LA-UR-16-25378

Title: DARHT Multi-intelligence Seismic and Acoustic Data Analysis

Author(s):   Stevens, Garrison Nicole
              Van Buren, Kendra Lu
              Hemez, Francois M.

Intended for: Internal Report

Issued: 2016-07-21

DARHT Multi-intelligence Seismic and Acoustic Data Analysis

Garrison N. Stevens, Kendra L. Van Buren, François M. Hemez

Summary: The purpose of this report is to document the analysis of seismic and acoustic data collected at the Dual-Axis Radiographic Hydrodynamic Test (DARHT) facility at Los Alamos National Laboratory for robust, multi-intelligence decision making. The data utilized herein is obtained from two tri-axial seismic sensors and three acoustic sensors, resulting in a total of nine data channels. The goal of this analysis is to develop a generalized, automated framework to determine internal operations at DARHT using informative features extracted from measurements collected external of the facility. Our framework involves four components: (1) feature extraction, (2) data fusion, (3) classification, and finally (4) robustness analysis. Two approaches are taken for extracting features from the data. The first of these, generic feature extraction, involves extraction of statistical features from the nine data channels. The second approach, event detection, identifies specific events relevant to traffic entering and leaving the facility as well as explosive activities at DARHT and nearby explosive testing sites. Event detection is completed using a two stage method, first utilizing signatures in the frequency domain to identify outliers and second extracting short duration events of interest among these outliers by evaluating residuals of an autoregressive exogenous time series model. Features extracted from each data set are then fused to perform analysis with a multi-intelligence paradigm, where information from multiple data sets are combined to generate more information than available through analysis of each independently. The fused feature set is used to train a statistical classifier and predict the state of operations to inform a decision maker. We demonstrate this classification using both generic statistical features and event detection and provide a comparison of the two methods. Finally, the concept of decision robustness is presented through a preliminary analysis where uncertainty is added to the system through noise in the measurements.

1. Introduction

The Dual-Axis Radiographic Hydrodynamic Test (DARHT) facility at Los Alamos National Laboratory (LANL) executes hydrodynamic tests that produce multi-degree of freedom images of implosion tests using x-ray beams generated by two linear accelerators on perpendicular axes (Scapetti 2008). As a part of the LANL multi-intelligence grand challenge problem, acoustic, network, triggered radiofrequency, seismic, and video data were collected around DARHT over the summer of 2014. The goal of this data collection effort is to fuse information from these streams through a multi-intelligence analysis for determining the state of operations at DARHT. In this context, "state of operation" refers to activities related to hydrodynamic tests at the facility.

This report details the development of a framework for performing a general automated analysis of large data sets to inform decision-makers of facility operations considering uncertainty in the measurements. The framework developed herein (Figure 1) is composed of four parts: (1) feature extraction, (2) data fusion, (3) classification and (4) robustness analysis, and is applied to classify the state of operations of DARHT using the seismic and acoustic data collects. The analysis works by taking data at known operation states to train a statistical classifier, which then produces a nominal prediction of the state of DARHT. Uncertainties in the data are then propagated at varying levels to assess how this nominal decision changes.
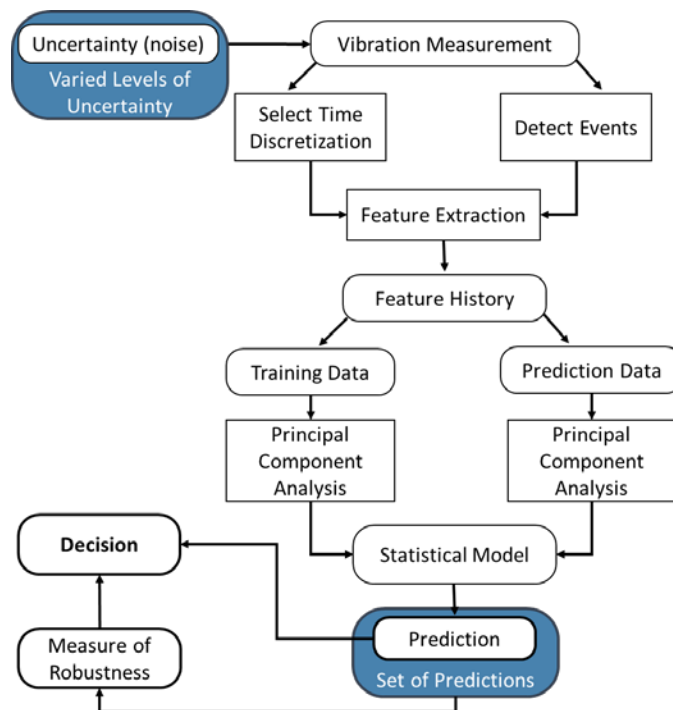

Figure 1. Framework for decision making through analysis of seismic and acoustic measurements.

The first component of this framework is ***feature extraction***. Feature extraction applies the concepts of structural health monitoring, which evaluate features of measurements continuously collected from a structure in an effort to detect when changes in these features indicative of damage occur (Farrar and Worden 2007). Two types of feature extraction are evaluated herein: (1) generic statistical feature extraction and (2) specific event detection, with the goal of comparing the two approaches to evaluate the level of detail needed in feature analysis to accurately classify activity at DARHT. Once features are extracted from each of the individual data streams, ***data fusion*** is applied to combine these features into a feature history

which can provide information from all of the streams together for classification. ***Classification*** is the process by which features extracted during times with a known state of operation are used to train a statistical model. The statistical model can then be used to predict states of operation using features obtained during a different time period. The prediction of this statistical model can be provided to a decision maker who needs details on the operations of a facility. However, uncertainty is inevitable in the development of a statistical model. The data may be imperfect, for example, stronger winds on different days may affect the acoustic measurements without being indicative of change in the facility. Therefore, the performance of the statistical classifier is dependent upon the quantity and quality of measurements used for training. As such, ***robustness analysis*** is wrapped around the entire framework to assess the effects of these uncertainties on the predictions obtained by the statistical model. Therefore, a decision maker may have access not only to the nominal prediction, but also some notion of how robust that prediction is to sources of uncertainties, which can be implemented in the decision-making process. The premise for assessing robustness is simple: the more robust predictions are to sources of uncertainty in the analysis, the more reliably they can be used by the decision-maker.

The remainder of this report is organized as follows. Information regarding the data collection at DARHT, specifically the seismic and acoustic measurements, is provided in Section 2. Next, details of the feature extraction process are provided, with Section 3 focusing on extraction of generic statistical features and Section 4 presenting the method developed for extracting specific events. Section 5 presents the development of a statistical model for predicting the state of operations at DARHT, accompanied by the results of classification using the two unique feature sets. A preliminary robustness analysis, where uncertainty in the measurements is propagated through the feature extraction and classification, is presented in Section 6. Finally, conclusions and recommendations for future work are provided in Section 7.

