2013

Regression-based Statistical Change Point Analysis for Damage Localization

Mallory Beth Nigro
Lehigh University

Regression-based Statistical Change Point Analysis for Damage Localization

by

Mallory Nigro

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Structural Engineering

Lehigh University

September 2013

This thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science.

_____
Date

_____
Shamim N. Pakzad
Thesis Advisor

_____
Sibel Pamukcu
Chairperson of Department

Acknowledgments

This research was conducted at Lehigh University's Advanced Technology for Large Structural Systems (ATLSS) in Bethlehem PA. Research funding is partially provided by the National Science Foundation through Grant No. CMMI-0926898 by the Sensors and Sensing Systems Program, and by a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA). This financial support is gratefully acknowledged.

I would like to thank Dr. Shamim N. Pakzad for his help and support through my undergraduate and graduate studies. His faith in my abilities was a driving force for me to finish my degrees. I would also like to thank Siavash Dorvash and Golnaz Shahidi for their guidance as mentors and co-researchers throughout the process. The computer technical support provided by Peter Bryan is much appreciated. I would also like to thank Andrew Thorsen for his help in making the Graphical user Interface to combine all of my research efforts.

Finally thank you to my loving parents and sister for their continued financial, emotional, and *technical* support throughout my education. And finally thank you to Kristen, for being my partner in crime for the past year. (We did it!!)

Table of Contents

LIST OF TABLES

TABLE OF FIGURES

Abstract

Structural health monitoring (SHM) research has become a vital tool in maintaining the integrity of structures that has been refined over the years. There are numerous methods for damage detection and localization; yet some are not efficient. For example, researchers have used dynamic properties as damage features to monitor a structure because they change in the presence of damage; however, these methods are global in nature. Research in improving them (i.e. having automated, statistical monitoring techniques) is critical to the advancement of the civil engineering field. This thesis presents the implementation of damage detection methods using an experimental structure. Damage features are created from linear regression models and are utilized in control charts to localize damage because they represent the changing properties of a structure in the event of damage. Therefore, this thesis evaluates the performance of different damage features and change point analysis methods in detecting and localizing damage.

1. Introduction

1.1 Overview

Damage detection is a fundamental element of SHM practice and is one of the most challenging research tasks for the civil engineering community. Literature shows numerous techniques for damage detection which are classified in different ways. Some approaches involve non-destructive evaluation (NDE) techniques. Many of these rely on the skill and experience of a trained inspector and also require a *priori* knowledge about the structure's properties; on the other hand, some are data-driven techniques which rely on measurement and monitoring and can be classified into model-based and model-free methods (Dorvash et al. 2012). Many approaches use modal parameter identification with data in time and frequency domain, to establish damage sensitive features such as modal frequencies (Yao and Pakzad 2013), mode shapes (West 1984) or mode shape curvatures (Pandey et al. 1991). A review of these methods can be found in Ewins (1984), Doebling (1998) and Salawu (1997). However, these methods for damage detection are generally unable to detect local damage (Alvandi and Cremona 2006); these features are not sensitive enough to the changes in local elements of the structure hence they require high signal to noise ratio of the measurement data as well as moderate damage levels (Farrar et al. 1994).

As an answer to the shortcomings of these damage detection techniques, model-free approaches can be employed which use local responses of the structure through sensor networks. These localized networks have been used on various types of structures such as large-scale steel beam-to-column moment connection, (Dorvash et al. 2013), large-scale moment connections (Labuz et al. 2011), long-span bridges (Pakzad 2010), concrete

structures (Nishikawa et al. 2012), composite and reinforced concrete bridges (Cruz et al. 2008), and highway bridges (Hu. et al. 2013). Considering the recent technological advancements in sensing technology, affordable sensors (such as wireless sensors) can address the need for higher number of outputs. Hence, with an increase in the number of output channels, the accuracy and resolution of the detection of change can increase without a loss for efficiency in time for computation. With an increase in the amount of data from these dense sensor networks, output only algorithms can be developed so there is no need for input parameters such as excitation. Furthermore, this data could be used in damage detection techniques in which there would be no need for a *priori* knowledge of the structure's properties or suspected location of damage.

In this thesis, data is collected from a dense sensor network and used in an algorithm to create damage features to use in control chart analysis. These features are produced from linear regression models including single variate, collinear, Auto Regression (AR) and Auto Regression with Exogenous term (ARX). They can either be the influence coefficients, parameters derived from these influence coefficients, or the residuals of the different types of linear regression models. These features can be correlated to a location on the structure through the dense sensor network. Hence, they can be used to localize damage on a structure. While localized model-free approaches, like the one used in this thesis, are easy to implement and effective in reflecting the changes in the structure's behavior, they are dependent on statistical analysis to determine the significance of the changes in the data.

Literature presents different statistical approaches which are developed for detection of changes in observations for different applications and disciplines including structural

damage detection such as in Amiri et al. (2011) and Nishina (1992). However, despite their importance in damage detection, control charts and statistical frameworks are not properly utilized in this area. On the other hand, most of the existing change point analysis techniques were originally developed for applications in the stock market, as in Taylor (2011), or industrial engineering processes, as in Zamba and Hawkins (2006); therefore utilizing them in structural damage detection needs performance evaluation and may also require particular customizing for adaption. Additionally, there are different control statistics based on the parameter being monitored. For example, some are sensitive to the changes in the mean of a data set and some are more sensitive to the change in the variance of a data set. These different types of control charts can be used to monitor different damage features. For instance, it is assumed that the mean of the influence coefficients will change from an undamaged set of data to a damaged set of data; meanwhile the variance of the residuals is subject to change in the presence of damage.

For this thesis, the response of a scaled steel frame is recorded from two states: the first is a baseline healthy state of the structure, and the second is an unknown state. This procedure creates two sets of acceleration data that would be taken from a structure pre- and post- a damaging event. After the data is collected, damage features from univariate and multivariate linear regression models can be estimated and used to localize damage. To achieve this, control charts are used. The damage indicators from the univariate regression (i.e. single variate and collinear regression) are used in different control schemes than the multivariate indicators (i.e. ARX and AR regression parameters); yet both are used to make a distinction, if any, between an undamaged and unknown state of

the structure using the data from a sensor network in order to correlate locations with acceleration.

1.2 Research Objectives

The contribution of this paper is in establishing and comparing the effectiveness of different control statistics, damage sensitive features, and threshold methods in detecting damage for the application of structural health monitoring. Performance of different control statistics are evaluated through implementation on data obtained from laboratory frame. The following objectives are established:

1. To develop a damage detection algorithm using linear regression parameters as damage features. Three different damage features are used: the influence coefficients of the regression, the regression residuals, and the angle between an undamaged and damaged linear regressed line. Four different types of regression models are used: single variate, collinear, Auto Regression (AR) and Auto Regression with Exogenous term (ARX).
2. To analyze and compare the effectiveness of different control statistics for detecting a significant change in the mean and variance of the damage features. Ten different control statistics are used based on the damage feature utilized.
3. To use statistical frameworks for generating confidence level threshold boundaries to determine the significance and location of damage. Bootstrapping is employed as well as simple Shewart thresholds that are used in many statistical process control techniques. These threshold types are used with different control statistics depending on applicability.

4. To instrument a scaled steel frame with a wired sensor network to validate the methods from different damage cases. A simulated model is also created to verify the response of the experimental set up.
5. To create a Graphical User Interface (GUI) so that the control charts used in this thesis can be available to researchers monitoring a process in different fields of work. The control charts should be able to be used in industries besides civil engineering structural health monitoring (most control analysis is used in industries other than SHM); therefore, in creating a GUI, these methods can be available to the public. It should be user interactive so that the researchers will be able to define their own confidence levels and charting properties such as title, axes labels etc.

1.3 Scope of Research

In order to accomplish these research objectives, the following work was completed. A damage detection algorithm was produced based on linear regression models and control chart statistics. Damage features were extracted from the linear regression models. These were utilized in control statistics that were generated and customized for the particular adaption to use for damage detection. To obtain initial damage features, a simulated model was produced in SAP2000. Then an experimental two-bay moment frame was developed and instrumented with a dense network of wired accelerometers. Data collected from both the simulated and experimental model were used in the algorithm to obtain damage features. These features were compared to assure the accuracy of the simulated model and overall effectiveness of the algorithm. Then, the results from the

experimental structure were used in control statistics. This thesis analyzes the effectiveness of the damage sensitive features in the control charts.

1.4 Organization of Thesis

This thesis contains 6 chapters as follows.

- Chapter 1 presents an introduction to the study providing a general overview, research objectives, scope and organization of thesis.
- Chapter 2 introduces the development of the localized damage detection algorithm in which relevant literature is reviewed. Structural theory is also introduced which forms the basis of using linear regression parameters as damage features. Afterwards, the linear regression models are presented and the features are derived.
- Chapter 3 is an introduction to the change point detection statistics in which relevant literature is reviewed. Bootstrapping is also explained as a method for threshold generation. Simple Shewart bounds are presented along with the specific control statistics for which they will be utilized.
- Chapter 4 presents the test specimen used to validate the damage detection methods. The simulated model is also presented along with the model updating. This model has the same physical properties as the experimental specimen. Testing procedures, as well as equipment used, are discussed.
- Chapter 5 shows the results from the experimental model for each damage feature and control statistic. The effectiveness of each statistic and damage feature pairing is analyzed.

- Chapter 6 includes a summary, conclusions and possible future work that can be dedicated to the subject of this thesis.

2. Development of Localized Damage Detection Algorithm

2.1 Introduction

As stated, there are many different parameters that can be used as damage sensitive features in control charts. These parameters should be sensitive to the changing properties of a structure in the event of damage. For example, Ghosh et al. (2006) uses the ratio between two different univariate linear regression coefficients while Lou et al. (2004) presents a novel time-series model in the form of a prediction model of auto-regressive with exogenous input. In this case, the standard deviation of the residual error is found to be the damage sensitive feature. Additionally, many researchers have used different methods for estimating these damage features. In order to find and use dynamic characteristics of a structure as damage features, Huang (2001) proposes a procedure that uses the ARV model for numerical simulations of a six-story shear building subjected to white-noise and low-pass filtered white-noise input. Similarly, He et al. (1997) uses M-ARMA and M-AR(P) models to find the modal parameters of a water transmission tower. Furthermore, Hung et al. (2004) identifies modal parameters from measured input and output data using a vector backward auto-regressive with exogenous model.

The damage features studied in this thesis come from linear regression coefficients produced by an algorithm called Influenced-based Damage Detection Algorithm (IDDA) developed by Dorvash et al. (2012). These damage features are shown to be effective ways of detecting damage in a structure because they are sensitive to the changing properties of a structure. It is essential in structural damage detection that the features used in the model reflect a unique pattern to the state of the structure (Qiao et al. 2012). This algorithm correlates the response of a structure at various locations by creating

influence coefficients from linear regression models based on output of a dense sensor network. When damage occurs, the relationship between responses at different locations changes, which will be reflected in the parameters and indicate the existence of damage. The position of damage can then be identified by correlating the location of each sensor to the damage feature.

2.2 Structural Behavior

2.2.1 Linear Behavior

A realistic scenario for implementation of the damage detection algorithm on real-life structures is to collect the ambient response of the structure before and after a catastrophic event. The algorithm introduced is only suitable for a linear damage defined by Doebling et al. 1998 as "the case when the initially linear-elastic structure remains linear-elastic after damage". Thus, it is acceptable to assume that the behavior of the structure is within the linear elastic range in which small angle theory applies which corresponds to negligible rotations (Dorvash et al. 2012). It is then also valid to use linear regression methods to model the behavior of the experimental model.

2.2 Linear Regression Parameters as Damage Indicators

2.2.1 Single Regression Model

The simplest linear regression is the single variate model. It relates the response of one location to another location at the current time step. This version of the model can be represented using Eq. (2.2) below.

$$y_j(n) = \alpha y_i(n) + \beta(n) + \epsilon(n) \quad \textbf{(2.2)}$$

Which correlates the response at node j to current response at node i through α at time step n with intercept β and residual ϵ. Since the effects of previous time steps are removed from the equation, the intercept (β) is added into Eq. (2.2) to account for the initial conditions. To verify the accuracy of the estimated coefficient, evaluation accuracy (EA) and estimation error parameter (γ) are used. These parameters ensure that the influence coefficients are correctly reflecting the properties of the frame. They are derived in Labuz 2011.

EA_{ij} is defined as the product of influence coefficients α_{ij} and α_{ji} as:

$$EA_{ij} = \alpha_{ij}\alpha_{ji} \quad \textbf{(2.3)}$$

The influence coefficient for location j regressed onto location i should be the reciprocal of the influence coefficient i regressed onto j if there is minimal residuals and a zero intercept. Therefore, a value for their product close to unity signifies a strong accuracy of estimation and a product of less than unity corresponds to progressively higher values of the noise and nonlinear behavior of the physical structure.

The normalized estimation error, γ_{ij} can be calculated by:

$$\gamma_{ij} = \frac{\sigma_{\alpha_{ij}}}{\alpha_{ij}} \quad \textbf{(2.4)}$$

As noted above, α_{ij} is the influence coefficient between nodes i and j, and $\sigma_{\alpha_{ij}}$ is the standard error of the influence coefficient estimates. $\sigma_{\alpha_{ij}}$ is estimated by Eq. (2.5):

$$\sigma_{\alpha_{ij}} = \frac{\sigma_e}{(\sum y_i^2)^{1/2}} \quad \textbf{(2.5)}$$

Here, σ_e is the standard error of the estimation residuals, (i.e. the standard deviation of the vector was obtained by subtraction of the estimated response from the true response) and y_i is the response at node i which is regressed with respect to the response at node j (y_j). Considering that the response has a zero mean, the denominator of Eq. (2.5) is simply the standard deviation of the response at node i. This equation is utilized to show that a low standard deviation of the influence coefficients, which corresponds to a more accurate estimate, produces a low estimation error. Hence, the closer γ_{ij} is to zero, the less error is associated with the estimated damage indicator α_{ij} (Labuz 2011).

The derivation and validation of this simplified mathematical model can be found in Dorvash et al. (2010) on a small scale beam-column connection. One of the focuses of this thesis is enhancing IDDA with the use of a larger structural system as opposed to a structural component. Additionally, since this damage feature has already been proven to detect and localize damage, it is used as a basis for comparison and derivation of the proceeding damage features discussed in this section.

2.2.2 Auto Regression

The single variate regression can be expanded to include past and present time steps. It is suspected that more information about the system can be included in this way by including dynamic effects. This Auto Regressive with Exogenous term (ARX) model can be written as:

$$\sum_{p=0}^{P} y_j(n-p) = \sum_{q=0}^{Q} \alpha_{iq} y_i(n-q) + \varepsilon(n) \quad \textbf{(2.6)}$$

where y_j and y_i are outputs at locations j and i respectively, a_{iq}'s are ARX coefficients, $\varepsilon(n)$ represents the residuals, n is the time index, and P and Q are orders of the

autoregressive and exogenous parts of the ARX model, respectively. Derivation and validation of this formulation can be found in Yao et al. (2012). These damage features are similar to those found in literature. Sohn et al. (2001) and Yao and Pakzad (2010) also use a similar AR and ARX model for feature extraction and pattern recognition. Similarly, Qiao and Esmaeily (2011) use an ARX regression model for time-history damage features.

This ARX model can be simplified to include just one location on a structure. In effect, this regression may produce more localized results if only one location is involved in the model. Response relationships at one location throughout time can be established using an Auto Regressive model as:

$$y_j(n) = \sum_{q=0}^{Q} \alpha_q y_j(n-q) + \varepsilon(n) \quad \textbf{(2.7)}$$

In this formulation, y_j $\bar{y}(n) = [y_1(n) y_n(n) \dots y_n(n)]$is the output at location j, a_q's are AR coefficients, $\varepsilon(n)$ $\bar{e}(n)$represents the residuals, n is the time index, and Q is the order of the autoregressive parts of the AR model. The Yule Walker method is used to produce these AR coefficients. This method multiples the linear equation by y_j $-k-1$, k being the time lag. Then the expectance is found and the results are normalized to form a square coefficient matrix with full rank and symmetry so that the inverse can be found.

The model order of the AR and ARX models must be determined before the linear regression parameters can effectively be used in damage control charts. The accuracy of the two regression models is dependent on the selected model order based on the raw data from the localized sensor networks. While higher model orders, in general, deliver more details of the system and reduce the estimation bias, it is always desired to keep the order at the minimum level to avoid over-parameterization. One way to establish the model

order is to minimize the Akaike's Information Criterion (AIC) which is used in Friedlander et al. (1984) and Figueriedo et al. (2011). And defined in Bozdogan (1987) as,

$$AIC(k) = -2log(\widehat{\theta_k}) + 2k \tag{2.8}$$

In Eq. (2.8), k is the number of parameters in the statistical model and $\hat{\theta}_k$ is the maximized value of the likelihood function.

2.3 Tri-variate Regression Model

The single regression model can then be modified even further to generate the tri-variate model. This model is used to correlate the response of three different locations at the current time step. It can be written as

$$y_k(n) = \alpha_{ik}^{j} y_i(n) + \alpha_{jk}^{i} y_j(n) + \beta_{ijk} + \varepsilon(n) \tag{2.9}$$

In this formulation, y_j, y_k and y_i are the outputs at the three locations involved in the regression. The α's are tri-variate coefficients, with β_{ijk} representing the intercept and $\varepsilon(n)$ representing the residuals. It will be shown later that this model is not robust to noise levels and may over parameterize the system. Therefore a different model, collinear regression, is presented in the next section as a different method for comparing three different locations within one regression model.

2.3.1 Collinear Regression Model

Another way to involve three different locations is to modify the single regression model even further to generate the collinear regression model. In order to correlate three locations on a specimen without over parameterizing the system, collinear regression can

be used in which the definition of the regressors is changed. There are many different types of regressors that can be used in collinear regression models. For this implementation y_i in Eq. (2.2) is changed to the average of two outputs. In effect, the mathematical model would be calculated from Eq. (2.10) below.

$$y_k = \alpha_{ijk} \frac{(y_i + y_j)}{2} + \beta + \epsilon \tag{2.10}$$

Here, an additional location's response output, y_k, can be included to create the new coefficient α_{ijk} . The effectiveness of collinear regression linear regression parameters is analyzed and compared to the other regression model parameters presented above.

2.3.2 The Damage Features

There are three main types of features that are analyzed. Two are used to test the null hypothesis that the mean of two populations are equal; the other is used to test the null hypothesis that the variance of two populations is equal. The first of these is the influence coefficient, α, obtained from all of the regression methods discussed above. This will be called the Alpha Coefficient. The second is called the Angle Coefficient. This coefficient does not capture the slope of a linearly regressed line as α does, but it measures the angle between two different lines. In other words, for damage detection methods, instead of measuring the difference in slope between a healthy state linearly regressed line and an unknown-state-line of a structure, the angle between the two lines can be compared to detect change as well. In effect, the Angle Coefficient can be written as

$$\Gamma = cos^{-1} \left| \frac{v * v\prime}{\|v\| \|v\prime\|} \right| = cos^{-1} \left| \frac{\alpha\alpha\prime + 1}{\sqrt{\alpha^2 + 1}\sqrt{\alpha\prime^2 + 1}} \right| \tag{2.11}$$

Here v and v' correspond to a vector $[-1, \alpha]^T$ for an undamaged state and an unknown state respectively. In this formulation, α is the respective influence coefficient from the linear regression models. For single and collinear regression models, α will be a scalar value. However, in AR and ARX regression models, the linear regression parameter is actually a set of coefficients. Hence, $v = [-1, \alpha_1, \alpha_2 \ldots \alpha_{2p+1}]^T$. Here, p is the model order chosen using the AIC parameter. These two coefficients, α and Γ, can be tested for a change in their mean.

However, the variance will be affected in the third feature, the residuals from all the linear regression models. The mean of these features is deemed to be zero and unchanged with damage; however, their variance can change because of a damaging event. The residuals are found by convolving the coefficients of an undamaged state with the data collected from the unknown, or damaged state of the structure, and the coefficients from the damaged state with the data collected from the undamaged state of the structure.

Because the alpha and angle coefficient's mean can change in a damaging event while the residual's variance change in a damaging event, these damage features lend themselves to different control statistics. The next chapter will discuss some control statistics are used to measure the change in the mean of a process and some are more sensitive to the change in variance of a process; hence, they are paired with the appropriate damage feature.

3. Statistical Framework of Control Charts

3.1 Introduction and Background

Control statistics can be used to monitor a change in a process. However, there are two causes of variation that could occur in a process: *common causes* and *assignable causes*. For instance, in civil engineering, a structure's properties could change slightly with passage of time (common causes) and is not damaged, or a structure could be damaged during an event (assignable cause). In order to distinguish between the two, control charts can be used to provide boundary limits for detecting a change in a process. Once the boundary is crossed, the change in the process can be denoted as *out of control* from an assignable cause and, in the case of civil engineering, a structure can be deemed damaged.

Literature presents several different types of control statistics that can be used for change point detection in different processes. Univariate and multivariate statistics can be used to detect a single change or multiple changes in the mean or variance of the data (Amiri et al. 2011). As an early effort in this area, the standard univariate Shewhart $\bar{X}$ control chart was introduced in 1924 by Walter Shewhart as simple tools to be used by workers in production lines (Wilcox 2003). Since then, control schemes have found widespread application in different disciplines including civil engineering. For example, in order to quantitatively address the uncertainty in measured response data, Sohn et al. (2000) applied Shewhart control charts to auto regressive coefficients and successfully detected the irregularities associated with damaged structures. Other control statistics have since been introduced.

Challenges in univariate quality control occur when one observes a set of quality characteristics that have components with the potential to be interrelated. One major flaw in using the mentioned univariate control statistics is that they can only monitor one variable at a time. It is common in civil engineering that more than one variable would need to be examined. Although it could be argued that univariate control charts could be applied independently to each component of the multivariate data, misleading results may be obtained in some cases due to failure to allow for the inherent relationship among the components of the multivariate data (Zamba and Hawkins 2006). Therefore, this thesis also explores multivariate control statistics.

This thesis aims to provide a better understanding of the performance of different change point detection methods in civil engineering damage detection through implementation and validation of different approaches. A flow chart of the certain damage features that can be utilized in the specific control charts is shown in Figure 3.1.

3.2 Exponentially Weighted Moving Average (EWMA)

First introduced by Roberts (1959), the Exponentially Weighted Moving Average (EWMA) is easy to implement and interpret and has optimal properties in forecasting and control applications; yet, only recently has its value in other detection schemes been recognized. Lucas and Saccuci (1990) used and enhanced this statistic in order to compare it to the CUMSUM statistic. Steiner (1999) used EWMA statistics because of its ability to detect small persistent process shifts and advanced its use by producing time-varying control limits. Additionally**,** Macgregor and Harris (1990) identified two uses of the invariant EWMA statistic.

The EWMA statistic (Z) is an exponentially weighted average of all previous observations shown in Eq. (3.1):

$$Z_i = \lambda Y_i + (1-\lambda) Z_{i-1} \quad 0<\lambda<1 \tag{3.1}$$

where Y_i are components of the data set and typical values of λ are between .05 and .25 for quality monitoring techniques. This statistic is used with upper control limits and lower control limits to create a threshold of change. The control limits that are pre-derived for this statistic from Lucas and Saccuci (1990) are based on the asymptotic standard deviation of the control statistics as

$$LCL = \mu - L\sigma \qquad UCL = \mu + L\sigma \tag{3.2}$$

where σ is the standard deviation of all of the Z_i's, μ is the mean, and L is the factor of standard deviations away from the mean. In practice, Shewhart control limits are designed to limit the acceptable range of values within three standard deviations of the mean (Lucas et al. 1990). This value correlates the point in a standard normal distribution plot (of mean zero and standard deviation of one) in which 99.7% of the area under the curve is included. It will be shown in Chapter 5.1.1 that this control limit does not suit all data. Therefore, bootstrapping will be used for comparison of the EWMA statistic with some of the control statistics discussed that do not have previously specified thresholds.

3.3 Cumulative Sum

The Cumulative Sum (CUSUM) statistic is one approach which is commonly used for detecting change points in data sets. It is a very flexible method which can be applied to different types of data (Lucas et al. 1990). The Cumulative Sum (CUSUM) indicator, was

used by Taylor (2011) in analyzing change in U.S. trade data in order to see if and when a trade deficit occurred during the late 1980s. The CUSUM indicator is optimal in detecting shifts in the mean, yet Hawkins (1981) advanced its use for monitoring a process' variance.

All historic data is used to create the statistic meaning each value of the plot is a function of all previous values. In this study, it is implemented in damage detection schemes to distinguish if there is a change between an undamaged data set and an unknown data set. The cumulative sums S_0, S_1, S_2… S_n can be calculated by the following equations where X_1, X_2… X_n are the components of the data set (Taylor 2011).

$$S_0 = 0 \tag{3.3}$$

$$S_i = S_{i-1} + (X_i - \bar{X}_i) \tag{3.4}$$

where, $\bar{X}_1$ is the mean of all the known in control data. Shown in Eq. (3.4), the cumulative sums are not simply sums, but the sums of the differences between the values of data and the average of the data. For this reason, it should start at zero and eventually make its way back down to zero. A segment of the CUSUM chart with an upward slope indicates a period where the values tend to be above the overall average. Similarly, a segment with a downward slope indicates a period of time when the values tend to be lower than the average. It is based on the theory that as each consecutive data point is added into the cumulative sum, the distance to the average should be relatively small and forever monotonic to the actual average if there is no significant change in the data set. Therefore, in a damage detection scheme, a drastic change in the slope, or inflection point, indicates a possible damaged state of the structure. In the case where two segments of data are considered from a pre and post-damaging event, there should only be one

change point in the data: when the data switches from an undamaged set to a damaged set. This change point would occur where the magnitude of the CUSUM chart was furthest away from zero and is called the Cumulative Sum Indicator. This indicator can then be investigated by comparing it to threshold conditions in order to verify the state of the structure.

3.4 Mean Square Error Indicator

Taylor (2011) introduces the Mean Square Error (MSE) indicator, to analyze the U.S. trade deficit data in order to synchronize the results with the CUSUM. It is a scheme that splits the data into two segments, from one to *m* and *m+1* to *n* ($1 \leq m \leq n-1$). This procedure of splitting the data in this way will be referred to as Procedure 1 in this thesis. The data is then analyzed to see how well it fits the two estimated averages of those segments (Taylor 2011). This is based on the idea that if successive values in a vector are close in magnitude to one another, they will also be close to the average of themselves. However, if successive values are not close to each other, some of them will be much further away from the mean than others and the value of MSE, or distance to the mean, will be larger. As a result, this method groups the coefficients that are closest to each other and their mean. In this respect, the point *m*+1 indicates the first point after the value that minimizes MSE, a possible change point in the data. MSE can be defined as:

$$MSE(m) = \sum_{i=1}^{m}(X_i - X_1)^2 + \sum_{i=m+1}^{n}(X_i - X_2)^2 \quad \textbf{(3.5)}$$

$$X_1 = \frac{\sum_{i=1}^{m} X_i m}{m} \quad \text{and} \quad X_2 = \frac{\sum_{i=m+1}^{n} X_i}{n-m} \quad \textbf{(3.6)}$$

where X_i is a member of the data set and X_1 and X_2 are the means of the two segments of the data. The MSE is dependent on the variability within the data, which will be shown in

Chapter 5.1.3. Consequently, the statistic must be modified to expunge the variance between the two data segments produced while still maintaining their independent means. Therefore, a new statistic, ModMSE is introduced as:

$$ModMSE(m) = \sum_{i=2}^{m}\left(\frac{X_i - X_1}{std(X_{k|k=2\ldots i})}\right)^2 + \sum_{i=m+1}^{n}\left(\frac{X_i - X_2}{std(X_{k|k=m+1\ldots i})}\right)^2 \quad \textbf{(3.7)}$$

where variables are defined the same as in Eq. (3.7); however, the value is normalized by the variance (i.e. std(X_k^2)) of the two independent vectors. Results for the verification of this statistic are shown in Chapter 5.1.3. It is noted the statistic starts at the second value in the vector in order to avoid using the standard deviation of a single number which would produce numerical instability.

3.5 Bootstrapping for Threshold Construction

Bootstrapping is used to create a threshold for the three univariate control statistics discussed above. It creates multiple iterations of new data generated by randomly resampling the original data with replacement. Since these new data vectors are randomly sampled, the properties of the statistic they create will have similar properties to an undamaged data set and therefore can be used as a basis for comparison. It is common in practice to create at least 1000 bootstraps for comparison (Taylor 2011). Damage detection methods try to statistically prove that a significant change has occurred in a data set. Once a statistic crosses the threshold bounds produced by the bootstrap, the change in the data can be deemed from an assignable cause and the properties of the system have changed significantly to become out of control. In damage detection schemes, this means damage has occurred.

The confidence level produced by the bootstraps depends on the amount of bootstraps produced. This threshold can be defined in Eq. (3.8) as:

$$\frac{X}{N} = \frac{\#\ of\ Ranges\ for\ S_{diffzero} < S_{diff}}{\#\ of\ total\ Bootstrap\ samples\ performed} \quad \textbf{(3.8)}$$

In this equation $S_{diffzero}$ is the difference between the maximum and minimum value of each bootstrapped data set, S_{diff} is the range of the original data, N is the number of bootstrap samples performed and X the number of bootstraps for which $S_{diffzero}$ was less than the original S_{diff} of the data. A confidence level calculates the degree of certainty to which a change has occurred in the data.

3.6 Mahalanobis Distance using the T^2 Statistic and Fisher Criterion

The original and best known work in multivariate control charts are those described in Hotelling (1947), using the T^2 statistic. This statistic is a direct multivariate equivalent to the Shewart $\overline{X}$ control statistic and is used to create Mahalanobis distances which extract damage features by measuring the amount of variation from a reference condition. This method uses two assumptions: the samples are taken from a population that has a multivariate normal distribution, and that successive samples are independent over time. Zamba and Hawkins (2012) use these distances in their work, in which the quality of a smelter feed in an aluminum smelting process is checked, to analyze change in the mean vector and compare the results of univariate and multivariate methods. Wang and Ong (2012) use Hotelling T^2 to detect damage using simulated data on a progressively damaged reinforced concrete frame subjected to lateral loading. Pompe and Feelders (1997) used the Mahalanobis distance in order to determine the significance of the change in their selection criterion. Gul and Catbas (2009) use autoregressive models in

conjunction with a Mahalanobis distance-based outlier detection algorithm when implementing structural damage detection in a laboratory simply supported beam and highly redundant steel grid structure. In order to alleviate the need for initiating assumptions, Verdier and Ferreira (2010) propose another distance that is nonparametric. This new distance can be applied without the postulation of normal probability distribution and is used for fault detection in semiconductor manufacturing.

When significant changes in more than one variable are to be identified, multivariable models are applicable. As a measure for quantification of changes, assume that there is a matrix of damage indices associated with a certain condition and a certain location on the structure. The Mahalanobis distance gives the distance between selected damage features corresponding to a condition of interest and those corresponding to a reference condition. It is used in autoregressive techniques because it is a multivariate potential outlier method that correlates all the damage indices with a certain location.

Zamba and Hawkins (2012) explain that many different scenarios can exist when using this statistic: the mean vector could change from μ to μ_1 while the covariance structure remains unchanged; the covariance structure could be perturbed from Σ to Σ_1; both the mean and covariance could have a step change; one or both of these parameters could drift; and lastly, the distribution could change (for example, from normal to some other forms). In this study, structural damage detection cases are considered in which the feature mean undergoes a step change but the covariance remains unchanged. This does in fact represent the realistic condition when a property of the structure is changed due to damage because the change does not affect the estimation uncertainty of the damage feature. To unify all of the localized damage features, the covariance of damage feature

matrix is taken into account. Without using the covariance matrix, the distance would ignore the possibility of correlated sets of multivariate measurements that usually form and compromise the superior performance of multivariate methods (Zamba and Hawkins 2012).

The Mahalanobis distance $D_m(x)$ can be computed by using Eq. (3.9) from Mosavi et al. (2011):

$$D_m(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)} \tag{3.9}$$

where μ is the mean of the damage feature x, and S is the covariance. In detection schemes, a larger Mahalanobis distance indicates that the location is closer to damage.

While a Mahalanobis distance, $D_m(x)$, reflects the changes in a set of damage indices, a criterion is needed to statistically investigate the significance of this change. To address this need, the Fisher Criterion can be utilized (Mosavi et al. 2012). The Fisher Criterion measures the actual deviation of Mahalanobis distance under the damage condition of interest versus those for the healthy condition. Mosavi et al. (2011) used the Fisher Criterion on a two span continuous steel beam to detect damage in the structure. Damage is simulated on an idealized steel bridge frame using saw cuts in the bottom flange. This criterion, f, can be obtained using Eq. (3.10):

$$f = \frac{(m_D - m_h)^2}{\sigma_D{}^2 + \sigma_h{}^2} \tag{3.10}$$

where m_D and m_h correspond to the mean values of the Mahalanobis distances for healthy and damaged conditions, and σ_D and σ_h represent the standard deviations of those vectors, respectively. This statistic is calculated for all of the sensor parings on a structure and is the largest for locations closest to the damage. A threshold value of T = $\mu + 1.96\sigma$ is used. Here, μ and σ are the mean value and standard deviation of the Fisher

Criterion statistic from all of the locations on the structure. This threshold is chosen because 95% of the area under a normal curve lies within 1.96 standard deviations of the mean (Mosavi et al. 2011).

3.7 The Normalized Likelihood Ratio Test

There are many types of Likelihood Ratio Tests. Srivastava and Worsley (1986) propose a form of the likelihood ratio test that is more effective in detecting a shift involving only the mean vector. On the other hand, Zhang et al. 2009 presents a control chart used for detecting mean, along with variance, shifts. It, like the similar method proposed by Zhou et al. (2010), integrates the generalized likelihood ratio test and the exponentially weighted average statistic. Both of these studies use Average Run Length (ARL) to compare the effectiveness of the proposed control charts and are variations of an ELRT chart. On the other hand, Zou et al. (2006) presents a control chart based on change-point models for monitoring linear profiles and names the proposed method the standardized likelihood ratio test. Additionally, Hawkins et al. 2012 proposes a variation of the GLRT. For this thesis, a normalized ratio test from Sullivan et al. 1996 is used in which the population is tested for change in its mean or variance.

The Normalized Likelihood Ratio test can detect a shift in the mean and/or variance of a data set. It assumes that there are m independent observations that are normally distributed with mean μ and standard deviation σ. If a process is in control, at any partition of the data, the two sets would have similar means and variances. However, if there was a change in the process, the means and variances of the two subgroups would

vary substantially from one another. The objective of using a control chart is to find the significance of this variation by creating a confidence bound for comparison.

The log of the likelihood function for the first m_1 observations can be written as

$$l = \frac{-m_1}{2} log[2\pi\sigma^2] - \frac{m_1 \widehat{\sigma_1}^2}{2\sigma^2} log\left[\widehat{\sigma_1}^2\right] - \frac{m_1 (\overline{x_1} - \mu)^2}{2} \quad \textbf{(3.11)}$$

Here, $\overline{x_1}$ and $\widehat{\sigma_1}^2$, represent the mean and variance of the first m_1 observations; σ^2 represents the variance of all n observations. This function can be maximized to generate l_1 presented below.

$$l_1 = \frac{-m_1}{2} log[2\pi] - \frac{m_1}{2} log\left[\widehat{\sigma_1}^2\right] + \frac{m_1}{2} \quad \textbf{(3.12)}$$

This procedure can be repeated for the remaining $m_2 (= n - m_1)$ observations to create the maximum likelihood function, l2. In this way of partitioning the process into m1 and m2, there is an assumption that there is a change in the data at point m1+1. However, if this were not the case and the process was in control for all m observations, the likelihood function would be maximized using $\bar{x}$ and $\widehat{\sigma^2}$, the mean and variance of the entire set of observations. This would generate lo, the maximum of the likelihood function for an assumed in control process (Sullivan et al. 1996). This procedure of splitting the data in this way is the same as Procedure 1. The data points are split from 1 to m1 and m1+1 to n and the likelihood function is maximized for each split.

