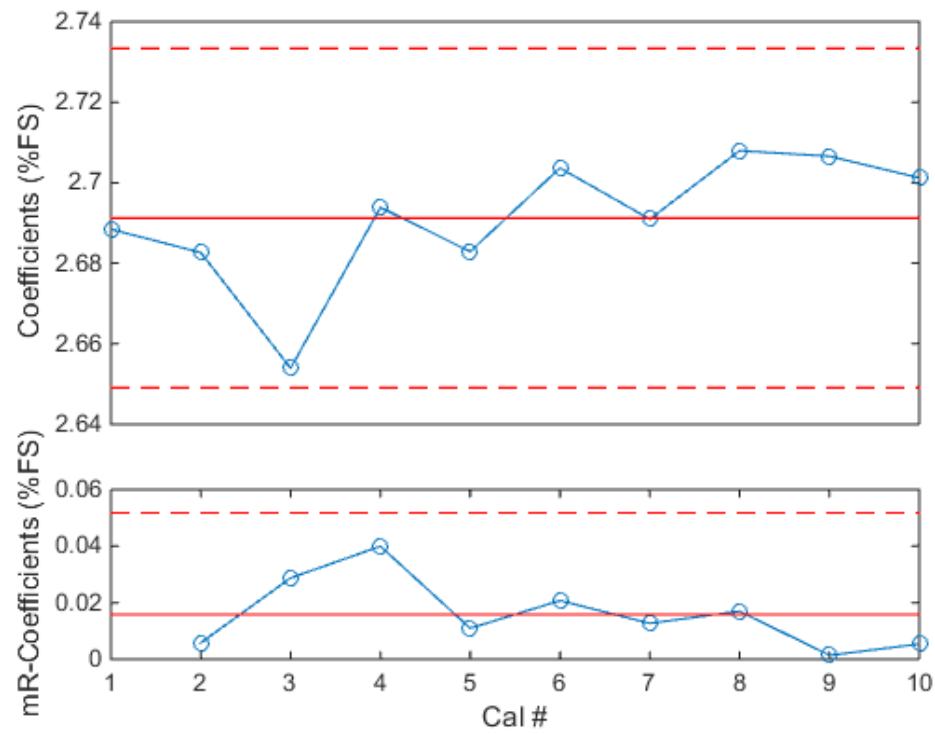


Figure 143: Normal Force Model - PM First Order Coefficient (MC-60E)

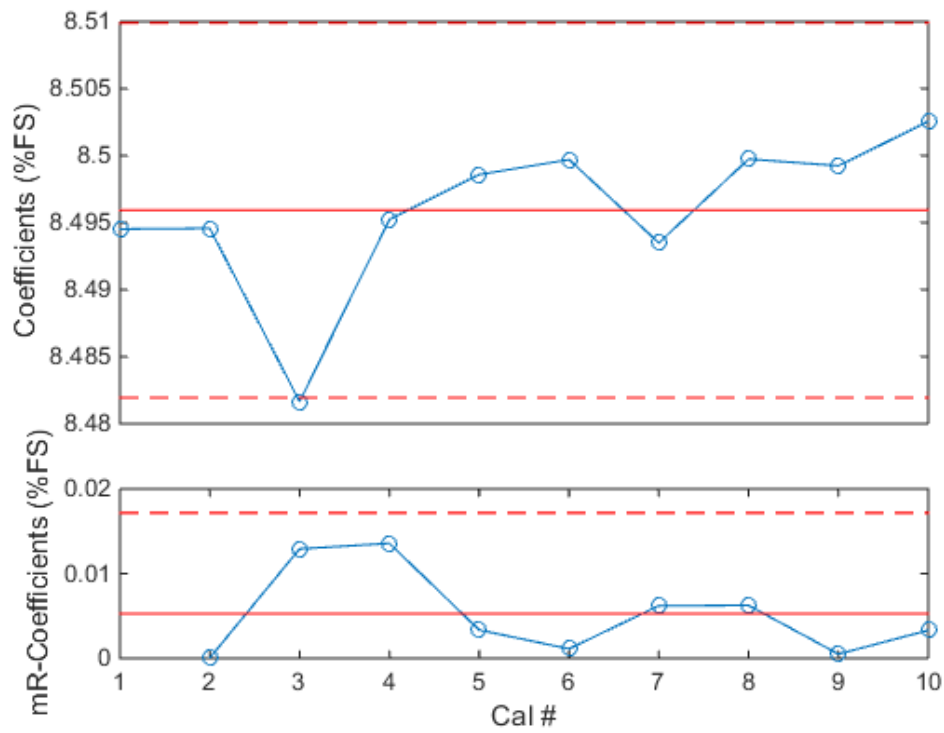


Figure 144: Normal Force Model - RM First Order Coefficient (MC-60E)

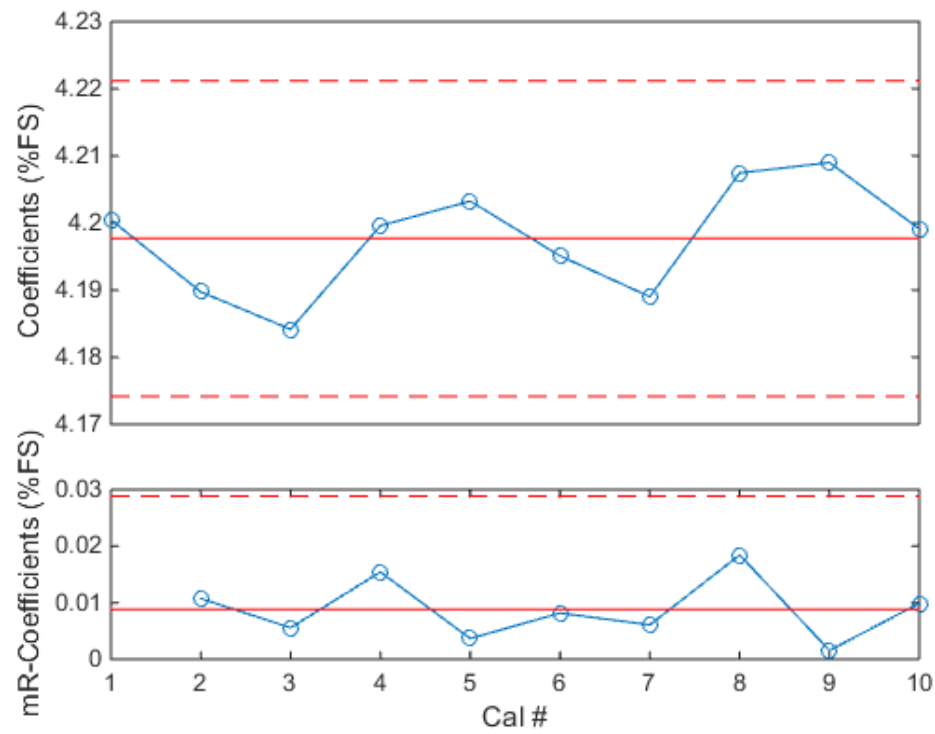


Figure 145: Normal Force Model - YM First Order Coefficient (MC-60E)

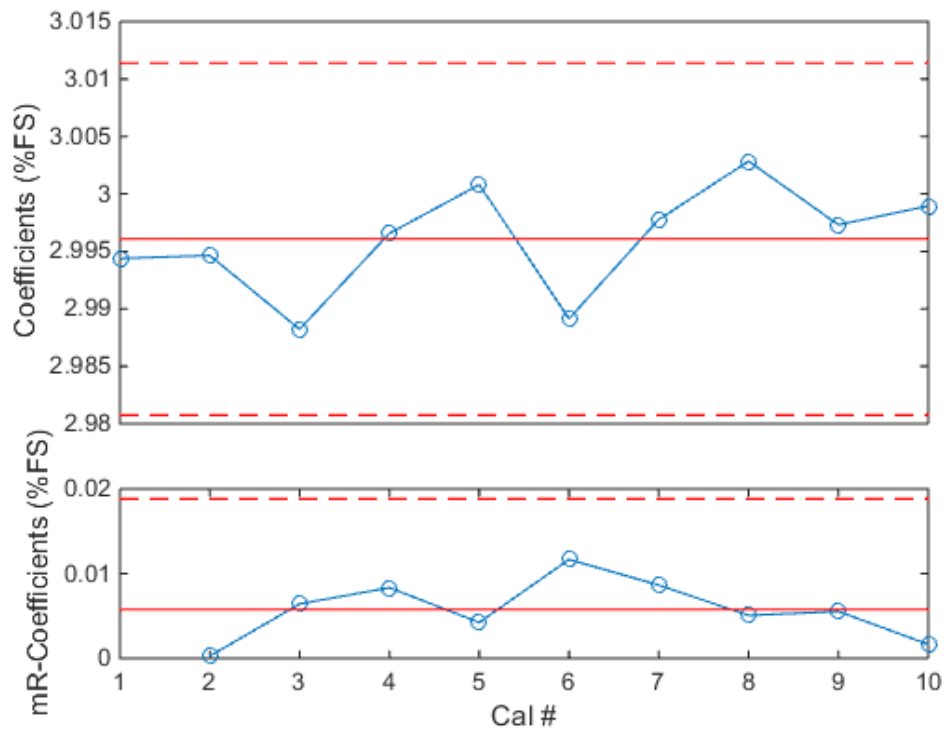


Figure 146: Normal Force Model - SF First Order Coefficient (MC-60E)