2. Data Collection

As the focus of the LANL multi-intelligence challenge, five different types of data were collected around the DARHT facility. These data sets include two triaxial seismic sensors, three acoustic sensors, radiofrequency measurements, two video feeds, and cyber logs, as shown in Figure 2. Ultimately these thirteen data sets will be combined for determining the state of operations at DARHT (Van Buren 2016). This report will focus solely on the seismic and acoustic data sets. Three hydro-shots took place during the four months of summer 2014 in which seismic and acoustic data were collected: May 8, July 30, and August 28.
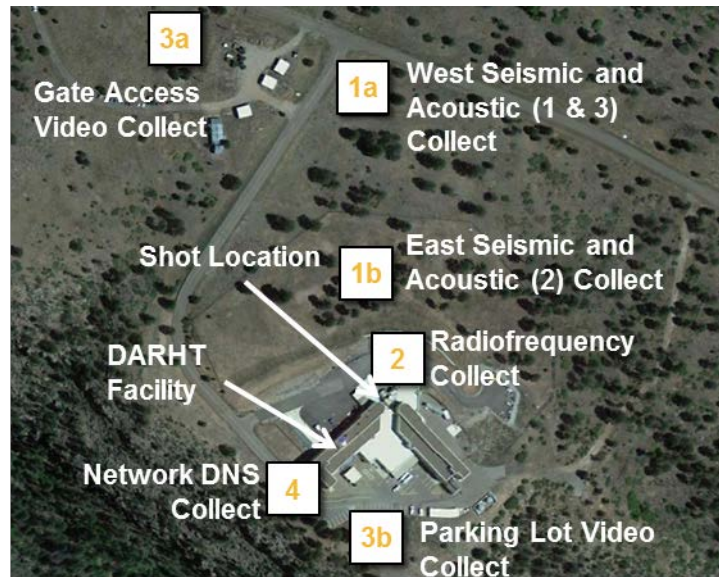

Figure 2. Locations of DARHT multi-intelligence challenge data collection.

2.1 Seismic Data

Seismic measurements are collected from two sensors located in the vicinity of DARHT. In an effort to maintain sufficient ground-to-sensor coupling, sensors are placed on the ground, buried with gravel, and covered by a protective housing. Both seismic sensors are tri-axial, measuring ground vibrations in the vertical, north, and east directions at a sampling rate of 250 samples per second. The first sensor (referred to as SPwest) is located beside the entrance road to DARHT, across from a facility housing heavy vehicles and equipment, and collects measurements from April 28 to September 14, 2014. The location of this sensor relative to the road allows for traffic patterns to be clearly picked up, as seen in a sample of the raw data stream shown in Figure 3. The east direction of this sensor encountered an issue half way through May 22, causing the signal for this direction to drop off for the remainder of the data collection period. The second sensor (referred to as SPeast) is located closer to the DARHT facility and collects data from April 28 to September 14, however, measurements are not collected during working hours until May 16. A sample week of velocity measurements from this sensor are shown in Figure 4. While this sensor also demonstrated traffic measurements, the signal strength relative to car activity is lower than that of the SPwest sensor.
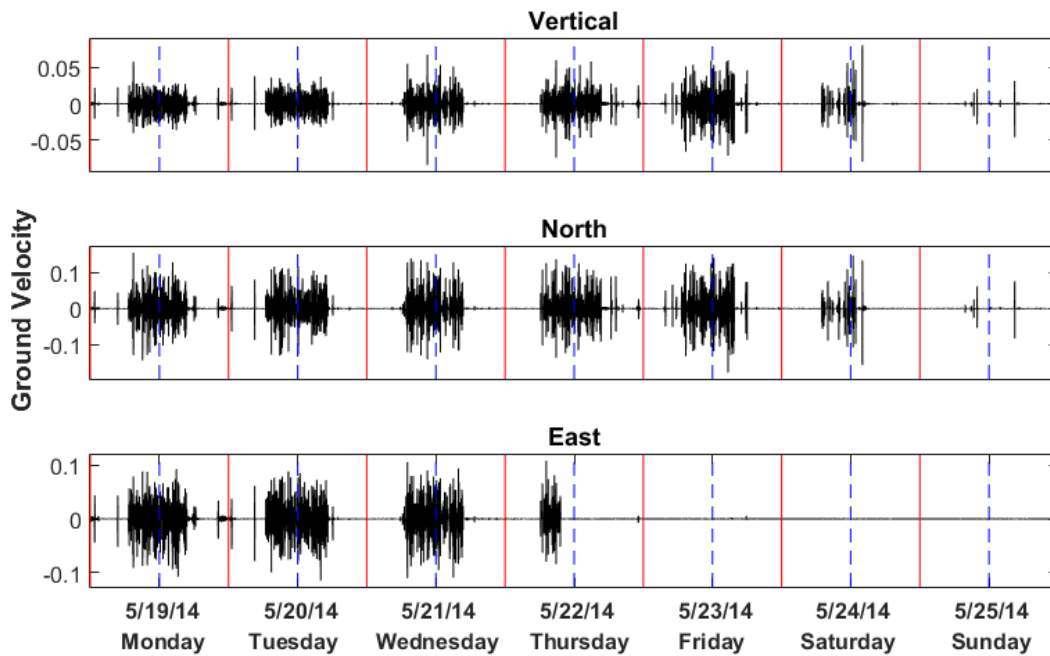
Figure 3. Raw data streams for SPwest seismic sensor located by the DARHT entrance road.
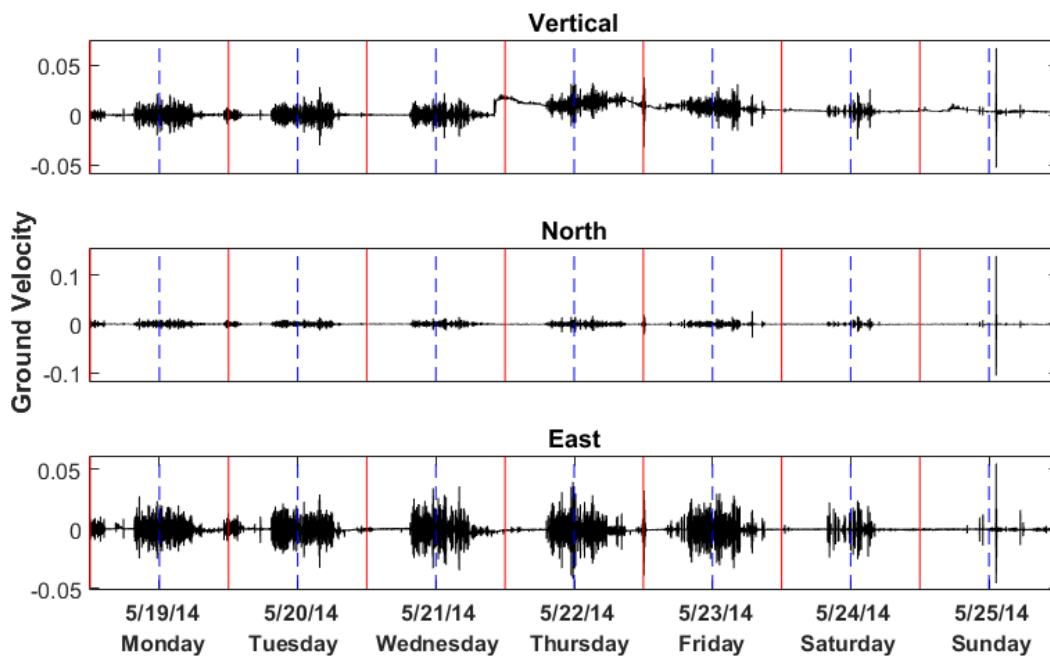