If l_a, the sum of l_1 and l_2, is much larger than l_o, the process is deemed to be out of control. For this reason, the likelihood ratio test detects the significance of the difference between the two. It is defined as $lrt[m_1, m_2] = -2(l_o - l_a)$ and has an asymptotic chi-squared distribution (χ^2) with two degrees of freedom.

This statistic is normalized to create the NLRT with a threshold value of unity. In damage detection schemes, any value of the likelihood ratio for a damage feature that is

above one represents an out of control feature. This can then be correlated to a location on a structure if the damage feature originated from data taken from a localized sensor network. In order to normalize the statistic, it is divided by its expected value (E), based on the dimensionality of the observations, p, and an upper control limit based on a desired overall in-control false alarm probability ω. As explained in Sullivan et al. (1996), the in control expected value is not the same for all values of m_1. If m_1 and m_2 are small, the expected value is larger than when both are the same. Therefore, if the model order is 1, the expected value can be approximated by simulation or

$$E = 2\left[\frac{m_1+m_2-2}{(m_1-1)(m_2-1)} - 1\right] \quad \textbf{(3.13)}$$

This value will change as m_1 and m_2 change. If the model order were to be greater than one, the approximated expected value follows three different best fit lines. If m_1<p+1, E follows the first equation in Eq. (3.14). If (m-m_1) <p+1, E follows the second equation in Eq. (3.14). Otherwise, E will follow the last equation in Eq. (3.14).

$$ev = a_p + m_1 b_p \quad \text{or} \quad ev = a_p + (m - m_1)b_p \quad \text{or} \quad ev = \frac{m-2p-1}{(m_1-p)(m-p-m_1)} \quad \textbf{(3.14)}$$

In all three cases, the expected value is then calculated using Eq. (3.15).

$$E = p(p+3) * ev \quad \textbf{(3.15)}$$

Derivation of this formulation can be found in Sullivan et al. (1999) for which the values of m_1 from two to p, fit a linear estimation. The line takes the shape of slope and intercept:

$$a_p = \frac{-.08684(p-14.69)(p-2.036)}{p-2} \qquad and \qquad b_p = \frac{.1228(p-1.839)}{p-2} \quad \textbf{(3.16)}$$

The statistic is also normalized using an upper control limit. It is usually set to give a specified in control average run length. Based on *m* and p, the upper control limit can be approximated. Its value has been tabulated in Sullivan et al. (1996) using Eq. (3.17):

$$UCL = \frac{1}{1.7} F^{-1}[(1-\omega)^{\frac{1}{k^*}}] \quad \textbf{(3.17)}$$

Here k* is the best fit number of independent variables and F is the CDF of the χ^2 distribution. In situations where the damage feature is scalar, the NLRT can directly be used on the feature itself. However, in cases where the damage feature is a matrix with dimensionality greater than one, the features are first condensed using the Mahalanobis distance and then used in the NLRT statistic.

3.8 t – Distribution Tests

The t-distribution is similar to the normal distribution in that it is symmetric about zero and bell shaped. Its main difference from the normal distribution is in the tails. A larger degree of freedom creates smaller tails and therefore the distribution is closer to the normal distribution. A statistic can be generated that should have a t-distribution and then tested to determine if it does fit the t-distribution. Because the statistic generated is to have a t-distribution, its peak should be in the center. If the statistic falls outside of the threshold boundary, it will be at this center point. Therefore, in structural health monitoring schemes, the timing of damage cannot be indicated using a t-test. It is only the magnitude of the maximum point that will be used to determine if damage has occurred. If the statistic falls outside the range of the threshold created at its midpoint, the null hypothesis is rejected and it is concluded that the population means are different. If the statistic stays within the threshold bounds, the null hypothesis is neither rejected nor

accepted; all that can be concluded is that it is not rejected. The Satterthwaite-Welch method is used with the assumption that the variance of the two populations is unknown and unequal. Rossi uses this type of t-test as a solution to the Behrens-Fisher problem to confidence bounds. For this thesis, we will use the well-known and trusted Student's t-test, and another statistic with the t-distribution generated Pakzad (2008), and used by Labuz (2011), here called a Bayesian hypothesis test.

3.8.1 Student's t-test

The Two-sample t-test is a form of statistical hypothesis testing to distinguish significant differences in measurements. Montegomery et al. (1987) shows the applicability of the t-test for detecting trends in water quality variables. Additionally, Limentani et al. 2005 shows the advantages to the two one-side t test and two-sample t-test. The t-test is a very common procedure for testing the differences between the means of two samples with normal distribution and length M. It has been used to statistically show the correlation between the mean of two sets of data. Hawkins and Zamba et al. 2013 uses the t-test in conjunction with the GLRT in order to distinguish between a shift in the mean and the variance in a gold mining quality control example.

There are three assumptions that the Student's t-test follows: 1) samples come from a parent population that is normally distributed, 2) the two sample groups have equal variances with N-2 degrees of freedom (N being the combined length of the two vectors, $2M$ 3) sample observations are independent. It is given by

$$t = \frac{X - \bar{X}}{Sp\sqrt{\frac{1}{n1} + \frac{1}{n2}}} \tag{3.18}$$

Where X is the grouped population, $\bar{X}$ is the mean and n_1 and n_2 are the size of each population. The variable S_p is the pooled variance defined by:

$$Sp = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} \tag{3.19}$$

Here S_1 and S_2 are the standard deviation of each population and the degrees of freedom are equal to $n_1 + n_2 - 2$. This is the approximate t-test for a change in the mean using the Student's t method. This method, used for cases in which the variance is assumed to be unchanged, can be used with linear regression parameters. This is because it represents the realistic condition when a property of the structure is changed due to damage: the change does not affect the estimation uncertainty of the damage feature and therefore the variance remains unchanged. Control limits are calculated using the Student's t inverse cumulative distribution function, k, at a certain confidence level and degree of freedom. This k value is then used to create the upper and lower control limits using Eq. (3.20). The confidence level can be user defined and ranges from 0 to 1.

$$UCL = (\bar{x} - \bar{y}) + k$$

$$LCL = (\bar{x} - \bar{y}) - k \tag{3.20}$$

Here, $\bar{x}$ and $\bar{y}$ are the means of the two populations being considered. In effect, the bounds represent the acceptable spread in the difference of the two means of the two populations.

3.8.2 Bayesian Hypothesis Testing

Bayesian Hypothesis Testing is a type of control chart that was previously used in Labuz 2011 to verify the use of linear regression influence coefficients from a simple beam column connection with harmonic excitation for damage detection. For this thesis,

it is used to verify the use of linear regression influence coefficients from a more complex structure with impact excitation for damage detection. It is also used to prove whether the tri-variate linear regression influence coefficients are valid representations of the frame properties and should not be used in damage detection schemes. It is a change point detection tool that, like the Student's t-test measures the variation in the mean of a population. It is very similar to the Student's t-test in its formulation but its application is a little different. It tests the hypothesis that the mean of the population for each successive index is equal to the mean of the entire set of N values

$$H_0 = \alpha = \alpha_1 = \alpha_2 \ldots = \alpha_N = \bar{\alpha} \tag{3.21}$$

against the one-sided alternative that the mean of the population after a certain change point, denoted as *k*, is greater than the mean prior to k by a significant amount.

$$H_a = \alpha = \alpha_1 = \alpha_2 \ldots = \alpha_k < \alpha_{k+1} = \alpha_N \tag{3.22}$$

The change point and mean of the population are all unknown values. Because the standard deviation is unknown, it is estimated as $\hat{\sigma}$. The statistic can be written as

$$t = \frac{S_N}{\hat{\sigma}\sqrt{\frac{N(N-1)(2N-1)}{6}}} \tag{3.23}$$

Where S_N is the Bayesian statistic

$$S_N = \sum_{i=1}^{N-1} i(\alpha_{i+1} - \bar{\alpha}). \tag{3.24}$$

The test statistic has a t-distribution with N-2 degrees of freedom. The hypothesis are tested at a 90% confidence level. The value of the CDF is found at 90% with N-2 degrees of freedom and its positive and negative value (because a t-distribution is symmetric, the value of the CDF at 90% and 10% are equal) is plotted with the statistic itself to create the upper and lower control limits respectively. If the statistic remains inside the

confidence bound, one can say that the properties have not changed; however, if it crosses the bounds, one cannot say that the properties have not changed.

To implement this statistic a baseline population is used to create the statistic. Then each index is added to the population one by one and the statistic is generated for that index. As shown in Eq. the Bayesian statistic is multiplied by the index *i*. This will have a drastic effect on its magnitude if the process becomes out of control. In effect, as each index of the feature is added into the statistic, it will have more of an effect on the magnitude of the statistic. If $(\alpha_{i+1} - \bar{\alpha})$ is not large, multiplying it by *i* will not have a major effect on the magnitude of the statistic. However, if $(\alpha_{i+1} - \bar{\alpha})$ is large, multiplying it by *i* will have an exponential effect on the magnitude of the statistic.

3.9 Moving Range

The moving range control chart will be used to determine the difference in variance of the regression residuals. This chart is used because it is a known and trusted method that has been used in many applications. Radson shows the behavior of the moving range statistic in the presence of autocorrelation. Additionally, Marks et al. (2009) as well as Amin et al. (1998), evaluates the design of the moving range chart along with the individuals control chart.

It is sensitive to changes in variability of a data set and is used for a control chart of the linear regression residuals. The Moving Range (MR) statistic is calculated below as:

$$MR_i = (x_{i+1} - x_i) \tag{3.25}$$

The x's in Eq. (3.22) refer to members of the data. As shown, the moving range is simply the difference between consecutive points. Each point on a moving range plot is only

affected by x_{i+1} and x_i. Splitting the data in this way will be referred to as Procedure 2. The process can be deemed as out of control when the process reaches a control threshold. There are two questions that should be answered upon choosing a threshold value. One is how often will there be false alarms; two is how quickly a change will be detected. There are many different means of computing a control limit for a desired in control threshold. For this implementation, an upper control limit (UCL) and lower control limit (LCL) are defined in Eq. (3.26) below.

$$UCL = 4.65\overline{MR} \text{ and } LCL = 0 \quad \textbf{(3.26)}$$

Here, $\overline{MR}$ is the average of all moving range statistics and 4.65 is chosen as an in control multiplier for the desired average run length. The average run length is defined as the number, on average, of how many successive plots will be generated before a point is detected beyond the control limits.

3.10 F-Test

The t-test monitors the change in the mean of a population by using the t-distribution function. Similarly, the F-test distinguishes between the variances of two populations using the F-distribution function. It is used here to be compared to the well-known Moving Range control chart. Similar to Rossi, Kabaila uses the F Test for the solution to the Behren-Fisher problem. Diaz et al. (1998) uses the F test for data with the Weiner Stochastic process pattern in the covariance matrix. However, for this thesis, the F-test will be used to determine the significance of the difference between variances of linear regression residuals.

The F test is used to test for a change in variance because it is sensitive to non-normality. The F distribution is zero for [-∞, 0], the total area under the curve is 1 and the curve as a value of zero at x=0. It is positive for x>0 and extends infinitely to the right to then approach zero as x goes to infinity. It is asymmetric, or right-skewed, so that its CDF will have a peak to the right of zero. If the populations used have multiple dimensions, a test statistic is generated having an F distribution. For this reason, similar to the t-test, if the statistic generated by the F-test is to have a F-distribution, the timing of damage in structural health monitoring schemes cannot be indicated using the F-test. The plot will have its maximum value at an index depending on the F-distribution, not depending on the timing of the damaging event. In effect, only the magnitude of the maximum point will be used to determine if the structure is indeed damaged. The F-test uses the CDF of the F-distribution to create a confidence bound for the statistic. The null hypothesis of the test is that the variances of the two populations are in fact equal. If the statistic crosses the confidence bound, the null hypothesis can be rejected. If the statistic does not cross the confidence bound, the null hypothesis is not rejected but it is also not necessarily accepted. For this implementation, the F Test is utilized with the regression errors from the linear regression models because the mean is zero yet the variance may change after a damaging event. The test statistic is given as

$$F = \frac{{s_1}^2}{{s_2}^2} \tag{3.27}$$

Here ${s_1}^2$ and ${s_2}^2$ are the variances of the two populations in question. The closer F is to unity, the less likely the null hypothesis is rejected. A confidence interval is found using the inverse of the F function with numerator and denominator degrees of freedom equal to one less than the size of each population ($F_{\alpha,df1,df2}$). Standardly, α is the confidence

level and df1 and df2 are the degrees of freedom of the numerator and denominator respectively.

3.11 Graphical User Interface (GUI)

A Graphical User Interface was produced in order to make the investigations in this thesis accessible to other researchers. Two different GUIs can be generated. One for generating the damage features in Chapter 2, and another for producing the control charts described above in Chapter 3. At this current time, only the second GUI has been created. The first GUI can be included in the future work in this area of study. Because many control charts are used in industrial settings, these GUIs could be used by researchers in other industries who are monitoring a process or wanting to create property sensitive features.

Upon entering the first GUI the user would be asked to input the raw data. This data can be any response data taken from a structure or industrial process. Each column should be a different sensor location's response (as in structural health monitoring) or a different characteristic that is being monitored. Once the data is loaded, it can be graphed to easily view inconsistencies. This preprocessing check allows the researcher to find invalid data such as data that would be taken from broken sensors. After the data is checked, it can be used in the different regression models discussed in Chapter 2. Additionally, the user should be able to choose which feature he/she wants to generate (whether it is the alpha coefficient, angle coefficient, or residuals). Upon creating these indicators, the user can export the results to MatLab or continue onto the second GUI.

This second GUI can be used to create control statistics and produce the control charts. A layout of the initial screen is shown in Figure 3.2. The user may have their own feature to use in the control schemes and therefore this GUI is completely separate from the first one. The user can choose different charts based on the type of data used. If the feature is a vector, having a dimensionality of one, the cumulative sum, exponentially weighted moving average, and mean square error, normalized likelihood ratio, t-test, Bayesian hypothesis test, Moving Range and F-test can be used. However, if the feature is a matrix, having dimensionality greater than one, the Mahalanobis distance, can be used to condense the feature in order to use it in any univariate scheme along with the fisher criterion. The charts that are available to the user would depend on whether the mean or the variance of the data set was being monitored. For example, the Moving Range and F-tests are used to monitor a change in the variance, the Normalized Likelihood Ratio test is used to monitor either the mean or the variance, and all the other methods are used to monitor a change in the mean. Addititonally, a feature is added so that if the user creates the CUSUM, MSE, or EWMA statistic and uses bootstrapping as the threshold generating method, he or she can choose to see the bootstraps or just see the minimum and maximum value. Once the charts are plotted and the title and axis labels are added, the user can save the figure as a ".fig" and then open it in MatLab.

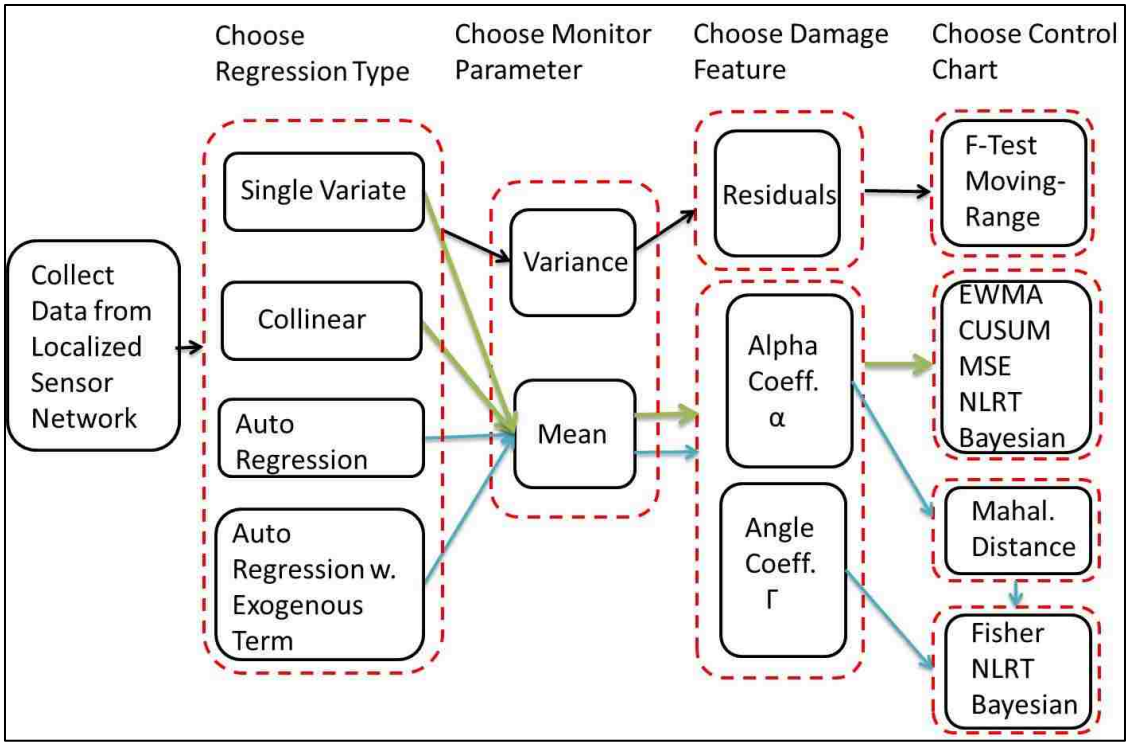


Figure 3.1: Flow Chart of specified pairs of monitor parameter, damage feature and control chart

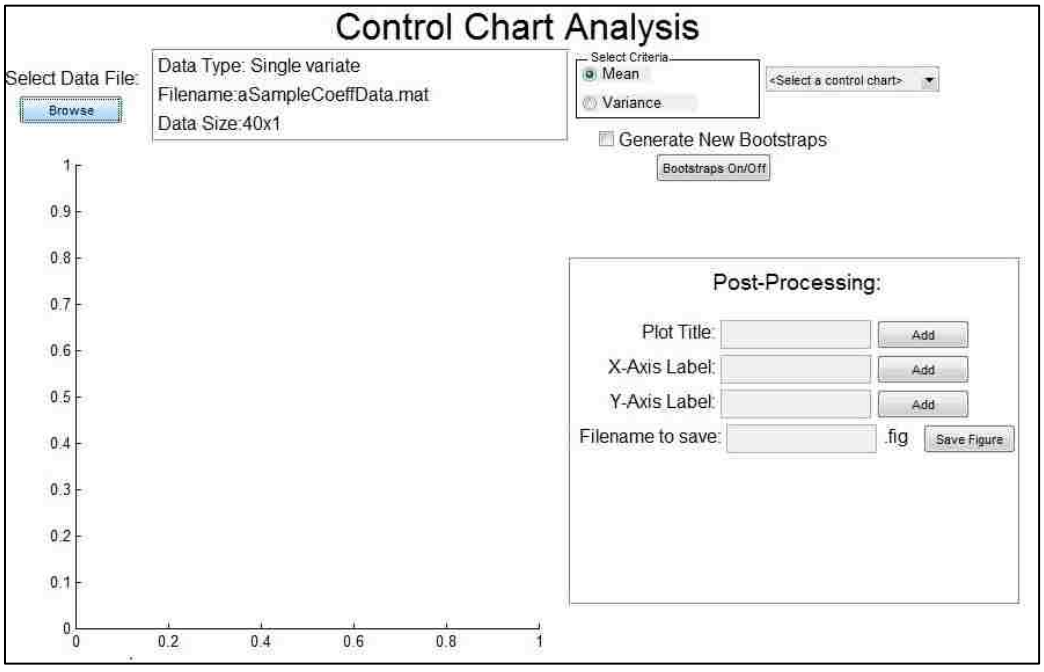


Figure 3.2: Preliminary GUI set up for the second GUI for control Chart Analysis

4. Application to Two Bay Frame Test-bed

4.1 Introduction

To test the performance of the damage features, a simulated model of a two bay frame test bed was constructed using SAP2000. Additionally, a two-bay steel frame with the same physical properties was constructed at the laboratory of Advanced Technology for Large Structural Systems (ATLSS) at Lehigh University. The first three natural frequencies as well as vibration amplitudes due to random vibration are compared between the simulated model and experimental frame in order to achieve a model that correctly represents the response of the specimen and to verify the algorithm. Then, both models are used to create damage features using single variate regression model. The results are compared. Once the simulated model is updated and the algorithm is verified, the experimental frame can be used for damage detection and localization.

4.2 Properties of the Structure

The simulated and experimental models are both made of steel tube sections. A sketch of the structure is shown in Figure 4.1. In order to simulate damage, there are nine sections that can be changed throughout the frame. These interchangeable sections are 0.19 m long shown in Figure 4.2 and have different cross sectional properties than the healthy state (shown in Table 4.1) which correspond to a 20% reduction in member moment of inertia (only for the length of the interchangeable piece). It is important to note that these property changes only correspond to less than a 0.5% decrease in the overall stiffness of the frame when one piece is replaced at a time (the global lateral stiffness of the frame

reduces by up to this amount due to simulated damage at different locations). This change in stiffness may be insignificant on the global scale but may have a larger impact on the response of the structure locally. A small change in stiffness can be from a small crack, corrosion, or any other damage mode representing a loss in cross sectional area resulting in the loss of stiffness. Damage has been considered as such stiffness reduction in Farrar et al. (1999) and Nichols et al. (2003). The interchangeable sections could be used in order to test the effectiveness of the damage features and control charts discussed in Chapter 2 and 3 respectively.

4.2.1 Assumption of Mass

It is assumed that the mass of the structure undergoing univariate damage detection using this algorithm is negligible. In this way, dynamic effects are not considered. Pakzad (2008) and Chang (2010) have presented simulated examples that support this assumption reasoning that if the stiffness of a structure is much larger than the mass, then the influence of the mass term becomes negligible and the dynamic equation of motion can be reduced to a static one. This is only valid, however, within a local joint where nodes are close to each other. Therefore, this thesis only creates damage features for nodes that are within the same local joint of the experimental structure.

To consider the dynamic effects in the frame multivariate methods can be used. Two removable lumped masses (lifting weights) are attached to the frame at the middle of the spans. The weights are selected in the possible range of loads that such a frame would be subjected to. The stress in the two beam spans due to bending from the two lumped masses should not exceed 50ksi in order to maintain linear elastic properties,

assuring that the frame will not yield. Considering the smallest moment arm of the added mass and the moment of inertia of the steel tube section, the largest load that should be placed on the beams can be calculated using the simple stress calculation

$$\sigma = \frac{M*y}{I} = \frac{PL/8*y}{I}. \quad \textbf{(4.1)}$$

Using this formulation the maximum load was found to be 1.74 kips. Therefore a load with a magnitude of less than ten percent of this maximum load was placed at midspan of both beams to ensure linear elastic behavior. It should be noted that the addition of the mass to the frame is limited to the tests where Auto Regressive and Auto Regressive with Exogenous term (multivariate) methods are applied. This is because single variate and collinear (univariate) methods are not proper for modeling the system with dynamic effects as validated above.

4.3 Simulated Model

In order to make sure that the model correctly portrays the response of the actual frame, the first three natural frequencies along with acceleration response amplitudes were compared. These features were chosen because they are of the most basic features of a structural system. The frequencies of the actual frame were found using the Structural Model Identification Toolbox for MATLAB. The acceleration response was found by exciting the frame with ambient vibrations and recording the response using sensors discussed in Chapter 4.4.3. The first three natural frequencies were found to be 13.66Hz, 21.95Hz and 31.58Hz shown in Figure 4.3**.** Additionally, two different locations on the frame were chosen for acceleration amplitude investigation, L4 and R4. These nodes are the closest to the beam and their responses are relatively large, 1.18g, for both sensors.

Initially, the simulation was modeled to have a fixed connection at the end of each column. However, the actual frame was bolted to the ground in the laboratory which did allow vibration to pass through the joint; hence, it was not completely fixed. Therefore, a fixed connection modeled in SAP2000 would not correctly represent the frame. It was assumed that because the three bases are bolted in the same manner, they would have the same support stiffness. Hence, connections were replaced with hinges and moment springs. From the results shown in Figure 4.4 it can be shown that the value of the first two frequencies gets closer to the actual value when the stiffness of the supports is reduced. However, they never reach the actual value. Therefore, more updating was necessary.

Each interchangeable section has a moment connection, shown in Figure 4.2. These connections should not be modeled as completely fixed due to laboratory conditions. As a result, these features of the frame needed to be updated as well. However, not all of these connections could be assumed to be the same. Therefore, the stiffnesses were divided into four groups: the left beam, right beam, columns and supports. These rigid connections were replaced with hinges with moment springs and changed until the results converged. It is shown in Figure 4.5- Figure 4.7 that the first mode shape is mainly dependent on the vibration from the right beam, the second mode is dependent on the horizontal vibration of the columns, and the third mode is dependent on the left beam vibration response. The stiffness of each connection was modified and the results and error are shown in Table 4.2. It is shown that the error is very minimal and the first three frequencies are correctly generated using the updated model.

Using this information, the acceleration responses were then used as convergence criteria. The accelerations responses of the updated simulated model were initially reviewed and shown in Table 4.3. Here is it shown that even though the first three natural frequencies are consistent with the experimental specimen (a conclusion from Table 4.2), the response accelerations were not similar to values obtained from the actual specimen. The error in Table 4.3 is too high to conclude that the simulated model is correctly portraying the properties of the experimental structure. Hence, it is shown in Figure 4.8 that even though the springs on the column don't have much effect on the natural frequencies, they do alter the acceleration amplitudes. Therefore, by changing the stiffness in the column connections, the percent error for the acceleration amplitudes was in an acceptable range shown in Table 4.4. More updating could be done using different convergence criteria; however, the results are in a tolerable percent error for the complete simulated model.

4.4 Experimental Model

A two-bay, steel tube frame testbed was constructed at the laboratory of Advanced Technology for Large Structural Systems (ATLSS) at Lehigh University for the implementation of the damage detection algorithm and the verification of the performance of different change point detection methods.

4.4.1 Experimental Procedure

To collect data, the specimen was instrumented with 21 wired accelerometers, labeled in Figure 4.1.for L, C, or R for left, center, and right sections of the frame. These sensors

were spaced throughout the two-span frame as shown. Initially, data was collected on a healthy frame for multiple test runs. Then, a damaged section was replaced and data was collected again. For all procedures, the amount of healthy state runs is the same as damaged, or so called unknown-state runs. In effect, a change point in the data should be found at the midpoint of the test samples. Additionally, damage should only be detected where the damaged piece was inserted and the healthy section was replaced.

Additionally, as stated in Chapter 4.2.1, to consider the dynamic effects in the frame response two removable lumped masses (lifting weights) are attached to the frame at the middle of the spans. Therefore in multivariate methods discussed in this thesis there will be two sets of results for the two sets of data.

4.4.2 Specifications of Excitation

A realistic scenario for implementation of the damage detection algorithm on real-life structures is to collect the ambient response of the structure before and after a damaging event. Thus, it is acceptable to assume that the behavior of the structure is within the linear elastic range in which small angle theory applies (Dorvash et al. 2012). Hence, in this implementation of the damage detection algorithm on the laboratory specimen, the excitation amplitude of the laboratory specimen was limited as not to contradict this assumption. It should be noted that other forms of excitation can be used with this algorithm. This is shown in Dorvash et al. (2010) in which simulations are used to prove that theoretically, in a stiff system with small mass, IDDA should detect damage regardless of excitation type.

4.4.2.1 For comparison of simulated model and experimental specimen

Both the simulated and experimental models were subjected to a harmonic excitation of 18Hz. This frequency was chosen as to reduce resonant effects. Once the data was collected the alpha coefficients from single variate regression were calculated to compare their effectiveness from both the simulated and experimental models. For this study, the damage case consisted of replacing a healthy section with one of less stiffness at the locations corresponding to sensors R2 and R5. Sixty test runs were collected for the experimental frame (i.e. 30 undamaged runs and 30 damaged runs). For each run, the sampling rate was 500Hz and 10,000 samples were recorded so that each test lasted a total of 20 seconds. The coefficients were created for each run number for each pairing of locations using the experimental specimen. The values for the undamaged and damaged runs were averaged and compared to the simulated coefficients. It should be noted that in Labuz 2011, the excitation frequency was chosen to be 12Hz. Even though these results, as well as the results shown in Chapter 4.4 using a forcing frequency of 18Hz, are consistent with the damage case, the results are too dependent on the forcing frequency chosen. Therefore, a new excitation method was needed so that resonant effects and variability with forcing frequency could be eliminated altogether. The next section describes the impact excitation used for the rest of the damage detection schemes.

4.4.2.2 For experimental specimen and change point detect schemes

In order to take the dynamic effects into consideration, impact load is chosen as a means of excitation. Thus the excitation does not have frequency content of any particular range. To apply impact load, the frame is tapped with a light hammer on the

beam-column connection at the right side of the frame with a magnitude less than 10% of the load required to cause yielding (using the simulated model, this linear elastic limit was found to be 1500 pounds). For this study, the damage case consists of replacing a healthy section with one of less stiffness at the location corresponding to sensors R5. Location R2 was not involved to see how localized the algorithm could detect damage. In total 40 tests were conducted. Half of them were from the undamaged state of the structure and half were taken after R5 was switched with its damaged counterpart. According to Tague 2004, only when there are at least 20 sequential points from a period when the process is in control, is the basis for comparison complete. For each run, the sampling rate was 500Hz and 1,000 samples were recorded so that each test lasted a total of 2 seconds. Based on the simulated model analysis, the first three frequencies of the frame are 13.66Hz, 21.95Hz and 31.58Hz. These frequencies are much lower than the sampling rate as to cause aliasing in the data. Since the data used in this study was collected from the real testbed structure it already contains measurement noise as opposed to the simulation results in which artificial simulated noise is often added to the signal.

4.4.3 Wired Capacitive Accelerometers

The accelerometers used in this experiment are model number 2210-002 produced by Silicon Design. Inc. (2010) shown in Figure 4.9**.** This is a capacitive accelerometer and type of small micro electro-mechanical systems (MEMS) with an anodized aluminum case that is epoxy sealed and easily mounted using two screws. The sensor produces two analog voltage outputs, which vary with acceleration. The sensitive axis is perpendicular

to the bottom of the package, with positive acceleration defined as a force pushing on the bottom of the package.

The signal outputs are fully differential about a common mode voltage of approximately 2.5 volts independent from the supply voltage which could range from +9 to +32 volts. At zero acceleration the output differential voltage is nominally 0 volts DC; at ±full scale acceleration the output differential voltage is ±4 volts DC respectively. Analog output voltages are proportional to acceleration. An analog-to-digital converter (ADC) reads the change in capacitance and translates this to a voltage that corresponds to a specific acceleration.

4.4.4 Data Acquisition System

Acceleration responses were collected from the wired sensors using the CR9000, shown in Figure 4.10**,** a modular, multiprocessor data acquisition (DAQ) system and PC9000 support software by Campbell Scientific, Inc. (2005; 2009). The CR9000 has a 16-bit analog-to-digital converter and a peak sampling rate of 100 kHz.

4.5 Comparison of Simulated and Experimental Model

The simulated and experimental models were both subjected to an 18Hz vibration and acceleration data was collected during excitation. This data was then used in the damage detection algorithm to generate single variate alpha coefficients. This regression type and damage feature is chosen because Labuz 2011 has already proven that these indicators correctly portray the changing properties of the frame and correctly localize damage. As stated, the damage case in question involves an interchangeable section at R2 and R5. For

this reason, the percent change in the alpha coefficient between an undamaged and damaged state of the frame should be largest at these locations.

As shown in Table 4.5 the experimental results do show damage in this general location. The percent change on the entire left side of the frame is very small compared to that of the right side. Additionally, sensors R2 and R1 have the greatest percent change. These results are satisfactory in that they generally locate damage to the right side of the frame and indicate R2 as the most damaged location. Even though R5 does not have one of the highest percent changes, it does have a higher percent change than any location on the left side of the frame. Additionally, it is expected that the location's responses close to the actual damage would also be affected by the damage. Therefore, the large percent change at R1 can be expected and these results are consistent with the damage case.

The results for the simulated model are shown in Table 4.6. The results are similar to that of the experimental frame in that the right side has a higher percent change than the left side in general. However, these results pinpoint R3 and R4 as the most damaged locations. Even though these results are not consistent with the actual damage case, these locations are the closest to the damaged locations R2 and R5. Therefore, these results would indicate the right side as a damaged joint and lead an observer to the correct general location of damage. These results could become more localized if the simulated model was updated to completely mimic the response of the frame. However, the influence coefficients do detect damage at the correct joint and therefore can be deemed valid features for damage detection.

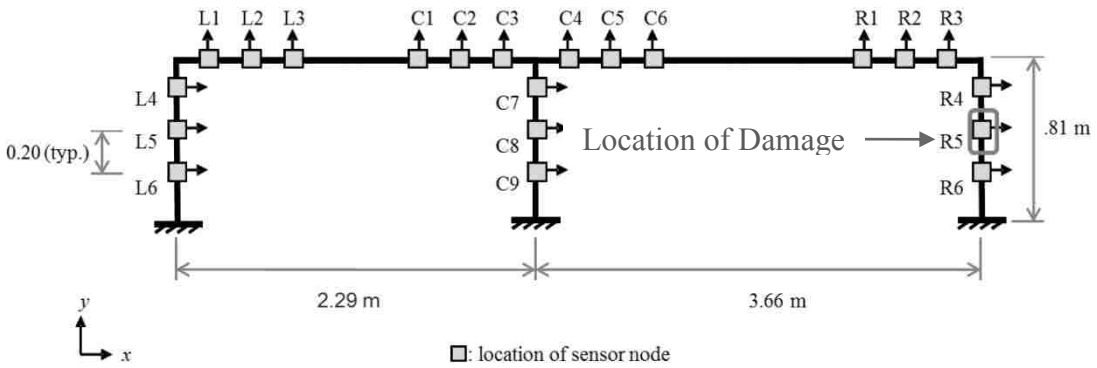


Figure 4.1: Sketch of the specimen and the location of the introduced damage.