b. Axial Force Model

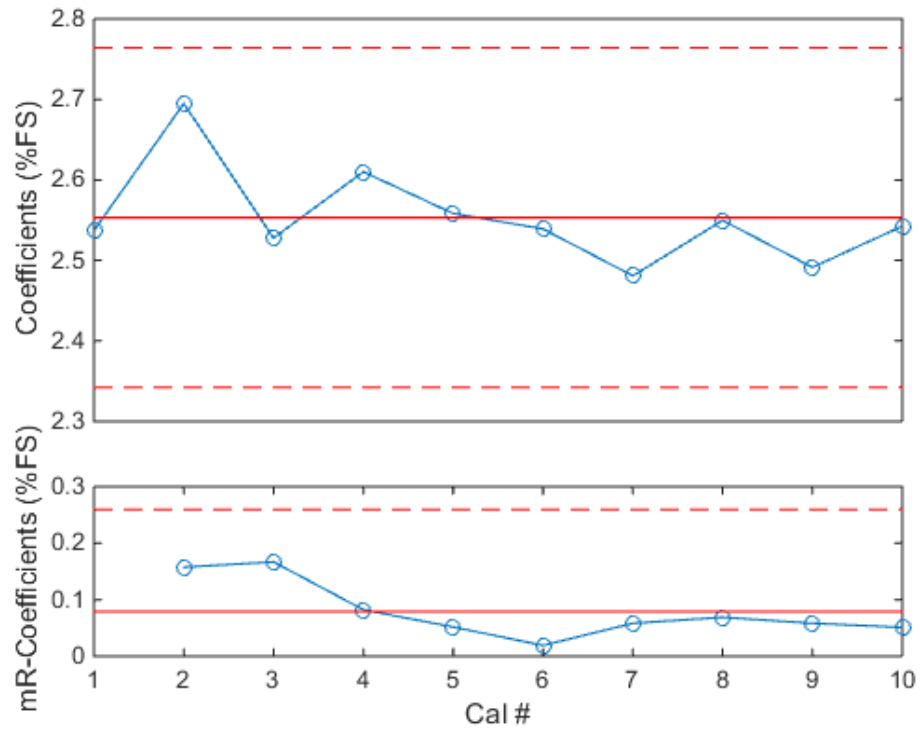


Figure 147: Axial Force Model - NF First Order Coefficient (MC-60E)

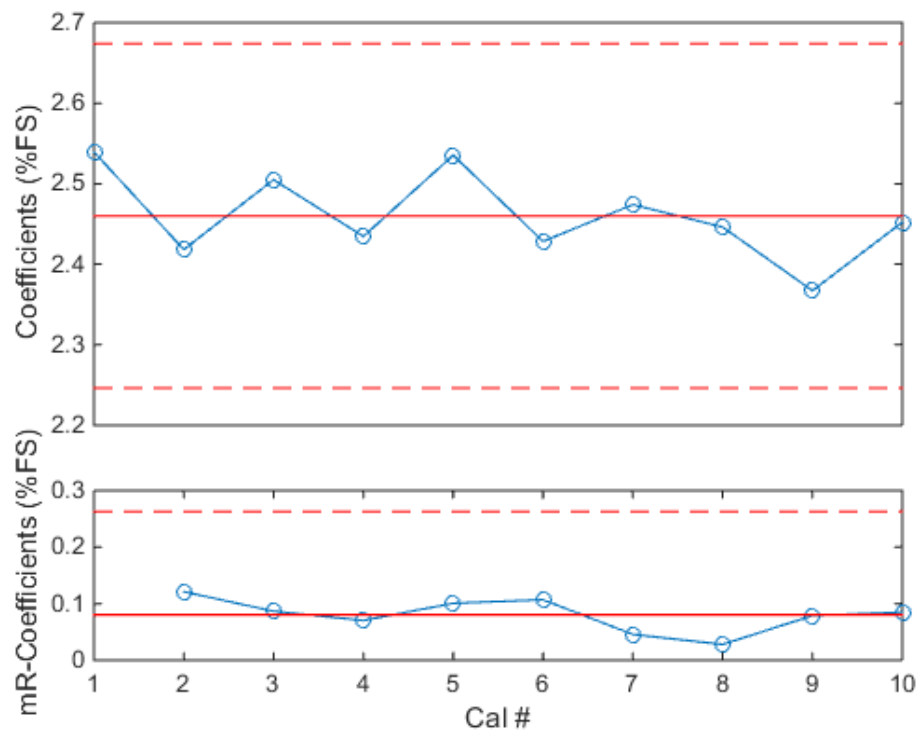


Figure 148: Axial Force Model - PM First Order Coefficient (MC-60E)

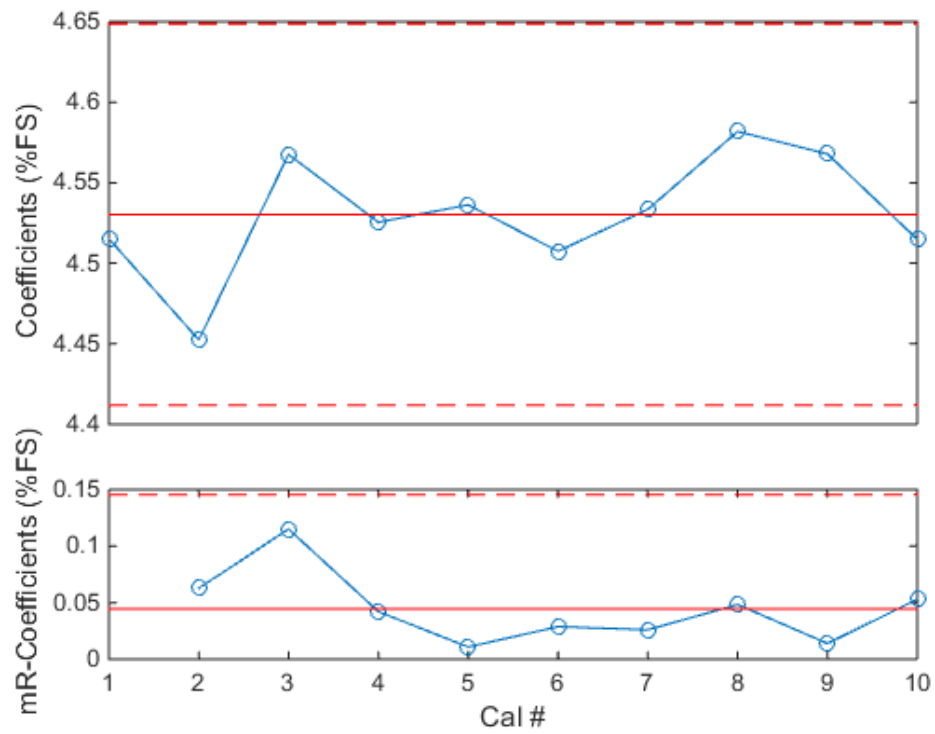


Figure 149: Axial Force Model - RM First Order Coefficient (MC-60E)

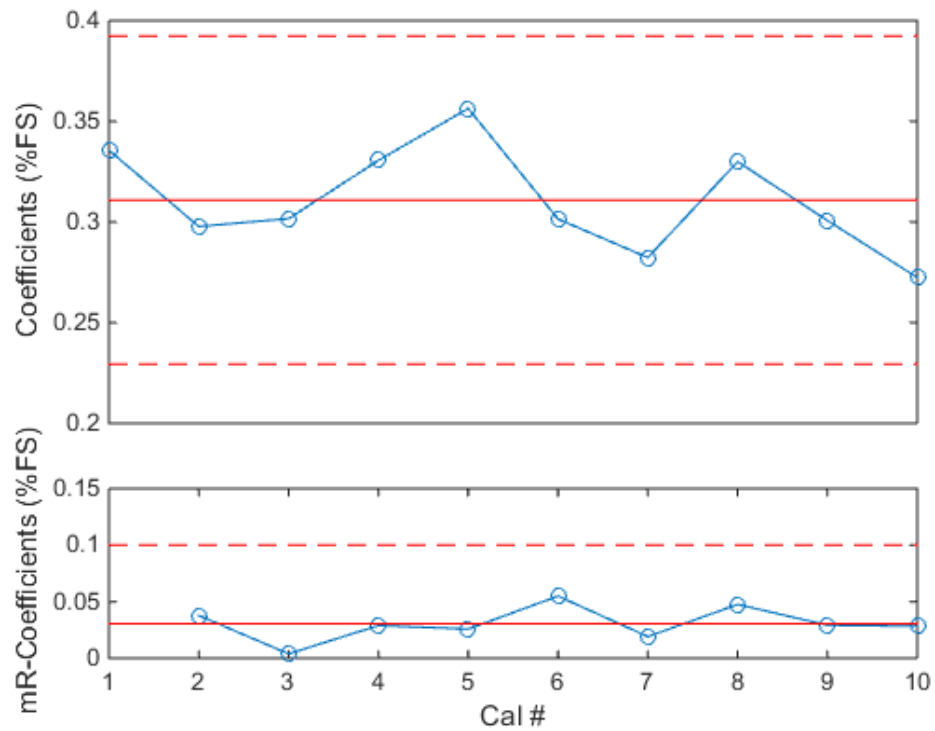


Figure 150: Axial Force Model - YM First Order Coefficient (MC-60E)