Figure 4. Raw data streams for SPeast seismic sensor located outside the DARHT fence.

2.2 Acoustic Data

Acoustic sensors are also placed at three locations around DARHT, each equipped with different filtering mechanics. Measurements are collected at a sampling rate of 250 samples per second, from April 28 to September 14. The first acoustic sensor, co-located with the SPwest seismic sensor by the entrance road, is simply furnished with a high frequency shroud. The second acoustic sensor, co-located with the SPeast seismic sensor, is connected to four garden hoses extended away from the sensor for noise reduction. The third acoustic sensor is located approximately 30 feet up the entrance road from the SPwest seismic sensor. The noise reduction set up for this sensor is more experimental than the first two. This sensor is equipped with a high frequency shroud and covered with a protective cage coated in a NASA developed material. Sample measurements for a week of data collection on all three of these sensors is shown in Figure 5. Acoustic sensors contain signatures of car activity like the seismic sensors. However, a wider range of events may be present in acoustic measurements such as wind, thunder, and distant horns from train crossings.
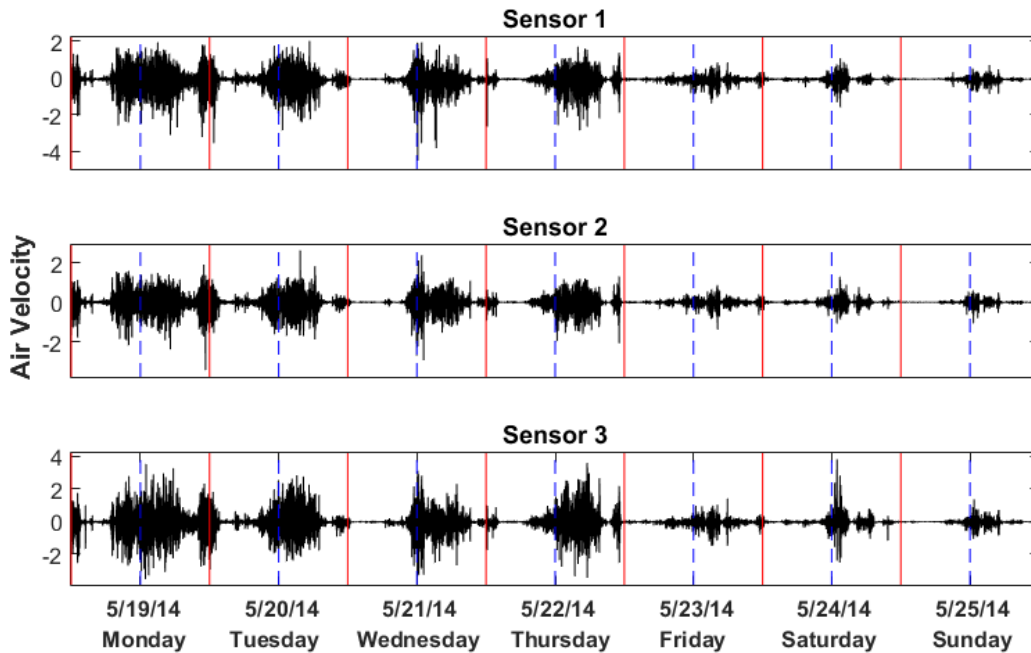

Figure 5. Raw data streams for three acoustic sensors.

3. Generic Statistical Feature Extraction

With each of the nine channels having over 3 billion samples total for the duration of the measurement collection, there is a need to reduce the information contained in these signals to a size that is computationally feasible to utilize for training a statistical model. Feature extraction presents a means for reducing the dimensionality of the data while still maintaining useful information contained in the raw measurements. For example, the seismic signals are dominated by car traffic. Therefore, simply considering the raw velocity measurements will not provide much information beyond the presence of car traffic. Extracting statistical features of these velocity measurements, however, can provide insight into the trends of traffic entering and leaving the DARHT facility, which may prove to be more informative in determining the state of operations.

Features may be extracted on any desired time scale, which should be selected based on the resolution at which the classification is to be completed. Figure 6 illustrates the reduction of data dimensionality for seismic data collected from May 19 until May 25, 2014, where the number of samples per week has been reduced from 151,200,000 samples in the raw data to 168 samples for the hourly root mean square (RMS) which indicates relative activity level, and 7 samples for the daily RMS value of the raw measurement. It is noted that visually, the traffic patterns in the raw measurements are maintained in the hourly features shown in Figure 6, as indicated by the higher RMS values during working hours and lower RMS values during night hours and the weekend. The daily features are useful for characterizing the variation in traffic patterns on different days, as demonstrated by the considerably lower values obtained for the weekend days in Figure 6.
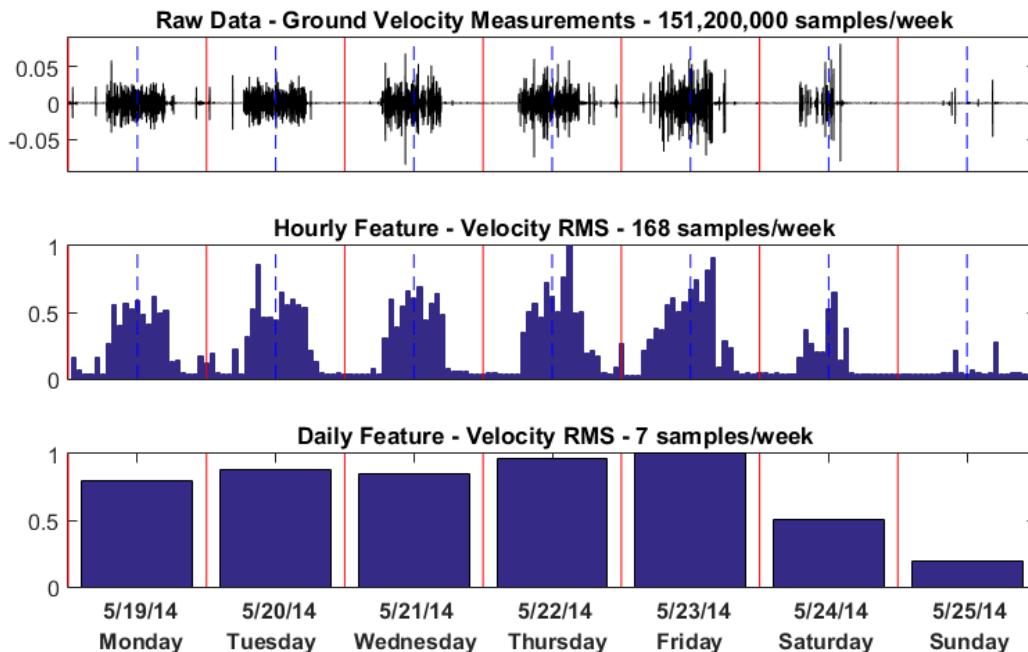