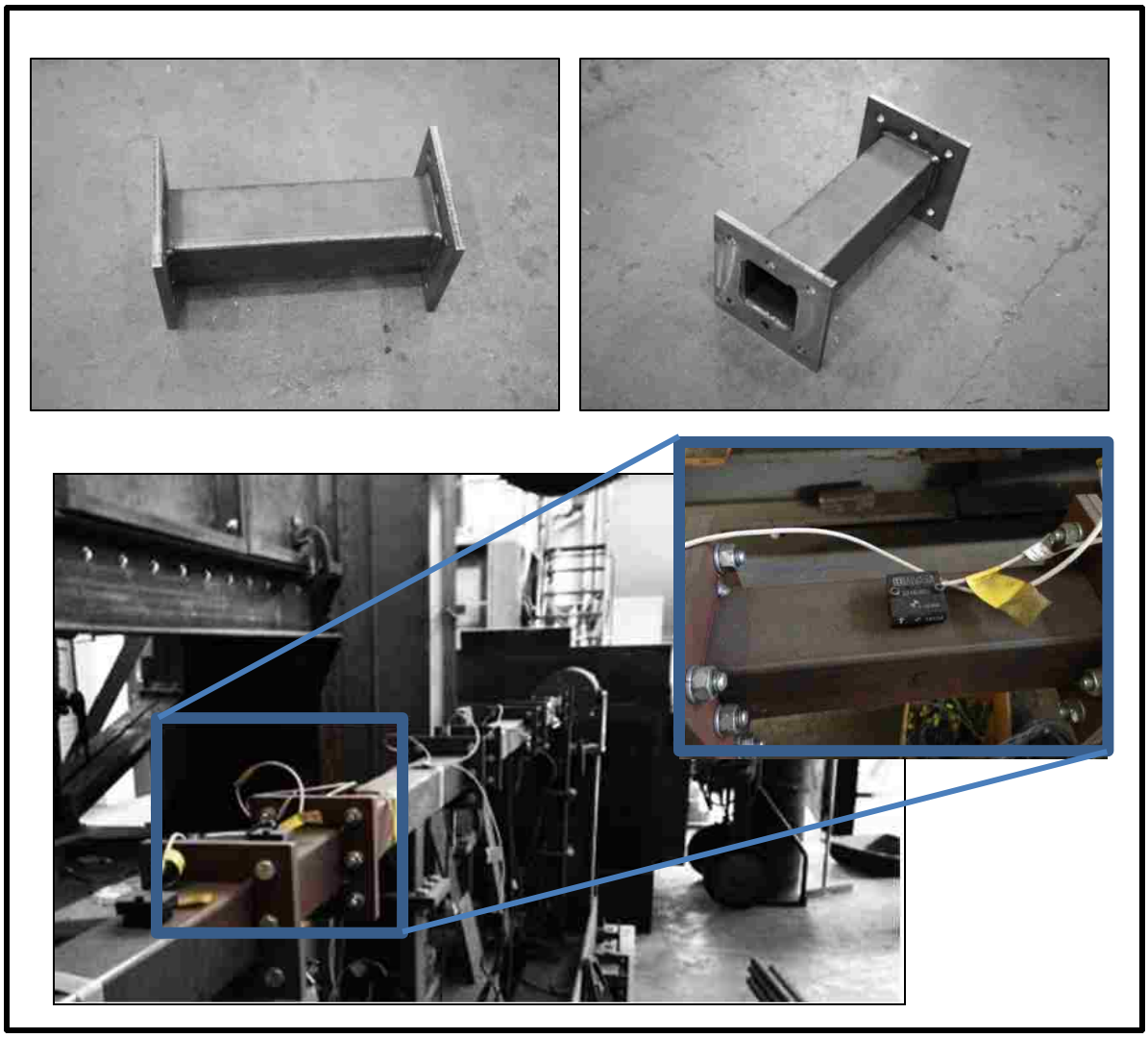


Figure 4.2: Different Views of the Interchangeable sections. (Top two Figures courtesy of E. Labuz 2011)

Table 4.1 Geometry of Baseline and Interchangeable sections

Feature	Baseline Sections	Interchangeable "Damaged" sections
Outer Dimension of Hollow Cross Section	0.05m	0.05m
Tube Thickness	2.16 mm	1.65mm
Cross Sectional Area	410.57 mm^2	324.57 mm^2
Moment of Inertia	162526 mm^4	130811 mm^4

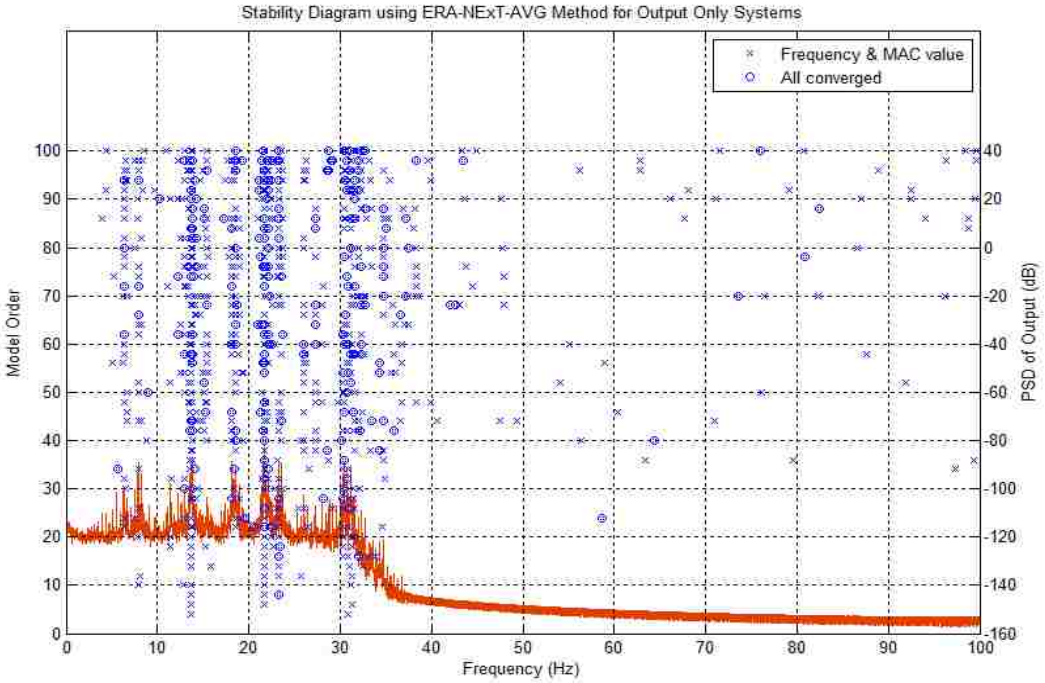


Figure 4.3: SMIT output of the first three natural frequencies of the experimental frame

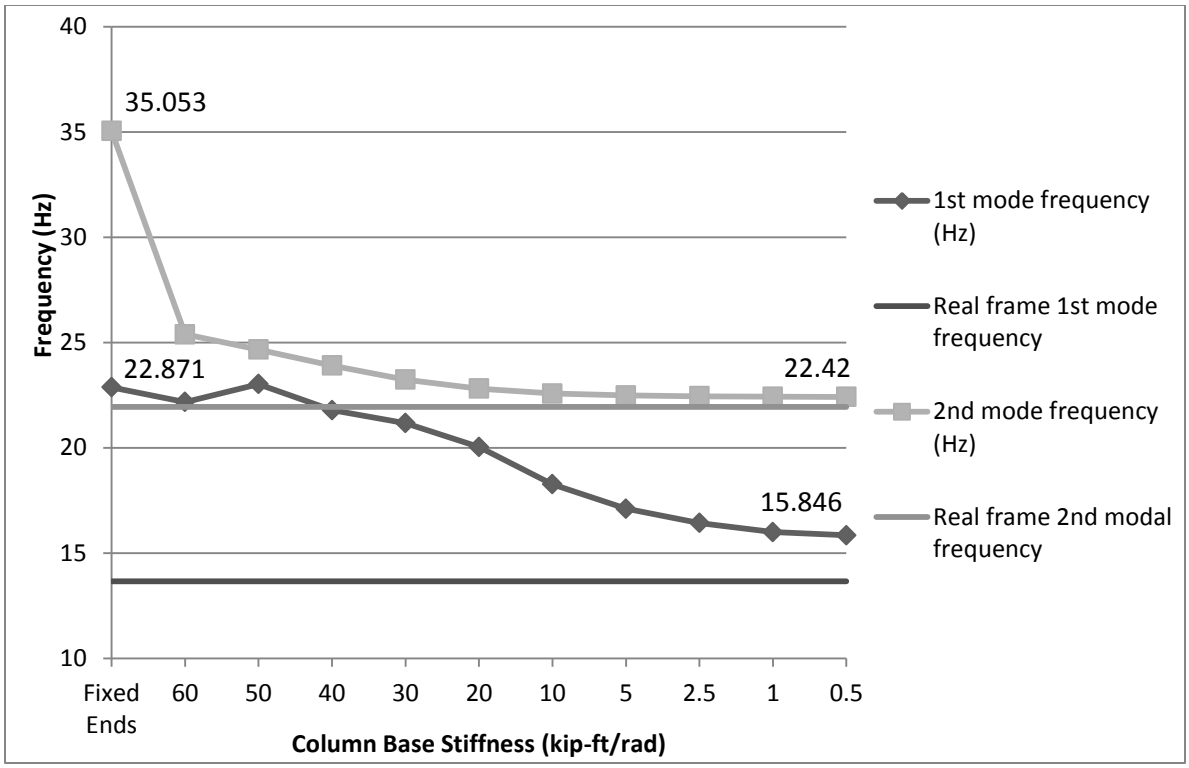
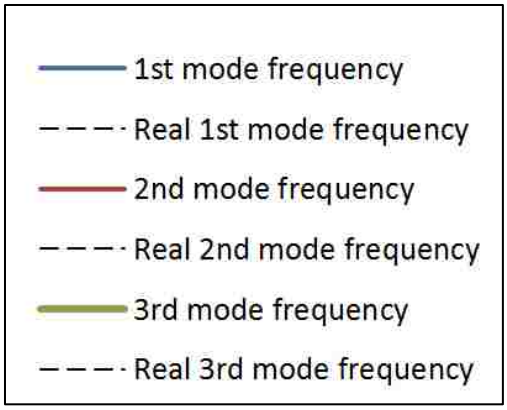


Figure 4.4 First and second mode frequency convergence generated from base Column Stiffness changes



Legend for Figures 4.5-4.7

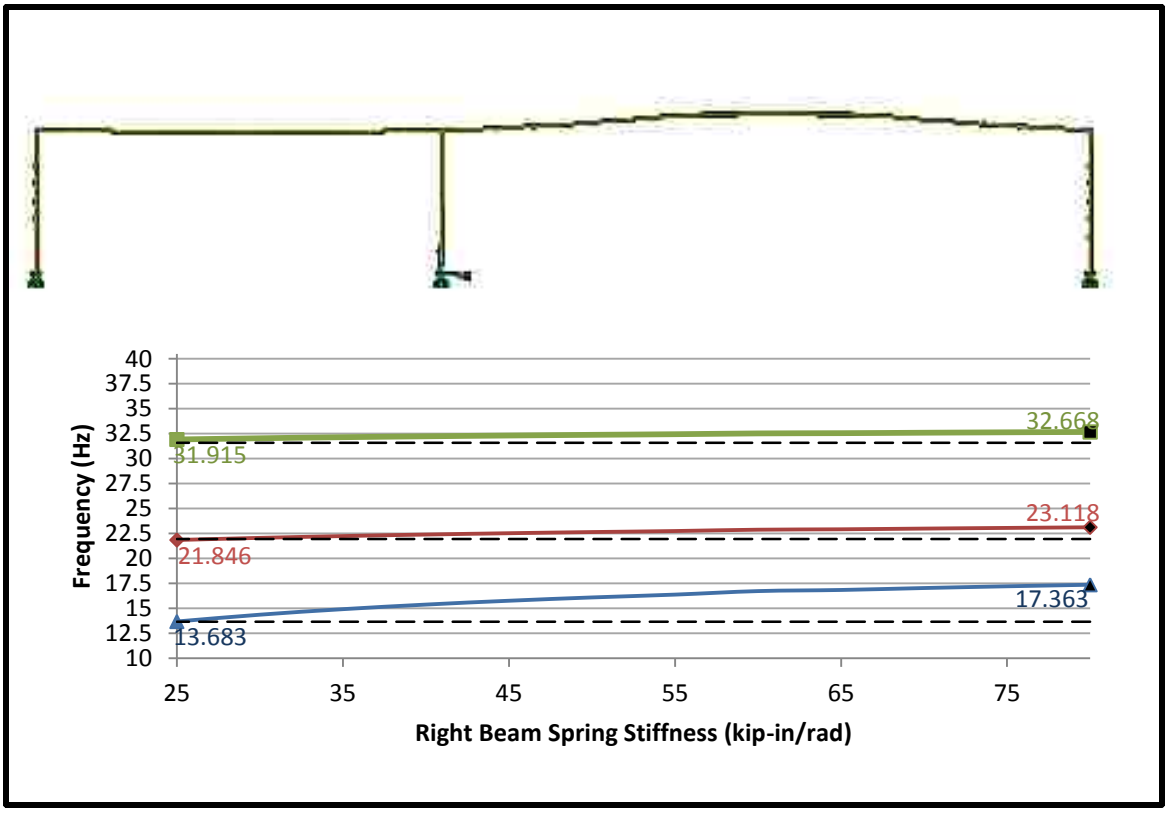


Figure 4.5: Right Beam Spring Stiffness vs Frequency and First Mode Shape

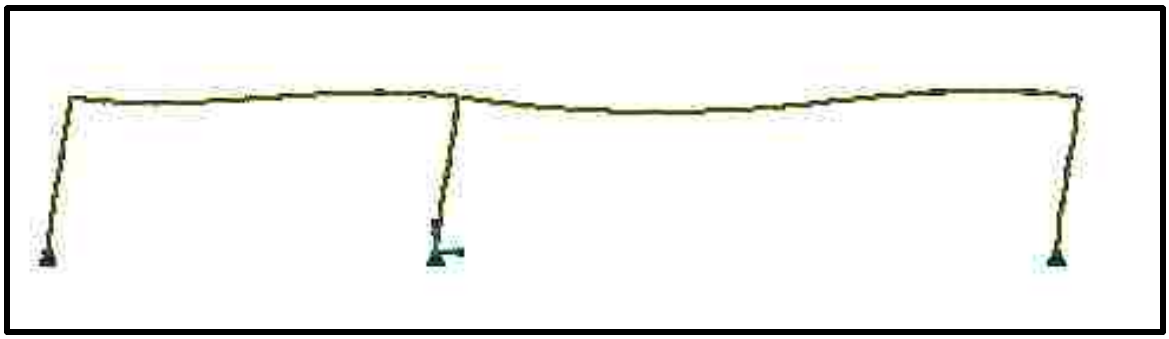


Figure 4.6: Visual of Column Spring Stiffness Effect on Second Mode Shape

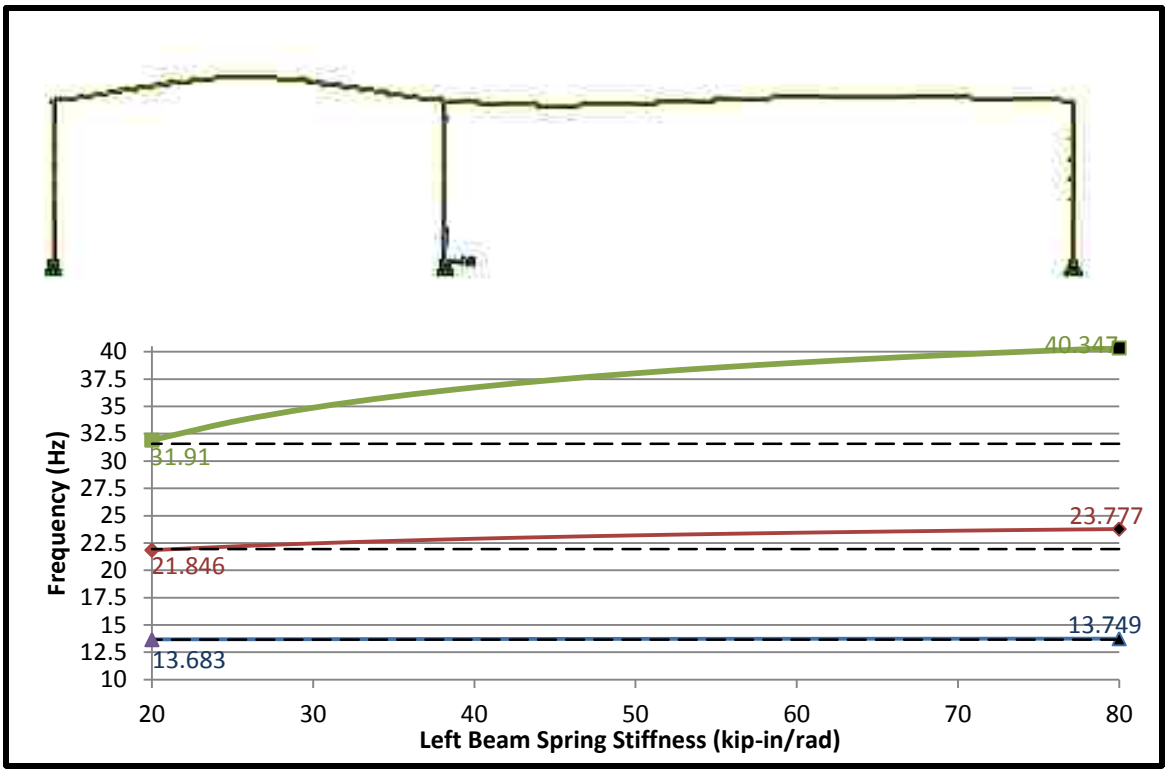


Figure 4.7: Left Beam Spring Stiffness vs Frequency and Third Mode Shape

Table 4.2 Simulated vs Experimental results of the first Three Natural Frequencies

	First Mode Frequency	Second Mode Frequency	Third Model Frequency
Simulated Model	13.68	21.812	31.906
Experimental Frame	13.660	21.945	31.581
Percent Error	.307%	.606%	1.029%

Table 4.3: Simulated and Experimental Acceleration Response Amplitude

	Acceleration L4 (g)	Acceleration R4 (g)
Simulated Model	.966	.860
Experimental Model	1.18	1.18
Error	27.11%	18.14%

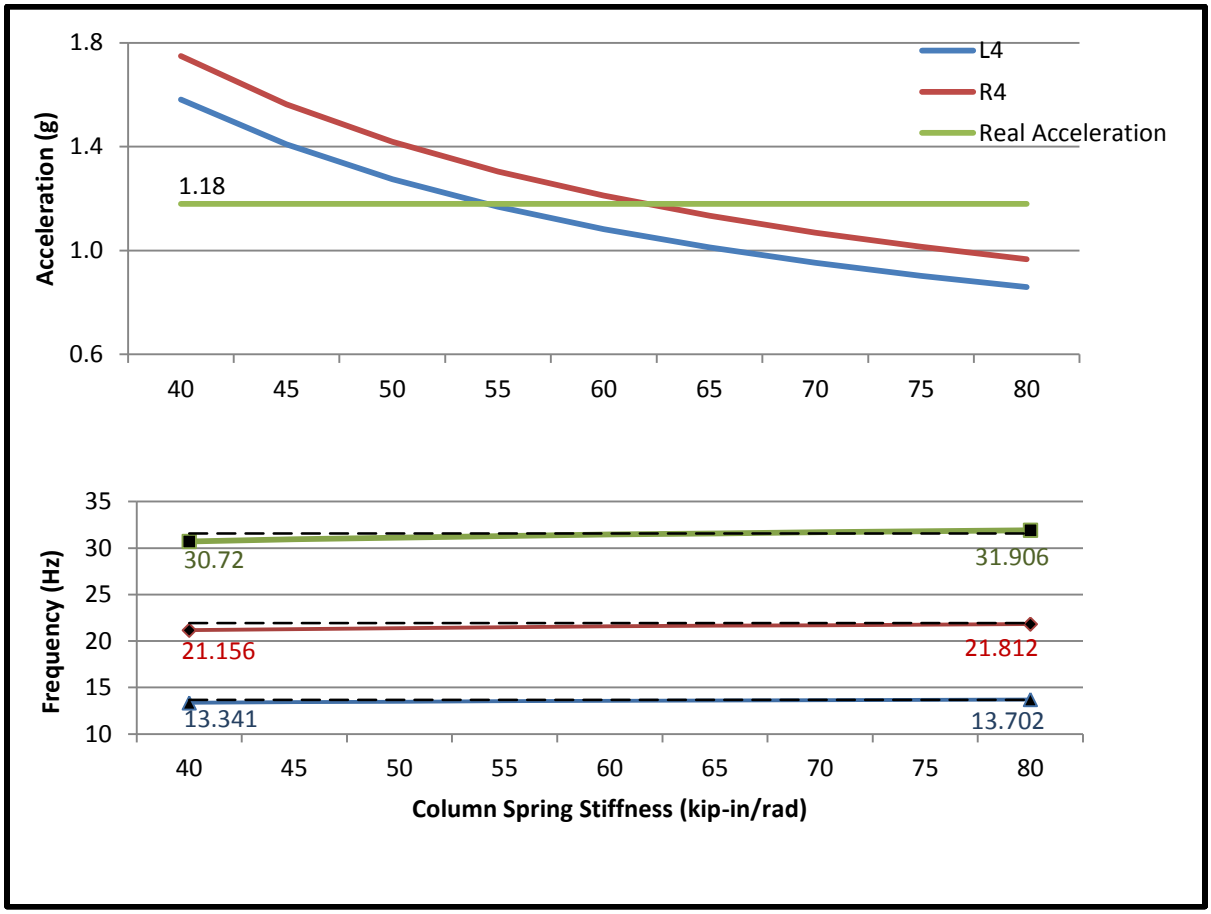


Figure 4.8: Column Spring Stiffness vs Acceleration response amplitude and Column Spring Stiffness vs Frequency

Table 4.4: Final Simulation vs Experimental Models

	First Mode Frequency (Hz)	Second Model Frequency (Hz)	Third Mode Frequency (Hz)	Acceleration L4(g)	Acceleration R4 (g)
Simulated Model	13.568	21.512	31.545	1.11	1.12
Experimental Model	13.660	21.645	31.581	1.18	1.18
Percent Error	.67%	1.97%	.11%	5.84%	3.80%

Figure 4.9: Model number 2210-002 produced by Silicon Design. Inc. (2010) Courtesy of E. Labuz 2011

Figure 4.10: MODAL 50A Actuator Courtesy of Y. Pan 2012

Table 4.5: Experimental Specimen Percent Change Of Alpha Coefficients from Single Variate Undamaged to Damaged State

Side/Node	1	2	3	4	5	6
Left Side	3.60%	3.89%	6.32%	3.85%	3.96%	5.04%
Right Side	53.3%	241%	X	28.2%	27.6%	27.7%

Table 4.6: Simulated Percent Change of Alpha Coefficients from Single Variate Undamaged to Damaged State

Side/Node	1	2	3	4	5	6
Left Side	.491%	.599%	.770%	1.22%	.619%	.325%
Right Side	2.55%	2.72%	2.87%	4.61%	1.21%	2.595%

5. Results of Change Point Analysis

Because the damage case being considered is a section on the right side column, the results for each method should detect damage at or near this location. Only the results on the left and right side of the frame are compared and presented for simplification. Considering there are six sensors on both sides of the frame, there are many different combinations of sensors that can be paired in the different linear regression models. However, an assumption of negligible mass only holds true within local joints of a structure. Therefore, only sensors within the same side of the frame will be used to create damage features. In effect, for cases in which two sensor nodes are compared to one another, thirty pairs can be made without comparing a sensor to itself. This occurs in single variate and ARX linear regression. However, in collinear regression, 120 different combinations can be made and in the AR model 6 coefficient can be made with each sensor location regressed onto itself. Coefficients are calculated for each sensor pair combination which are then analyzed using the change point detection methods. First, the statistical methods are examined to see if they correctly identify occurrence of the change point, at the 20^{th} run, when the damage is introduced. Then, the results are further inspected to determine if they can provide where damage occurred.

Upon investigation of the data, the sensor corresponding to location L1 on the frame was found to be damaged. A picture of a damage sensor can be found in Figure 5.1 Additionally, the normalized error associated with this sensor was large. For this reason, this sensor is eliminated as a faulty sensor in all data sets and control charts. If this location is indicated as a damaged section, the results are disregarded.

5.1 Single Variate Regression Results

Some of the methods use the assumption that the variables used in the control charts are normally distributed. It is shown, using normal probability plots in the Figure 5.2 (a) Figure 5.2 (b), that the alpha and alpha coefficients R56 are normally distributed. Their purpose is to graphically assess whether the data has come from a normal distribution. If the data is indeed of a normal distribution, these plots would be linear. The coefficients are used to create the blue "+" markings. These are compared to a red linear line superimposed on the graph joining the first and third quartiles of the coefficients. This line is extrapolated out to the ends of the sample to help evaluate the linearity of the data. It is shown here that the coefficients are indeed from a normal distribution because they do not stray far from the red line.

Figure 5.3 shows typical single variate linear regression coefficients and their corresponding Evaluation Accuracy and Estimation Error parameters for the influence coefficient α_{ij} in Eq. 2.2, for locations 5 and 6, namely α_{56}, for both the left side (shortened to α_{56-L}) and the right side (α_{56-R}) of the frame. This coefficient will be used as an example plot for some of the single variate control statistics being investigated. It is crucial to use a damage feature that changes with the properties of the structure. Even though the timing of damage would be unknown, a comparison between the undamaged and damaged features was generated to evaluate the effectiveness of these damage features. If no difference is found between a location near and far from damage, these features would be deemed less effective in identifying the location of damage. The first graph in each plot in Figure 5.3 shows the magnitude of the coefficient for all 40 runs. For the left side coefficient, α_{56-L} shown Figure 5.3(a) the mean of the first 20

(undamaged) runs is different from that of the last 20 (damaged) runs by 1.25%. On the right side (α_{56-R}), however, this variation increases to 5.50% as shown in Figure 5.3(b). This indicates that the right side properties of the frame have changed more than the left side. In using a control chart, the timing and location of the damage should be indicated since these parameters are previously unknown in real damage detection scenarios. Note that the accuracy parameter does not deviate far from one on both of these graphs which indicates that these parameters should not be disregarded as false alarms. Additionally, the estimation error can be used to further justify these coefficients as correct representations of the properties of the frame.

5.1.1 Estimated Weighted Moving Average

The process is considered out of control when the Z value in Eq. (3.1) falls outside the range of the control limits. As stated previously, the conventional threshold from Lucas and Saccucci, (1990) is three standard deviations away from the mean. However, when this method is implemented, damage wasn't detected anywhere on the frame as plots from both the left and right side of the frame never crossed the boundaries. An example is shown in Figure 5.4 where the left and right side coefficient α_{56} plotted using the Shewhart control threshold. As a result, this way of determining thresholds is deemed insufficient. It is reasonable to decrease the value of L for it corresponds to a 99.7% confidence, creating a very large confidence region. However, a bootstrap is used in order to compare the effectiveness of some of the EWMA statistic to other univariate methods in this thesis.

The results of the EWMA bootstrapped plot of the left and right side alpha coefficient α_{56} are shown in Figure 5.5 (a) and Figure 5.5 (b) respectively. It is important to note that

there is a well-defined change in the EWMA slope for the original data at the 21st run, as soon as the damage cases are included in the statistic. This indicates a change from a baseline state of the frame at the correct time of damage. However, this could also be considered as trivial variation until the line reaches the threshold generated from bootstrapping. Once the threshold is crossed, the change can be considered assignable and it can be concluded that the frame was damaged near this location. In Figure 5.5 (a) the left side coefficient α_{56-L} does not cross the bounds created by the bootstrap. This indicates that the frame's properties do not change significantly near these sensors. However, in Figure 5.5 (b) the EWMA for α_{56-R} does cross the threshold at the 24th run, a fairly quick detection of damage. Therefore, the properties of the right side of the frame near sensor locations R5 and R6 do change significantly with a 91% confidence level using Eq. (3.8). These results are consistent with the damage case and occurrence of damage. Every sensor pairing is analyzed in a similar fashion and the results confirm that there is damage on the right side column of the frame. As a result, the EWMA correctly identifies the correct time and location of damage.

Similarly, the results for the angle coefficient are generated for the same location pairing, Γ_{56} and the results are shown in Figure 5.6. These results are very similar to the alpha coefficient results described above. There is no damage indicated on the left side because the original data EWMA plot does not cross the threshold generated by the bootstrap. Additionally, the right side coefficient EWMA does cross the confidence bound at the same run as the alpha coefficient indicating a damaged location on the frame. This time, however, the confidence level here was determined to be 92.2%. Therefore the results can be deemed very trustworthy which indicates the certainty to

which there is a significant change in the data at the 23rd run when the original EWMA crosses the threshold boundary. These results are very comparable to the results for the alpha coefficient. Both damage features seem to indicate the correct timing and location of damage to confidence level within 91%.

5.1.2 Cumulative Sum Chart

The CUSUM plots that were created for the left and right side of the specimen are shown in Figure 5.7 (a) Figure 5.7 (b) respectively. Each line in these plots represents a sensor pairing. Both plots exhibit a change in slope at the 20th run thus indicating a possible change point at the 21st run. Therefore the CUSUM correctly identifies when the change occurred (there were 20 undamaged and 20 damaged cases); however, this does not help differentiate the left side from the right side and thus does not provide a location of damage. Because the entire structure's response is affected by the damaged section, the left side CUSUM shows a change in slope as well. However, the change should not be as significant as the right side considering the damage case in question. It is noted that in Figure 5.7 the ranges of the cumulative sum are a lot smaller for those sensor locations on the left side of the frame which would indicate that the properties of the structure on the right side have changed more compared to the left side. Yet, no conclusion can be made about the significance of the change.

In order to create the necessary threshold bound, bootstrapping is used to show if the change of properties on the right side of the frame is indeed significant. A bootstrap is able to depict the behavior of the frame for the case of no damage by statistical simulation. Therefore, the left side cumulative sum charts should remain within the range

of bootstraps; yet, the cumulative sum charts for the right side coefficients should cross the bounds of the bootstrap as close to the 21st run as possible at locations near damage. The bootstrapped plots for α_{56-L} and α_{56-R} are shown in Figure 5.8 (a) and Figure 5.8 (b). In Figure 5.8 (b) the cumulative sum chart for alpha coefficient α_{56-R} crosses the confidence bound produced by the bootstrap. Using Eq. (3.8), this indicates damage is detected at this sensor pairing with a 98.8% confidence level. Additionally, this method detects damage at the 21st run, which is a very quick detection scheme. On the other hand, the bootstrap threshold is not exceeded for coefficient α_{56-L}, indicating no damage at this location. Charts are created for all sensor pairings on the frame and it is very evident that there is a significant change on the right column of the frame. The results show damage at sensor locations R4 and R6. These locations are not directly at the damage, but because of their proximity to sensor R5, they show a more significant change than others. Conversely, the left side cumulative sum plots have a range very similar to the bootstraps. Because the bootstraps are supposed to simulate data that has no damage, the left side of the frame can be characterized as so. In effect, the cumulative sum for the alpha coefficient correctly identifies the occurrence and location of damage.

The results for the angle coefficient are shown in Figure 5.9 where the cumulative sum for the angle coefficient has a peak at the 21st run. This is the correct timing of damage. However, this peak cannot be deemed significant until it crosses the confidence bounds. This does not occur for the left side coefficient Γ_{56-L}, yet does for the right side coefficient Γ_{56-R}. This indicates significant change in the data at the 21st run and a damaged location. Additionally, the confidence level was found to be 100%.; none of the ranges for the bootstraps were bigger than the actual data set range; all of the bootstrap

ranges were below the range of the actual damage. This directly shows that there is damage at this location. Including all of these factors, the cumulative sum for the angle coefficient correctly identifies the occurrence and location of damage.

5.1.3 Mean Square Error

The MSE algorithm separates the coefficients from all 40 runs into two segments and finds the MSE associated to each segment. This testing method follows Procedure 1. For instance, the first point on an MSE plot is the MSE calculated for the first run and the next 39 runs paired together. The second point on the plot is the MSE for the first and second runs paired together and then the next 38 runs grouped together etc. An example of the alpha coefficient results is shown in Figure 5.10 (a) and Figure 5.10 (b) where the MSE indicator is plotted for all 30 coefficients on the left and right side of the frame respectively. The 20th run marks the lowest value for the MSE implicating the 21st run as the first possible index of the damaged state.

This is the correct timing of damage; however, this plot, like the CUSUM chart, does not allow the observer to detect the location of damage. As a result, a bootstrap was generated to establish a threshold and the results are shown in Figure 5.11 (a) and Figure 5.11 (b) for coefficient α_{56-L} and α_{56-R}, respectively. It is expected, as shown in the CUSUM and EWMA plots previously, that the α_{56-R} coefficient cross the bounds near the 21st run and the α_{56-L} coefficient not cross at all. However, the MSE plot of this sensor pairing crosses the bounds on both plots shown in Figure 5.11. This implies that both α_{56-L} and α_{56-R} undergo a significant change in the data and therefore signify a damaged section at L5, L6, R5, and R6. Additionally, the MSE plot crosses the bound

around the 13th run which is well before the damage occurs. Hence, both the timing and placement of damage from these plots are misleading and the statistic must be changed in order to detect the correct location and time of damage. As stated earlier, damage detection using control charts relies on the change in the mean and constant variance. Therefore the ModMSE can be used, shown in Eq. (3.7), which expunges the variance between the two segments produced and maintains their independent means.

Before a bootstrap is conducted, the difference between the original MSE and the ModMSE is inspected using a similar approach to the CUSUM indicator discussed at the end of Section 3.3; this indicator is the point furthest away from zero which can be a possible damaging point. Because the MSE may not start or end at zero, the ranges of the statistic were recorded instead of the largest magnitude. As a result, all the coefficients, from the left and right side of the frame, can be examined at one time. On the plots in Figure 5.11, the ranges of the MSE and ModMSE respectively are shown for all 60 coefficients. The first 30 coefficients are from the left side and last 30 coefficients are from the right side. Additionally, a boundary condition of three standard deviations away from the mean (a normal procedure of a Shewhart control chart from Lucas et al. 1990) is used for a simple comparison. Figure 5.11 (a) shows the MSE range plot using the original MSE coefficients. This model detects damage at sensor locations R1, R2, R3, and R4 which are located on the right beam and column close to the joint. Although these locations are on the right side of the frame, the actual location of damage, R5, is not included or detected. This procedure proves that the original MSE statistic does not always produce localized results. On the other hand, the normalized MSE statistic, ModMSE, was then used to create the same plot and the results are shown in Figure 5.11

(b) where the damage is correctly identified. Sensors R3, R4, and R6, along with R5, show a significant change in properties because of their proximity to the damage because the ModMSE ranges cross the confidence bound for the coefficients α_{54-R}, α_{43-R}, α_{36-R}, α_{56-R}, and α_{61-R}.

Additionally, false alarms can be eliminated by using the evaluation accuracy and estimation error parameters. If these numbers are low and high respectively, the coefficient should be disregarded because it is not correctly representing the properties of the frame in accordance with the assumptions made. As shown in Figure 5.12 (b), coefficient α_{35-L} seems to cross the bounds, signifying a damage location. However, it is at this point that the accuracy and error parameters associated with the sensor pairing, shown in Figure 5.13, can be used to eliminate it as a false alarm. The accuracy and error parameters are very high and low respectively; hence, this coefficient can be disregarded.

Now that the ModMSE has shown the correct location of damage in Figure 5.12 (b), a bootstrap can be generated to analyze its results in order to compare it to the EWMA and CUSUM plots previously discussed. The results of the bootstrap are shown in Figure 5.14 (a) and Figure 5.14 (b) for the left, $\alpha_{56-L}$and right, α_{56-R} coefficient respectively. Damage is correctly detected at sensor pairing α_{56-R} because the right side coefficient crosses the bound at the 21st run (i.e., as soon as the damage occurs). Additionally, the left side coefficient α_{56-L} does not cross the boundary, indicating an undamaged location. This demonstrates that the ModMSE can be used to correctly find the timing and location of damage.

The original MSE results for the angle coefficient were very similar to that of the alpha coefficient. The original data crossed the threshold value on both the left and right

sides at around the 13th run number. As a result, only the ModMSE results will be discussed in the proceeding section. The results for the angle coefficient can be seen in Figure 5.15. It can be shown that the timing of damage is found at the correct run. The original ModMSE data plot does not cross the bounds for the left side coefficient but does for the right side at the 21st run.

The plots from both the alpha and angle coefficient have indicated that the ModMSE is an effective method for damage detection and localization. The timing and location of damage are correctly indicated using this pair of damage feature and control chart. Therefore, the use of the ModMSE to eliminate the MSE's sensitivity to variance is validated.

5.1.4 Normalized Likelihood Ratio Test

The coefficients made using single variate regression were used in the Normalized Likelihood Ratio Test. This type of regression produces coefficients that are already scalar and can be directly used in the NLRT statistic; hence, a possible damage point should be detected at the 21st run of testing. As stated in Chapter 3.7, the statistic is normalized using the expected value in Eq. (3.13)-(3.16). This value is supposed to simulate the mean of the statistic and eliminate the variability brought with different dimensionalities. However, when these simulated values of E were compared with the actual statistic, they were not capturing the mean of the statistic. For example, in Figure 5.16**,** the expected value obtained in Eq. (3.13) is plotted with the actual likelihood statistic for all the coefficient pairings on the left side of the frame (the black dotted lines). These locations are of the undamaged, healthy frame and should not indicate any

damage. As the figure shows, the expected value of the statistic is far lower than the statistic itself; the mean of the actual statistic for all locations is 15.9; yet the value for E ranges from about 4 to 2.5. Since E, as defined in Sullivan and Woodall (1996, 1999) is only derived using the dimensionality of the data, it cannot fully capture the mean of the likelihood statistic in this case. Therefore, the statistic was instead divided by the mean of all the coefficient pairings from both the left and the right sides. Therefore, the expected value is properly indicated the average of the likelihood statistics. At locations of damage, the right side likelihood ratio statistics should be larger than those at an undamaged location. Hence, in using this method of normalization, the relative difference between the right and left sides of the frame can be established. The results for the alpha coefficient are shown in Figure 5.17 where the correct timing is indicated. Both plots show peaks at or around the 21st run. The results can then be analyzed for damage localization.