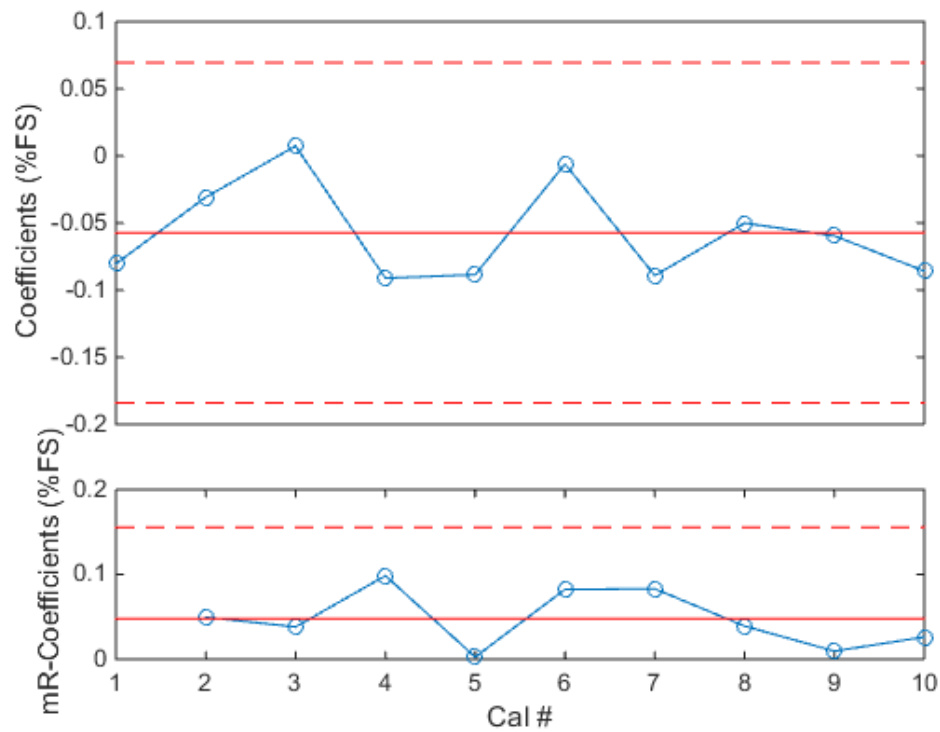


Figure 151: Axial Force Model - SF First Order Coefficient (MC-60E)

c. Pitching Moment Model

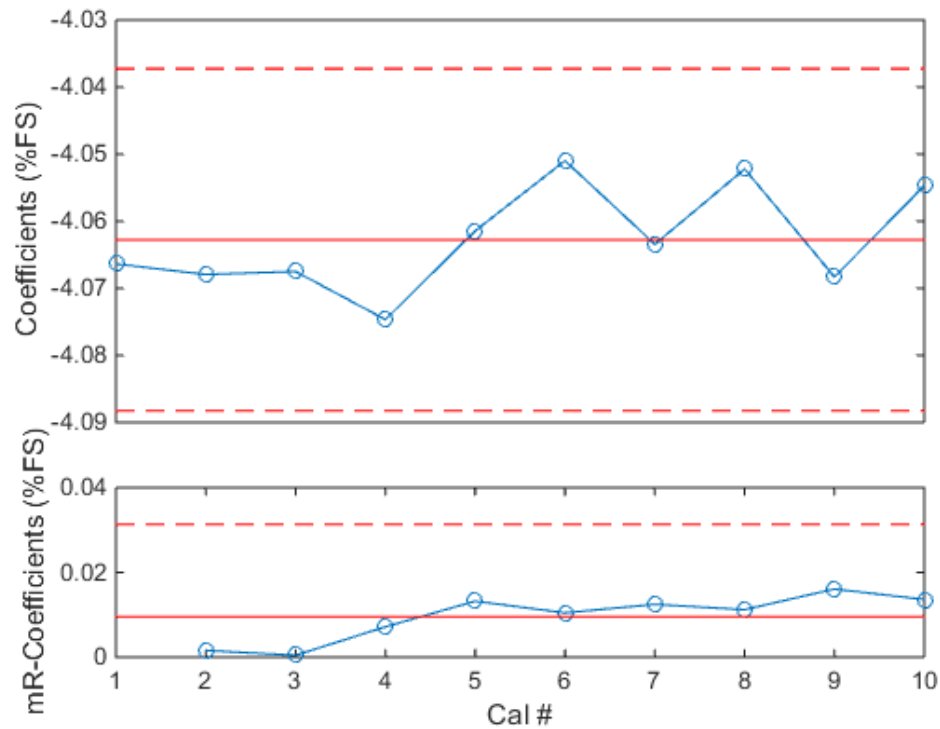


Figure 152: Pitching Moment Model - NF First Order Coefficient (MC-60E)

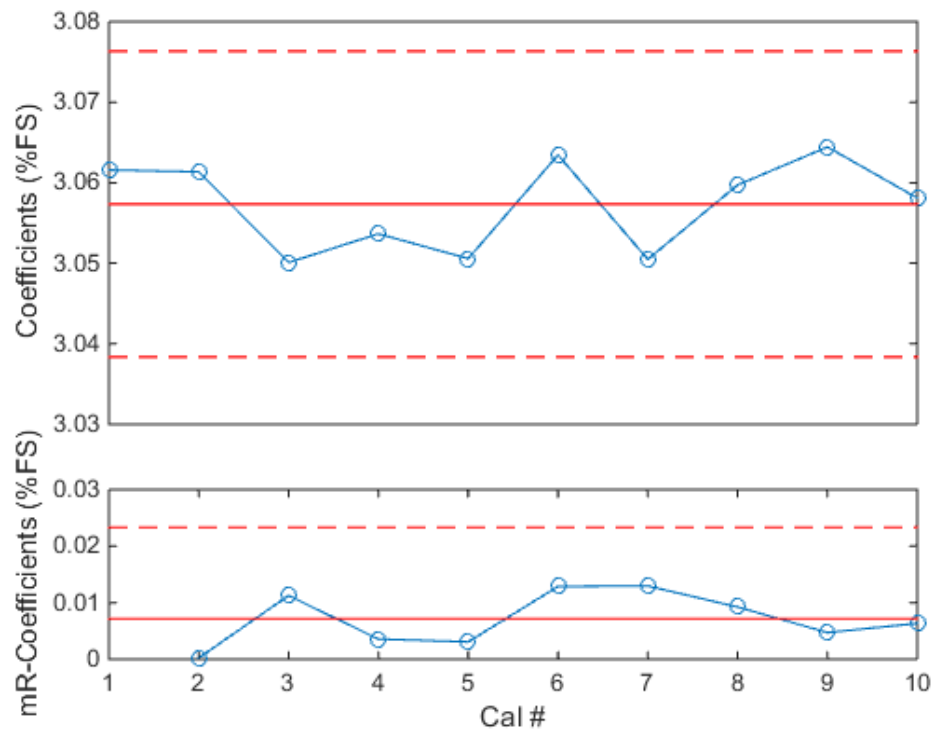


Figure 153: Pitching Moment Model - AF First Order Coefficient (MC-60E)

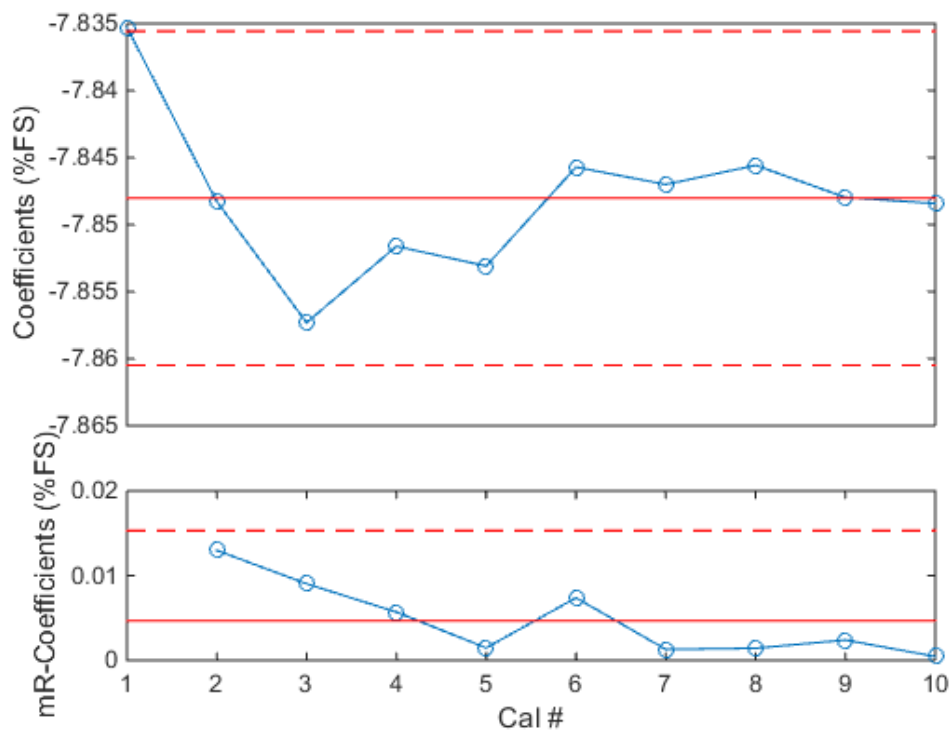


Figure 154: Pitching Moment Model - RM First Order Coefficient (MC-60E)