Figure 6. Measurement dimension reduction through extraction of hourly and daily features.

When an analyst knows the event or activity they are interested identifying, features with known relationships to this activity should be selected. For example, in structural health monitoring, natural frequencies and mode shapes of a structure are often used as damage indicating features, as they are likely to change as an effect of the damage (Doebling et al. 1996). The goal in this application, however, is to detect change in operations without in-depth knowledge of the activity that may be the cause of this change. Therefore, the feature extraction is kept generic with the goal of extracting information from the raw

measurements without needing inside knowledge of the activities at the facility that would cause this change. The signals are evaluated in both the time and frequency domain to gain the most knowledge possible. Sohn et al. (2001b) demonstrates the potential of statistics of vibration signals in these domains to be used as condition related features.

3.1 Time Domain Features

Signals from the nine channels are velocity measurements as a function of time. Therefore, analyzing statistics of the signal in the time domain is the natural first step. Table 1 lists the statistics that are extracted at both hourly and daily rates. While these statistics are general, each is capable of indicating a specific characteristic of the signal. For example, maximum velocity can indicate the presence of a large, abnormal event while root mean square and mean of the velocity can provide information in relation to the overall activity level during an hour or day.

Table 1. Features extracted from the raw seismic and acoustic measurements.

Feature	Daily	Hourly
Mean Velocity	X	X
Maximum Velocity	X	X
Number of Velocity Peaks	X	X
Root Mean Square Velocity	X	X
Variance of Velocity	X	X
Zero Order Temporal Moment	X	
First Order Temporal Moment	X	
Central Temporal Moment	X	

Though these statistics are useful for both hourly and daily rates, temporal moments can be particularly useful for daily rates, as these moments indicate the distribution of activity throughout a day. Definitions for these moments are provided in Eq. 1 and Eq. 2, where *TM* is the temporal moment of order *k*, determined by integrating time *t* and velocity *v(t)* at that instance in time. Therefore, these moments are able to provide information in regards to changes in activity patterns. For example, a change in temporal moments may indicate that employees are coming in earlier than usual, such as what is observed on shot day. Additionally, these moments can indicate days where access to the facility is limited and therefore, activity levels decrease during times where activity would usually be present. Figure 7 and Table 2 illustrate the ability of temporal moments to indicate a difference between three unique days, including a normal working day, a weekend day, and a shot day.

$$\text{TM}_k = \int_{-\infty}^{+\infty} t^k \big(v(t)\big)^2 dt \qquad\qquad (Eq.\ 1)$$

$$\text{Central TM} = \frac{\text{TM}_1}{\text{TM}_0} \qquad\qquad (Eq.\ 2)$$
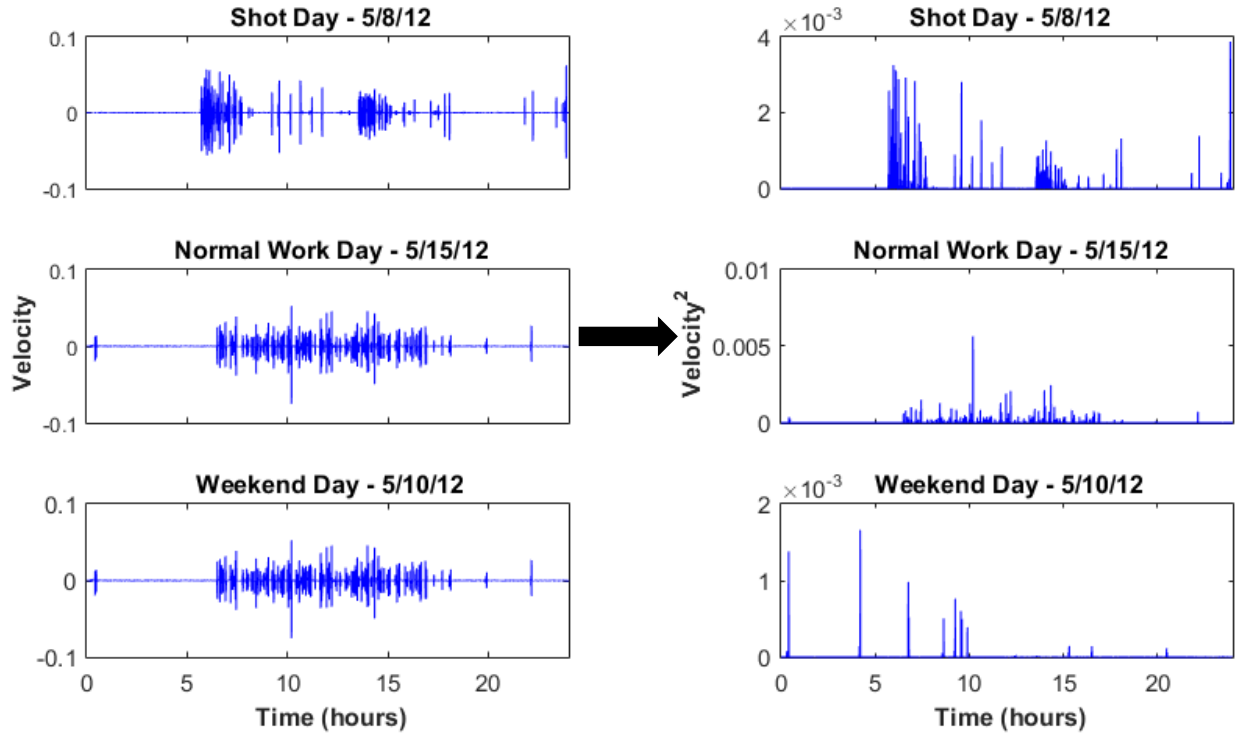
Figure 7. Change in temporal moments indicating variation in daily activity for veritcal measuremnets collected at the SPwest sensor.