Because the entire frame's response is changing with the switch of the damaged section, it is expected that the damage features do peak on the left side as well as the right side of the frame. However, the change should not be as significant on the left side. Hence, the coefficients on this side do not cross the confidence boundary of unity. Additionally, damage is detected at the exact location: the middle of the right column. Coefficients that cross the bounds are those involving sensor pair R56 (α_{56-R}), and R65 (α_{65-R}) only. Since a location is not compared to itself in this type of regression, these results are very localized.

The results from the angle coefficient show a very similar pattern in Figure 5.18. No coefficients on the left side of the frame cross the confidence bounds; yet the same pairs

do on the right side with an addition of coefficients Γ_{51-R} and Γ_{61-R}. These coefficients are still on the right side of the frame and therefore can be considered a correct location of damage because sensors around the actual location of damage are expected to be more affected by the damage as well. It is noted that the Normalized Likelihood Ratio Test can be used in conjunction with the alpha and angle coefficient from single variate regression to correctly indicate the timing and location of damage.

5.1.5 Student's t-test

The Student's t-test can be used with the alpha, angle coefficients from the single variate regression model. As explained, the t-test is a type of hypothesis test. If the statistic falls outside the range of the confidence level, the null hypothesis can be rejected. However, if the statistic stays within the bounds of the threshold, the null hypothesis can neither be reject nor accepted. On the plots for the Student's t-test there are 40 numbers. Each of these correspond to a different split of the data using Procedure 1. For example, the first data point on the graph is based on splitting the data between the first run and the next thirty nine runs. The second point indicates the t-statistic when the first and second runs are paired together and the next thirty eight runs are paired together. This would continue until the first 39 runs are paired together and the 40^{th} run is by itself.

When the pre-derived threshold for the t-test, using the inverse T-distribution function, was used on the coefficients, the first run was already outside the threshold boundary. This would indicate that from the beginning, the hypothesis can be rejected. This would mean that the mean of the two segments was not the same; which is to be expected. The mean of the first 20 undamaged runs is projected to be different from the

next 20 damaged runs. For this reason, when pairs of runs are grouped in such a way as to include both sets of data, it would not be surprising that the means are not the same. However, this thesis is interested in when and where on the frame the biggest spread in the means occurs. The "when" cannot be answered using the t-test threshold. Due to the testing procedure, it would be expected that when the data is split into 20 and 20, the difference in the means would be the greatest. However, as stated in Chapter 3.8.1, the t-statistic has a t-distribution. Therefore, its maximum value is always going to be in the middle. Hence, the timing of damage cannot be formally indicated by the t-test hypothesis threshold. For this reason, a threshold similar to the one used in the Normalized Likelihood Ratio test is used to find the "where". The t statistic is normalized so that the threshold is unity. In summary, it would be expected that the t-test null hypothesis would be rejected because the means of the undamaged and damaged data are different. In effect, no matter where the test runs are segmented, the means are going to be different. Therefore, the widest spread in the difference of the two means is investigated.

The results for the alpha coefficient are shown in Figure 5.19. Here is can be seen that α_{12-R}, α_{13-R}, α_{51-R}, α_{52-R}, α_{53-R}, α_{54-R}, α_{61-R}, α_{62-R}, α_{63-R}, α_{64-R}, and α_{65-R} cross the confidence bound of unity at the 21st run. This would indicate these locations having the biggest spread in the difference of their means when the runs are separated into two populations of 20 tests. This is the correct timing of damaged; however, this can be attributed to the t-distribution. It cannot be concluded that this maximum value is because of the data, but because of the distribution of the statistic itself. However, the coefficients

that do cross the unity bounds are mainly from R5 and R6 paired with every other location on the right side of the frame. This is consistent with the damage case.

Similarly, the results for the angle coefficient are shown in Figure 5.20. These results are even more localized. The only coefficients that cross the bound of unity are Γ_{56-R}, and Γ_{65-R}. These results only include locations R5 and R6 which localize the damage to the exact location. It is noted that the t-test can be used to correctly detect where damage has occurred on the frame using both the alpha and angle coefficient from the single variate regression model. It cannot be used to detect the timing of damage because the t-distribution is always going to have its maximum value in the center.

5.1.6 Bayesian Hypothesis test

The Bayesian Hypothesis test can be used to detect a change in the mean of two populations. As stated in Chapter 3.8.2, an initial population is needed as a baseline. For this procedure, the first 15 runs were used. Therefore, if there is damage that a feature indicates, the statistic should cross the confidence bound after the 5th run on the hypothesis test figures in this thesis. This point represents the first damaged coefficient being added to the statistic (the 20th run of testing). The results are shown in Figure 5.21 for the single variate regression alpha coefficient α_{45}. These results show that the left side of the frame, near locations L4 and L5 are undamaged because the statistic does not cross the threshold boundary. This is consistent with the damage case. Additionally, the right side coefficient does cross the threshold boundary right after the 5th run with a 90% confidence level. Therefore, this test indicates the correct timing and location of damage.

The results for the angle coefficient also show a similar pattern. The result for Γ_{56} is shown in Figure 5.22**.** The plot for the Bayesian statistic does not cross the threshold for

the left side coefficient. This indicates an undamaged section of the frame and is consistent with the damage case. Contrastingly, the right side coefficient does cross the bounds of a 90% confidence level immediately after the first damaged run is introduced into the statistic. Therefore, these results comply with the correct timing and location of damage and can be used as a viable damage detection method.

5.1.1 Moving Range Chart for Residuals

The Moving Range chart can be used to test the variance of the residuals. This type of chart is a univariate method. In order to use the residuals as damage features they need to be condensed into scalar values for each run number. Instead of using the Mahalanobis distance to do so, the variance of each residual vector at each run number was found and used as the feature for the Moving Range plots. The results for the single variate regression are shown in Figure 5.23 for coefficient α_{54}. It is evident that the variance changes at the 21st run on both the left side and the right side because both plots have a higher statistic magnitude at this point. However, only the statistic on the right side plot crosses the confidence bound. This indicates damage on the right side column, the correct placement of damage. Other pairs of coefficients were analyzed and it is clear that there is damage on the right side column of the frame. The plots cross the confidence bound for α_{15-R}, α_{16-R}, α_{25-R}, α_{26-R}, α_{35-R}, α_{36-R}, α_{56-R}, α_{45-R}, α_{46-R}, α_{53-R}, α_{54-R}, α_{63-R}, α_{64-R}. These results generally localize damage to the right side joint of the frame. Although it can be noted that these coefficients are a combination of R5 and R6 with the other four coefficients on the right side of the frame. This would indicate R5 and R6 as the most damaged locations on the right side of the frame. Since

these locations are on the right side column, closest to the actual damage location, it can be deduced that this way of detecting damage does actual localize the results to the right side column.

5.1.2 F-Test for Alpha Coefficient Residuals

The residuals for the single variate regression model can be used in an F-test for the change in their variance. The F-test has a null hypothesis that the variance of two populations in question is the same. However, if the statistic crosses a confidence bound, this null hypothesis is rejected and the variances have changed significantly enough to have come from an out of control process. For the single variate regression, the residuals,$\varepsilon(\mathrm{n})$, were found using Eq. 5.1 below.

$$\varepsilon(\mathrm{n}) = y_j - \alpha y_i \tag{5.1}$$

Here y_j and y_i are the responses of two locations at time step *n* correlated through α. In effect, the residuals represent how well the alpha coefficient fits the single variate model and represents the true relationship between the two outputs. The mean of the residuals should be zero and remain unchanged through a damaging event; however, the variance is subject to change.

As stated in Chapter 3.10, in the populations used in the F-test are matrices, the statistic that is generated will have an F-distribution and the timing of admage will not be able to be determined. Therefore, only procedures that create a vector of populations are used in the F-statistic so that the test produces a scalar value at each run number and the distribution is determined by the properties of the data and not the properties of the F-statistic.

Initially, Procedure 1 was used to split the data. In order to condense the matrices of residuals, the variance of each run test was calculated. This created a vector of length 40 that could be used with Procedure 1 and put in to the F-statistic so that only a scalar value is formed. However, when the first statistic was generated using the first run as one population and the next 39 runs together as another population, the statistic fell outside the confidence bounds produces using the cumulative distribution function. These results are similar to the t-test and would be expected. The variances are not going to be the same when the data is split in this way. The main concern of damage detection is finding where the largest difference of the variance occurs in time and on the frame and to test the significance of this difference. Therefore, the statistic was normalized using a a threshold generation similar to that of the Normalized Likelihood Ratio to make the boundary of unity. The results are shown in Figure 5.24 for the alpha coefficient of the single variate regression. Here, the statistic immediately changes when the runs are separated into 20 and 20. This shows the correct timing of damage; however, nothing can be said about the significance of the change until a threshold is used. It is shown that two coefficient pairings cross the threshold bound. These include α_{64-R} and α_{46-R}. These locations are the closest to the actual damaged location R5. Even though this sensor location is not indicated as damaged, it is between the two that are indicated. Therefore an observer would be led in the correct direction to locate the damage. This method of normalizing the statistic and using the variance instead of the entire set of residuals only showed correct damage localization for the single variate regression residuals. The variance of other regression types produced an F-statistic with very high magnitude that could not be normalized for a threshold of unity using Eq. (3.14-3.17). Therefore a new

procedure was adapted. In order to use all of the residuals without condensing the vectors, Procedure 2 can be used similar to that of the trusted Moving Range chart.

In using Procedure 2 is may also be possible to detect the timing of damage. With each run consisting of 1000 data points, in order to use these full vectors in the F- test each run is used as a separate population. In this procedure, the test runs were split into consecutive pairs and put into the F statistic. The first point in this procedure would be the F-statistic for the first and second runs as the two different populations. Then the second point would be the second run tested against the third run and so on until all 40 runs were tested. In this way, only consecutive runs numbers are tested against one another and the statistic may not have the F-distribution so that the timing of damage can be detected. The threshold was found using the maximum and minimum of the F-distribution CDF at 95% confidence.

The residuals for the single variate regression are put into the F-Test and the results are shown in Figure 5.25. It is noted that this distribution does not match the F-distribution. This may be attributed to the way the test runs are split using Procedure 2. The timing of damage is investigated first to see if the plot shows the correct occurrence of the change in the frame. Shown in Figure 5.25 the plot peaks at the 20^{th} run; this indicates the correct stage of damage for this statistic was created when the data was split using the 20^{th} and 21^{st} run of testing. The results are then further analyzed for the localization of damage. It can be shown that α_{46-L}, α_{45-L} and α_{36-R} cross the confidence bounds produced for a 95% confidence level. Therefore, these locations, R3, R4, R5, and R6, are indicated as locations of damage. This is the correct location of damage for these locations are on the right side column of the frame (R4, R5, and R6) and the closest

location to the right side column but on the beam (R3). Therefore, the results show that the F-Test is a valid method for detecting and localizing damage. It should be noted that the F-test can be used to detect the correct timing and location of damage using Procedure 1 and 2. However, only Procedure 2 will be used for the rest of the regression types.

5.2 Auto Regression with Exogenous Term Results

In order to use many control charts, there is an assumption that the variables used are normally distributed. It is shown in Figure 5.26 that the ARX alpha and angle coefficients are normally distributed through the use of a normal probability plot. The ARX alpha coefficient is actually a matrix as described in Chapter 2.3.2. Therefore, there are many coefficients depending on the order number. It will be shown that an order number of four is sued and therefore there are nine coefficients in the ARX alpha model. These are individually compared to a line to show that they are normally distributed. However, in the angle coefficient, there is just a vector of coefficients because from each run test, a scalar value is found. In Figure 5.26 (b) the blue "+" marks indicate the actual coefficients and the red line is a regressed line for comparison. The marks do not stray far from the red line and therefore the coefficients can be deemed from a normal distribution.

5.2.1Mahalanobis Distance with the Fisher Criterion, Normalized Likelihood Ratio Test and Bayesian Test

In order to use the Mahalnobis distance to condense the AR and ARX influence coefficients, it is necessary to know the model order of the data. However, the model

order is limited based on the amount of run numbers. This is explained as follows: In the matrices that are used to calculate the Mahalanobis distance, each column represents a coefficient from the ARX or AR model and each row represents a run number. However, in order to use these matrices to generate a distance, there cannot be more columns than there are rows in the influence coefficient matrix. Therefore, the amount of coefficients produced, which depends on the model order, cannot exceed the amount of test runs used to create the Mahalanobis distance. In the next section, the AIC is derived for both data sets in order to calculate the Mahalanobis Distance.

5.2.1.1 Mahalanobis Distance

In order to condense the ARX alpha coefficients, Mahalanobis distances are calculated between coefficients from the first 10 healthy state runs and the last 10 healthy state runs. This step creates a baseline distance. Then, the first 10 healthy runs and the 20 damaged runs are used to create a distance to compare to the reference. The distances calculated in the latter coefficients should be bigger than the baseline condition at areas of damage. Through the procedure discussed above, the smallest amount of runs used to create the Mahalanobis distance is ten. Therefore the maximum model order for both of the data sets used in this paper has to be four (i.e. there would be four columns from y_j ARX coefficients and five columns from ARX y_i coefficients. The total number of coefficients cannot exceed ten).

Histograms are used as a means to plot the Mahalanobis distances because coefficients with the largest distance are those that are closest to a location of damage and stray furthest from the baseline state. As discussed in Section 2.2.2, two testing

conditions are used for the AR and ARX data: a system without added mass and a system with mass. The two sets of results obtained from these two testing conditions for the alpha coefficient and are shown in Figure 5.27 (a) and Figure 5.27 (b) respectively. To analyze these results, the sensor pairings with the larger Mahalanobis distances are deemed closer to a possible damaged location because they deviate more from the reference condition. In Figure 5.27 (a), the results are inconsistent. Sensors from the left side of the frame, α_{43-L} and α_{61-L}, as well as sensor pairing α_{56-R} have large Mahalanobis distances. On the other hand, using the impact test data with added mass, shown in Figure 5.27 (b), the Mahalanobis distances have the greatest magnitude for coefficients α_{45-R}, α_{54-R} and α_{64-R}. Overall coefficient α_{45-R} has the greatest distance and this result properly indicates the location of damage; despite these observations, without a threshold it cannot be concluded that any location is necessarily undamaged or damaged. Hence, the next sections introduce the Fisher Criterion, Normalized Likelihood Ratio Test and Bayesian Hypothesis test to minimize the discrepancy of damage localization by creating a threshold boundary.

The results for the angle coefficient are analyzed and shown in Figure 5.28. The procedure for these coefficients differs slightly than the procedure for the alpha coefficient. The major difference is that the angle coefficient is a scalar value for each run number (making a vector including all 40 runs) for each sensor pairing. Therefore, instead of dealing with a matrix of coefficients as in the alpha coefficient, the Mahalanobis distance is used to find the distance between values in a vector for the angle coefficient. However, the same number of baseline distances and so called "damage distances" are used as a comparison making the possible change point at the 11^{th} run. A

histogram was made with these angle coefficients for tests without mass and the results are shown in Figure 5.28 (a) and Figure 5.27 (b) respectively. Figure 5.28 (a) indicates Γ_{43-R}, and Γ_{65-R} as coefficients with larger Mahalanobis distances. As indicated above, there is no threshold value for these distances. Hence, the Fisher criterion, Normalized Likelihood Ratio Test and Bayesian Hypothesis Test are utilized in the next section to distinguish between significant differences in the data.

The data with mass was also used to create the Mahalanobis histogram shown in Figure 5.28 (b). The coefficients with the largest distances include Γ_{45-R}, Γ_{65-R}, and Γ_{56-R}. These results a more localized than the results without mass because only those coefficients involving the right side column of the frame are included. Still no conclusion can be made about the significance of these distances until the results from the Fisher Criterion, Normalized Likelihood Ratio Test and Bayesian Hypothesis Test are analyzed. These results are presented in the next sections.

5.2.1.2 Fisher Criterion

The Fisher Criterion is used to create the boundary threshold for the Mahalanobis histogram that is necessary to distinguish the location of damage and the state of the structure. Figure 5.29 (a) and Figure 5.29 (b) show the Fisher criterion for the alpha coefficient without and with added mass, respectively. This statistic, as explained in Section 3.6, includes the mean and variance of the distances in order to statistically determine the significance of the deviation of the larger distances from the baseline distances. The results for tests without added mass show the correct location of damage while also eliminating the discrepancies from the Mahalanobis distance results. The only

sensor coefficients that cross the bounds are α_{45-R} and α_{54-R} corresponding to sensor pairings that are located on the right side column of the frame. Although the histogram in Figure 5.27 (a) shows large values for coefficients α_{43-L} and α_{61-L} for the impact data without added mass, the Fisher Criterion plot eliminates these as false alarms of damage detection because these values remain well below the threshold. This verifies that even though the Mahalanobis distances may have had some larger magnitudes at locations on the left side of the frame, these locations did not experience any significant change in properties and can be considered undamaged.

Similarly, when the Fisher Criterion is analyzed using the tests with mass added to the system, the results correctly point to the location of damage. The coefficients that cross the threshold indicating damage are at $\alpha_{43-R,}\alpha_{53-R}$, and α_{63-R}. Although there is no damage associated with sensors R4, R6 and R3, these locations are the closest to sensor R5 where the damage is located. It is reasonable that the coefficients corresponding to these sensor pairings also cross a damaged boundary.

The results for the angle coefficient without mass are shown in Figure 5.30 (a). The Mahalanobis distances are largest for Γ_{43-R}, and Γ_{65-R} as stated above; however, the fisher criterion can use a threshold value to distinguish between significant distances and those that are not. The fisher criterion for the data without mass crosses the confidence bound for Γ_{21-R}, Γ_{42-R}, Γ_{43-R}, Γ_{56-R} and Γ_{65-R}. (From the figure α_{61-L} also crosses the confidence bound. However, as stated earlier, the sensor L1 results are misleading because the error is very high. Therefore this is disregarded as a false alarm). Hence, the coefficients include only locations on the right side beam and column. They are not localized to the right side column, but the right side of the frame in general. Even though

these results are not very localized, they do show the correct vicinity of damage. No coefficient on the left side crosses the bounds indicating an undamaged section.

The results for the data with mass can be analyzed and compared to these results. Figure 5.30 (b) shows the plots for the angle coefficient fisher criterion using the data with mass. The coefficients that cross the confidence bound include Γ_{46-R}, Γ_{54-R}, and Γ_{56-R}. These results only include coefficients from those locations on the right side column of the girder: those that are closest to the actual damaged location. Therefore, the fisher criterion specifies the right location of damage using the data with mass. In all of these cases, both the alpha and angle coefficient show the correct general location of damage. Therefore, they are valid methods for damage detection using the fisher criterion as a control statistic.

These results indicate that the Mahalanobis distance statistic can be used in conjugation with the Fisher Criterion in order to properly indicate the location of damage. It can be noted that the timing of a possible damaging event is assumed to be known; the distances are calculated using the two separate sets of data from pre and post damaging event.

5.2.1.3 Normalized Likelihood Ratio Test

The ARX regression alpha coefficients can only be used in the NLRT after they have been condensed using the Mahalanobis distance. As stated, the first 10 distances are baseline healthy distances; hence there should be a possible damage point at the 11^{th} run. The results for the alpha coefficient without mass will be analyzed first and are shown in Figure 5.31. The damage features for ARX regression do show the correct timing of

damage. Both plots peak at the 11th run. The results can then be analyzed for their effectiveness in localizing the damage. Similar to the single variate regression results, the results for the ARX coefficients using the data without mass only show coefficients on the right side of the girder crossing the confidence bounds. In effect, damaged pairings are correctly found in the vicinity of the actual damage on the girder. Coefficients that cross the threshold on the right side of the frame include α_{34-R}, α_{43-R}, and α_{63-R}; however, no coefficients cross the threshold on the left side.

Although there is no damage on the right side beam, these locations are close to the actual damaged section and therefore it is reasonable that these locations would experience more of a significant change in response due to a damaging event. These results do show the correct general location of damage; however, the results can only localize the damage to the right side of the frame, not specifically the middle of the right column. Yet, the correct general location of damage is detected; hence the normalized likelihood ratio test using alpha coefficients as damage features from ARX regression for the data without mass can be deemed a viable method for damage detection but not necessarily localization.

The results for the alpha coefficient with mass, however, show misleading results. Even though the correct timing of damage is indicated by both plots having peaks at the 11th run, as shown in Figure 5.32, no damage features from the left and right side of the girder are indicated as damaged. This does not comply with the damage case and should not be used a valid results. Therefore, the NLRT statistic should not be used with the ARX regression test data with mass in order to test for a change in the alpha coefficients.

5.2.1.4 Bayesian Hypothesis test

The Bayesian Hypothesis test can be used to monitor the mean of the alpha influence coefficients made with the ARX model. First, the matrix of coefficients must be condensed using the Mahalanobis distance. The Bayesian Hypothesis test is a univariate method and can only be used with a population with dimension of one. As stated, using the Mahalanobis distance creates a vector of length 30 and a change point should be detected at the 10^{th} run. The first five runs are used as a baseline population. Therefore, in the Bayesian plots in this thesis for the ARX regression types, the statistic should cross the bounds as close to the 6^{th} run as possible. This corresponds to the 11^{th} Mahalanobis distance in which the damage features are introduced into the statistic. The results for the alpha coefficient without mass are shown in Figure 5.33 where neither the left side nor the right side coefficient cross the 95% confidence bounds. More pairs of coefficients were analyzed and the results don't indicate damage anywhere on the frame. This would indicate that the entire frame is in an undamaged, healthy state. However, this is not the case and therefore these results are not consistent with the actual damage. Contrastingly, the results for the tests with mass for coefficient α_{56} are shown in Figure 5.34. It is shown here that the left side at these locations is undamaged because the statistic does not cross the confidence bounds. However, the statistic for alpha coefficient α_{56} does cross the bounds indicating a damaged section at sensor locations R5, and R6. These locations are on the right side of the frame. These results localize the damage to the right side column of the frame which is exactly where the actual damage is located.

The results for the angle coefficient ARX model for the data without and with mass respectively were not as conclusive and are not shown in this thesis. Neither of these

results comply with the damaged location and therefore this damage feature should not be used in the Bayesian statistic to localize damage to the correct section of the frame.

5.2.2 Normalized Likelihood Ratio Test using the Angle Coefficient

The angle coefficient from the ARX model does not need to be reduced using the Mahalanobis distance. For this reason, a possible change point is at the 21st run when the damaging event occurs. The Normalized Likelihood Ratio Test results for the angle coefficient using ARX regression without mass can be analyzed and compared to the alpha coefficient. These results are shown in Figure 5.35 using the data without mass. As seen in the plots, the timing of damage is correctly identified. Each plot shows peaks at the 21st run of testing. The results are then analyzed for the localization of damage. Like the alpha coefficient ARX results using the same data set, damage is generally located to the right side of the frame. The plots for the coefficients from the left side of the frame, shown in Figure 5.35 (a), do not cross the boundary threshold indicating no significant change or damage to these locations. However, damage is detected on the right side of the frame because the NLRT statistic for coefficients Γ_{52-R}, Γ_{56-R}, Γ_{61-R}, Γ_{65-R}, Γ_{25-R}, Γ_{26-R}, and Γ_{51-R} cross the unity confidence bound. However, these results are not localized to the right side column.

The results for the angle coefficient using the data with mass show damage at locations R1 and R3 as shown in Figure 5.36 because Γ_{13-R} crosses the confidence bound. These locations are on the right side of the frame; however, the coefficient pairing only involves locations on the right side beam. Since the damage is on the right side column, these results would lead an observer to the correct vicinity of damage, but would

not lead to the exact location of damage. It can then be noted that ARX regression is a viable method for correct timing of damage and general damage localization when using data without mass. Both the alpha and angle coefficient results do correctly show an undamaged left side and a damaged right side of the frame with peaks in the graph at the correct point of the damaging event. However, the data with mass is proven to show inconsistent results. Additionally, more tests would need to be done to truly localize the damage to the center of the right side column.

5.2.3 Student's t-test

Instead of condensing the ARX alpha coefficients using the Mahalanobis distance in order to use them in the univariate Student's t-test, the residuals were used instead. For each run number there are 1000 data points for the residuals because that is equal to the length of each data test collected. The variance of all 1000 data points was calculated and then used in the Student's t-test to test for a change in the mean. It is assumed that the variance of the residuals would change due to a damage event. Therefore, the variance of the first 20 runs should be different from the variance of the second 20 runs. In effect, the change in mean of the variance of the residuals can be tested using the Student's t-test. Procedure 1 was used to split the data. The results for the ARX regression model are shown in Figure 5.37 and Figure 5.38 for the tests without and with mass respectively. Shown in Figure 5.37, most of the coefficients cross the unity confidence bound on the right side of the frame. The only ones that do not cross the threshold are α_{13-R}, α_{15-R}, α_{16-R}, α_{31-R}, α_{35-R}, α_{36-R}, α_{42-R} and α_{45-R}. This would indicate damage on the entire right side joint. However, one coefficient, α_{32-L} also

crosses the confidence bound on the left side of the frame. Yet, so many more coefficients cross the bound on the right side of the frame that an observer would be more inclined to check the right side of the frame first for damage. Therefore, this would correctly indicate the location of damage.

Similarly, the results for the data with mass in Figure 5.38 shows coefficient , α_{62-L} crossing the confidence threshold on the left side of the frame. However, more coefficients cross the right side. Those coefficients that do cross the bound, α_{16-R}, α_{52-R}, α_{56-R}, α_{61-R}, α_{62-R} and α_{65-R}, dictating damage on the right side mainly near locations R5 and R6. This is the correct location of damage. Therefore it can be noted that the change in mean of the variance of the residuals of the ARX model can be tested using the Student's t-test. Even though both data sets, without and with mass, indicate a one damaged coefficients on the left side, the damaged features are densely found on the right side of the frame. Therefore, the results can be used to correctly indicate damage.

5.2.4 Bayesian Hypothesis Test using Angle Coefficient

The angle coefficient from the ARX model does not need to be condensed using the Mahalanobis distance in order to be used in the Bayesian statistic. This statistic is univariate, it uses a vector and generates another vector. The angle statistic is scalar. When all the test runs are put together a vector is formed. Therefore, this coefficient can be directly used in the hypothesis test in order to detect damage. The results for the angle coefficient using the test data without mass for the left and right side coefficient Γ_{45} are shown in Figure 5.39 (a) and Figure 5.39 (b) respectively. It is shown that the right side of the frame at locations R4 and R5 are deemed damaged sections because the statistic

crosses the 95% confidence level. These locations are on the right side column of the frame and therefore this complies with the damage case. Additionally, the left side coefficient does not cross the bounds and is therefore deemed undamaged and healthy. However, the damage occurred at the 21st run of testing which corresponds to the 6th run on this plot. Therefore it would be expected that the statistic cross the threshold soon after this run was included in the statistic. Yet, the plot crosses the confidence bound at the 23rd run on this plot which corresponds to the 38th run of testing. This is a delayed detection of damage. Still, the correct location of damage is found.

The results for the data with mass are not as conclusive. As shown in Figure 5.40, the left and right side coefficient Γ_{65} crosses the confidence level indicating a damaged section. However, the left side is not damaged in this case. Other pairs of coefficients were analyzed and the results indicate that the entire frame was subject to damage. These results are not consistent with the damage case and therefore this pair of test data, damage feature and control chart should not be used together in order to correctly located damage.

5.2.5 Moving Range Chart for Residuals

The moving range control chart can be used with the ARX regression residuals in the same way that the single variate regression errors were used. The results using the data without mass are compared to those with mass. In Figure 5.41 (a) the results are shown using the data without mass for α_{53-R}. This coefficient, along with α_{24-R}, α_{25-R}, α_{34-R}, α_{52-R}, and α_{62-R} cross the confidence bounds. These results comply with the actual damaged location.

Similarly, the results for the data with mass generally location the actual damaged location. Shown in Figure 5.41 (b) the results are shown for α_{53-R}. The plots that also cross the bounds are α_{23-R}, α_{32-R}, α_{34-R}, α_{42-R}, and α_{63-R}. These results would lead an observer to the right side of the frame for damage. It would not indicate the right side column any more than the right side beam. As a result, the moving range control chart generally identifies the right side of the frame as a damaged section. It does not pinpoint the exact location of damage; however, does indicate the left side as an undamaged section which is consistent with the damage case.

5.2.6 F-Test for ARX residuals

The residuals for the ARX regression model can be used as damage features that can be tested for a change in their variance in the occurrence of damage. The residuals can be found using the equation below

$$\varepsilon(n)= \sum_{p=0}^{P} y_j(n-p) - \sum_{i=1}^{k}\sum_{q=0}^{Q} \alpha_{iq} y_i(n-q) \quad \textbf{(5.2)}$$

This is the equation for ARX regression shown in Chapter 2 rearranged to solve for the residuals. In effect, the alpha coefficients are tested for how well they estimate the relationship between the two locations involved within time. The data without mass and that with mass were used to generate the F-test for the ARX regression. The results are shown for a comparison in Figure 5.42 (a) and Figure 5.42 (b) respectively using Procedure 2. For the single variate regression residuals, Procedure 2 was able to find the placement and timing of damage. Therefore, it is now tested for the ARX regression residuals. For both plots in Figure 5.42, the correct timing of damage was found after the 20th run of testing because both plots peak at this point. The location of damage can then

be further analyzed. Using the data without mass, the F-test indicates a damaged location solely at R5 and R4 because α_{54-R} is the only coefficient that crosses the bounds of a 95% confidence level. This is a very localized results for damage considering the actual damage is located at R5.

The results for the test data with mass were not as conclusive. Four coefficients, α_{25-L}, α_{26-L}, α_{35-L} and α_{52-L} crossed the confidence bound for the same confidence level. These damage indices are from the left side of the frame. This joint was not subjected to any damage and should not be indicated as damaged locations. Therefore, these results are deemed inconclusive. The data with mass should not be used in conjunction with the F-Test for detecting a change in the variation of ARX residuals.

5.3 Auto Regression Results

In order to use many control charts, there is an assumption that the variables used are normally distributed. It is shown in Figure 5.43 that the AR alpha and angle coefficients are normally distributed through the use of a normal probability plot. Like the ARX alpha coefficient, the AR alpha coefficient is actually a matrix as described in Chapter 2.3.2. These are individually compared to a line to show that they are normally distributed. Contrastingly, in the angle coefficient, there is just a vector of coefficients because from each run test, a scalar value is found. In Figure 5.43(b) the blue "+" marks indicate the actual coefficients and the red line is a regressed line for comparison. The marks in both the alpha and angle coefficient stray from the line at the bottom of the plots. This may be why some of the methods do not precisely show the timing and location of damage shown in this section.

5.3.1 The Mahalanobis Distance used in the Fisher Criterion and Normalized Likelihood Ratio Test

5.3.1.1 Mahalanobis Distance

The Auto Regression alpha and angle coefficients can be used to create a Mahalanobis distance histogram. In this way, the damage features with the highest distance are those that are closest to the damage. First the alpha coefficient results will be examined. As seen in Figure 5.44 (a) the results for the data without mass are shown. The coefficient with the highest Mahalanobis distance is R1. This location is not very close to the actual damage although it is on the same joint; hence the results are not localized to the actual location of damage but would lead an observer to the correct vicinity of damage. Next the results for the data with mass can be analyzed and compared to those of the data without mass for the alpha coefficient. As seen in Figure 5.44 (b), those features with the highest distances include R5, R6 and L4. These coefficients on the right side indicate the correct location of damage on the right side column; however, the left side should not experience any damage. For this reason, the fisher criterion and normalized likelihood ratio test can be used to eliminate the discrepancies in these results by using a threshold value.

The results for the angle coefficient are similar to that of the alpha coefficient. In Figure 5.45 (a) the Mahalanobis distance histogram is shown for the angle coefficient using data without mass. R1, R2, R5 and R6 are associated with the highest distances. Therefore damage is localized to the correct joint of the frame; even though R1 and R2 are located on the right side beam, these locations are close the actual damaged location

and are within the same local joint. Contrastingly, the results for the angle coefficient for the data with mass in Figure 5.45 (b) show L5, R1 and R2 as damaged locations. These results are misleading for coefficients on the left side should not indicate damage. However, R1 and R2 are on the same local joint as the actual damage.

In all of these cases, nothing can be said about the significance of the larger distances because there is no threshold to use for a comparison. Therefore, the Fisher Criterion and Normalized Likelihood Ratio Test can be used to create this confidence bound and are described in the next sections. There should be no damage on the left side of the girder considering the damage case in question.

5.3.1.2 Fisher Criterion

The Fisher Criterion can be used to monitor the significance of the Mahalanobis distance magnitudes. Initially the alpha coefficient results will be examined. The results for the data without mass are shown in Figure 5.46 (a). This indicates damage at R1 (L1 crossing the confidence bound is neglected). However, the results for the data with mass are shown in Figure 5.46(b). This plot indicates that only R4 is a damaged section. This eliminates the coefficient α_{4-L} as a damaged feature. R4 is on the right side column of the frame near the damaged location and R1 is on the right side beam; however, the actual location of damage is not shown. Therefore these results would lead an observer in the correct vicinity of damage but not to the correct actual location of damage.

Similar to the results for the data without mass for the alpha coefficient, the results for the angle coefficient also indicate R1 as the damaged section as shown in Figure 5.47 (a). These results are not localized enough to lead an observer to the correct section of

damage; however, R1 is on the same local joint as the actual damage. The results for the data with mass, shown in Figure 5.47 (b) do not indicate any damaged location. None of the sensor's Fisher Criterion statistic crosses the boundary threshold for damage. This would lead to the conclusion that the entire frame is undamaged and in a healthy state which is not consistent with the damage case. Therefore, these results are deemed inconclusive and this pairing of damage indicator and control chart should not be used for statistical process control.

It should be noted that the Mahalanobis distance and Fisher Criterion results for the Auto Regression model do not give very localized or trustworthy results. As a whole, these results are not consistent with the actual damaged section and are misleading. This combination of damage feature and control chart should be used in conjunction with other control charts in order to solidify the results.

5.3.1.3 Normalized Likelihood Ratio Test

The AR alpha coefficients can be used in the NLRT control chart to test for a damaged location only after they are condensed using the Mahalanobis distance. The alpha coefficient results using the data without and with mass are shown in Figure 5.48 and Figure 5.49 respectively. The plots using the data with mas show a completely healthy frame; yet the results for the data without mass show a damaged frame on the left side of the frame because coefficient α_{2-L} crosses the unity bounds. Both of these results are not consistent with the actual damage case. Therefore the NLRT statistic should not be used with the alpha coefficient to detect and localize damage.

It should be noted that all the control charts based on the Mahalanobis distance of the AR model alpha coefficients give inconsistent results. This would lead to the conclusion that the Mahalanobis distance should not be used to the condense AR alpha coefficients in order to use them in univiariate methods. Since the AR model regresses the same location response onto itself, it may be more sensitive to manipulation.

5.3.2 Normalized Likelihood Ratio Test for the Angle Coefficient

The angle coefficient does not need to be reduced using the Mahalanobis distance in order to be used in the Normalized Likelihood Ratio Test. The results are shown in Figure 5.50 for the data without mass. Here Γ_{1-R} is indicated as a damaged location. This location is on the right side beam of the frame. Therefore these results would lead an observer to the correct vicinity of damage; however, not to the actual damaged location. Additionally, the results for the data with mass show a damaged section at L5 shown in Figure 5.51. This location is on the completely opposite side of the frame as the actual damaged location, R5. Therefore these results are very misleading As a result, the AR coefficients should not be used with the NLRT statistic to detect and localize damage.

It should be noted that neither the alpha nor the angle coefficient from the AR model should be used in the Normalized Likelihood Ratio Test to detect damage. The results do not comply with the actual damaged location.

5.3.3 Student's t-test

The same procedure was used for the AR residuals than for the ARX residuals. The variance of the residuals was tested for a change in the mean. The results for the data

without mass are shown in Figure 5.52. Here locations R5 and R6 are indicated as damaged sections of the frame because they cross the unity threshold. The results for the data with mass are not as conclusive. The entire frame is indicated as healthy and undamaged as shown in Figure 5.53. None of the coefficients from either side of the girder cross the threshold bound of unity. Therefore, this data set, damage feature, and control chart should not be used with each other in order to properly locate damage.