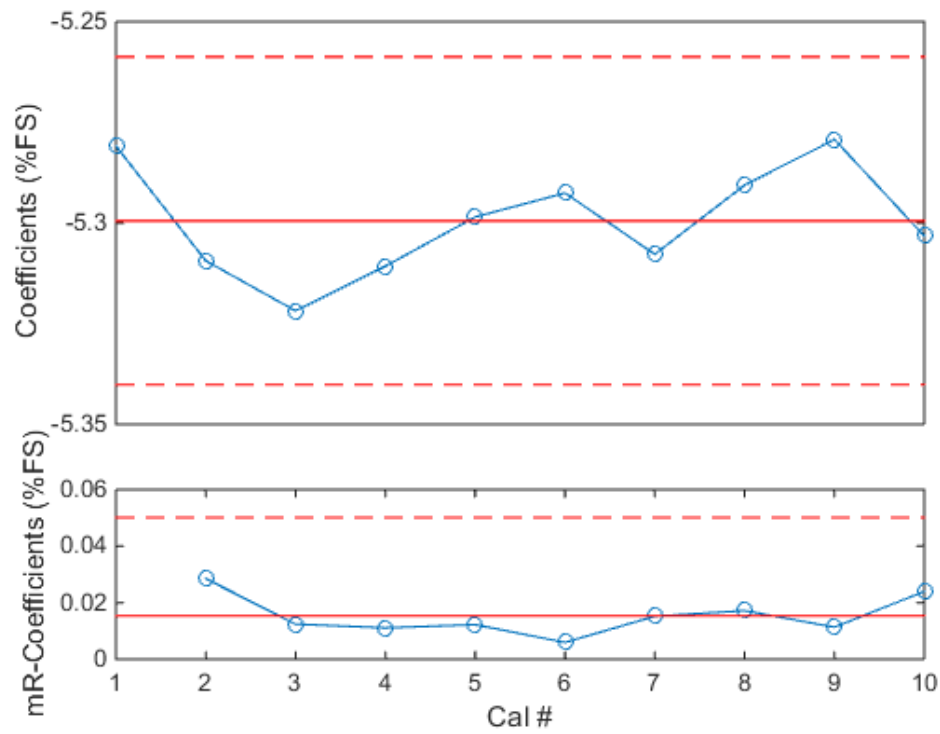


Figure 155: Pitching Moment Model - YM First Order Coefficient (MC-60E)

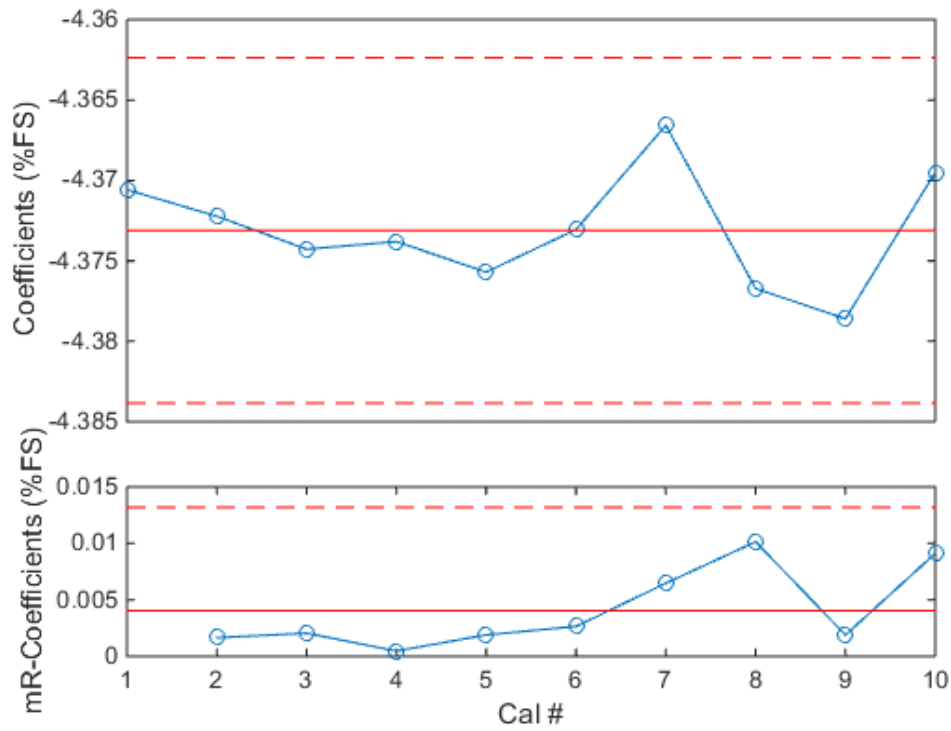


Figure 156: Pitching Moment Model - SF First Order Coefficient (MC-60E)

d. Rolling Moment Model

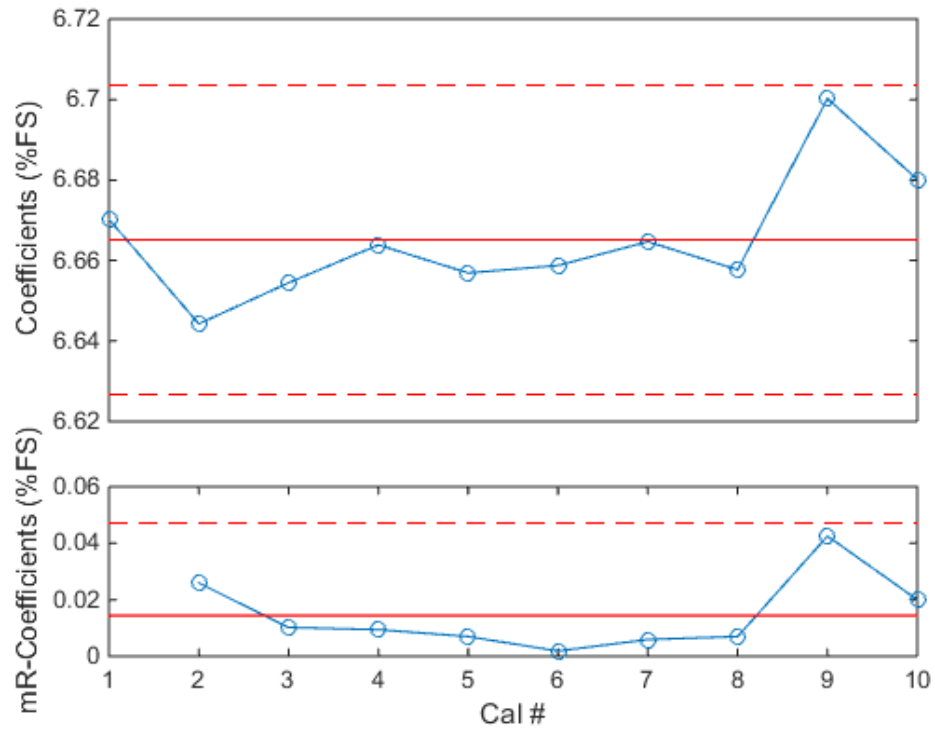


Figure 157: Rolling Moment Model - NF First Order Coefficient (MC-60E)

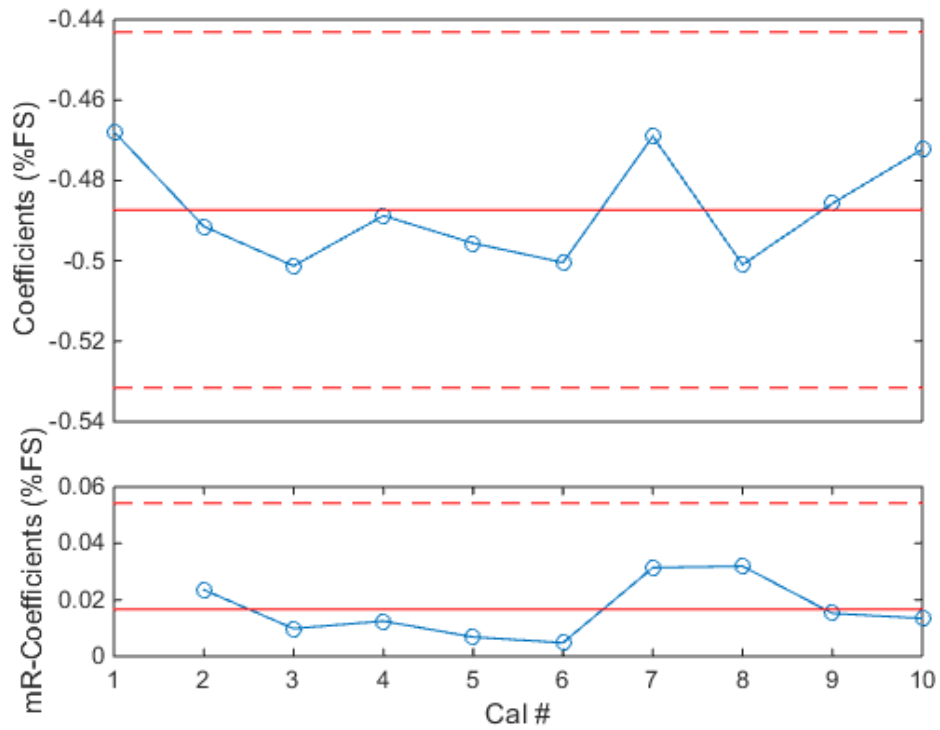


Figure 158: Rolling Moment Model - AF First Order Coefficient (MC-60E)