Table 2. Temporal moments calculated for varying daily classes.

	TM_0	TM_1	**Central TM**
Shot Day – May 8, 2014	0.0066	251.1	37,759.5
Normal Work Day – May 15, 2014	0.0068	294.1	43,284.4
Weekend Day – May 10, 2014	0.0012	41.2	34,750.7

3.2 Frequency Domain Features

After features are extracted in the time domain, the signals are converted to the frequency domain for further analysis. Vibration signals are commonly investigated in the frequency domain as characteristics of the signal that are not apparent in the time domain are brought to light once converted to the frequency domain. For the seismic and acoustic signals, the four features listed in Table 3 are extracted. Peak frequency is the frequency in the signal with the highest concentration of energy. Spectral moments, much like temporal moments, capture the distribution of energy across the frequency spectrum. Eq. 3 defines the spectral moments, where *SM* is the spectral moment of order *k*, calculated by integrating frequency *f* and the one-sided power spectral density (PSD) *S(f)*, which captures the energy at that frequency. Figure 8 and Table 4 illustrate the shift in spectral moments during the 11:00 AM to 12:00 PM hour of a shot day, where DARHT is locked down so there is no activity, a normal working day, where broadband excitation is present from cars coming and going consistently throughout the hour, and for a weekend, where very little activity is occurring.

9

Table 3. Features extracted from frequency domain.

Feature	Daily	Hourly
Peak Frequency	X	X
Zero Order Spectral Moment	X	X
First Order Spectral Moment	X	X
Second Order Spectral Moment	X	X

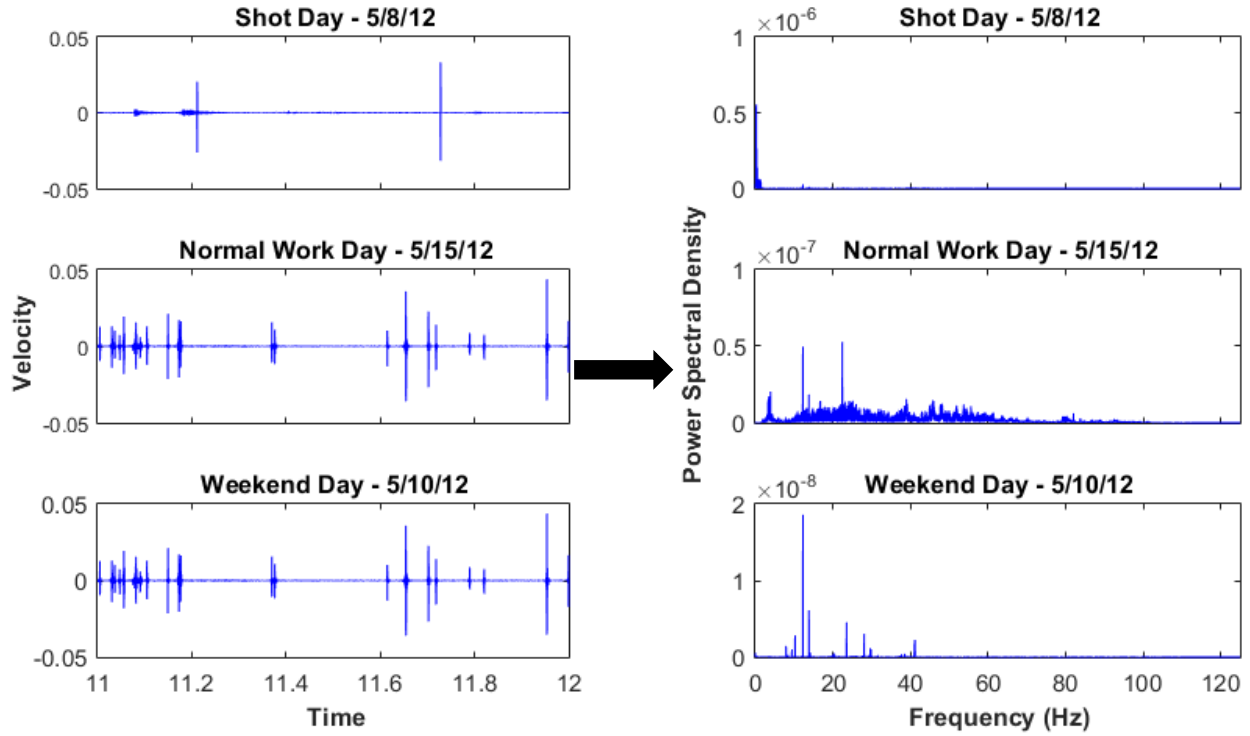$$SM_k = \int_0^\infty f^k |S(f)| df \qquad \text{(Eq. 3)}$$

Figure 8. Change in spectral moments indicating variation in hourly activity for vertical measurements collected at the SPwest sensor.

Table 4. Spectral moments calculated during 11:00 AM – 12:00 PM for varying daily classes.

|	SM_0	SM_1	SM_2
Shot Day – May 8, 2014	6.11×10^{-8}	6.67×10^{-7}	2.56×10^{-5}
Normal Work Day – May 15, 2014	1.95×10^{-7}	7.69×10^{-6}	3.97×10^{-4}
Weekend Day – May 10, 2014	1.39×10^{-9}	3.85×10^{-8}	1.62×10^{-6}

3.3 Data Fusion

Once hourly and daily features have been extracted from each data stream, these features are saved as time-stamped rate streams composing what is referred to as a "feature history" matrix. Additionally, a weekly feature history is built using a compilation of the daily features. This weekly feature history is composed by taking the twelve daily features of the five weekdays and making each a single feature of the weekly set. For example, rather than having one 'maximum velocity' feature, the weekly feature matrix will have five maximum velocities recorded, each labeled with its corresponding day (i.e. 'Monday maximum velocity', 'Tuesday maximum velocity', etc.). Creating a feature matrix using such a hierarchical approach is carried out with the goal of automatically identifying days of the week that are influential in predicting activity related to shots (such as low activity on the Tuesday before a shot when the high explosive is loaded into the vessel). Furthermore, in composing the weekly feature history, only daily features from Monday to Friday are used so that the overwhelming trend of weekday versus weekend activity may be eliminated and activity more relevant to the state of operations at DARHT can be seen. Figure 9 illustrates the feature history matrix for each of these time discretizations. Assessment of features at various rates is important for classification to be possible on different scales. For example, if the classification of interest is if the facility is in an operational state or dormant state for a given day, then a daily feature history is necessary. However, if the classification of interest is how many weeks away the facility is from carrying out a shot test, then a weekly feature history will likely be more informative.