5.3.4 Moving Range for Residuals

The moving range control chart can be used with the linear regression residuals to pinpoint damage. The results for the data without mass and the data with mass pinpoint the same locations for damage. Shown in Figure 5.54 (a) the moving range plot for the data without mass is shown for α_5 on both the left and right side of the frame. Both of these plots indicate damage on the right and left side of the frame. This does not comply with the damage case in question. The results for the data with mass show the same result as shown in Figure 5.54 (b). In effect, the AR residuals should not be used in conjunction with the moving range chart to test for a variance change and detect the correct location of damage.

5.3.5 F-test for AR Residuals

In order to be able to localize the damage to one sensor node, AR regression can be used. In this way, a set of coefficients can be made for each location separately. Instead of comparing the actual coefficients, the regression residuals are used as damage features. The assumption is that the residuals have a mean of zero; however in the event of

damage, the variance will change while the mean stays the same. For this reason, an F-test is used to test the change in variance. The AR residuals can be found using Eq. (5.4) below:

$$\varepsilon(n)= \sum_{p=0}^{P} y_j(n-p) - \sum_{q=0}^{Q} \alpha_{iq} y_j(n-q) \quad \textbf{(5.4)}$$

The results for the data set without mass will be analyzed using Procedure 2 and the results are shown in Figure 5.55 (a). The graph for the F-Tests crosses the 95% confidence level for coefficients α_{6-R}, and α_{5-R}, sensors that are on the center and lower end of the right side of the frame. It is at these locations that the null hypothesis can be rejected; the variance of the damage sensitive features has change significantly enough to indicate an out of control process. In the case of damage detection, this signifies a damaged location which is the correct location of damage. On the plot, coefficient α_{6-R} actually has a larger magnitude for the F statistic than coefficient α_{5-R} which leads to the misleading conclusion that sensor R6 is actually where the damage is. However, these results are more localized than the ARX and collinear results because the damage is concentrated in the column and not the general right side of the girder. These results are very similar to the single variate regression results in which the same two sensors are identified as damaged.

Secondly, the results for the test data with mass can be analyzed, shown in Figure 5.55 (b) and compared to that of the data without mass. The same plot is generated for this set of data and the results are shown in Fig. Here, the coefficients that cross the bounds are α_{6-R}, and α_{2-R}. These results would indicate damage on the right side column and beam. This does generally locate the correct damaged section although the actual damaged location, R5, is not indicated. As a result, the results without mass show a

more localized result. However, the results with mass are not necessarily wrong. Damage is not found on the left side of the frame but is found on the right side. Therefore, an observer would be led to the correct vicinity of damage.

5.4 Tri-variate Model

5.4.1 Noise Comparison

As stated earlier, this model has not been used previously in damage detection schemes. The coefficient $\alpha_{5:64}$ is plotted for the left and right side coefficient shown in Figure 5.56**.** This figure can compared to that of the single variate regression coefficient α_{56} shown in Figure 5.3. Initially, the error parameter is higher than the single variate model. Therefore, the model is tested for its robustness against noise using the normalized error parameter. The model is compared to the trusted single variate model. In effect, the average of each coefficient pairing was calculated over the 20 healthy runs and 20 damaged runs. Then the percent change was generated. The changes were grouped for each sensor pairing with the same *j* location. Noise was generated by creating a signal of normal distribution with a mean of zero and standard deviation of .001 and .008 for comparison. This noise signal was added to the data and used in the algorithm to create the influence coefficients.

The results for the single variate regression were very consistent even when noise was added to the system as shown in Table 5.1. The only coefficients that had a concerning amount of error associated with them were the ones dealing with the L1 location which is to be expected because this sensor was found to be defective. The results are shown for

an increasing amount of noise. In all cases, the sensors that have the highest percent change are sensor R5 and R6. This is consistent with the damage case.

However, the results for the tri-variate regression model were not as solid. Even without noise, more than half of the coefficients had large normalized error parameters associated with them. An X indicates that all of the coefficients associated with that sensor location have very high error and therefore were disregarded as false alarms. Even with the false alarms removed, the tri-variate model was able to find locations R2 and R4 as damaged sections. These results are consistent with the damage case. It would be expected that any section close to the actual damaged location could have a significant change in response due to damage. Therefore, pinpointing R2 as a damaged location is not necessarily wrong. Yet, as more noise is added to the system, the results do not remain consistent. In Table 5.2, the model with a .008 standard deviation noise level indicates L4 and R4 as damaged locations. This does not comply with the actual damage location. Section L4 should remain categorized as a healthy section; therefore this model is not robust to noise. It is therefore not used as a regression type for any control charts.

5.4.2 Bayesian Comparison

In order to further justify that the tri-variate regression coefficients were not correctly portraying the change in response of the frame, a Bayesian hypothesis testing statistic was used that was verified in Labuz 2011. This method was used to verify the influence coefficients from simple beam-column connection. This specimen was a simpler structure than the frame discussed in this thesis; however, the Bayesian statistic proved that the influence coefficients from harmonic excitation on a simplified beam column connection

could be used as damage features for damage detection methods. The damage was correctly found on the beam. Further discussion of those results can be found in Labuz 2011. For this thesis, the Bayesian statistics were used again in order to verify the use of single and tri-variate regression coefficients from impact data from a larger, more complex specimen.

The first fifteen runs were used as a baseline for the Bayesian statistic. As a result, the results for this test will show twenty five run numbers. The first five are the last set of undamaged runs; the next twenty are the damaged tests. Therefore, damage should be detected soon after the 5th run at locations close to the actual damage.

As previously shown, the results are shown in Figure 5.21 for the single variate regression coefficient α_{45}. These results show that the left side of the frame, near locations L4 and L5 are undamaged. The statistic does not cross the threshold boundary. However, the right side coefficient does cross the threshold boundary right after the 5th run. Therefore, these results pinpoint the exact timing and location of damage. However, the results for the tri-variate coefficients are not as conclusive. These are shown in Figure 5.57 for coefficient $\alpha_{4:65}$. Damage is detected on both the left or the right sides. This does not comply with the damage case and therefore these coefficients should not be used as correct representations of the properties of the frame.

5.5 Collinear Regression Results

It should be noted that this type of regression can only be used with data without mass because dynamic effects cannot be considered. Therefore only this data set is used and presented as results. Additionally, it should be shown that the collinear alpha and angle

coefficients are normally distributed. In Figure 5.58 the distributions are graphically compared to that of a linear regressed line. The coefficients in both plots do not stray far from the red line and therefore can be assumed to have a normal distribution. Additionally, from Eq. (2.10), the collinear regression coefficients are the same for when y_i and y_j are switched. For example, the coefficient α_{ijk} should be the same as coefficient α_{jik} because the regressor is simply the average of y_i and y_j and will be the same no matter the order of y_i and y_j. For this reason, this section will call $\alpha_{k:ij}$ (and $\Gamma_{k:ij}$) for the regression coefficient that involves y_k being regressed onto y_i and y_j.

5.5.1 Exponentially Weighted Moving Average

The exponentially weighted moving average statistic can be used with the collinear regression influence coefficient to detect and localize damage. The results for this statistic using the data without mass for the alpha coefficient is shown in Figure 5.59. These results are from $\alpha_{6:54}$ on both the left side and right side of the frame. This coefficient involves all of the sensor nodes that are the locations closest to damage: the right side column. The coefficient pairing should cross the bounds of the EWMA on the right side; however, they should stay within the range of the bootstraps for the left side. As seen from the figure, $\alpha_{6:54-L}$ does not cross the bounds while $\alpha_{6:54-R}$ does. This would indicate R4, R5, and R6 as damaged locations on the frame and L4, L5 and L6 as healthy locations on the frame. This is consistent with the damage case; the confidence bounds was found to be 94.2%.

Unlike the results from the alpha coefficient, the results for the angle coefficient Γ_{654} do not show damage on either side of the frame for this coefficient. However, damage is

found on the right side column of the frame using $\Gamma_{4:56}$ shown in Figure 5.60. This proves that the order of regressors does change the results. The confidence level is found to be 94.8%. Additionally, no damage is found on the left side using this coefficient indicating this section as an undamaged, healthy section. Therefore, damage is detected and localized to the correct location of the frame using the angle coefficient.

5.5.2 Cumulative Sum

The Cumulative sum statistic can be used with the collinear regression influence coefficients to test for damage. The alpha coefficient results are shown in Figure 5.61 (a) and Figure 5.61 (b) for $\alpha_{4:56}$ on the left and right side respectively. The left side coefficient crosses the bounds while the right side coefficient does not. This is the opposite of what is to be expected based on the damage case. There is damage on the right side of the frame, not the left side. Other coefficients from this regression model were analyzed which indicate that there is damage all over the frame on both the left and right sides. These results are deemed inconclusive.

On the other hand, the result for the angle coefficient $\Gamma_{4:56}$ is shown in Figure 5.62. Here the results are consistent with the damage case. The plot for the right side coefficient indicates damage by crossing the confidence bound; however, the left side is deemed as healthy because it does not cross the confidence bound. It would be expected that the CUSUM plot would cross the bounds as close to the 20th run as possible. Yet, the plot doesn't cross the threshold until about the 38th run. This is a delayed indication of damage; still the correct location of damage is detected to a 94.5% confidence level. The sensors R4, R5 and R6 are closest to the actual damaged location and are expected to

change in response significantly in the presence of a damaging event. Additionally, the left side is correctly indicated as undamaged. Other pairs of coefficients are analyzed and the results are consistent with the damage case. In conclusion, the CUSUM should only be used with the collinear regression angle coefficient to correctly identify the timing and location of damage.

5.5.3 Mean Square Error Indicator

Only the results for the ModMSE are presented. The alpha and angle coefficient can be used in the ModMSE to detect and localized damage. The results for the angle coefficient $\Gamma_{4:56}$ are shown for the left and right side in Figure 5.63(a) and Figure 5.63 (b) respectively. Here the ModMSE for the original data crosses the bounds on sides of the frame. This would indicate damage on both the left and right side. However, this is not consistent with the damage case. Additionally, other pairs of coefficients were analyzed and these results also show damage on both sides of the frame. Therefore, the ModMSE should not be used with the collinear regression alpha coefficients to detect and localize damage.

However, the results for the angle coefficient $\Gamma_{4:56}$, shown in Figure 5.64 (a) and Figure 5.64 (b) for the left and right side $\Gamma_{4:56}$ coefficient, give the opposite result. No damage is found at locations R4, R5, and R6 as well as L4, L5 and L6. Additionally, damage is not detected anywhere on the girder because all the coefficients from the left and right side of the frame do not cross the bounds of the threshold.

It is noted that the collinear regression alpha and angle coefficients give misleading results using the ModMSE control chart. As a result, this pairing should not be used in control processes.

5.5.4 Normalized Likelihood Ratio Test

Collinear regression involves three different locations. In effect, the results may be able to show a more information about the location of damage because the coefficients themselves include more locations. It is still expected that those coefficients with combinations of the locations on the right side column will show significant damage by crossing the unity confidence bounds. The results for the alpha coefficient are shown in Figure 5.65. The plots are initially analyzed for the timing of damage. Both plots in Figure 5.65 show a peak at the 21st run of testing. This is the correct time of the damaging event. The results can then be analyzed for their effectiveness in localizing the damage to the right side column of the frame. The collinear regression for the alpha coefficient shows damage at locations R1, R2, R4, R5 and R6 because $\alpha_{2:64-R}$ and $\alpha_{6:51-R}$ cross the boundary threshold. Even though there is no damage at sensor location R1 and R2, these results are still localized to the right side of the frame with more attention to the right side column (R5 and R6) where the damage is actually located. Additionally, no coefficient crosses the bounds from the left side of the frame indicating an undamaged state of the frame. These results are in compliance with the actual damage location.

The results of the angle coefficient can now be compared to the alpha coefficient and are shown in Figure 5.66. They are very similar to the alpha coefficient, except that more coefficients cross the bounds. These include $\Gamma_{1:52-R}$, $\Gamma_{21:54R}$, $\Gamma_{1:62-R}$, $\Gamma_{1:46-R}$, $\Gamma_{1:65-R}$,

$\Gamma_{2:53-R}$, $\Gamma_{2:54-R}$, $\Gamma_{2:63-R}$, $\Gamma_{2:64-R}$, $\Gamma_{5:63-R}$,and $\Gamma_{6:51-R}$. This result pinpoints not only R2, R5, and R6 as damaged locations, but also R1, R3 and R4. All of these locations make up the right side of the frame. Inspecting the magnitudes of the statistic, it would seem as though all of these locations have similar amounts of damage with R2, R6 and R5 with the largest influence on the statistics that cross the bounds. This is the same result as the alpha coefficient.

The collinear regression results for both the alpha and angle coefficient show similar results. The damage is not as localized as the single variate regression; however, it does detect damage in the correct vicinity. Coefficients from right side of the girder cross the bounds while no coefficients from the left side cross. These results can be compared to those of the ARX results in which damage is localized to the right side of the girder but not within that particular side.

5.5.5 Moving Range Control Chart for Residuals

The residuals of the collinear regression can be used in a moving range chart to test for a change in their variance. The same procedure for condensing the residuals was used on this type of regression as the single variate regression. The results are shown in Figure 5.67 for coefficient $\alpha_{5:46-R}$. The coefficient does not cross the bounds on the left side but does on the right. This indicates damage on the right side column of the frame. Other coefficient pairings were analyzed and damage can be detected on the entire right side of the frame. It is not indicated specifically at the right side column; however, it does exclude the left side from any damaged section.

5.5.6 F-test for Collinear Regression Residuals

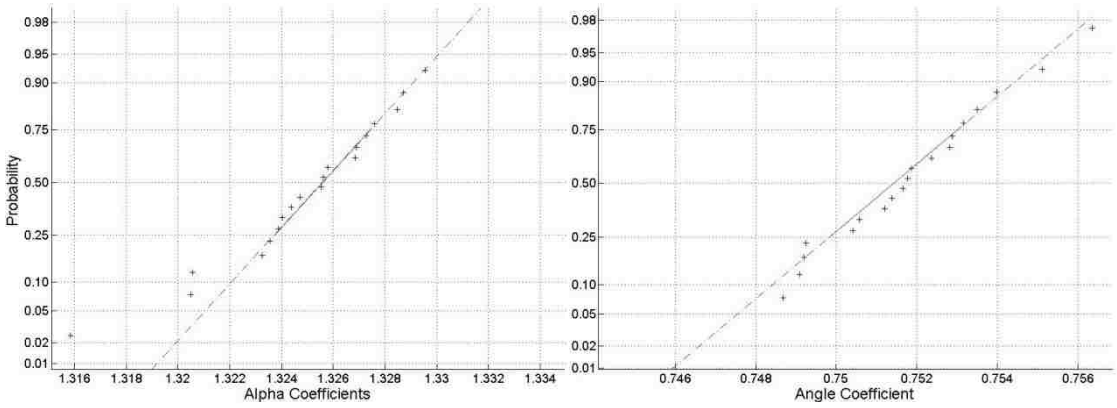
The residuals for the collinear regression can be used in the F-test to test the null hypothesis that the variance of the undamaged and damaged residuals stays the same. When the statistic crosses the bounds, the location associated with the residual can be deemed as a damaged location. The residuals for the collinear regression can be found using the equation below.

$$\varepsilon(n) = y_k(n) - \alpha_{ijk}\frac{y_i(n)+y_j(n)}{2} \quad \textbf{(5.3)}$$

Here y_k, y_j and y_i are the responses of three locations at time step *n* correlated through α_{ijk}.

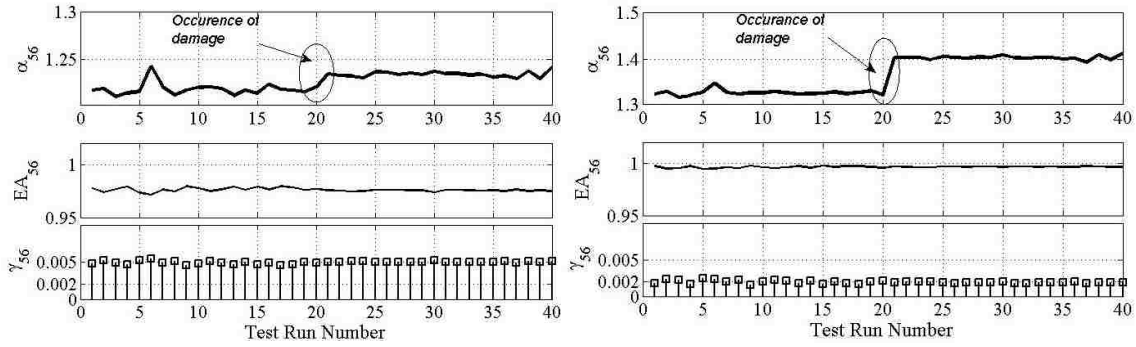
The results using Procedure 2 are shown in Figure 5.68. It can be seen that the correct timing of damage was found; the plots peak after the 20th run. Additionally, the results localize the damage to one sensor pairing: $\alpha_{6:52-R}$. This coefficient pinpoints the correct location of damage. Even though location R2 is on the beam, R5 and R6 are on the column very close to the actual damaged location. Therefore, these results are very localized and would lead an observer to the correct location of damage. These results are also very similar to the results for the angle coefficient using the normalized likelihood ratio test. In effect, the comparison of the results for both of these charts allows an observer to be more certain about the findings.

Figure 5.1: Damage on the back of a sensor that would impair it from working correctly



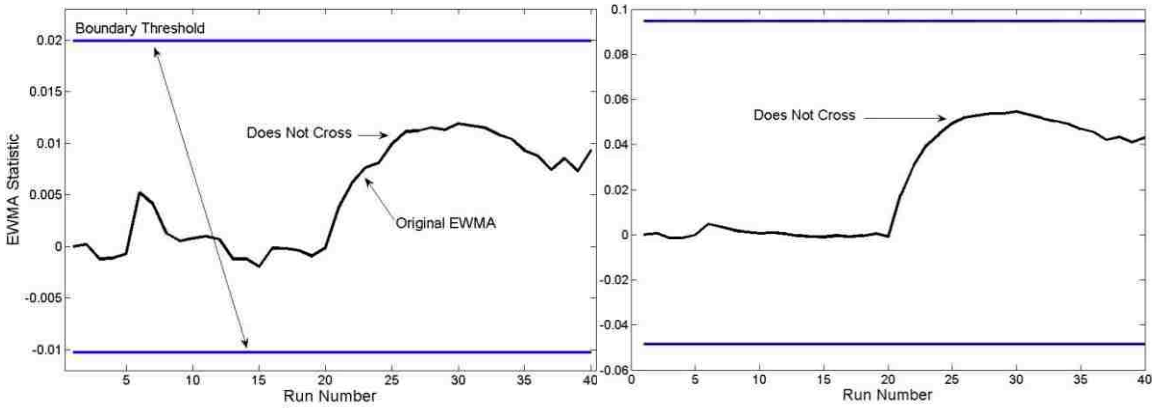
(a) (b)

Figure 5.2: Normal Probability plots for the (a) alpha coefficient α_{56-R} and (b) angle coefficient Γ_{56-R} for the single variate regression model



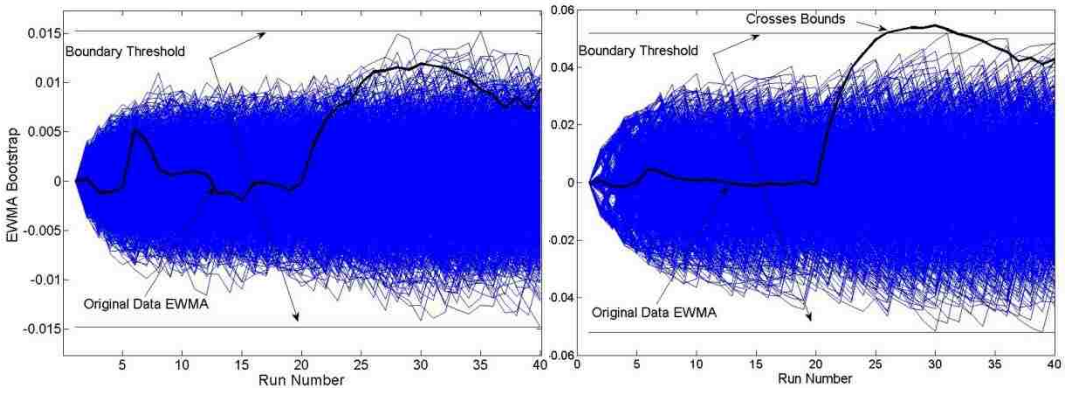
(a) (b)

Figure 5.3: Results for the coefficient values, accuracy and error parameters for (a) coefficient α_{56-L} and (b) coefficient α_{56-R}.



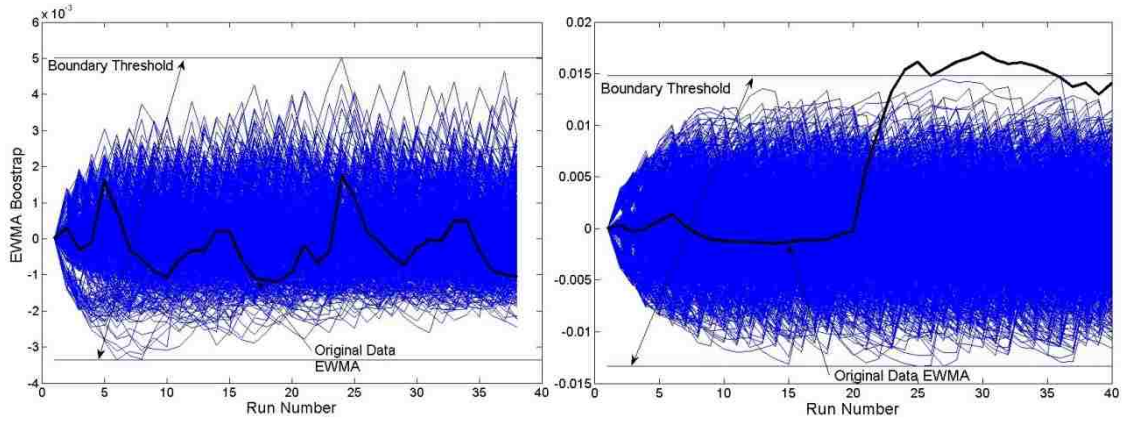
(a) (b)

Figure 5.4: Typical plots for the EWMA alpha coefficient with Shewhart threshold for the (a) coefficient α_{56-L} and (b) coefficient α_{56-R} .



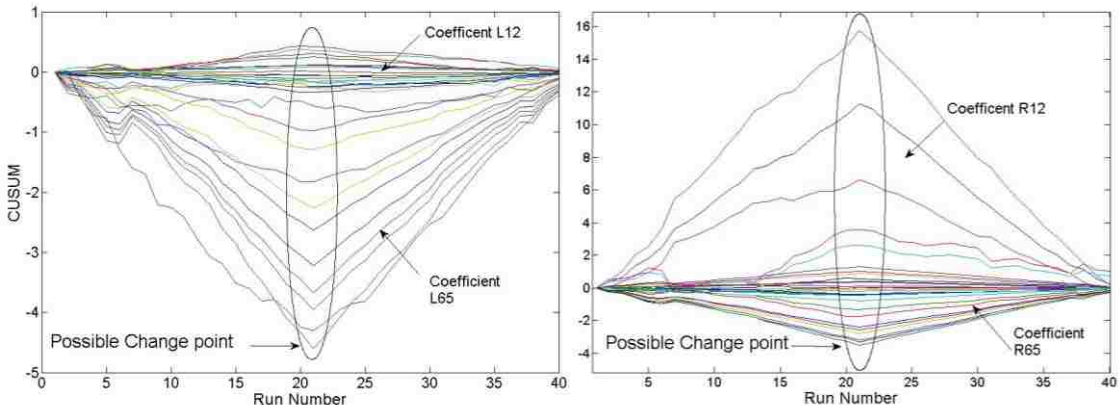
(a) (b)

Figure 5.5: Typical plots for the EWMA alpha coefficient with bootstrapping for the (a) coefficient α_{56-L} and (b) coefficient α_{56-R} .



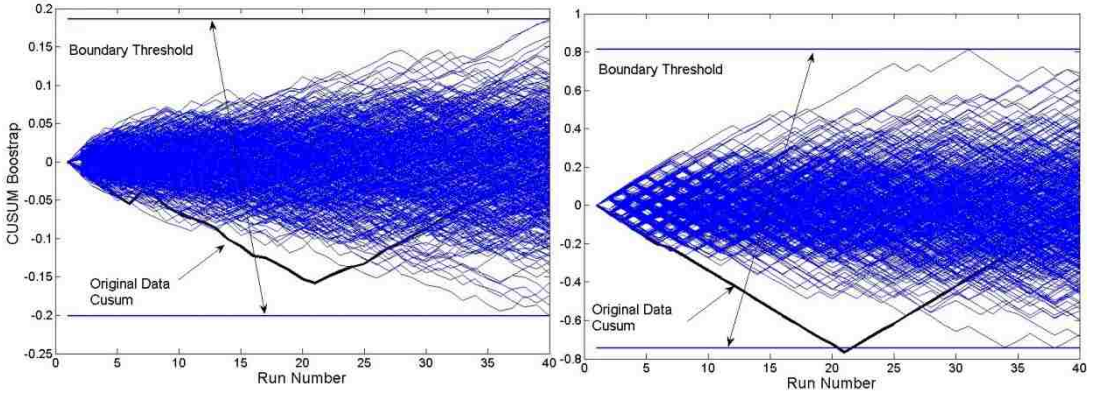
(a) (b)

Figure 5.6: Typical plots for the EWMA angle coefficient with bootstrapping for the (a) coefficient Γ_{56-L} and (b) coefficient Γ_{56-R}.



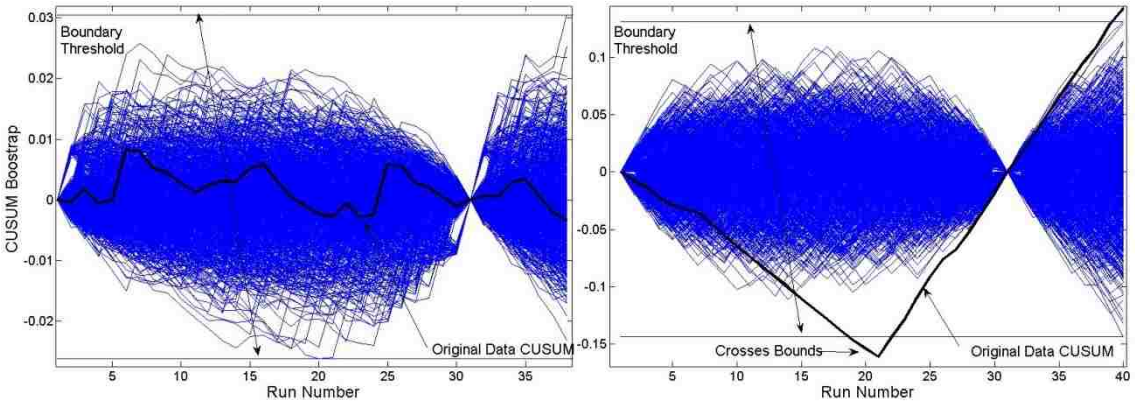
(a) (b)

Figure 5.7: Cumulative Sum Charts for (a) the left and (b) right side of the frame. Shows no boundary for detecting damage, just a possible change point



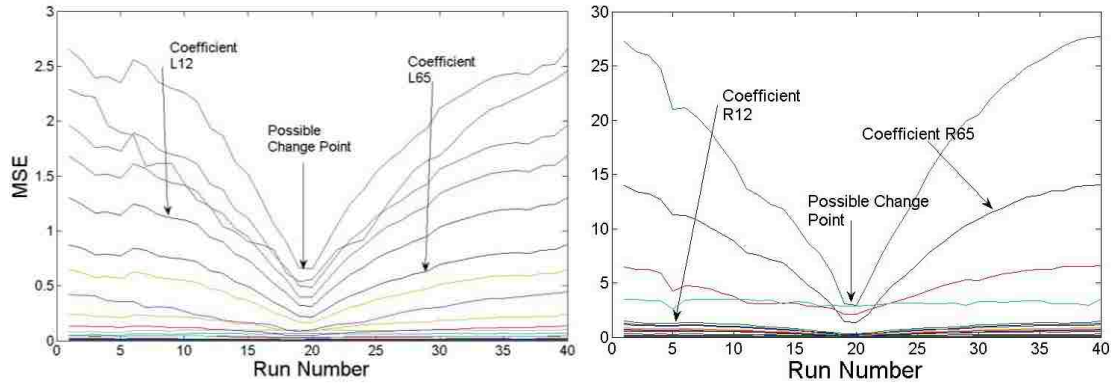
(a) (b

Figure 5.8: Typical plots for the CUSUM alpha coefficient with bootstrapping for (a) coefficient α_{56-L} and (b) coefficient α_{56-R}



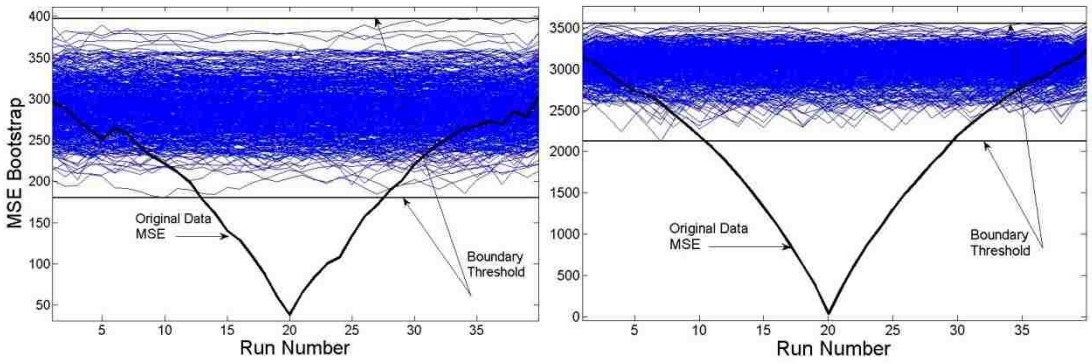
(a) (b)

Figure 5.9: Typical plots for the CUSUM angle coefficient with bootstrapping for (a) coefficient Γ_{56-L} and (b) coefficient Γ_{56-R}



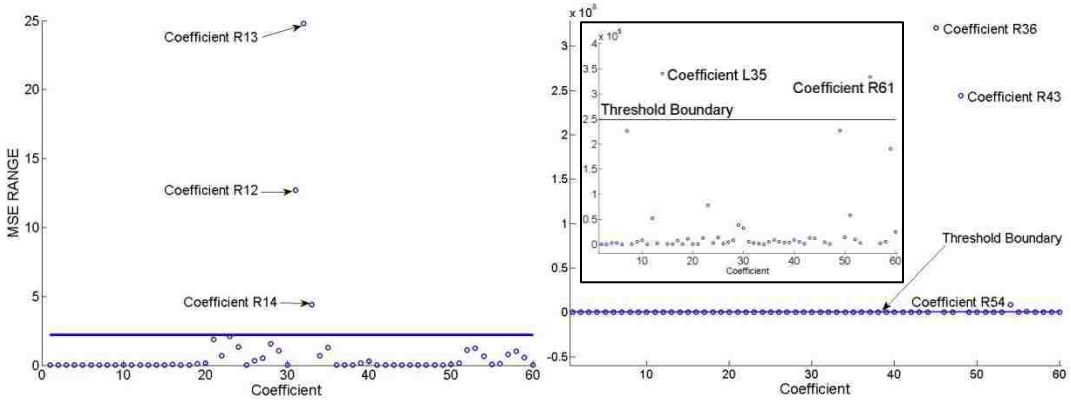
(a) (b)

Figure 5.10: Typical plots of Mean Square Error Indicator for alpha coefficient on the (a) left and (b) right sides of the frame



(a) (b)

Figure 5.11: MSE plots for alpha (a) coefficient α_{56-L} and (b) coefficient α_{56-R} . The threshold limit is crossed on both sides within the supposed undamaged runs and the bootstrap does not fit the data.



(a) (b)

Figure 5.12: The magnitudes of the (a) MSE and (b) ModMSE for all pairs of alpha coefficients.

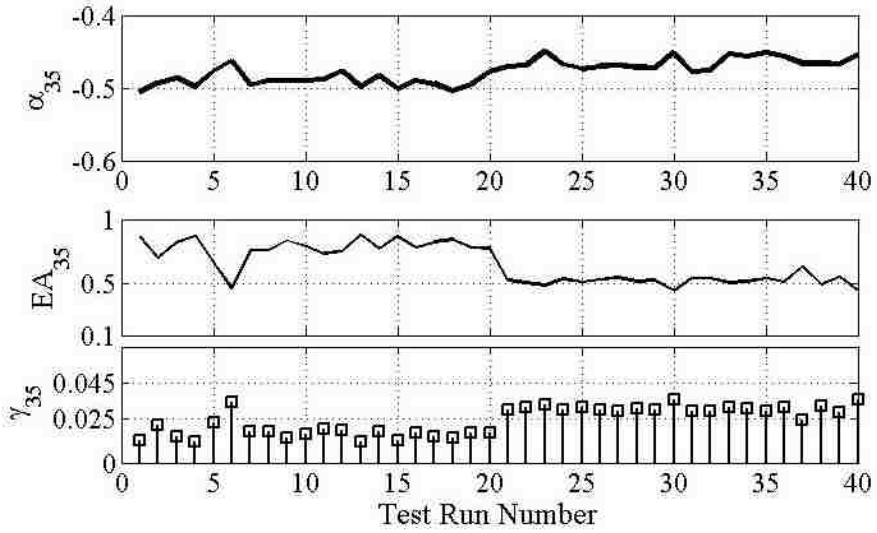
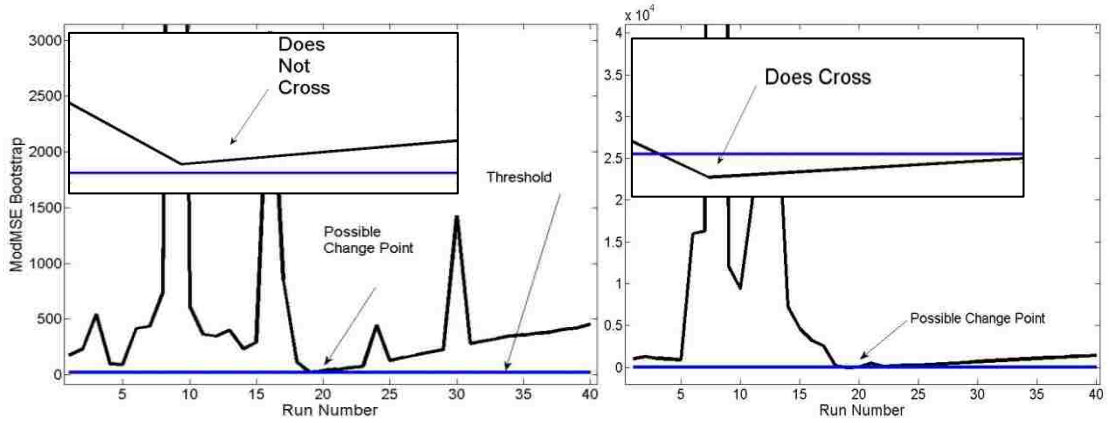
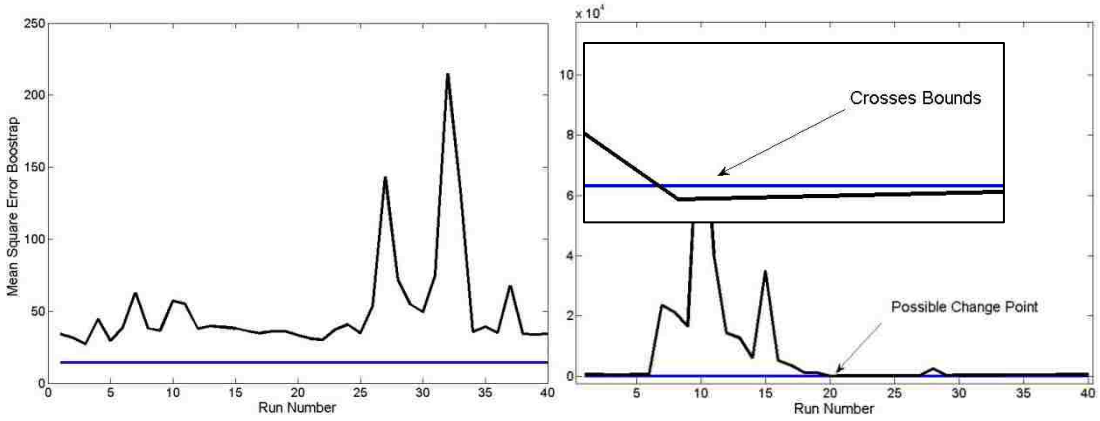


Figure 5.13: Eliminating False Alarms: The accuracy and error parameters associated with coefficient α_{35-L} can be analyzed to correct for any false results.