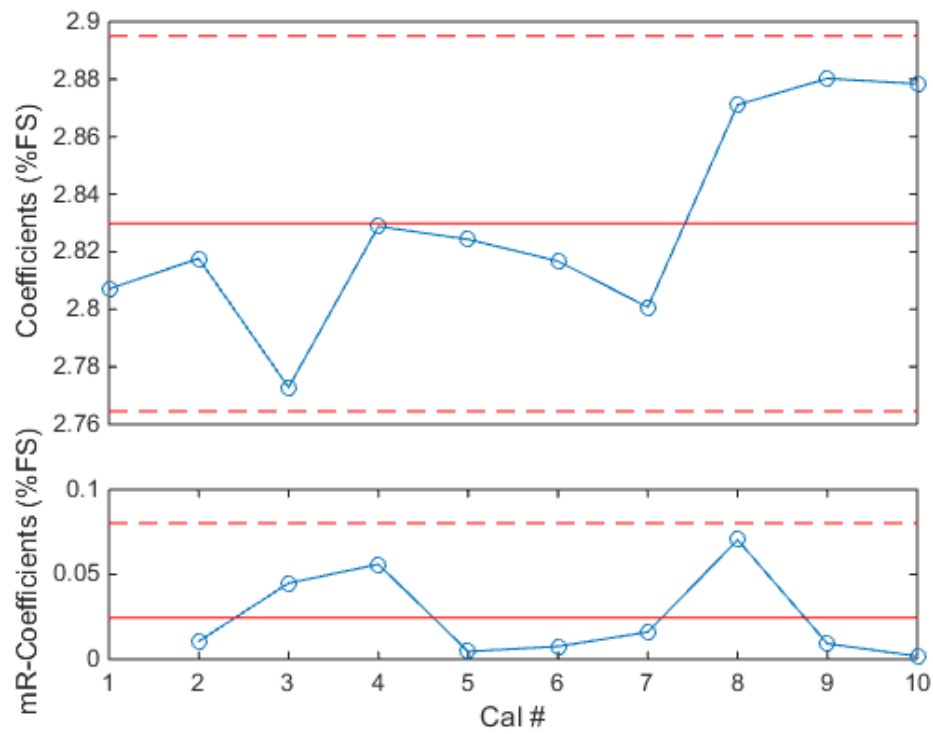


Figure 159: Rolling Moment Model - PM First Order Coefficient (MC-60E)

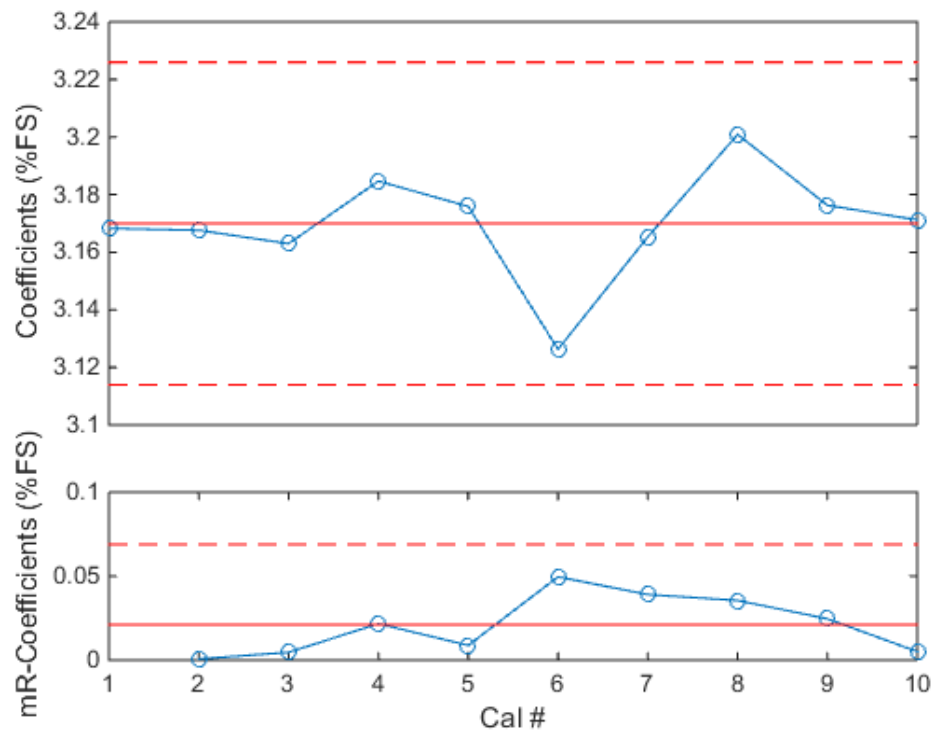


Figure 160: Rolling Moment Model - YM First Order Coefficient (MC-60E)

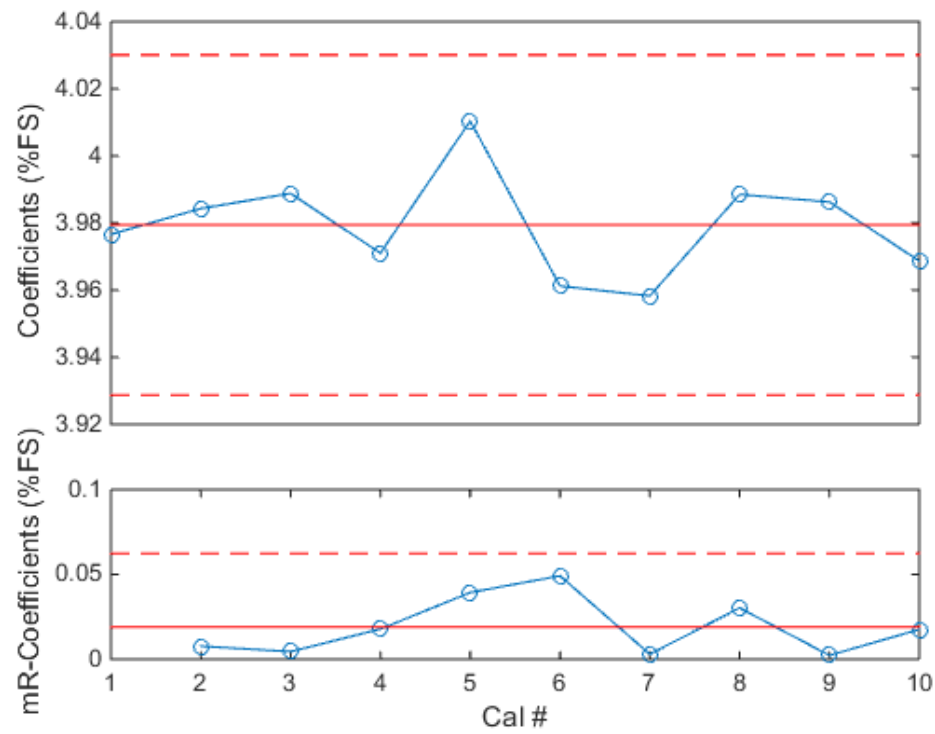


Figure 161: Rolling Moment Model - SF First Order Coefficient (MC-60E)

e. Yawing Moment Model

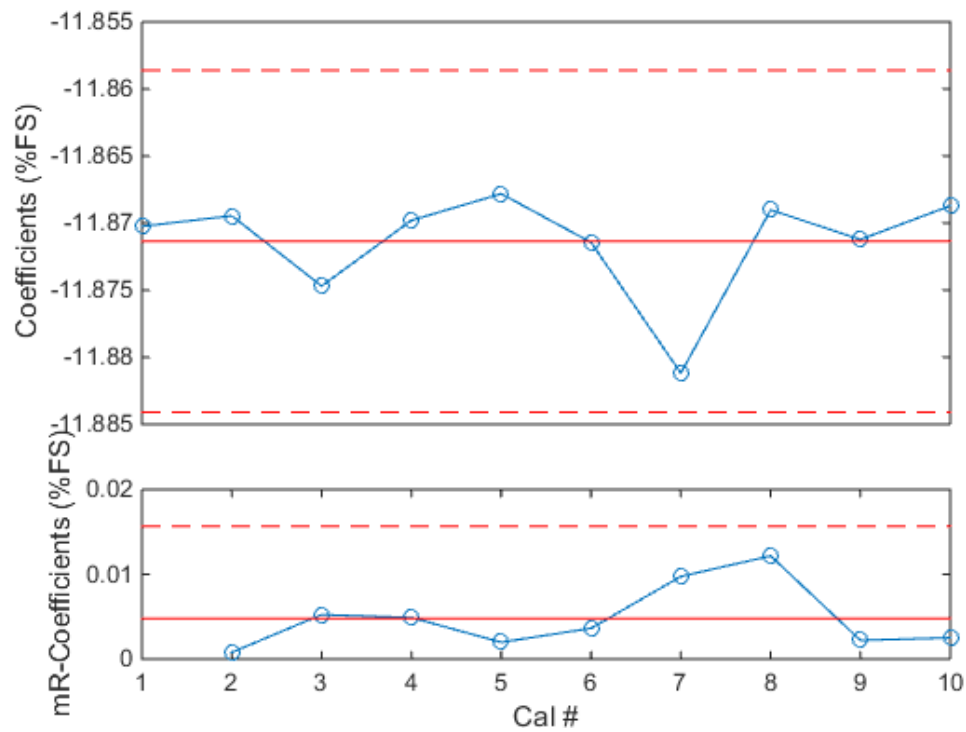


Figure 162: Yawing Moment Model - NF First Order Coefficient (MC-60E)