Reduction of the dataset through feature extraction is substantial, with the size of the raw measurements from nine streams of 106.8 GB being reduced to 3.1 GB for the hourly rate and 413 KB for the daily rate. Though the dimensionality of the data has been greatly reduced by feature extraction, trends in the data such as weekdays and weekends, can still be clearly distinguished. White areas in these figures indicate time periods with gaps in the data. For the hourly data stream, any hour or day that is missing at least 25% of the data for that time frame is removed from the set, so as to not cause extraction of misleading features. These gaps in the data are later inferred through imputation so that complete datasets can be used for classification (Casleton et al. 2016).
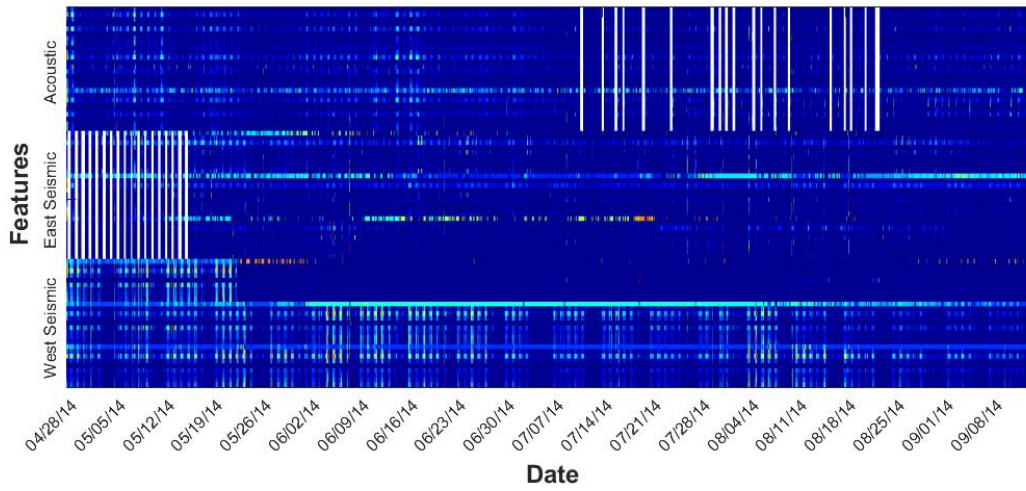
(a) Hourly feature history composed of 81 features x 3360 hours
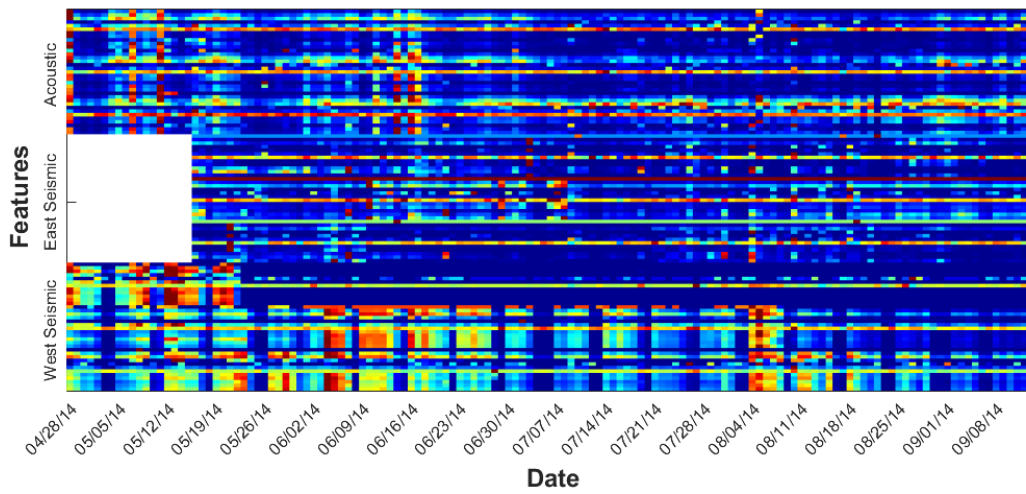
(b) Daily feature history composed of 108 features over 140 days
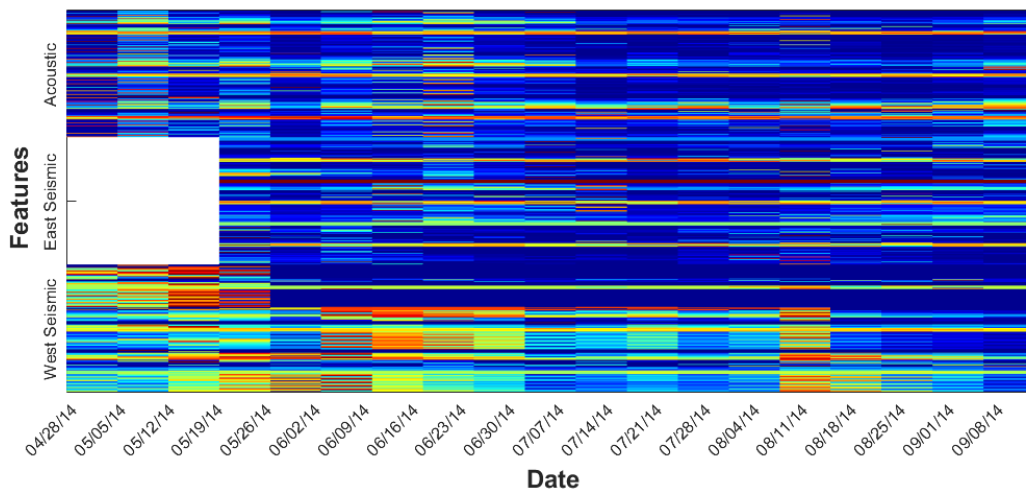
(c) Weekly feature history composed of 540 features over 20 weeks

Figure 9. Generic statistical features extracted on (a) hourly, (b) daily and (c) weekly time scales.