(a) (b)

Figure 5.14: ModMSE Bootstrap for alpha (a) coefficient α_{56-L} and (b) α_{56-R} coefficient .



(a) (b)

Figure 5.15: ModMSE Bootstrap for angle (a) coefficient Γ_{56-L} and (b) Γ_{56-R} coefficient.

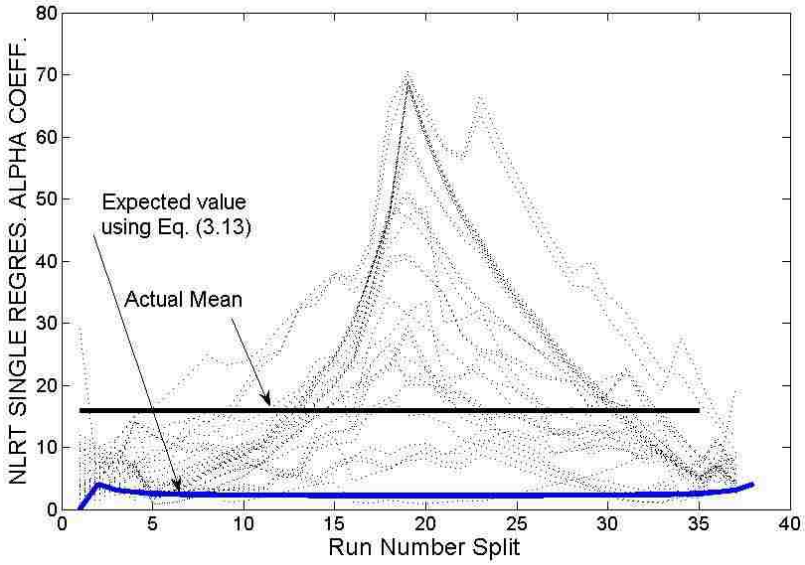
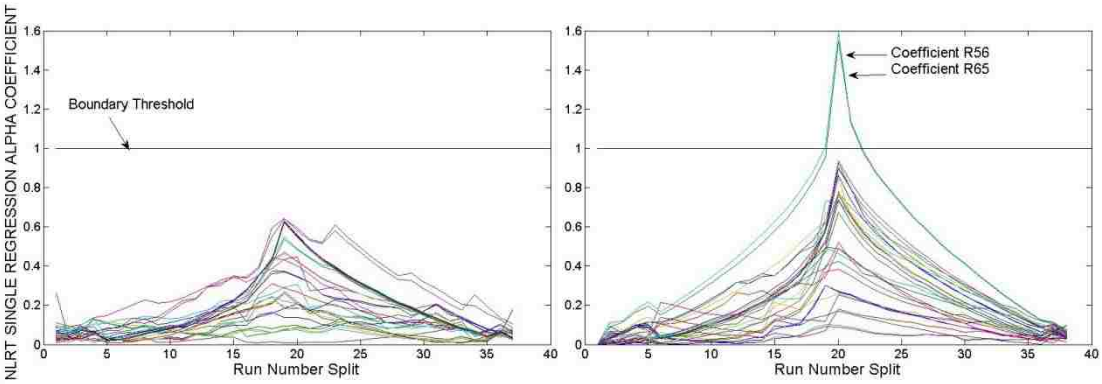
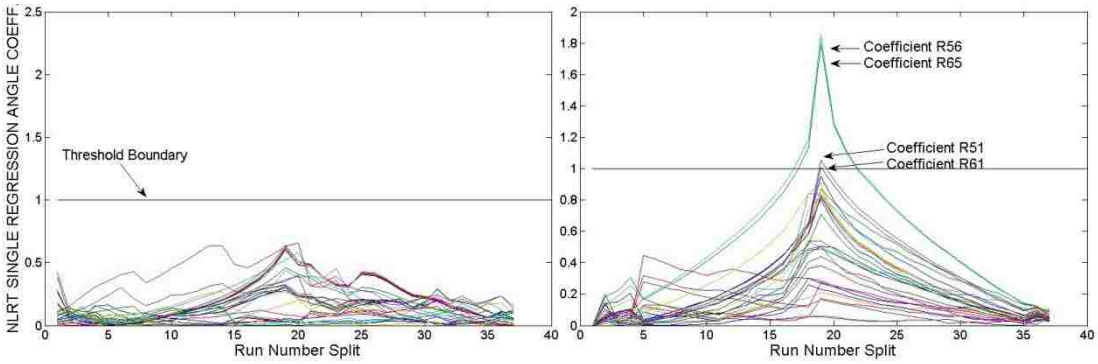


Figure 5.16: Likelihood Ratio Statistic plotted with actual mean and expected value obtained from Eq. (3.13).



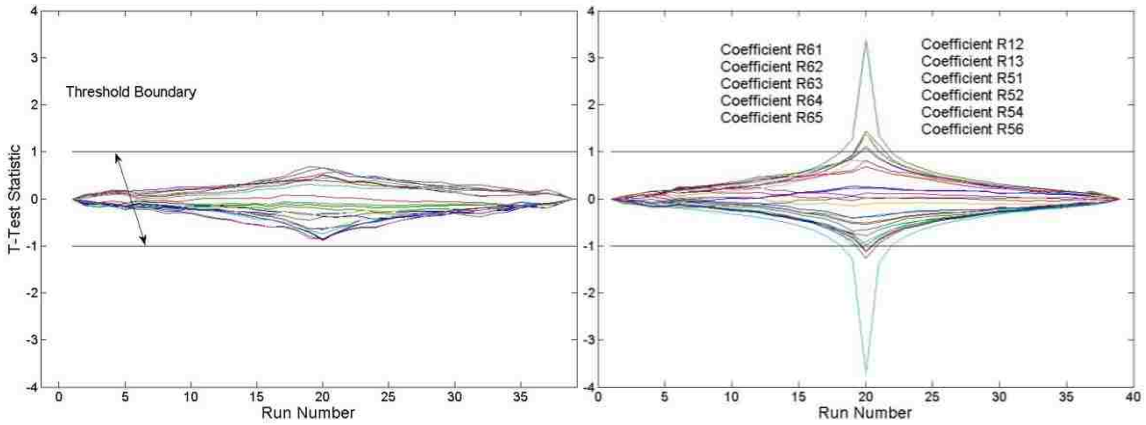
(a) (b)

Figure 5.17: The results for the Single variate regression using the alpha coefficient in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder.



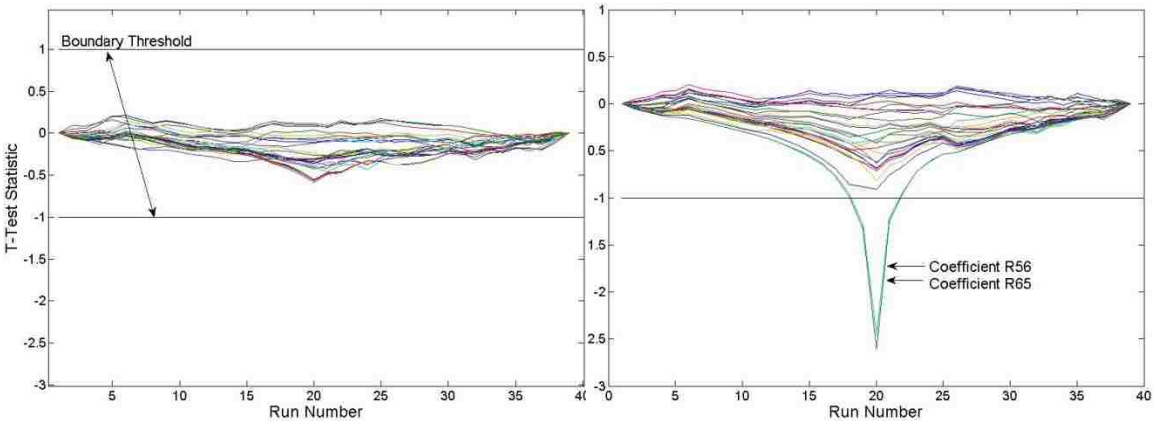
(a) (b)

Figure 5.18: The results for the Single variate regression using the angle coefficient in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder.



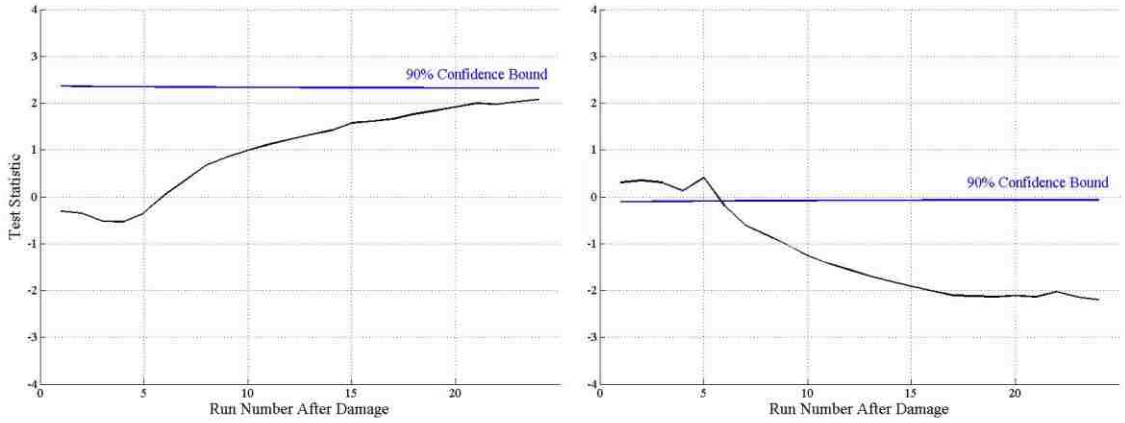
(a) (b)

Figure 5.19: Student's t-test for the Single Variate alpha coefficient for the (a) left and (b) right side of the frame respectively.



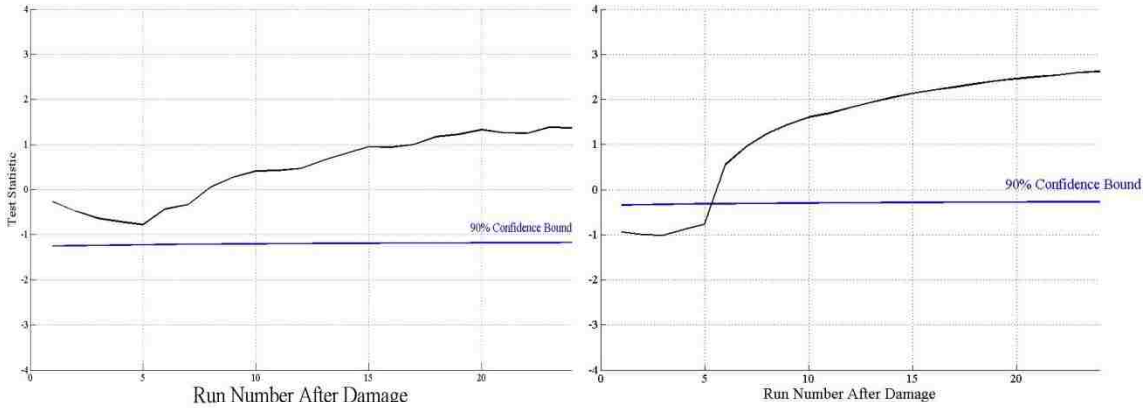
(a) (b)

Figure 5.20: Student's t-test for the Single Variate angle coefficient for the (a) left and (b) right side of the frame respectively.



(a) (b)

Figure 5.21: Single Variate Regression alpha coefficient using impact test data Bayesian Hypothesis test for the coefficient (a) α_{45-L} and coefficient (b) α_{45-R}



(a) (b)

Figure 5.22: Single Variate Regression angle coefficient using impact test data Bayesian Hypothesis test for the coefficient (a)Γ_{56-L} and coefficient (b) Γ_{56-R}

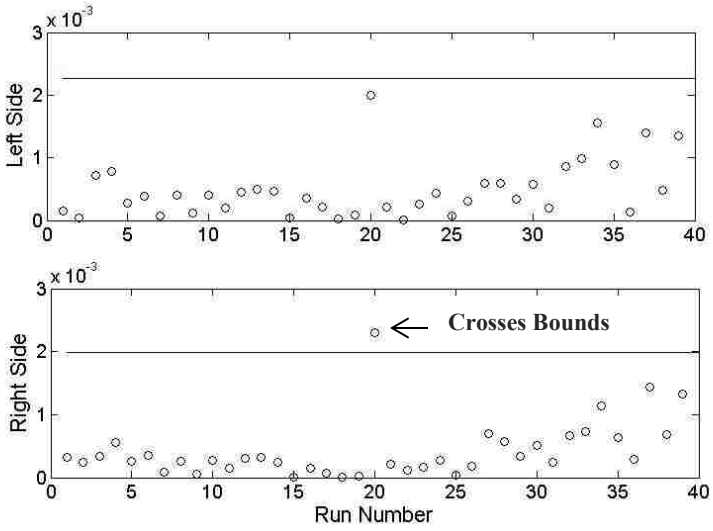


Figure 5.23: Moving Range results for the regression residuals of the single variate regression model for coefficient α_{54}

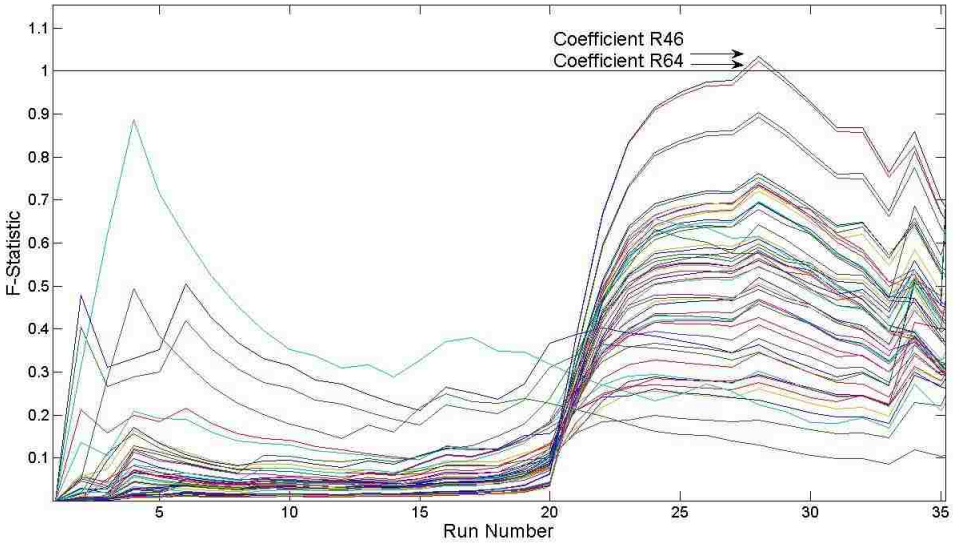


Figure 5.24: F-test results for the regression residuals of the single variate regression model using Procedure 1.

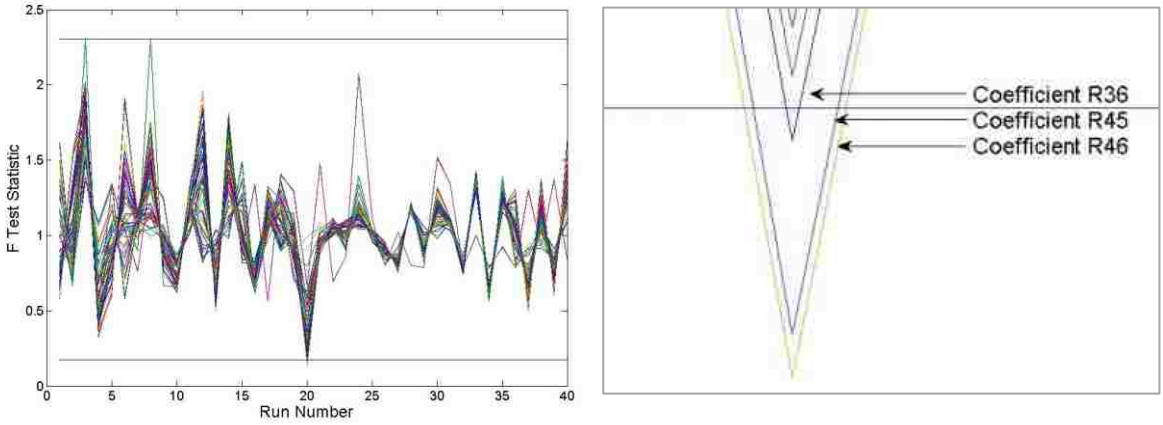
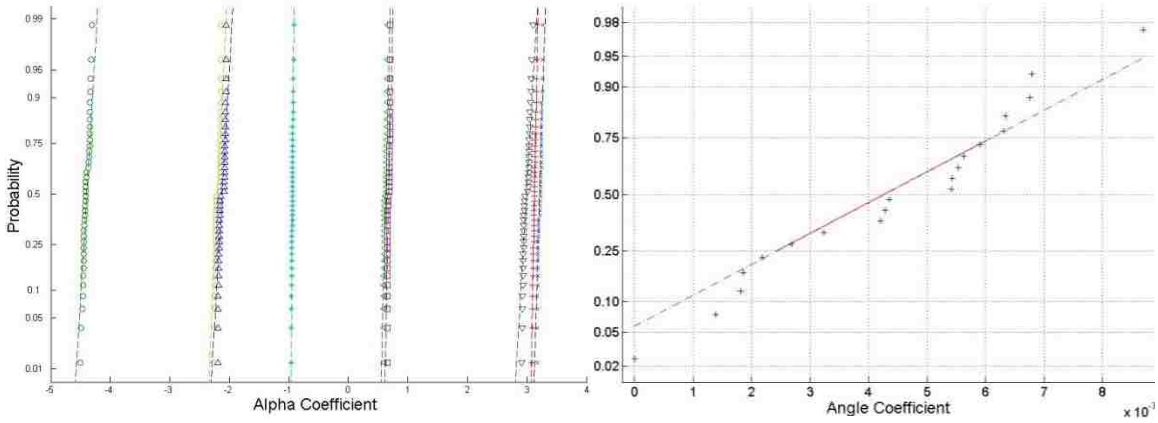
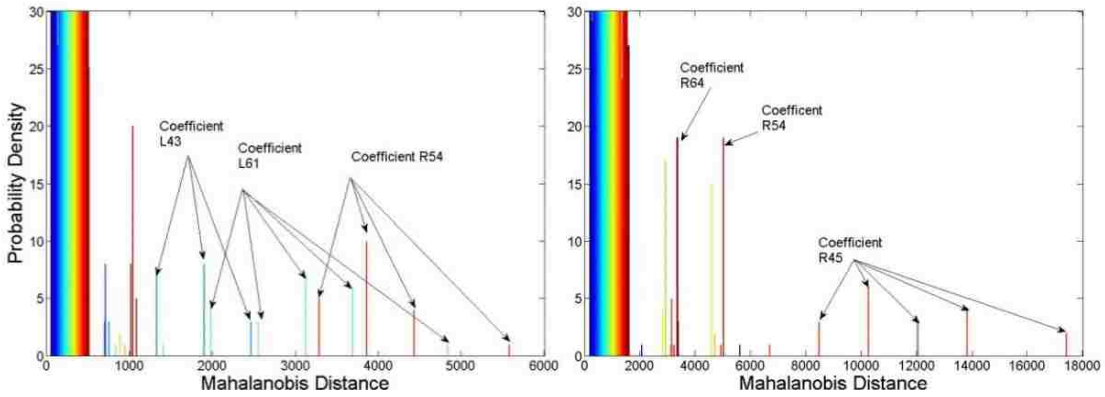


Figure 5.25: F-Test results for the regression residuals of the single variate regression model using Procedure 2.



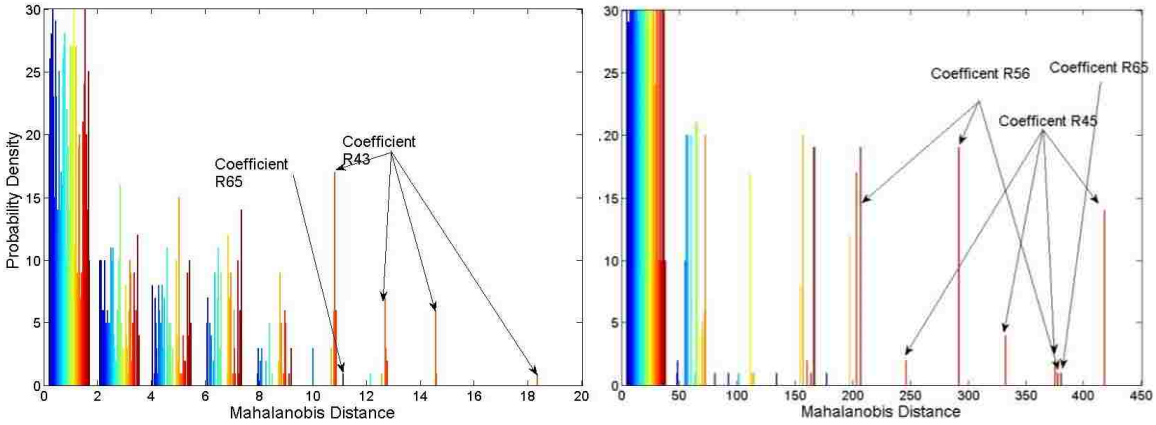
(a) (b)

Figure 5.26: Normal Probability Plot for the (a) alpha α_{456-R} and (b) angle coefficient Γ_{456-R} from the ARX regression model



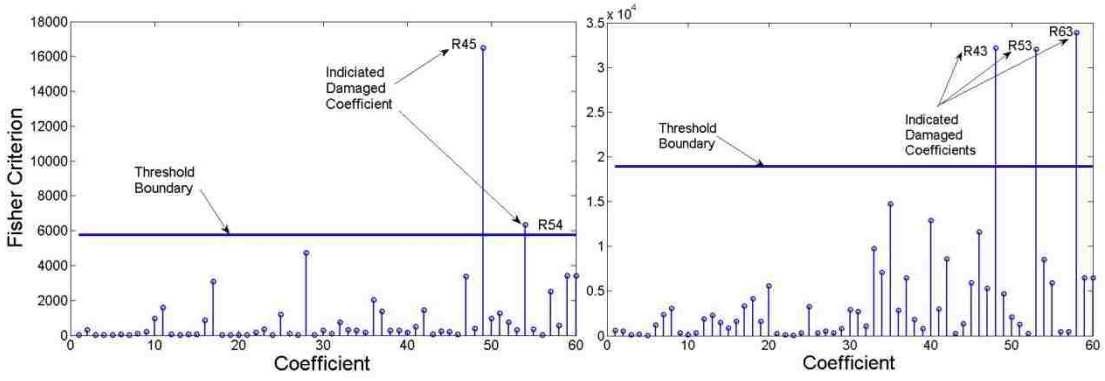
(a) (b)

Figure 5.27: Mahalanobis Distance Histograms for alpha coefficient using data (a) without mass and (b) with mass.



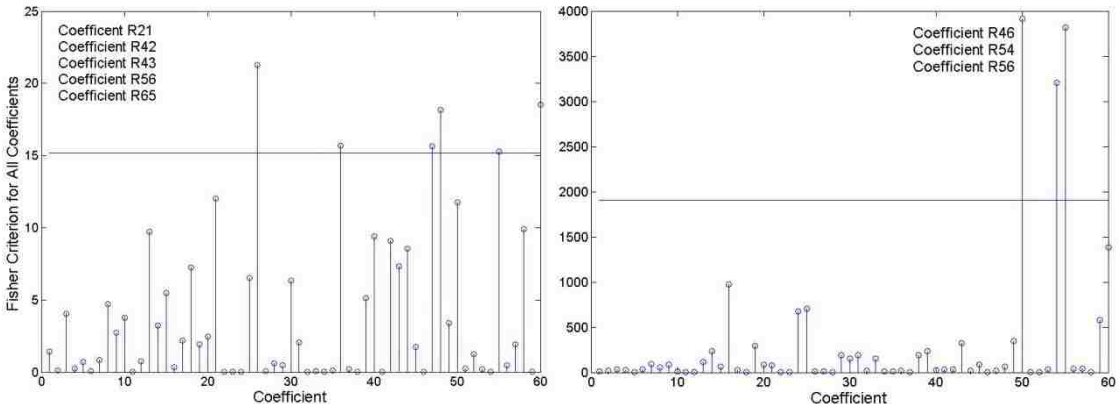
(a) (b)

Figure 5.28: Mahalanobis Distance Histograms for angle coefficient using data (a) without mass and (b) with mass.



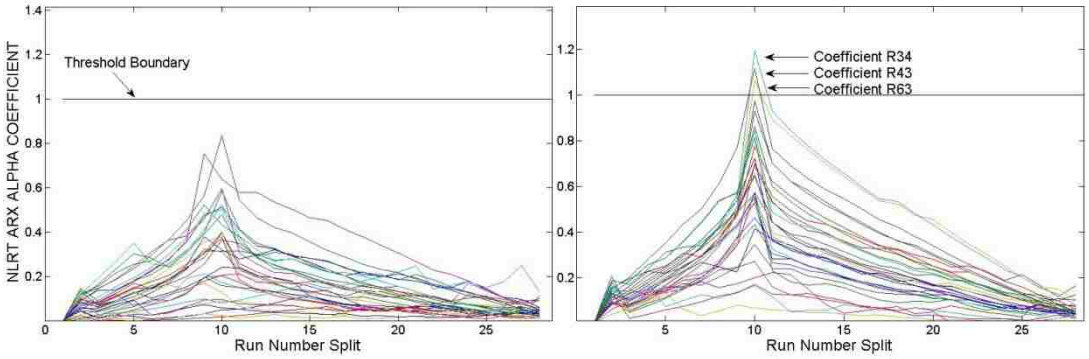
(a) (b)

Figure 5.29: Fisher Criterion for alpha coefficient (a) without and (b) with mass.



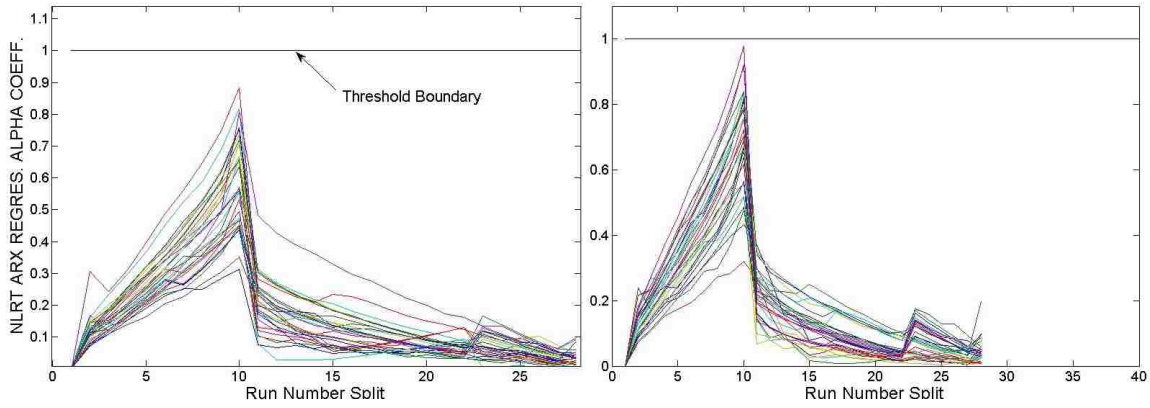
(a) (b)

Figure 5.30: Fisher Criterion for angle coefficient (a) without and (b) with mass.



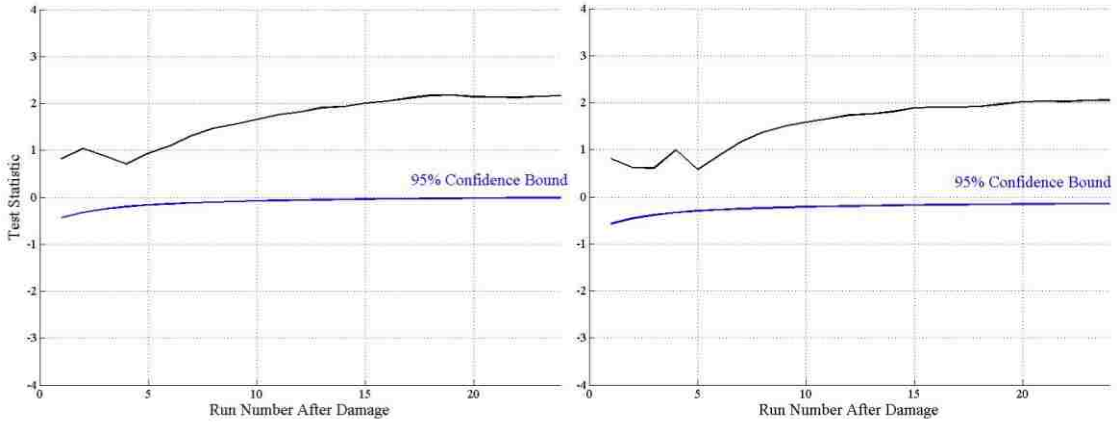
(a) (b)

Figure 5.31: The results for the ARX regression using the alpha coefficient without mass in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder



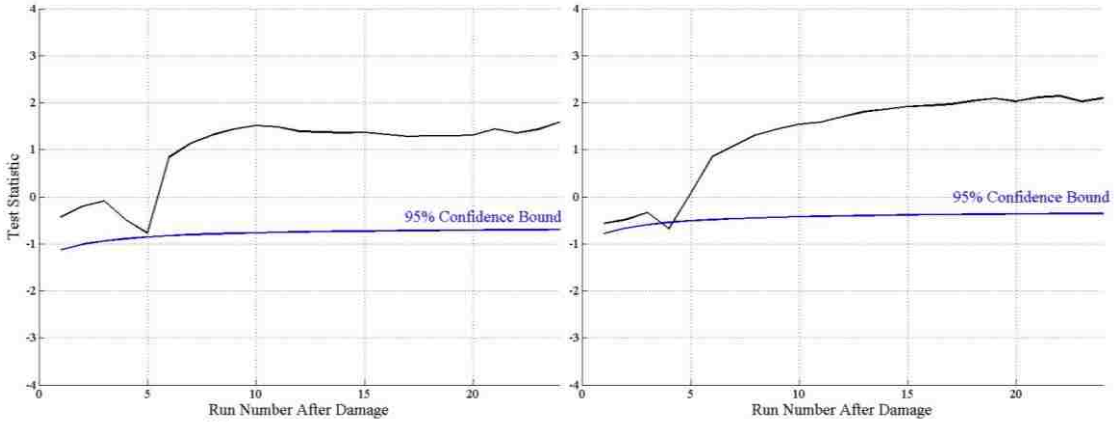
(a) (b)

Figure 5.32: The results for the ARX regression using the alpha coefficient with mass in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder



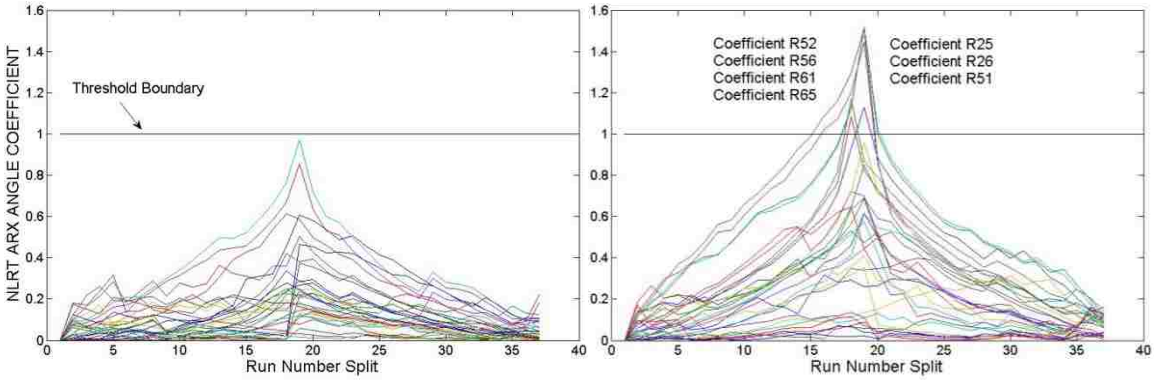
(a) (b)

Figure 5.33: Bayesian Statistic Hypothesis Test for ARX model alpha coefficient without mass for coefficient α_{65}



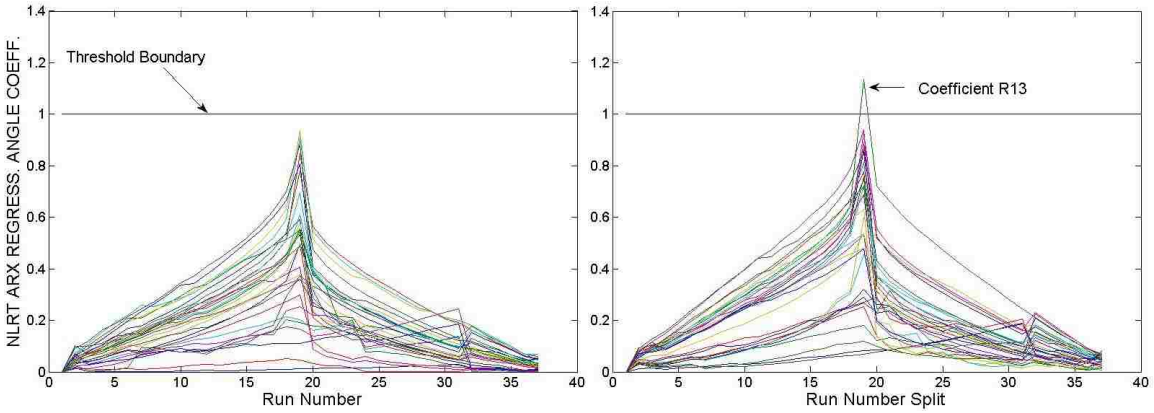
(a) (b)

Figure 5.34: Bayesian Statistic Hypothesis Test for ARX model alpha coefficient with mass for coefficient α_{65}



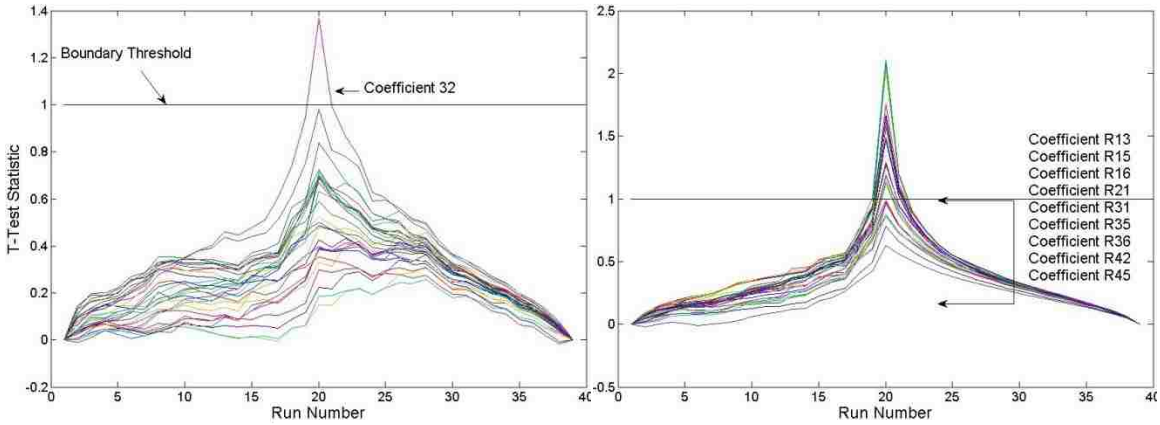
(a) (b)

Figure 5.35: The results for the ARX regression using the angle coefficient without mass in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder.