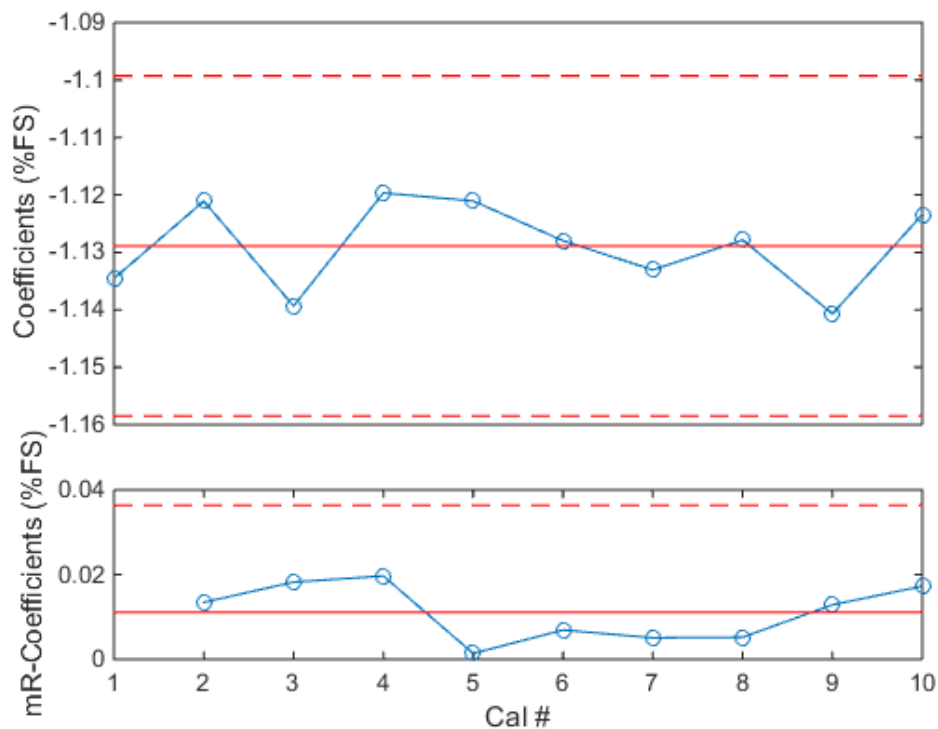


Figure 163: Yawing Moment Model - AF First Order Coefficient (MC-60E)

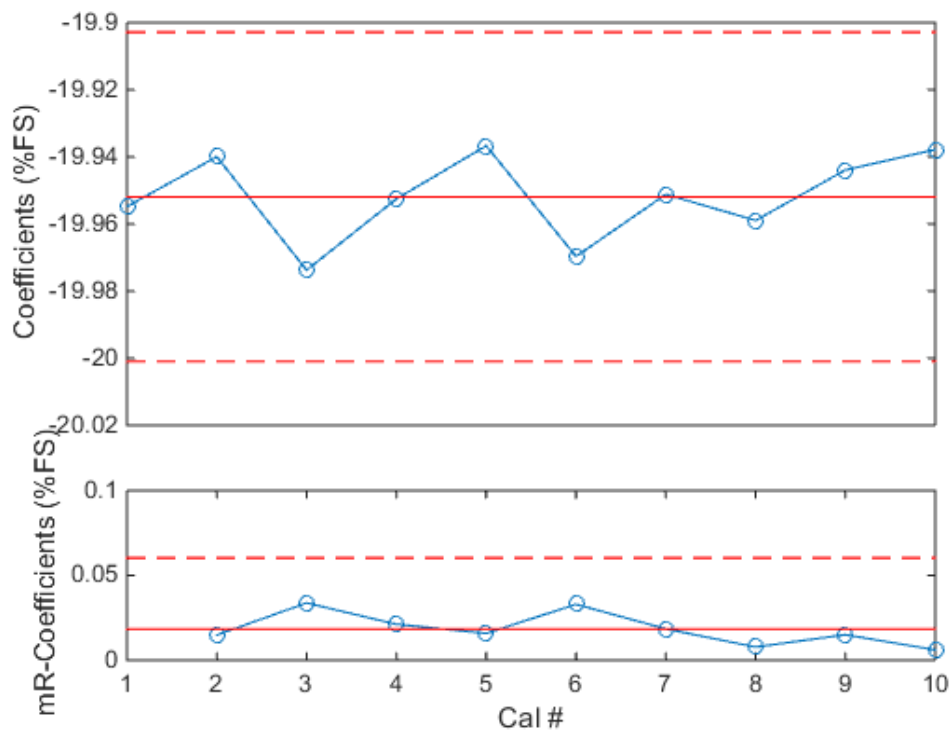


Figure 164: Yawing Moment Model - PM First Order Coefficient (MC-60E)

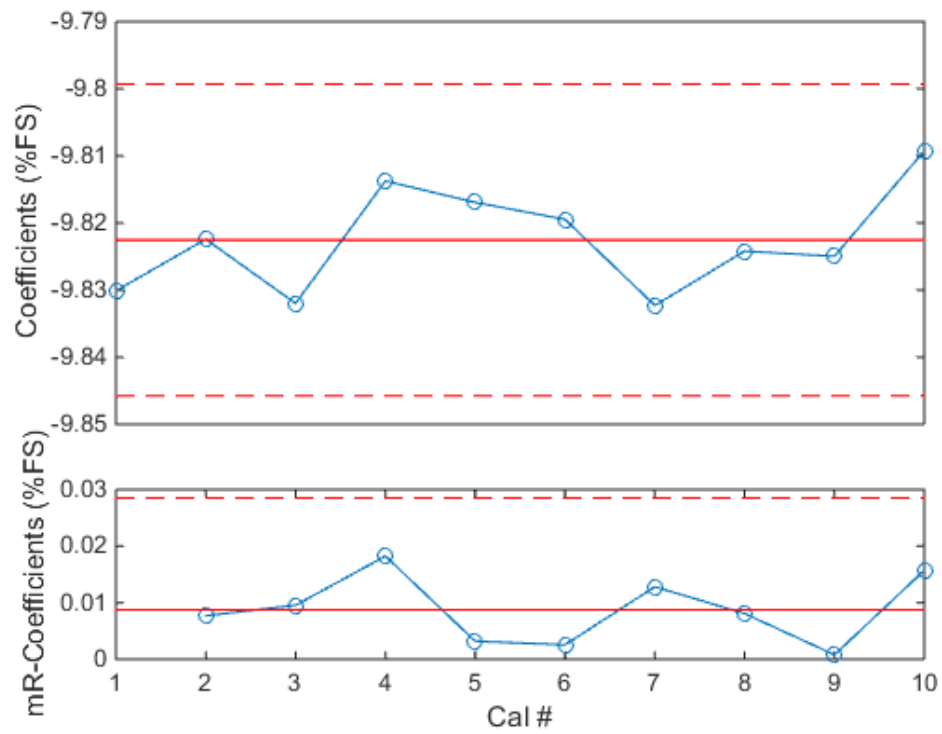


Figure 165: Yawing Moment Model - RM First Order Coefficient (MC-60E)

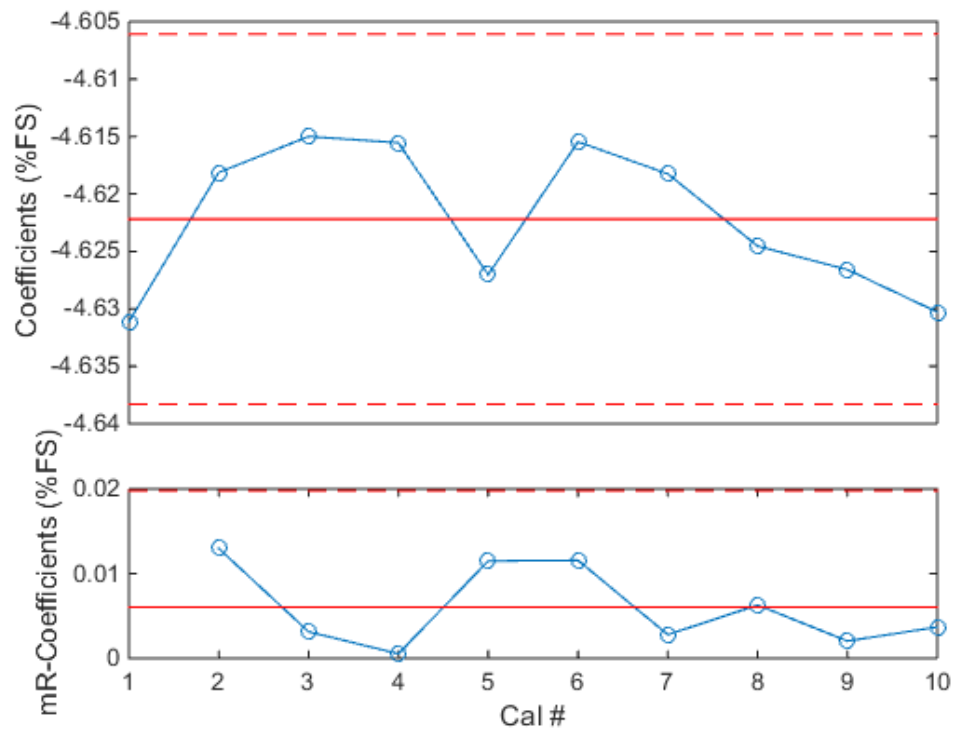


Figure 166: Yawing Moment Model - SF First Order Coefficient (MC-60E)