4. Event Extraction

Generic feature extraction demonstrates promise for assessing the amount of activity occuring during a particular day. Only limited information, however, can be gained from these features in relation to specific events occuring over small windows of time, such as explosive events. Therefore, event identification is also applied to the seismic data in an effort to detect specific events relevant to activities at DARHT. Rather than continously extracting generic features of the data streams, this method evaluates periods of time in which activity outside of "normal" operations is occurring and then determines features of measurements within that particular time frame. This type of extraction is based on the idea that individual events create their own unique signature. While this concept holds true in the DARHT seismic data, some unique events (such as a shot at the DARHT facility) have signatures in the velocity domain with amplitudes that are an order of magnitude lower than other events. Without obvious spikes in velocity data, these events may be masked by the more prominent events. It would be difficult to identify these events if a threshold value is chosen in the time domain. For this reason, the data is analyzed in both the time domain and the frequency domain, as unique signatures are apparent in both of these domains. As shown in Figure 10, car events, explosive testing performed at other test sites, and a DARHT shot, all provide unique signautes in both the time and frequency domains. Car signatures are identified by pairs of peaks in the time domain and the broadband frequency content (Wellman et al. 1997). The analysis implemented herein will combine outlier detection in the frequency domain with regression modeling in the time domain to detect events from the vertical seismic measurements collected by the entrance road (SPwest sensor).
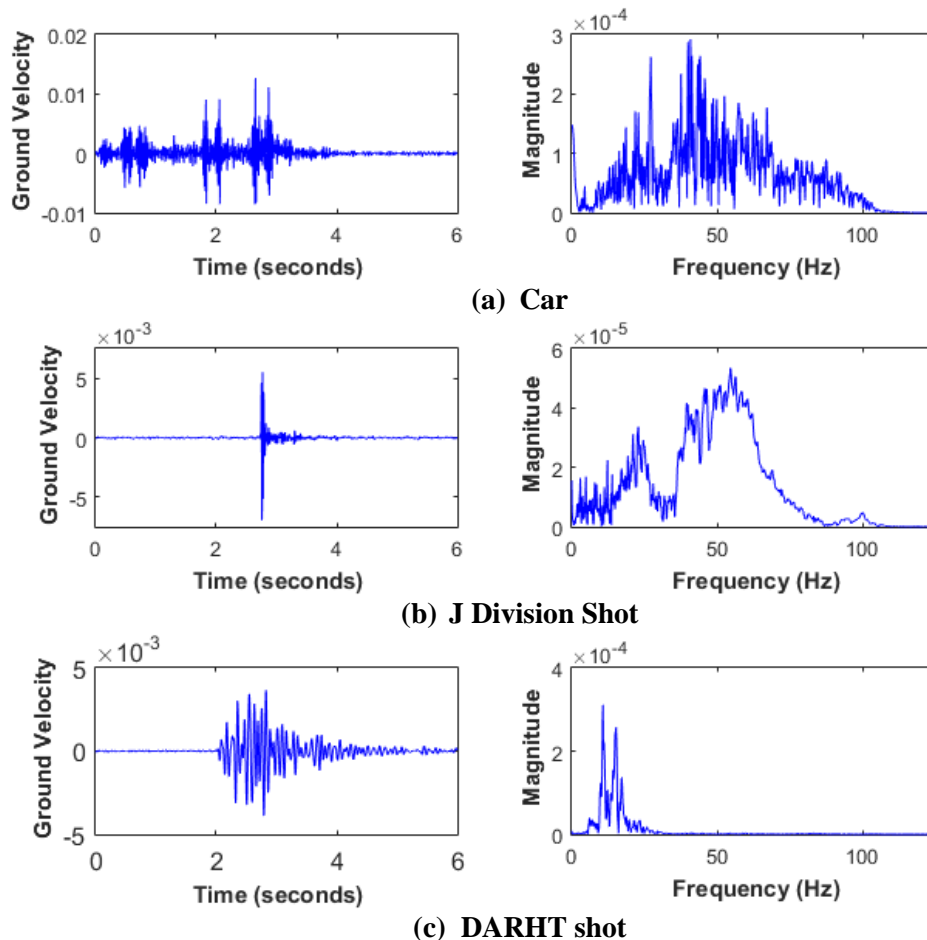

(a) Car

(b) J Division Shot

(c) DARHT shot

Figure 10. Time (left) and frequency (right) signature of cars, J Division shots, and DARHT shots from SPwest vertical seismic sensor.

13

4.1 Spectral Analysis

The first analysis completed for detecting events is an outlier analysis of the frequency content over a day. Outlier analysis, also referred to as novelty detection in the case of searching for a change in condition, seeks to identify portions of data from a large set which indicate some condition in which the system is operating in something other than a "normal" condition (Worden et al. 2000). Outliers in a spectrogram are an obvious indicator of abnormal occurrences in vibration data, as a signal remaining constant through time would result in horizontal lines across each frequency (Sohn et al. 2001a).

In our application, the constant stream of seismic measurements collects many time samples where "no activity" is occurring, which should be indicated by low energy in the frequency content as seen at 2:00 AM in Figure 11. When events occur, peaks are seen in the spectrogram at the frequencies being excited. For example, cars passing the sensor result in broadband excitation from approximately 10 Hz to 70 Hz with the most energy focused around 50 Hz (Figure 10a). Similarly, a shot at DARHT is seen with energy more concentrated around 15 Hz (Figure 10c). Therefore, higher energy at a particular frequency beyond the "normal" state of no activity indicates some event of interest has occurred. We evaluate the frequency spectrum using a spectrogram with five minute time intervals and five Hertz frequency bands (Figure 11) at frequencies ranging from 0 to the Nyquist frequency of 125 Hz. This spectrogram is computed on a daily basis and outliers are identified in time across individual frequency bands as a time having a power at a given frequency two standard deviations higher than the average power for that frequency. For example, on August 28, 2014 shown in Figure 11, the average intensity in the 10-15 Hz frequency bin is 14.3. However, around 20:57 PM the intensity in this frequency bin spikes to 130.7 indicating an event of interest, which happens to be a shot at DARHT.
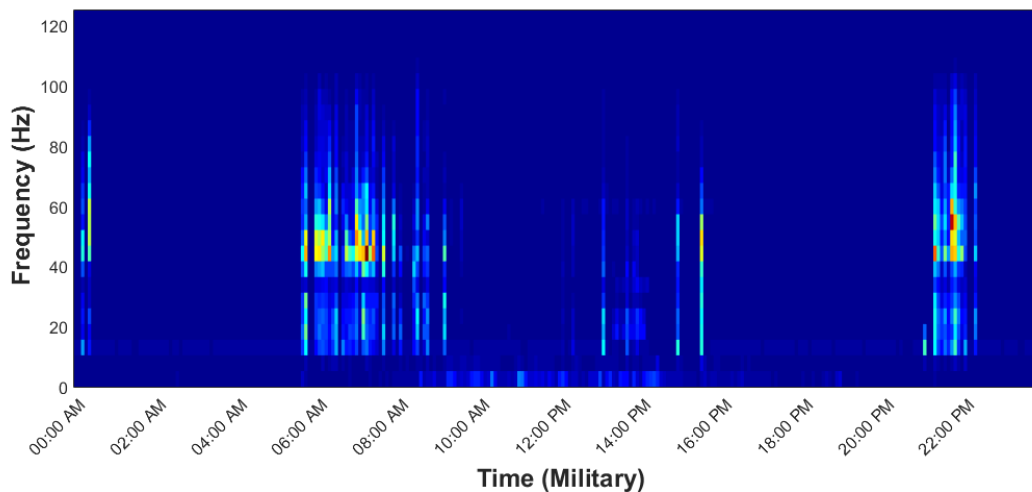

Figure 11. Spectrogram of SPwest vertical seismic data for August 28, 2014.