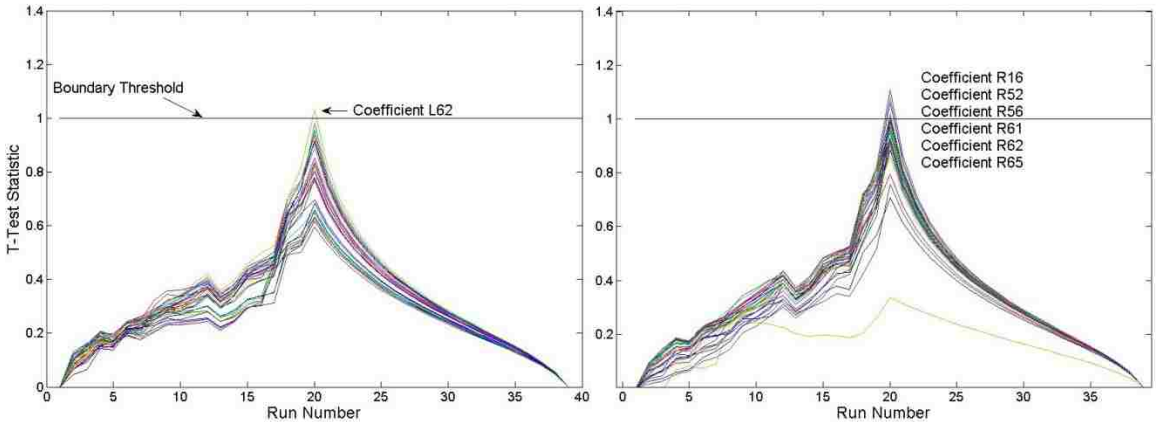
(a) (b)

Figure 5.36: The results for the ARX regression using the angle coefficient with mass in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder.



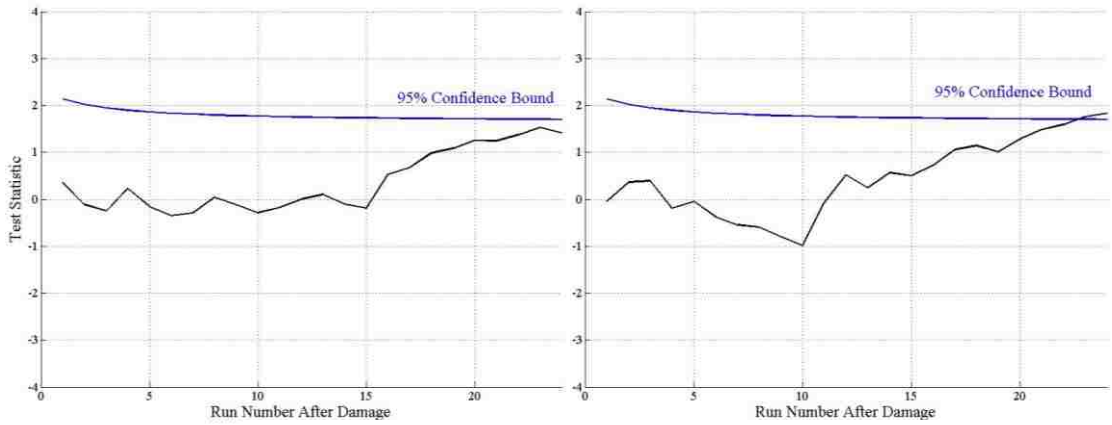
(a) (b)

Figure 5.37: Student's t-test for the ARX residuals using the data without mass for the (a) left and (b) right sides of the girder.



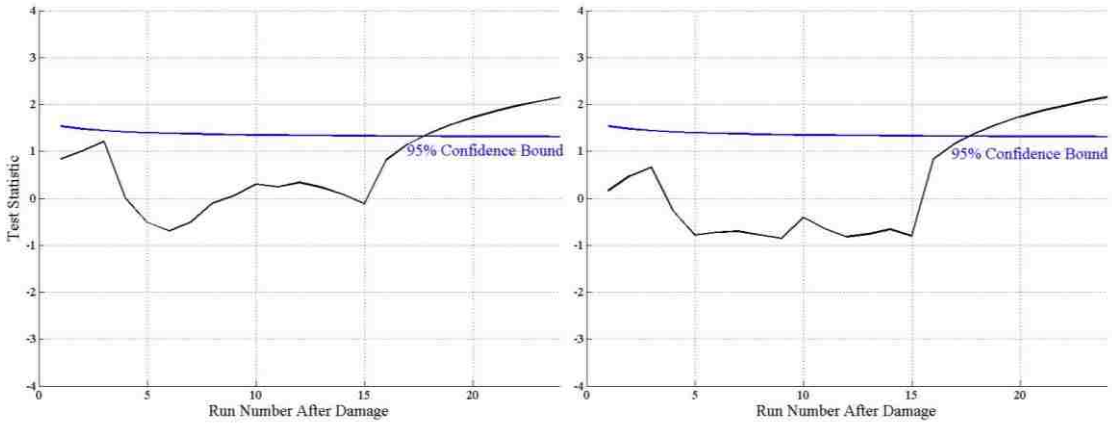
(a) (b)

Figure 5.38: Student's t-test for the ARX residuals using the data with mass for the (a) left and (b) right sides of the girder.



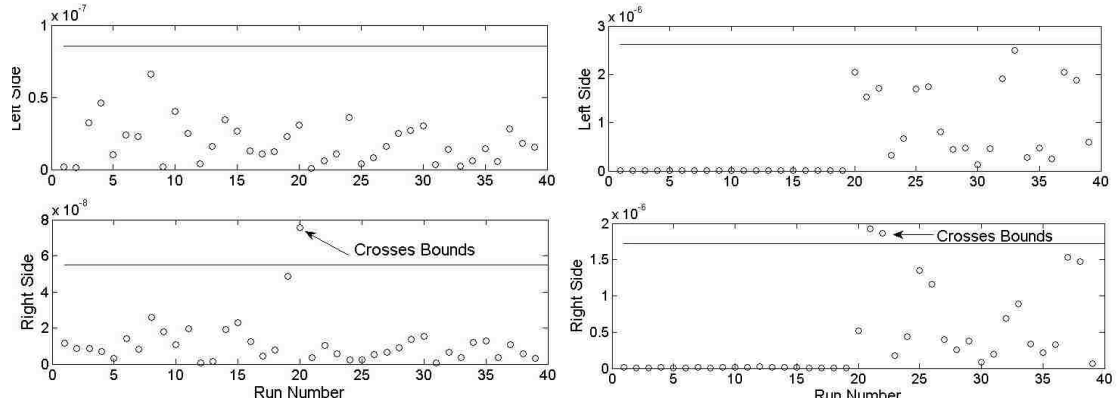
(a) (b)

Figure 5.39: Bayesian hypothesis test for the ARX angle coefficient without mass for (a) coefficient Γ_{45-L} and (b) coefficient Γ_{45-R}



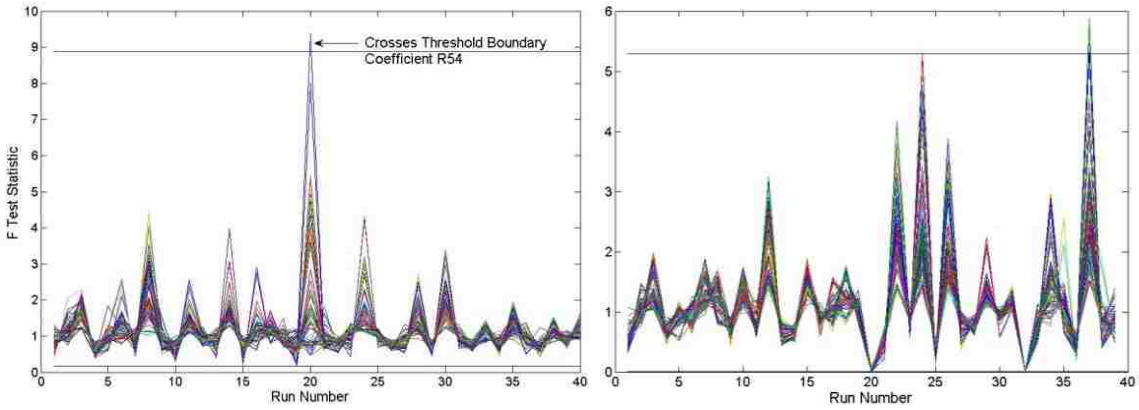
(a) (b)

Figure 5.40: Bayesian hypothesis test for the ARX angle coefficient without mass for (a) coefficient Γ_{65-L} and (b) coefficient Γ_{65-R}



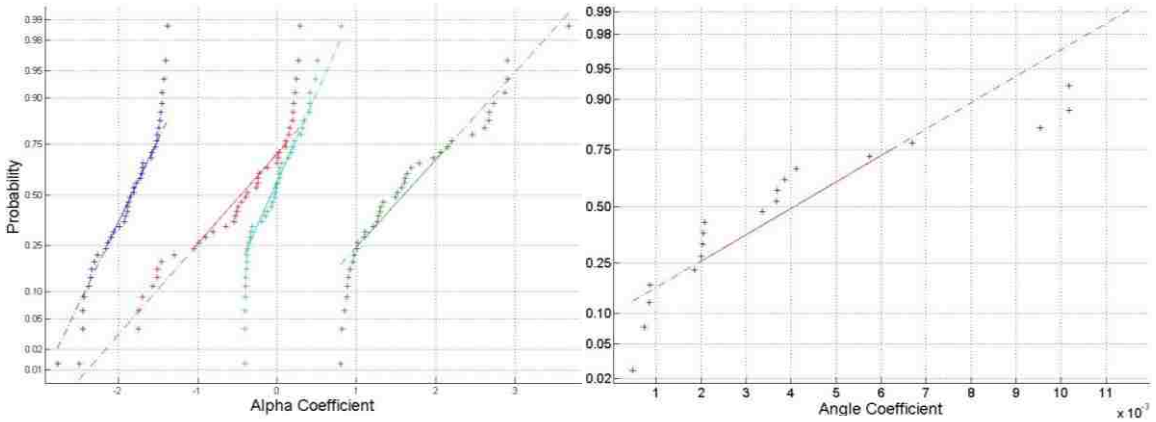
(a) (b)

Figure 5.41: Moving Range results for the ARX regression model coefficient α_{53} residuals (a) without mass and (b) with mass



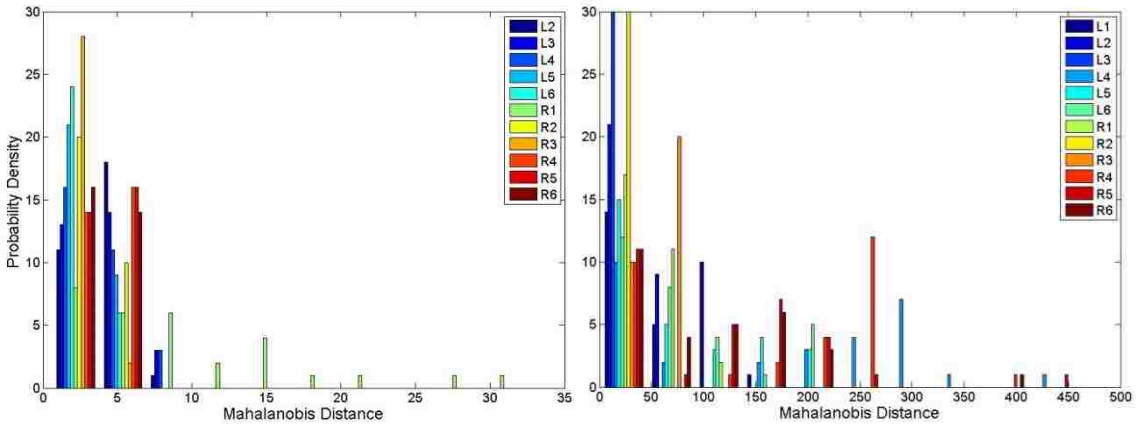
(a) (b)

Figure 5.42: F-Test results for the ARX regression model residuals (a) without mass and (b) with mass using Procedure 2



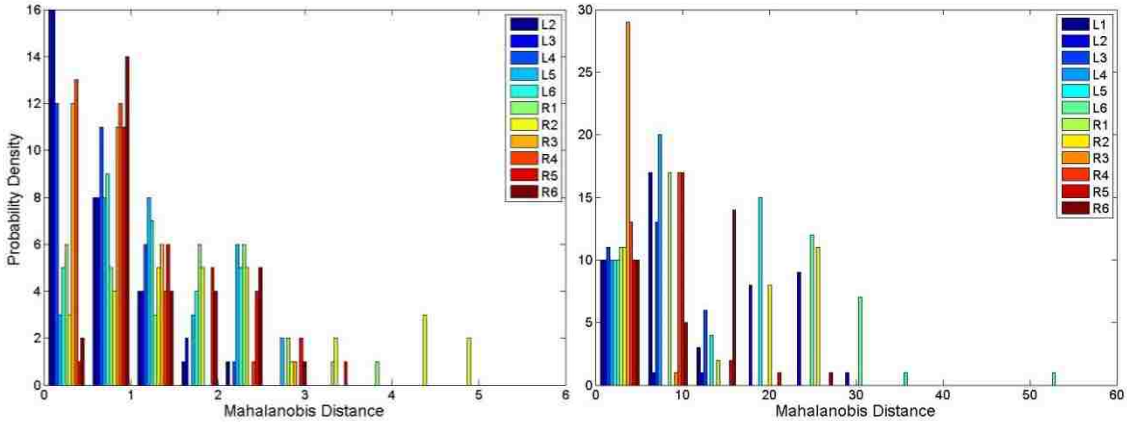
(a) (b)

Figure 5.43: Normal Probability plot for (a) alpha α_{5-R} and (b) angle coefficient Γ_{5-R} for the AR regression model.



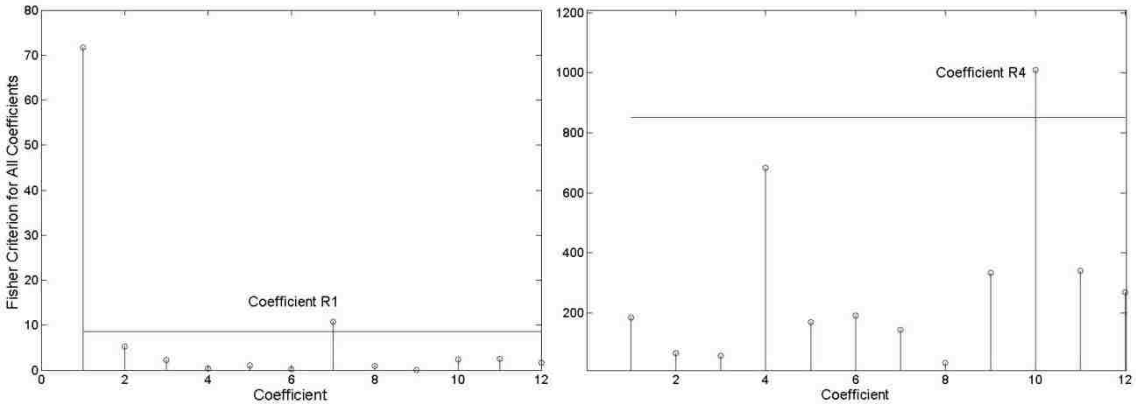
(a) (b)

Figure 5.44: Mahalanobis Distance Histograms for alpha coefficient for Auto Regression using data (a) without mass and (b) with mass.



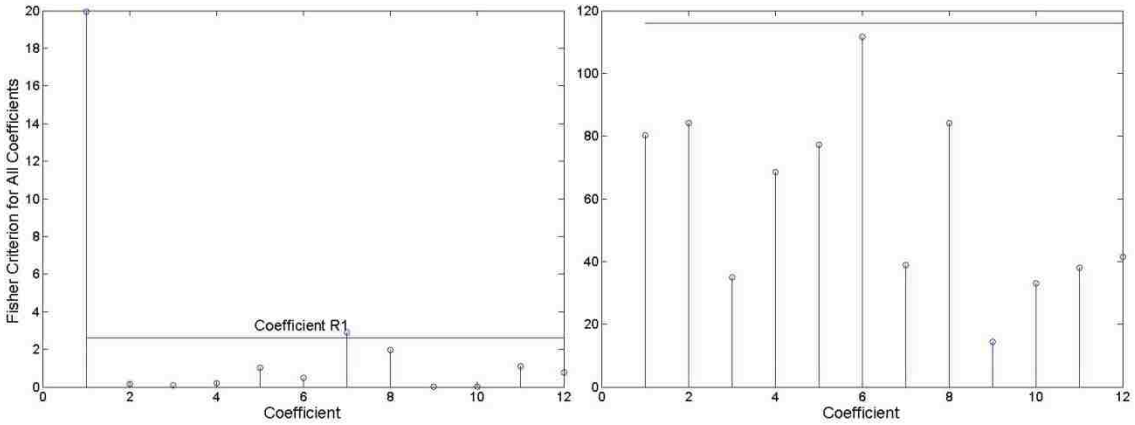
(a) (b)

Figure 5.45: Mahalanobis Distance Histograms for angle coefficient for Auto Regression using data (a) without mass and (b) with mass.



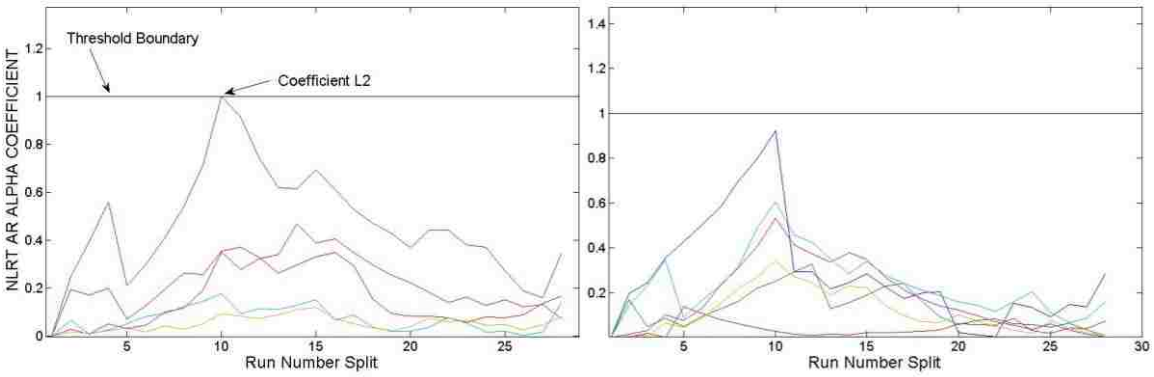
(a) (b)

Figure 5.46: Fisher Criterion results for the auto regression model for the alpha coefficient using data (a) without mass and (b) with mass



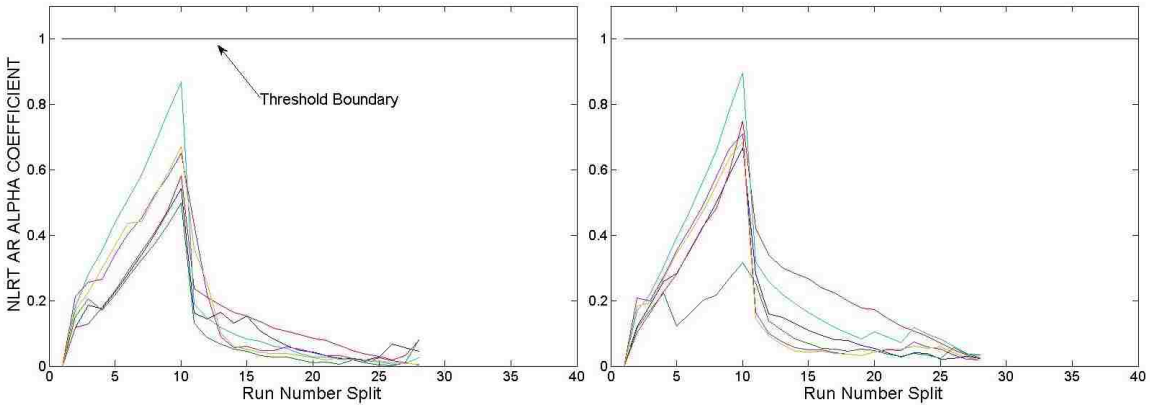
(a) (b)

Figure 5.47: Fisher Criterion results for the auto regression model for the angle coefficient using data (a) without mass and (b) with mass



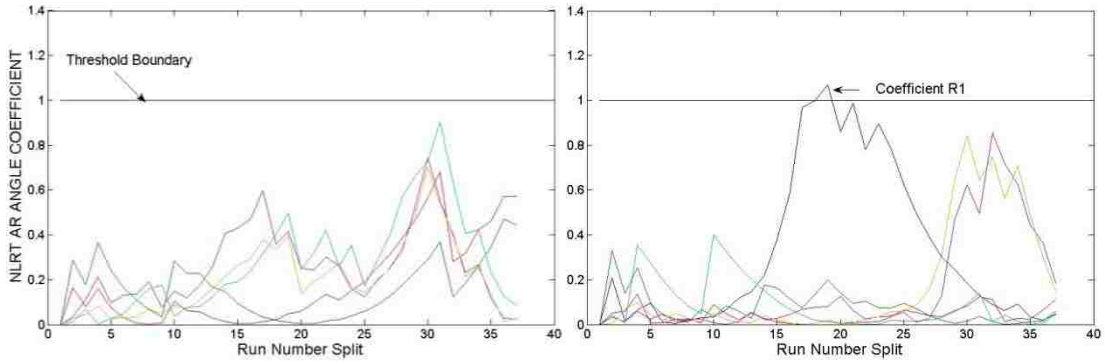
(a) (b)

Figure 5.48: NLRT results for the auto regression model without mass for the alpha coefficient on the (a) left side and (b) right side of the girder



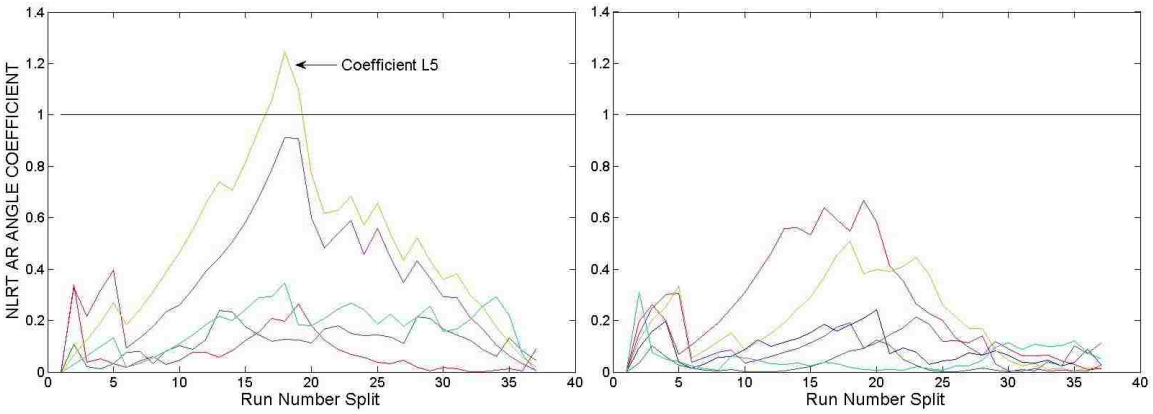
(a) (b)

Figure 5.49: NLRT results for the auto regression model with mass for the alpha coefficient on the (a) left side and (b) right side of the girder



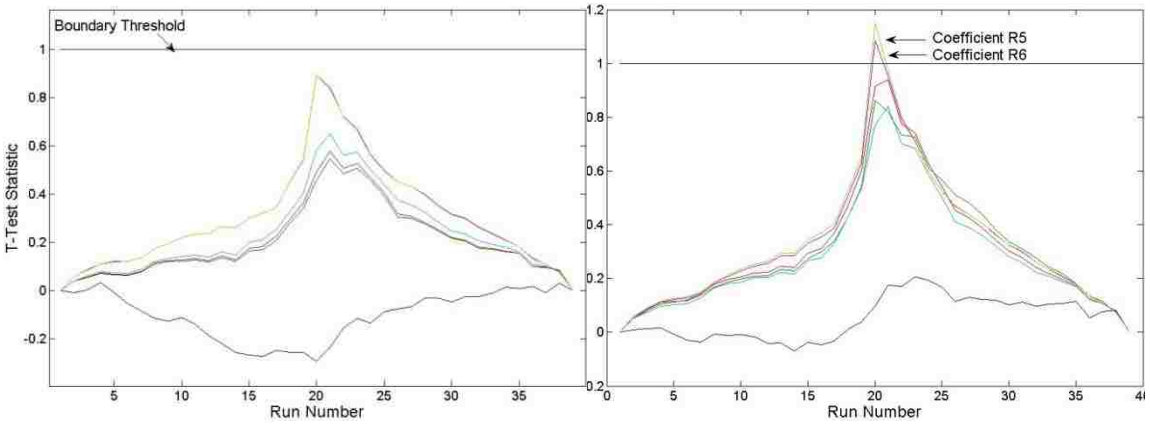
(a) (b)

Figure 5.50: NLRT results for the auto regression model without mass for the angle coefficient on the (a) left side and (b) right side of the girder



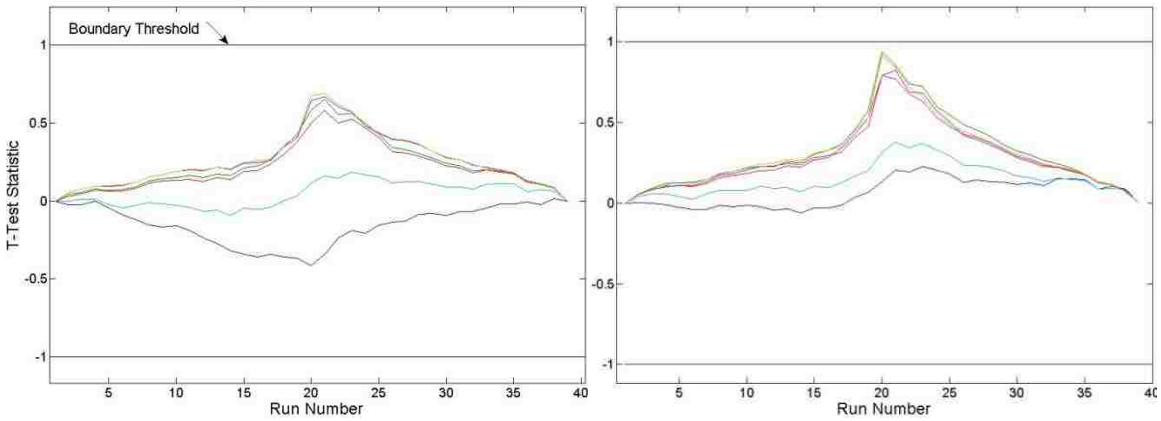
(a) (b)

Figure 5.51: NLRT results for the auto regression model with mass for the angle coefficient on the (a) left side and (b) right side of the girder



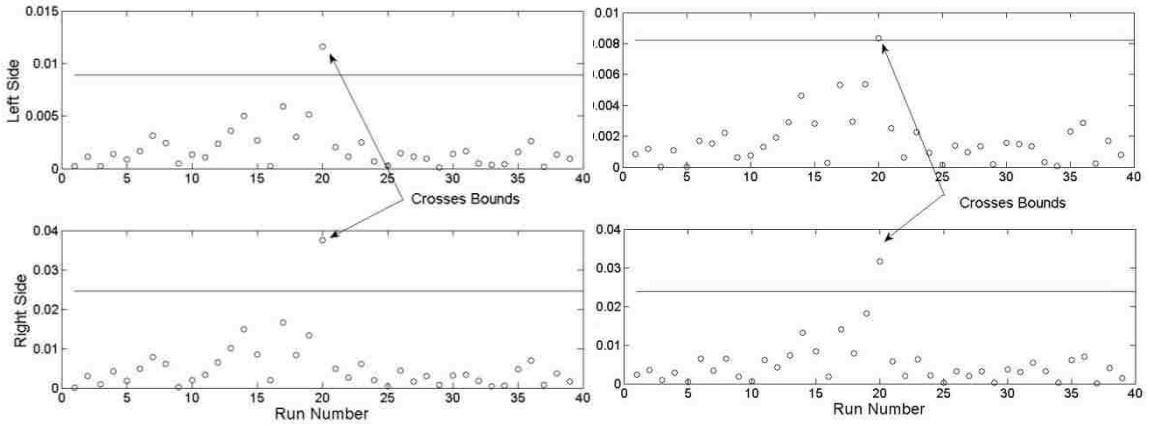
(a) (b)

Figure 5.52: Student's t-test for AR residuals using the data without mass for the (a) left and (b) right side of the girder



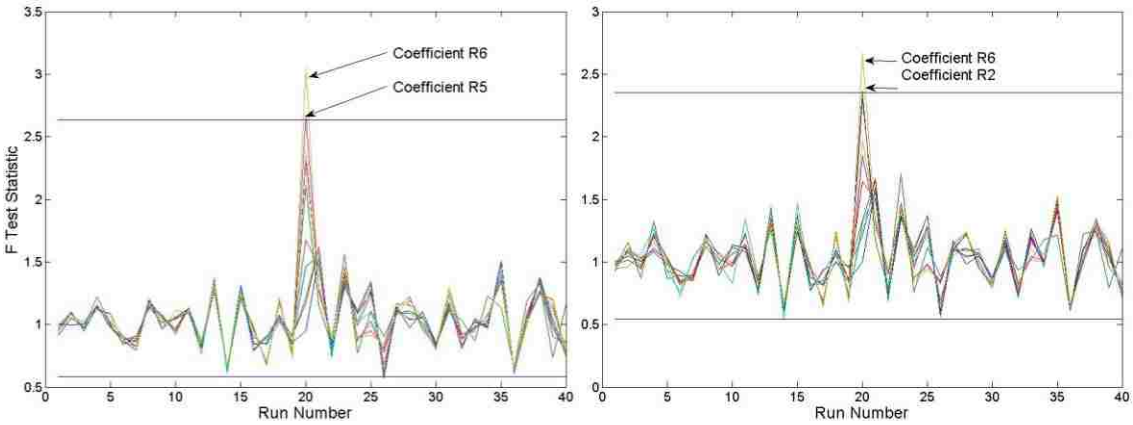
(a) (b)

Figure 5.53: Student's t-test for AR residuals using the data with mass for the (a) left and (b) right side of the girder



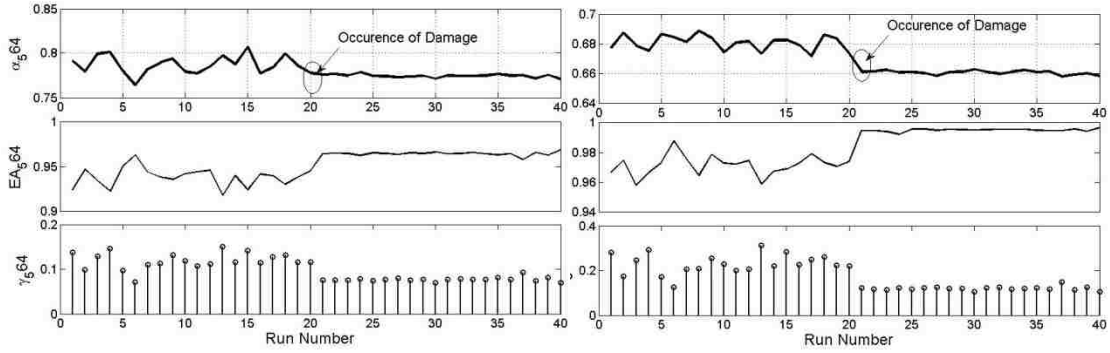
(a) (b)

Figure 5.54: Moving Range results for the auto regression residuals without mass for coefficient α_5 on the left and right side of the frame.