f. Side Force Model

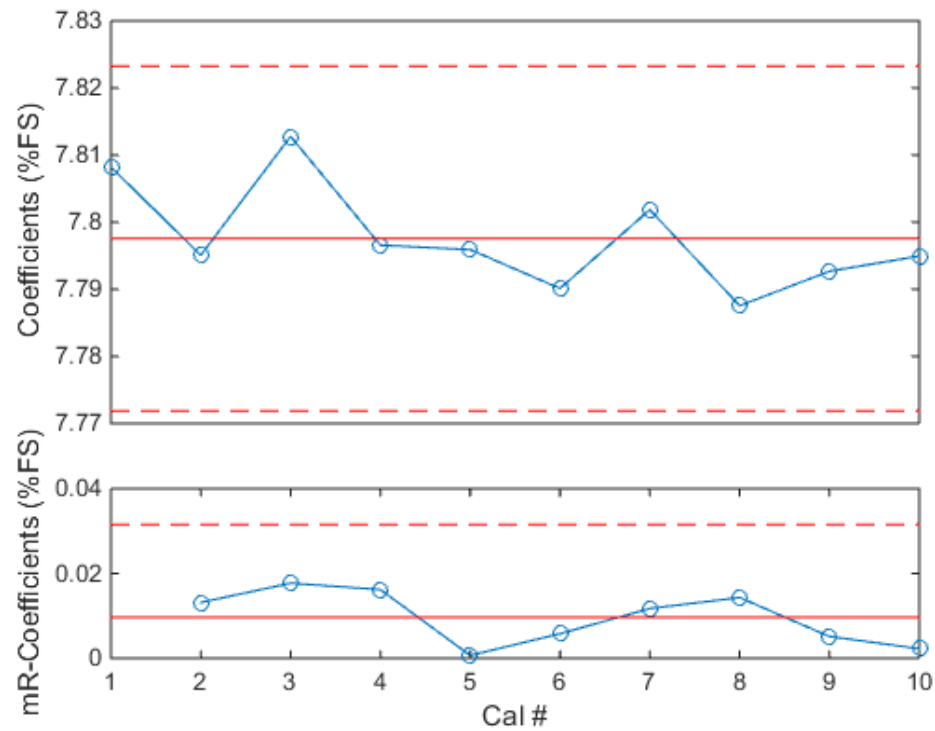


Figure 167: Side Force Model - NF First Order Coefficient (MC-60E)

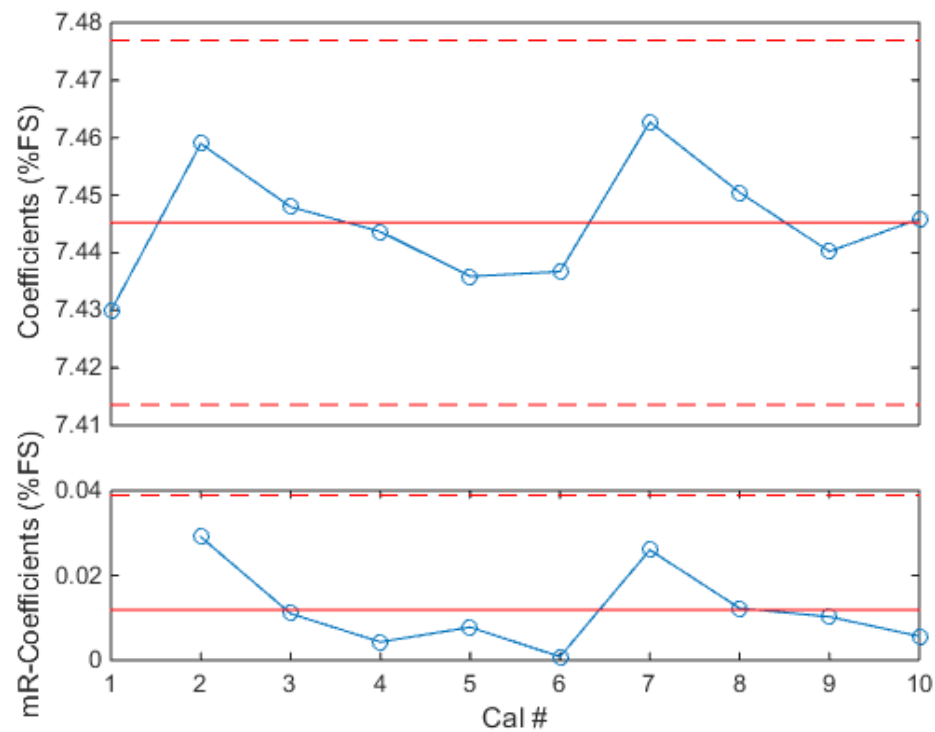


Figure 168: Side Force Model - AF First Order Coefficient (MC-60E)

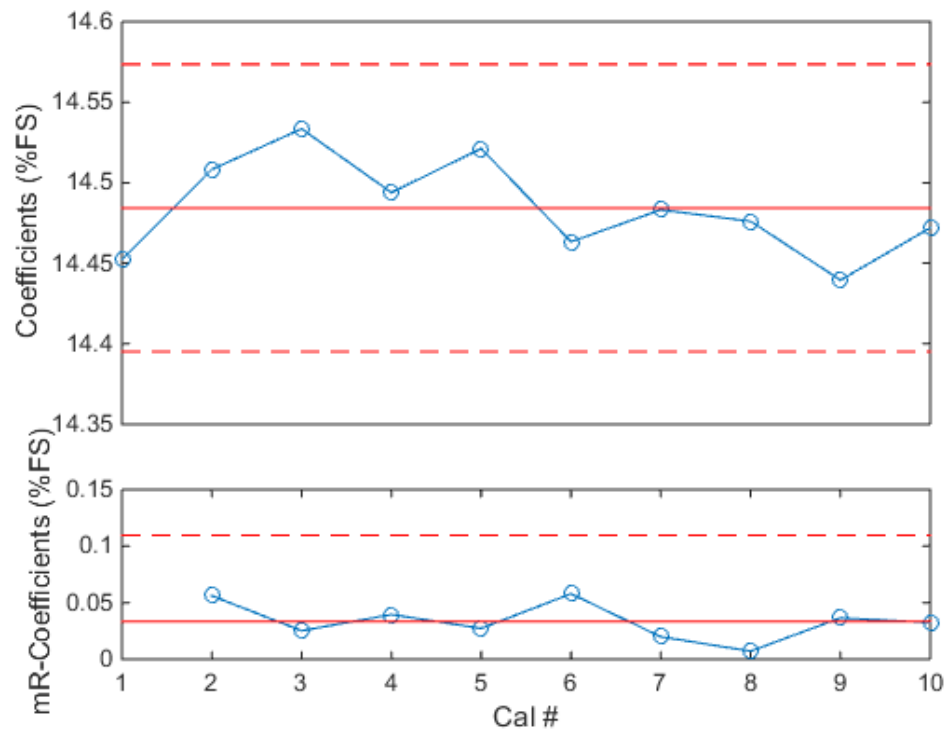


Figure 169: Side Force Model - PM First Order Coefficient (MC-60E)

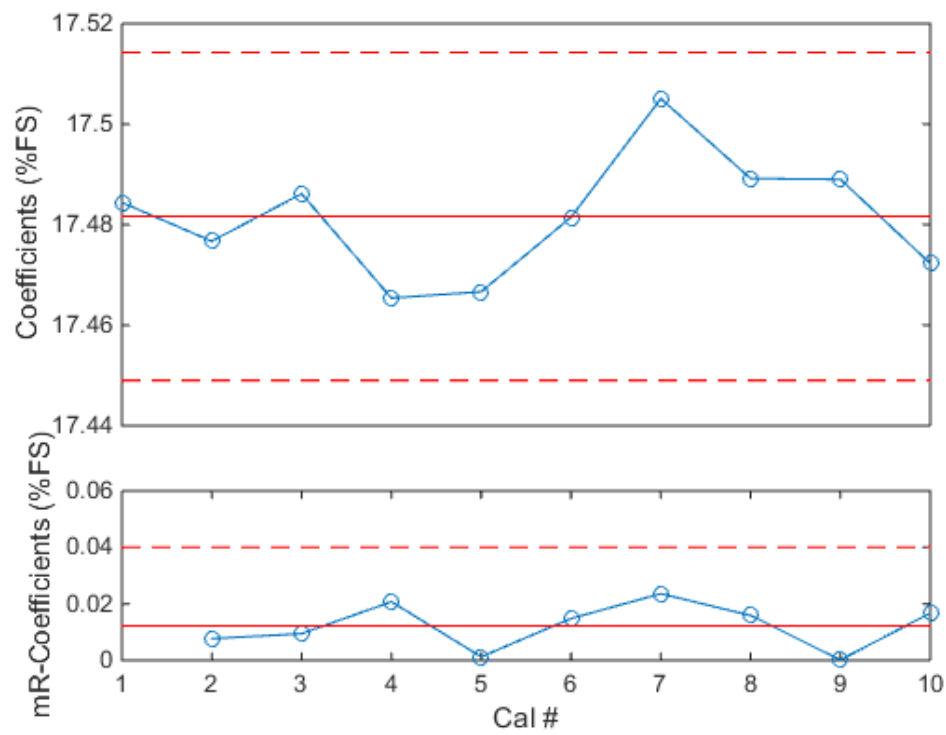


Figure 170: Side Force Model - RM First Order Coefficient (MC-60E)