(a) (b)

Figure 5.55: F-test results for the auto regression residuals using the data (a) without mass and (b) with mass using Procedure 2



(a) (b)

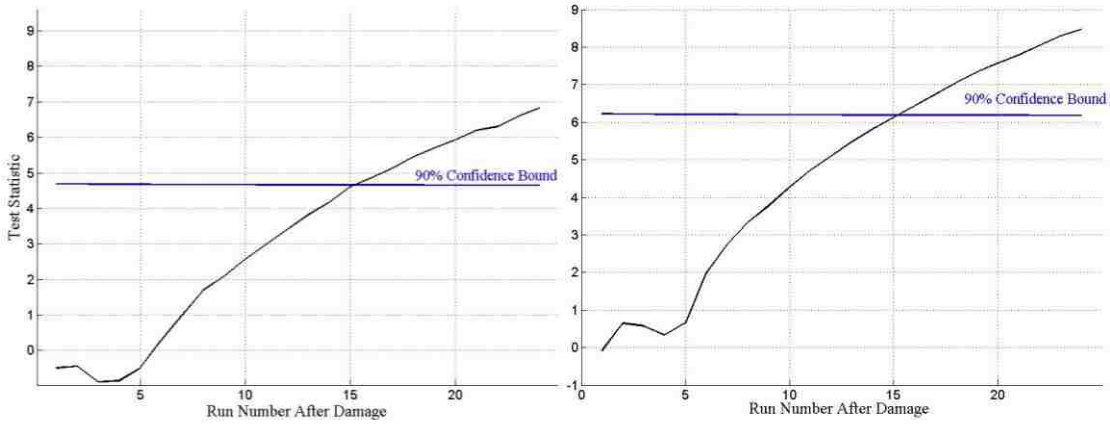
Figure 5.56: Coefficient, Accuracy and Error plot for coefficient (a) $\alpha_{5:64-L}$ and (b) $\alpha_{5:64-R}$

Table 5.1: Percent Change of Single Variate Coefficients for (1) No Noise, (2) Noise with .001 Standard Deviation, and (3) Noise with .008 Stand Deviation

(1)/Sensor Location	1	2	3	4	5	6
Left	X	0.024	0.037	0.110	0.223	0.242
Right	0.329	0.136	0.072	0.084	0.309	0.354
(2)						
Left	X	0.024	0.037	0.110	0.223	0.241
Right	0.330	0.136	0.072	0.084	0.309	0.354
(3)						
Left	X	0.036	0.036	0.105	0.217	0.229
Right	0.324	0.132	0.069	0.085	0.305	0.349

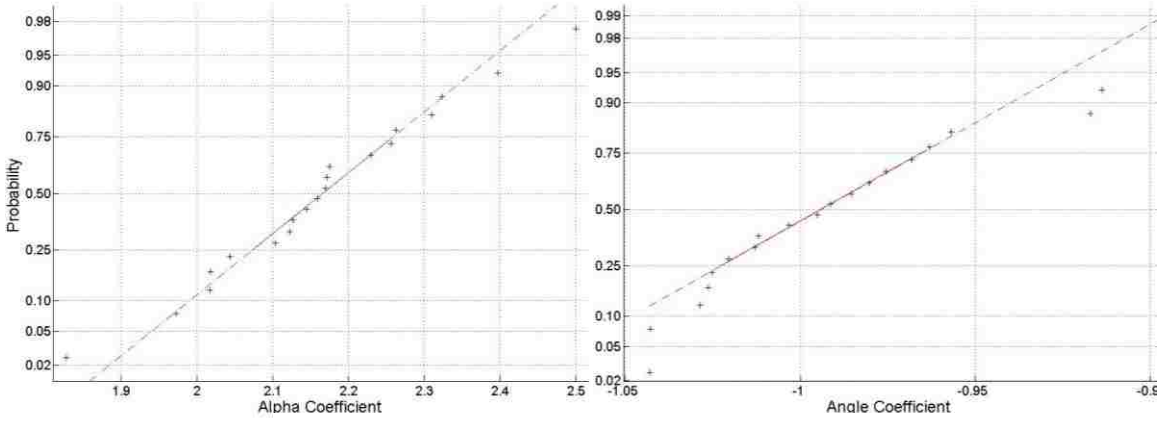
Table 5.2: Percent Change of Tri-variate Coefficients for (1) No Noise, (2) Noise with .001 Standard Deviation, and (3) Noise with .008 Stand Deviation

(1)/Sensor Location	1	2	3	4	5	6
Left	X	0.152	0.191	0.186	X	X
Right	X	0.262	0.225	0.248	X	0.227
(2)						
Left	X	.152	0.195	.232	X	X
Right	X	0.274	0.225	0.191	0.205	0.227
(3)						
Left	X	0.213	0.174	0.641	X	X
Right	X	0.168	0.148	0.227	.204	0.219



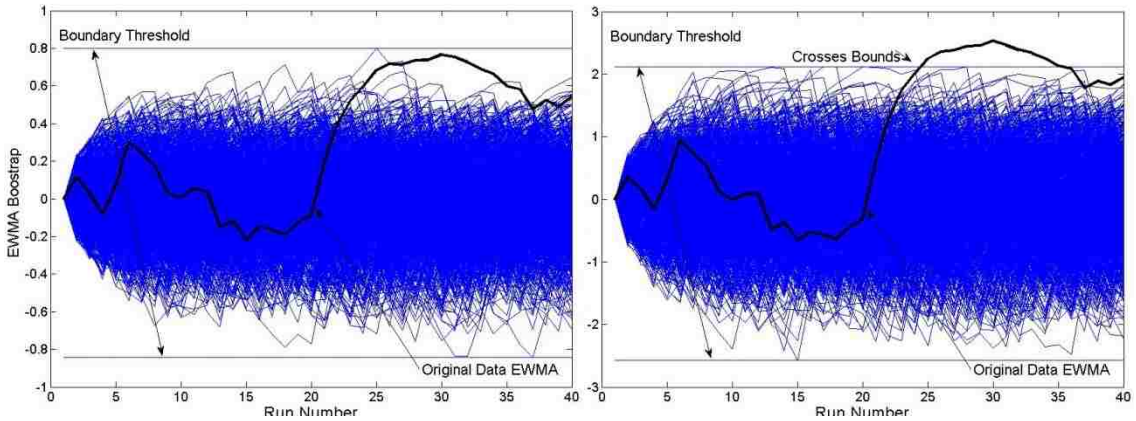
(a) (b)

Figure 5.57: Tri-Variate Regression coefficient using impact data Bayesian Hypothesis test for the coefficient (a) $\alpha_{4:65-L}$ and coefficient (b) $\alpha_{4:65-R}$



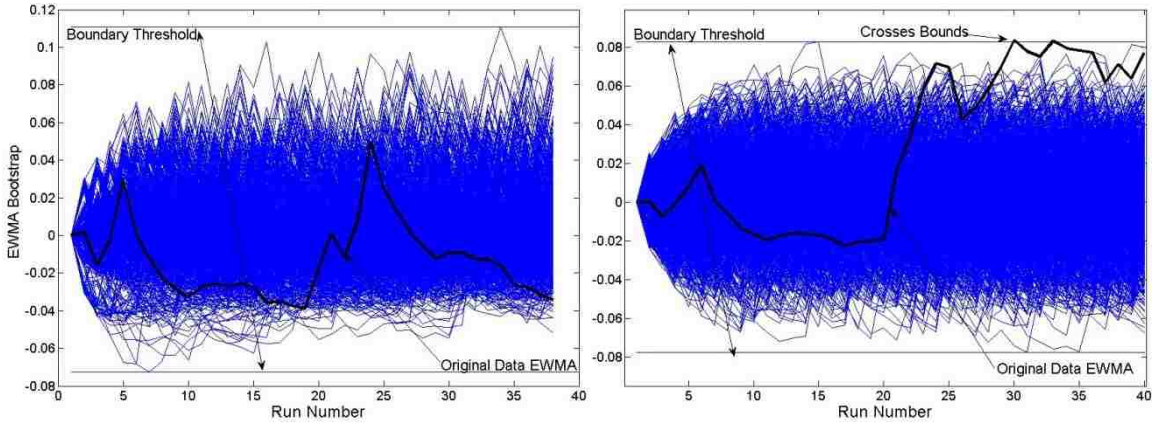
(a) (b)

Figure 5.58: Normal Probability Plot for the (a) alpha coefficient $\alpha_{4:56-R}$ and (b) angle coefficient $\Gamma_{6:54-R}$ for the collinear regression model



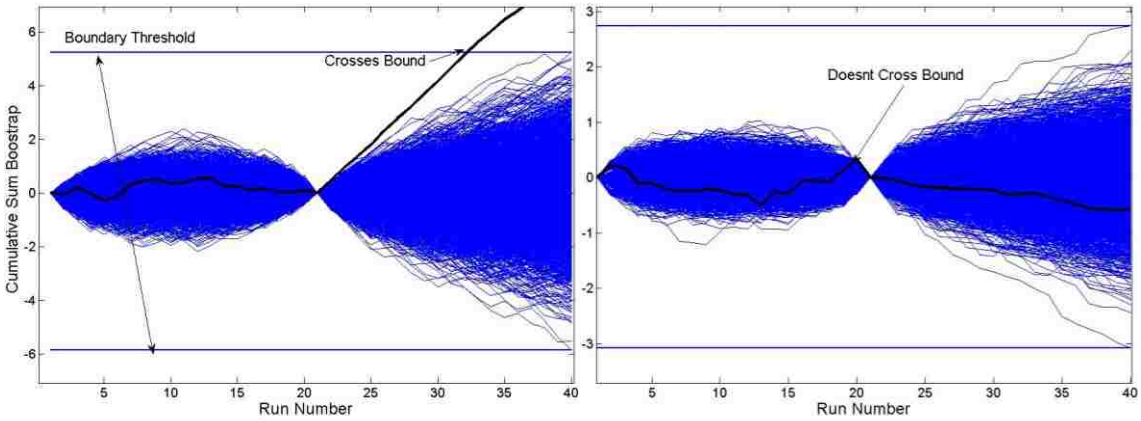
(a) (b)

Figure 5.59: Typical plots for the EWMA alpha coefficient collinear regression with bootstrapping using collinear regression for the (a) coefficient $\alpha_{6:54-L}$ and (b) coefficient $\alpha_{6:54-R}$



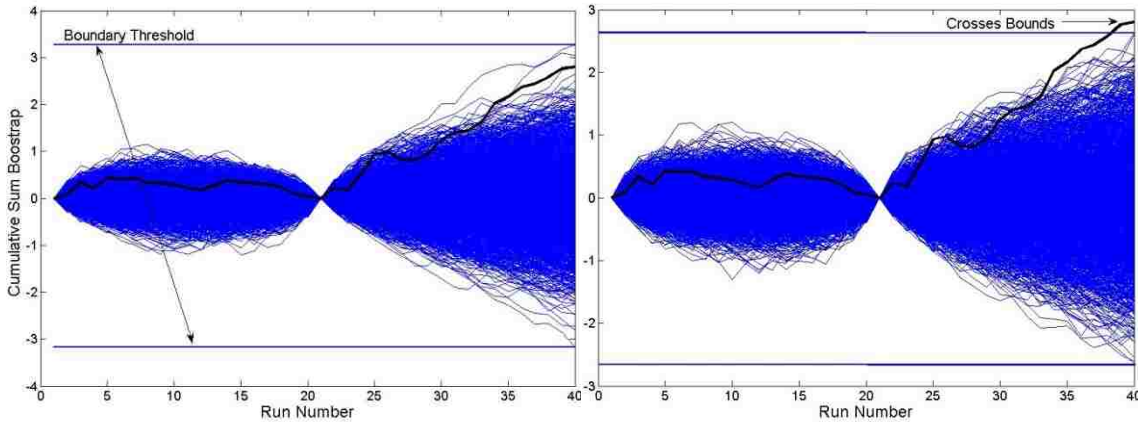
(a) (b)

Figure 5.60: Typical plots for the EWMA angle coefficient with bootstrapping using collinear regression for the (a) coefficient $\Gamma_{4:56-L}$ and (b) coefficient $\Gamma_{4:56-R}$



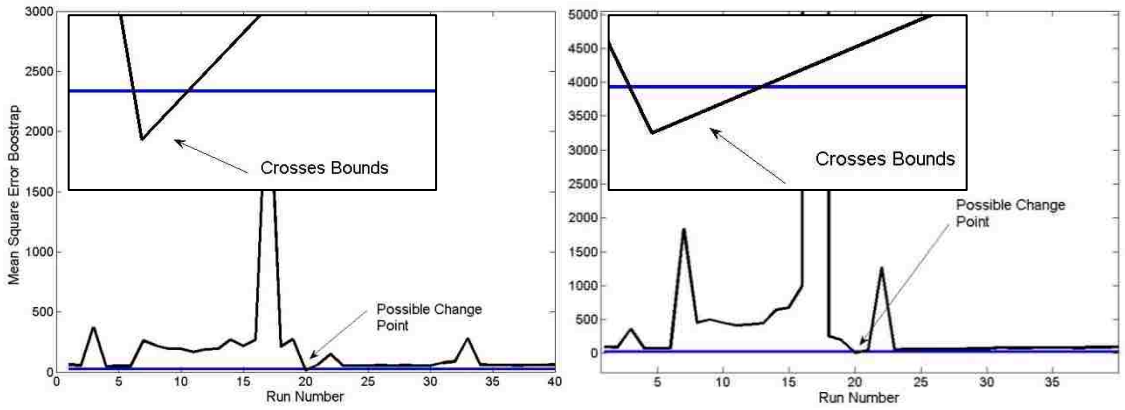
(a) (b)

Figure 5.61: Typical plots for the CUSUM alpha coefficient with bootstrapping using collinear regression for the (a) coefficient $\alpha_{4:56-L}$ and (b) coefficient $\alpha_{4:56-R}$



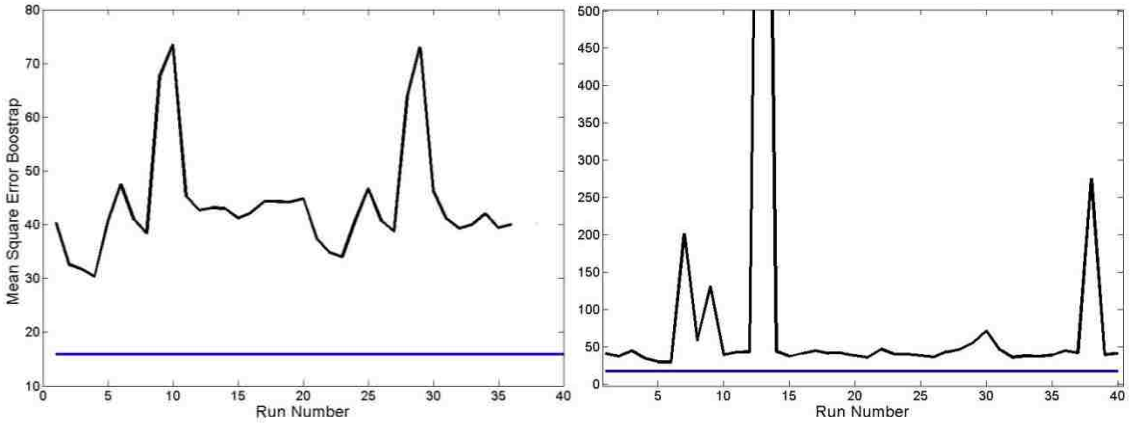
(a) (b)

Figure 5.62: Typical plots for the CUSUM angle coefficient with bootstrapping using collinear regression for the (a) coefficient $\Gamma_{4:56-L}$ and (b) coefficient $\Gamma_{4:56-R}$



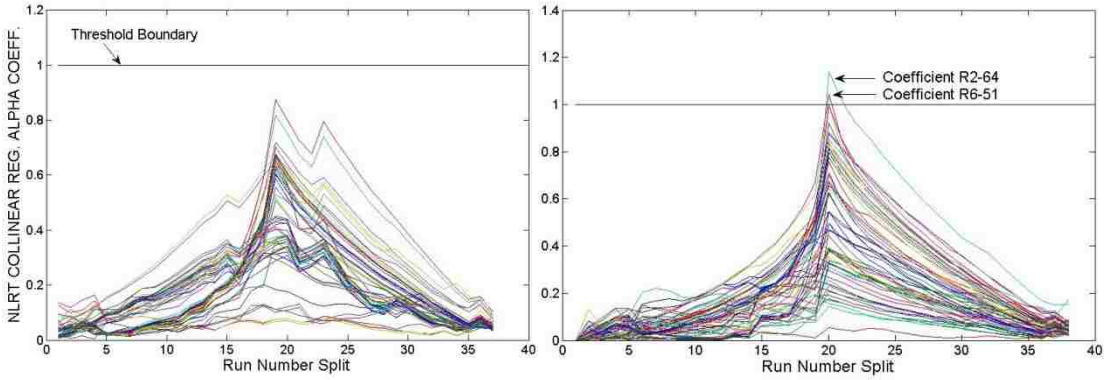
(a) (b)

Figure 5.63: Typical ModMSE plots for the alpha (a) coefficient $\alpha_{4:56-L}$ and (b) coefficient $\alpha_{4:56-R}$ using collinear regression



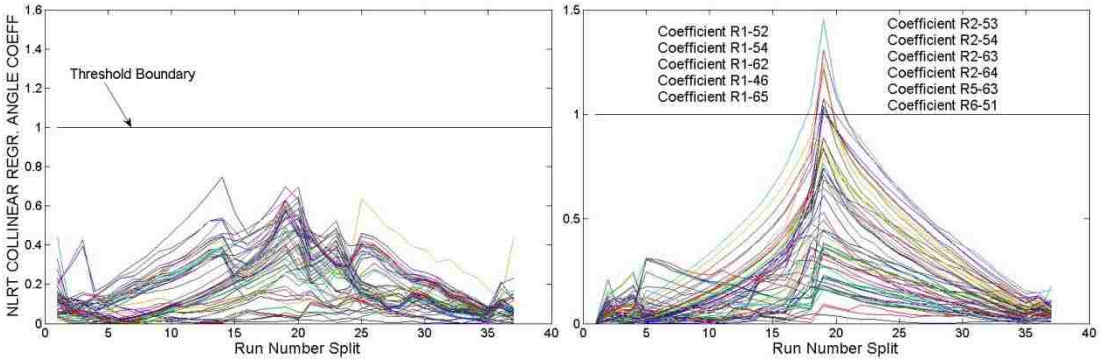
(a) (b)

Figure 5.64: Typical ModMSE plots for the angle (a) coefficient $\Gamma_{4:56-L}$ and (b) coefficient $\Gamma_{4:56-R}$ using collinear regression



(a) (b)

Figure 5.65: The results for the collinear regression using the alpha coefficient in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder.



(a) (b)

Figure 5.66: The results for the collinear regression using the angle coefficient in the NLRT statistic for the (a) left side of the girder and (b) right side of the girder.

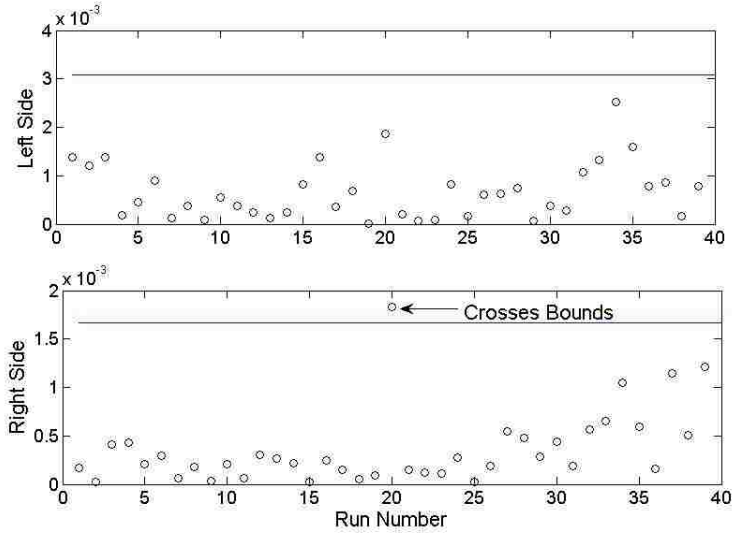


Figure 5.67: Moving Range for the collinear regression residuals for $\alpha_{5:46-R}$

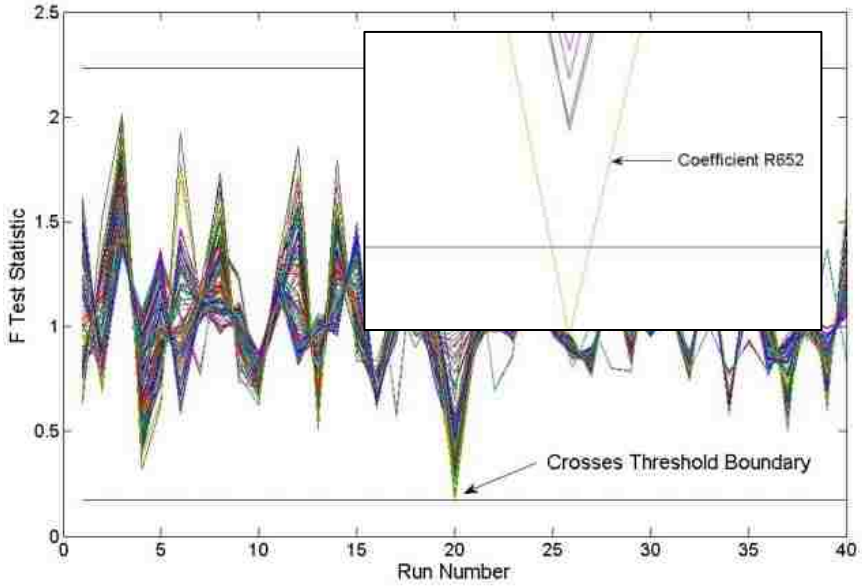


Figure 5.68: F-Test results for the collinear regression residuals using Procedure 2

6. Summary, Conclusions and Future Work

This thesis presents multiple damage features, control charts, and threshold generating methods that can effectively be used to detect the timing and location of damage. The algorithm introduced is an output only damage detection method that does not require any previous knowledge of the structure or damaged location. Furthermore, a modified Mean Square Error statistic is introduced which eliminates the susceptibility of the statistic to the variance within a data set. The enhanced algorithm, ModMSE, is introduced and its application and effectiveness is demonstrated.

The damage features used in this thesis are linear regression parameters. These indicators change with a change in the physical properties of a structure and can therefore be used to detect a change in a structure due to a damaging event. They include the slope of the linearly regressed line, the angle between two linear regressed lines, and the regression residuals. The slope and angle are subject to a change in their mean in the occurrence of damage and the regression residuals are subject to a change in their variance. Four different linear regression models are used which include the Single Variate, Auto Regressive with Exogenous term, Auto Regressive, and Collinear. Additionally, the effect of mass is investigated in this thesis. Univariate methods do not lend themselves to incorporate dynamic effects and therefore these effects are only considered in multivariate regression models: the AR and ARX linear regression techniques. Mass is added to the system in order to analyze its effect on these linear regression parameters.

Because the damage features are subject to different parameter changes (i.e. mean or variance) in the event of damage, various control charts are used with each damage

feature that are sensitive to the appropriate change. Additionally, different methods lend themselves to either single variate or multivariate parameters (i.e. scalar or matrix damage features). It is shown in this thesis that the single variate linear regression parameters (whether using the alpha, angle or residuals as damage features) are the most consistently correct damage features. The control charts used with these parameters consistently indicate the correct timing and location of damage. Generally, the ARX regression parameters localize damage to the correct local joint using the data without mass but cannot localize it further to the correct element in the structure. Similarly, the collinear regression parameters detect damage on the correct local joint of the frame. Most of the control charts detect damage in the correct vicinity of damage. However, the results for the Auto Regression are more inconsistent. This may be a result of the non-normality of the damage features themselves as shown in the normal probability plots for these features. Therefore, it may not capture the changing properties due to damage. Additionally, the superimposed mass has a significant effect on the results of the change point analysis. Many control charting methods, such as the Normalized Likelihood Ratio test, were not successful in indicating the correct location of damage using this data set. The vibration data more accurately depict the behavior of the structure when mass is not added to the frame.

Bootstrapping as well as simple Shewhart thresholds are all viable options for creating confidence bounds for control charts. These are all subject to certain confidence levels. Because each bootstrap is different, the results for the bootstrapping will be altered every time; additionally, they can be customized for user defined confidence levels which makes them suspect to inconsistencies. Unless enough bootstraps are used,

damage detection results will differ. On the other hand, confidence levels are pre-chosen for the Shewhart threshold. Therefore these methods may show more consistent results that can be repeatedly reproduced. However, it is also shown, in the case of the EWMA statistic that sometimes the simple Shewart thresholds do not lend themselves to all types of data sets. This threshold indicated a fully healthy frame. Similarly, distribution thresholds can also be valid methods for creating a threshold. These also use predetermined confidence levels. It is recommended that for civil engineering damage detection purposes, in using univariate and/or multivariate control charts a "shotgun approach" be utilized in which multiple charts be used on the same data set to coordinate damage localization.

However, more work may be done in this area. Other damage cases may be investigated to see how well the algorithm can detect and localize damage. Additionally, this thesis tries to evaluate the most well-known and trusted statistics in order to compare their results. Hence, other control charts may be analyzed as well. There are many different types of statistical process controls that are used in other disciplines. An example is the less common Grey relational coefficient. This statistic can be used to show a relationship among a series of data in order to detect a shift or damaged location in concrete crack patterns. Additionally, more work is needed to enhance the performance of the F-test; a more refined threshold boundary should be used in order to test for damage. Furthermore, the location and number of sensors can have an effect on the accuracy and resolution of damage detection. This investigation should be included in future work.

Nichols et al. (2003) define a successful scheme as one that "will both identify the damage and quantify the statistical significance of the result." It is shown in this thesis that IDDA, when combined with the control charts, is an efficient damage detection scheme. The implementation of change point detection methods with IDDA features that was presented in this thesid accurately locates the damage using statistically generated threshold values to determine the significance of change. A similar approach for other data-driven damage detection methods is recommended.

References

- Alvandi, A., & Cremona, C. (2006), Assessment of vibration-based damage. *Journal of Sound and Vibration* **292**, 179–202.
- Amiri, A. & Allahyari, S. *(*2011), Change Point Estimation Methods for Control Chart Post signal Diagnostics: A Literature Review *Quality and Reliability Engineering International. Wiley Library* Vol**28** Issue 7
- Bozdogan, H. (1987) Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions, The Psyometric Society Vol **52** No. 3 345-370
- Chang, M. (2010). Damage detection and modal identification of structural systems using sensor data.‖ MS thesis, Lehigh University, Bethlehem, PA.
- Chang, T.C., Gan, F.F. (2004) Shewhart Charts for Monitoring the Varaince Components *Journal of Quality Technology* **36** (3), 293-308
- Cruz, P.J.S. and Salgado, R. (2009), "Performance of vibration-based Damage Detection Methods in Bridges," *Computer-Aided Civil and Infrastructure Engineering*, 24:1, pp. 62-79.
- Del Rossi, J.A. (1975) An Application of Welch's Approximate t-Solution of the Behrens-Fisher Problem to Confidence Intervals, *Technometric* **17** (1) 57-60
- Diaz, Jr., C.J., Johnson, W.D. (1998) An F-Test for Multivariate Repeated Measures Data with the Wiener Stochastic Process Pattern in the Covariance Matrix, *Communications in Statistics Theory and Methods* **27** (2) 275-289
- Doebling, S.W., Farrar, C.R., & Prime, M.B. (1998) *A Summary Review of Vibration-Based Damage Identification Methods* The Shock and Vibration Digest, **30** 91-105
- Dorvash, S. Pakzad, S.N., and Labuz, E. (Submitted September 2012). "Statistics Based Localized Damage Detection using Vibration Response", *(IOP Publishing) Smart Structures and Systems, An International Journal,* (Under review)
- Dorvash, S., Pakzad, S.N., Labuz, E., Chang, M., Lib, X., & Chengb, L. (2010), Damage Detection and Localization in Structures: A Statistics Based Algorithm Using a Densely Clustered Sensor Network, *Proc. of 20th Annual Structures Congress* 53-64
- Dorvash, S., Pakzad, S.N., Labuz, E., Chang, M., Lib, X., Chengb, L (2010), Validation of a wireless sensor network using local damage detection algorithm for beam-column connections, *Proc. of SPIE* Vol **7674** 76719-1
- Dorvash, S., Pakzad, S.N., Labuz, E., Ricles, J.M., & Hodgson, I.C. (Submitted March 2013), Localized Damage Detection Algorithm and Implementation on a Large-Scale Steel Beam-to-Column Moment Connection, (under review *(EERI) Earthquake Spectra)*
- Ewins, D. J. (1984), *Modal Testing: Theory and Practice*, John Wiley and Sons, New York.
- Farrar, C. R., Baker, W. E., Bell, T. M., Cone, K. M., Darling, T. W., Duffey, T. A., Eklund, A., & Migliori, A., (1994), Dynamic Characterization and Damage Detection in the I-40 Bridge Over the Rio Grande, *Los Alamos National Laboratory Report* LA-12767-MS.
- Figueiredo, E., Figueiras, J., Park, G., Farrar, C.R., Worden, K. (2011), Influence

of the Autoregressive Model Order on Damage Detection, *Computer-Aided Civil and Infrastructure Engineering* ***26*** 225-238
- Friedlander, B., Porat, B. (1984) Modified Yule-Walker method of ARMA Spectral Estimation. *IEEE Transactions on Aerospace and Electronic Systems*, AES-20 (2), 158-173
- Fugate, M.L., Sohn, H., Farrar, C.R. (2001), Vibration-based Damage detection Using Statistical Process Control, *Mechanical Systems and Signal Processing* **15**, 707-721
- Gul M.,& Catbas, N.F. (2009), Statistical recognition for Structural Health Monitoring using Time Series Modeling: Theory and Experimental Verifications, *Mechanical Systems and Signal Processing* **23** 2192-2204
- Hawkins, D. M. (1981), A Cusum for a Scale Parameter*, Journal of Quality Technology,* **13**, 228-231.
- Hawkins, D.M., Zamba, K.D. Zamba (2005) Statistical Process Control for Shifts in Mean or Variance Using a Changepoint Formulation, *Technometrics*, 47:2, 164-173
- He, X., De Roeck, G., (1997) System identification of mechanical structures by a high-order multivariate autoregressive model, *Computers and Structures* **64** (1–4), 341–351.
- Hotelling, H. (1947), Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights, *C.Eisenhart et. al*.111-184
- Hu, X.Y., Wang, B., and Ji, H. (2013), "A Wireless Sensor Network-based Structural Health Monitoring System for Highway Bridges," *Computer-Aided Civil and Infrastructure Engineering*, 28:3, pp. 193-209.
- Huang, C.S. (2001) Structural identification from ambient vibration measurement using the multivariate AR model, *Journal of Sound and Vibration* **241** (3), 337–359.
- Hung, C.F., Ko, W.J., Peng, Y.T. (2004)Identification of modal parameters from measured input and output data using a vector backward auto-regressive with exogeneous model, Journal of Sound and Vibration **276** (3–5), 1043–1063.
- Kabaila, P. (2004) Assessment of a Preliminary F-Test Solution to the Behren-Fisher Problem *Communications in Statistics- Theory and Methods* **34,** 291-302
- Labuz, E., Chang, M., & Pakzad, S. N. (2010), Local Damage Detection in Beam-Column Connections Using a Dense Sensor Network, *Proc. Of Structures Congress* 3143-3154
- Labuz, E., Pakzad, S.N., Wurst, D. (2011), Localized damage detection in a large-scale moment connection using a strain gauge sensor network, *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring* 79813Q-79813Q-11
- Lai, W.C., Chang, T.P., Wang, J.J., Kan, C.W., & Chen, W.W. (2012) An evaluation of Mahalanobis Distance and Grey Relational Analysis for Crack Pattern in Concrete Structures*, Computational Materials Science* **65** 115-121
- Liang, J., Tang, M. (2009) Generalized F-tests For the Multivariate Normal Mean *Computational Statistics and Data Analysis* **53** 1177-1190
- Limentani, G.B., Ringo, M.C., Ye, F., Bergquist, M.L., McSorley, E.O. (2005) Beyond the t-test: Statistical Equivalence Testing, *Analytical Chemistry* 221-226

- Lucas, J.M., and Saccucci, M.S. (1990), Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, *American Statistical Association and the American Society for Quality, Technometrics* February **32** (1)
- Macgregor, J.F., & Harris, T.J. (1990), Discussion of 'EWMA Control Schemes: Properties and Enhancement' by Lucas and Saccucci, *American Statistical Association and the American Society for Quality, Technometrics* Vol **32** No.1 1-29
- Mongomery, R.H., Loftis, J.C. (1987) Applicability of the t-test for Detecting trends in Water Quality Variables, *American Water Resources Association* **23** (4) 653-662
- Mosavi, A.A., Dickey, D., Seracino, R., & Rizkalla, S., (2012), Identifying Damage Locations under Ambient Vibrations Utilizing Vector Autoregressive Models and Mahalanobis Distances. *Mechanical Systems and Signal Processing* **26** 254-267
- Nicols, J.M., Trickey, S.T., Todd, M.D., Virgin, L.N. (2003) Structural Health Monitoring Through Chaotic Interrogation, *Kluwer Academic Publisher*, **38** 239-250
- Nishikawa, T., Fujino, Y., Yoshida, J., and Sugiyama, T. (2012), "Concrete Crack Detection by Multiple Sequential Image Filtering," *Computer-Aided Civil and Infrastructure Engineering*, 27:1, pp. 29-47.
- Pakzad, S.N. (2008) Statistical approach to structural monitoring using scalable wireless sensor networks. PhD dissertation, University of California, Berkeley.
- Pakzad, S.N. (2010) Development and deployment of large scale wireless sensor network on a long-span bridge, *Smart Structures and Systems* Vol **6** Issue 5-6 525-543
- Pandey A. K. and Biswas, M. (1995), Damage Diagnosis of Truss Structures by Estimation of Flexibility Change, *Modal Analysis – The International Journal of Analytical and Experimental Modal Analysis*, **10**, (2), 104-117.
- Pelayo, A.M., Evangelista, F.S. (2003) A Statistical F test for the Natural Attenuation of Contaminants in Groundwater *Environmental Monitoring and Assessment,* **83** (1) 47-70
- Pompe, P.P.M., and Feelders, A.J. (1997), Using Machine Learning, Neural Networks, and Statistics to Predict Corporate Bankruptcy, *Computer-Aided Civil and Infrastructure Engineering **26*** 267-276
- Qiao, L., Esmaeily, A. and Melhem, G.H. (2012) Signal Pattern Recognition for Damage Diagnosis in Structures, *Computer-Aided Civil and Infrastructure Engineering 27* 699-710
- Roberts, S. W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *American Statistical Association and the American Society for Quality, Technometrics*, **1**, 239-250.
- Salawu, O. S. (1997), Detection of Structural Damage Through Changes in Frequency: A Review, *Engineering Structures*, **19** (9), 718-723.
- Sohn H., Czarnecki JA, & Farrar CR. (2000), Structural health monitoring using statistical process. *Journal of Structural Engineering* 1356-1363

- Sohn, H., & Farrar, C.R. (2000), Statistical Process control and Projection Techniques for Structural Health Monitoring, *European COST F3 Conferences on System Identification & Structural Health Monitoring*, 105-114
- Srivastava, M.S., & Worsley, K.J. (1986) Likelihood Ratio Tests for a Change in the Multivariate Normal Mean *Jounral of the American Statistical Association,* **81** (393) 199-204
- Steiner, S.H. (1999) Exponentially weighted moving average Control Charts and Time-Varying Control Limits and Fast Initial Response, *Journal of Quality Technology* **31** 75-86
- Sullivan J.H., Woodall, W.H. (1999) Change-point Dections of Mean Vector or Covariance Matrix Shifts Using Multivariate Individual Observations, *IIE Transactions* **32**, 537-549
- Sullivan, J.H. &Woodall, W.H. (1996), A Control Chart for a Preliminary Analysis of Individual Observations, *Journal of Quality Technology* Vol **28** No.3
- Tague, N.R. (2004) The Quality Toolbox, Second Edition ASQ Quality Press 155-158
- Talebinejad, I., Fischer, C., & Ansari, F. (2011) Numerical Evaluation of Vibration-Based Methods for Damage Assessment of Cable-Stayed Bridges *Computer-Aided Civil and Infrastructure Engineering* **26** 239–251
- Taylor, W.A. (2011), Change-Point Analysis: A powerful new Tool For Detecting Changes, www.variation.com/cpa/tech/changepoint.html
- Umesha, P.K., Ravichandran, R., Sivasubramanian K. (2009) Crack Detection and Quantification in Beams Using Wavelets, *Computer-Aided Civil and Infrastructure Engineering,* **24** (8), 593–607,
- Verdier, G., & Ferreira, A. (2011) Adaptive Mahalanobis Distance and k-Nearest Neighbor Rule for Fault Detection in Semiconductor Manufacturing, *IEEE Transactions on Semiconductor Manufacturing,* **24** (1) February
- Wang, Z., & Ong, K.C.G. (2008) Autoregressive Coefficients based on Hotelling's T^2 Control Chart for Structural Health Monitoring, *Computers and Structures* **86** 1918-1935
- West, W. M. (1984), Illustration Of The Use Of Modal Assurance Criterion To Detect Structural Changes In An Orbiter Test Specimen, in Proc. Air Force Conference on Aircraft Structural Integrity, 1–6
- Wilcox, M. (2003) The Philosophy of Shewhart's Theory of Prediction. Proceedings of the 9th Research Seminar: Deming Scholar's Program
- Williamson, P.P. (2009) Bayesian Alternatives to the F-test for Two Population Variances, *Communications in Statistics-Theory and Methods* 38:14 2288-2301
- Worden, K., Manson, G., & Fieller, N.R.J. (2000) Damage Detection using Outlier Analysis, *Journal of Sound and Vibration* **229** (3), 647-667
- Yao, R. & Pakzad, S.N. (2013) Performance comparison of different autoregressive damage features using acceleration measurements from a truss bridge, *International Workshop on Structural Health Monitoring*
- Yao, R. & Pakzad, S.N. (2010) , Autoregressive Statisical Pattern Recognition Algorithms for Damage Detection in Civil Structures*, Mechanical Systems and Signal Processing* **31** (2012) 355-368

- Yao, R. & Pakzad, S.N. (2012), Autoregressive statistical pattern recognition algorithms for damage detection in civil structures, Mechanical Systems and Signal Processing, Vol **31** 355-368
- Yao, R. & Pakzad, S.N. (2013), Time and frequency domain regression-based stiffness estimation and damage identification, *Structural Control and Health Monitoring d*oi: 10.1002/stc.1570
- Yin, T., Lam, H.F., & Chow, H.M. (2012) Bayesian Probabilistic Approach for Crack Characterization in Plate Structures *Computer-Aided Civil and Infrastructure Engineering* **25** 375–386
- Zamba, K.D. & Hawkins, D. M. (2006) , A Multivariate Change-Point Model for Statistical Process Control, *American Statistical Association and the American Society for Quality, Technometrics,* November **48** (4)
- Zhang, J., Zou, C., & Wang, Z. (2010) A Control Chart Based on Likelihood Ratio Test for Monitoring Process Mean and Variability, *Quality and Reliability Engineering International* Vol **26** 63-73
- Zhou, Q., Luo, Y., Wang, Z. (2010) A control Chart Based on Likelihood Ratio Test for Detecting Patterned Mean and Variance Shifts, *Computational Statistics and Data Analysis* **54** 1634-1645
- Zou, C., Zhang, Y., Wang, Z. (2006) A Control Chart Based on a Change-point Model Monitoring Linear Profiles *IIE Transactions* 38:12, 1093-1103

Vita

Mallory Beth Nigro, daughter of Rose and Scott Nigro, was born on November 23rd 1990, in Long Beach, NY. In May 2012, Mallory earned a Bachelor of Science in Civil Engineering from the Lehigh University in Bethlehem, PA. She began her undergraduate research studies under Shamim N. Pakzad her junior year. Once she graduated, Mallory then continued her graduate studies in the Department of Civil and Environmental Engineering at Lehigh University in Bethlehem, PA in July 2012. She will receive her Master of Science in Structural Engineering in September of 2013.