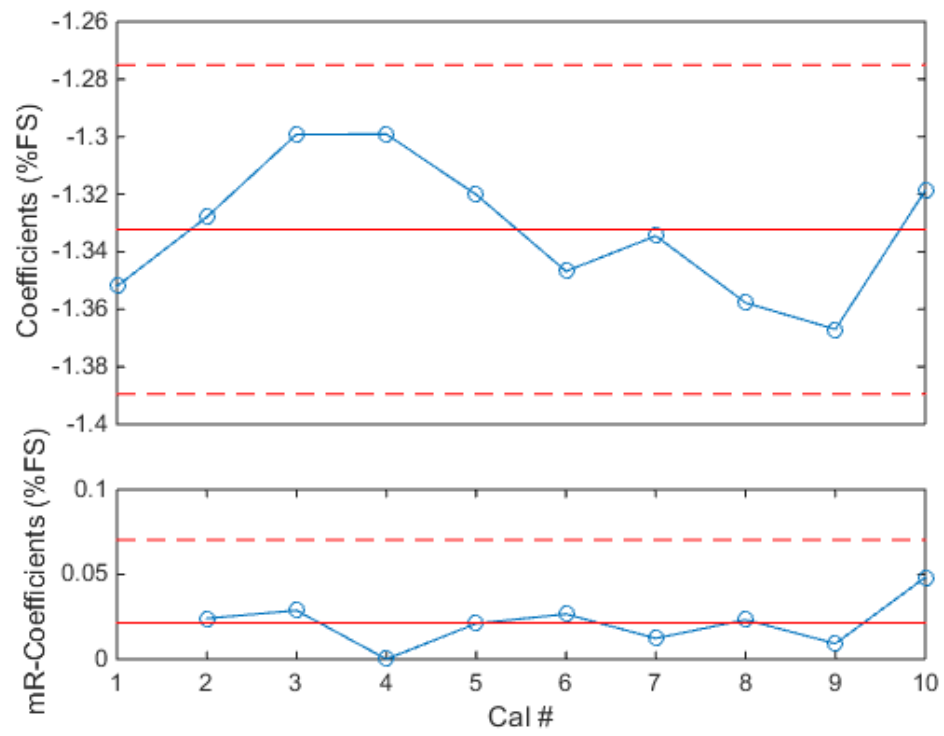


Figure 171: Side Force Model - YM First Order Coefficient (MC-60E)

VITA

EDUCATION

Old Dominion University - Norfolk, VA

Master of Science in Aerospace Engineering

May 2016

GPA: 3.850/4.0

Embry-Riddle Aeronautical University - Prescott, AZ

Bachelor of Science in Aerospace Engineering

May 2014

GPA: 3.444/4.0

ENGINEERING PROJECTS

NASA LANGLEY RESEARCH CENTER - Hampton, Virginia

Balance Calibration Study

August 2014 – Present

Conduct research into the current wind-tunnel balance calibration capabilities at NASA and evaluate the ability of multiple calibration systems to produce repeatable calibration results for a given style of wind-tunnel balance. Research is focused on understanding the long-term repeatability of a balance and identifying any possible improvements to current and future calibration systems and balance designs.

- Develop experimental load schedules for multiple wind-tunnel balance calibrations to minimize calibration time and cost while simultaneously increasing data quality
- Coordinate with test technicians regarding experimental setup, execution and troubleshooting
- Utilize and develop various data processing tools to analyze calibration data
- Apply statistical process control and other rigorous methods to track multiple calibration metrics
- Develop a comprehensive analysis to compare all historical data to fully understand the long-term repeatability of each balance on each of the calibration systems
- Actively evaluate and validate data, screen for anomalies and make informed decisions regarding the progress of the study and report findings to project leads

In-situ Load System (ILS) - ILS Validation Testing

June 2014 – July 2015

ILS addresses the issue of validating a wind-tunnel model system by providing a means to apply check loads to the balance while it is inside of the model. Multiple aspects of the system and equipment still require in-depth analysis before the system can be used confidently in a ground test environment. Current work is focused on integrating the ILS into a wind-tunnel model and is planned to be used in the wind-tunnel this year.

- Designed and executed an experiment to estimate the deflection of the ILS under all loaded conditions
- Estimated a response model to account for the deflection and added the compensation to the data acquisition software
- Lead integration of a load cell into the ILS which required integration of hardware and data acquisition software
- Derived a response model for the load cell which allowed accurate loads to be estimated within .05% full scale
- Eliminated the need for calibrated weights while simultaneously increasing the accuracy of the test data

OLD DOMINION UNIVERSITY – Norfolk, Virginia

SAE Aero East – Design/Build/Fly Competition

July 2014 – May 2016

Design, build and fly a remote-controlled aircraft at a competition with the goal of lifting maximum total weight. Points at the competition are awarded based upon successful flights and payload weight.

- Advised and collaborated with a multidisciplinary team of graduate and undergraduate students
- Designed aerodynamic and stability requirements for the aircraft to ensure stable and responsive characteristics
- Ensured quality of engineering, system-wide, by checking calculations, making design recommendations and validating the construction quality of the aircraft
- Participated in flight tests of the aircraft by preparing the payload configuration, reviewing post-flight data and comparing to pre-flight estimates of lift, stability derivatives, engine output, battery life and airspeed.
- ODU aircraft took 13th place overall at the 2015 competition out of 60 other teams, carrying just over 18 pounds.
- Further testing has proven the aircraft is capable of lifting at least 25 pounds
- Continued research and development is underway to prepare for 2016 competition

Manufacturing Influenced Design – Systems Engineering

August 2015 – December 2015

Along with Georgia Tech and NASA Marshall Space Flight Center, a systems engineering team provided support for developing a computer-based tool to generate cost and schedule estimates for design of new space launch systems. The tool will improve estimates of cost and schedule by taking advanced material and manufacturing considerations into account during the conceptual design phase and seek to reduce project overruns, thereby reducing the risk of project cancellation.

- Worked with an interdisciplinary team to generate systems requirements and architectures from functional requirements and customer requirements
- Developed a systems engineering management plan which will guide the project from concept phases, through development and operational phases to disposal of the system
- Implemented a complex systems engineering analysis, including a quality function deployment chart, overall evaluation criterion, morphological matrixes, etc. to determine the optimal specifications of the tool itself
- Collaborated with the customers to ensure satisfaction with the operational and interface specifications and results

EMBRY-RIDDLE AERONAUTICAL UNIVERSITY – Prescott, Arizona

General Atomics - Predator Modifications

August 2013 – May 2014

With a team of senior engineers and working with General Atomics, research was conducted into adding spoilers to the Predator B UAV for possible future applications of the aircraft and subsequent U-Class designs. The design was required to be entirely externally mounted and to have minimal impact on the current configuration of the aircraft.

- Lead coordination with General Atomics to ensure agreement for all system requirements and test activities
- Conducted flight and wind-tunnel tests to determine the best size and location of the spoilers
- Generated computational fluid dynamics models to match known configurations to predict future configurations

- Created a data base from the collected data and estimated response model for future reference in spoiler sizing
- Modifications were successful in doubling the power off descent angle of the aircraft in small scale flight tests

Study of Self-Contained Circulation Control

May 2013 – May 2014

Conducted wind-tunnel research into circulation control using a self-contained motor mounted inside of a wing. A jet of air was ejected out the trailing edge in an attempt to augment the lift while simultaneously eliminating the drag created by the separating jet by creating suction near the trailing edge.

- Designed a prototype wind-tunnel model to both contain fully contain the motor and produce maximum jet momentum
- Modeled the wing in CATIA and rapid prototyped the wing in Embry-Riddle's 3D printers
- Executed a series of experiments designed to characterize the behavior of the model with the trailing edge jet
- Analyzed data from multiple experiments at multiple Reynolds numbers and jet velocities to determine aerodynamic effects from the circulation control
- Experimental testing produced a maximum of 54% increase in lift and no appreciable increase in drag

EXPERIENCE

STUDENT ENGINEERING CONTRACTOR

June 2014 – Present

**NASA Langley Research Center/ National Institute of Aerospace - Hampton, Virginia
Engineering Directorate – Systems Engineering and Engineering Methods Branch**

- Work with multidisciplinary team of statisticians and engineers on projects related to experimental design, calibration, characterization, and rigorous data analysis
- Support research and develop new methodologies to further NASA's current calibration services
- Design experiments to efficiently and effectively characterize ground test instrumentation and other metrology devices
- Use rigorous statistical methods to analyze, understand and make informed decisions from data
- Develop procedures, processes, tools, algorithms and standards relevant to characterization of metrology devices
- Document, report and present findings in the form of internal papers, AIAA publications, colloquial presentations and a master's thesis specifically regarding the balance calibration study

STUDENT RESEARCH ASSOCIATE

May 2013 – May 2014

Embry-Riddle Aeronautical University – Prescott, Arizona

- Conducted research and hands-on wind-tunnel experimentation in the field of experimental aerodynamics
- Collaborated with other student researchers to efficiently meet research goals and deadlines
- Designed and fabricated 3D models and test articles

- Collected and processed wind-tunnel data
- Documented all findings in the form of internal research papers, presentations and journal publications
- Gained valuable experience in conducting experimental research and testing
- Developed extensive understanding of aerodynamic concepts as well as small scale wind-tunnel operations

CAMPUS ACADEMIC MENTOR

August 2011 – May 2014

Embry-Riddle Aeronautical University - Prescott, Arizona

- Planned curriculum and co-instructed a 101 course with a university professor
- Supported and helped freshman acclimate to college life, develop efficient study habits, and be successful in both curricular and non-curricular activities
- Developed strong leadership, instruction and constructive communication skills

PUBLICATIONS

- Long-Term Study of Wind-Tunnel Balance Calibration History using Statistical Process Control
AIAA Summer Conference 2015, Presenter
- Effect of Taper Ratio at Low Reynolds Number - AIAA Journal of Aircraft 2015, Vol.52