While this method shows promise for identifying times at which events occur and is particularly useful in identifying events that are not obvious in the time domain, there are downfalls calling for further refinement. Specifically, five minute time bins were found to work well for identifying the types of events of interest, however, several events such as cars passing can occur during a single five minute period making identification as independent events difficult. Separation of these events could be possible by refining the temporal discretization of the spectrogram, however, this finer discretization comes with the burden of significantly increased computational time.

4.2 Autoregressive Time Series Analysis

Following the spectral analysis, we seek a second approach capable of detecting events on a finer time scale. A common approach for such detection is computing residuals, where residual refers to the difference between a measurement and a model prediction at a normal or expected state (Sohn et al. 2001a). In such an analysis, the model is developed for a "normal" state and high residuals are therefore indicative of an abnormal state. This method is particularly useful if the model and measurements are both such that the residuals are sensitive to the state change of interest, such as damage in a structure (Basseville et al. 2004).

In our application, the measurements available for comparison to the nominal model are the raw ground velocity seismic measurements. Additionally, we know the events of interest for detection are those where anything other than "no activity" is occurring. Therefore, a model predicting the seismic response during a period of activity is needed. Autoregressive (AR) models are a common model for generating time series predictions. Eq. 4 defines a standard AR model where t is a time vector, α is the set of model parameters, p denotes the model order, and ε_t accounts for the white noise. The resulting prediction will be a vector X that is time dependent.

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t \qquad \text{(Eq. 4)}$$

There are many different forms of AR models, some more appropriate than others depending on the application. In this case, an Autoregressive exogenous (ARX) model, in which a known external input is also included in the model, may be useful. Eq. 5 defines an ARX model where Y is a known time dependent measurement for any given time t, β is an additional set of model parameters, and q denotes the model order of the external input.

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i Y_{t-i} + \varepsilon_t \qquad \text{(Eq. 5)}$$

When measurements representing the desired prediction are available, the vectors of AR and ARX parameters, α and β, may be trained using available measurements. Once these parameters have been trained the model is used to generate predictions (Figure 12). Finally, a new measurement is compared to the model prediction and residuals are evaluated to determine the similarity of these two signals. Here, events of interest, such as cars entering the facility or explosive events, are then identified based on an activity indicator derived from statistics of residuals between the seismic signal and the ARX model predictions, referred to as Shift in Residual Index (SIRI) and defined by Eq. 6. SIRI depends on the sum of squares statistic, *SS*, for both a nominal measurement, *N*, used for training as well as the new measurement for which a prediction is being made, *F*, referred to as the forecast. Eq. 7 defines the sum of the squares statistic, where μ is the mean prediction error, σ is the standard deviation of the prediction error, and r is the range of the response. With this metric, if the signal used for training is the same as that used for prediction the resulting SIRI will be equal to zero. Thus, the closer the SIRI is to zero, the better the fit of the new measurement to the training signal.

$$\text{SIRI} = 100 \times \left| \frac{1}{SS_N} \times SS_F - 1 \right| \qquad \text{(Eq. 6)}$$

$$SS = \frac{\mu^2 + \sigma^2}{r^2} \qquad \text{(Eq. 7)}$$

Due to the location of the SPwest sensor next to the entrance road, and particularly the fact that it is located near a crack in the road, car activity dominates the events picked up in these measurements. Therefore, the time series model parameters are trained using the velocity signature of a car. Though cars dominate the signal we still have the goal of identifying other events such as explosive testing and the DARHT shot, which should be possible with the model trained with a car signature due to the fact that the signatures of these events are more similar to a car passing than they are to no activity occurring. For this reason, the nine signals shown in Figure 12 are used for developing the model. The four car signals are used for both training and testing the model while the three DARHT signals and two explosives signals are only used for testing the model.
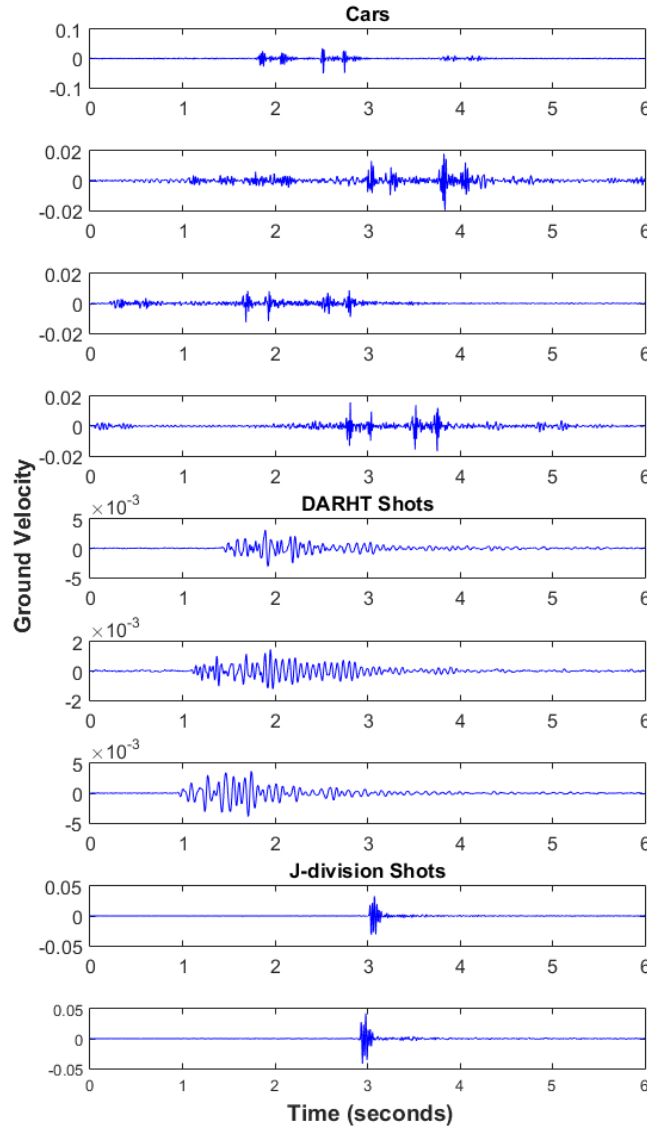

Figure 12. Nine signals used for training and testing autoregressive time series models.

Ultimately the desired outcome is a model with which the SIRI for these different event types is unique with low variability. Uniqueness is desired so that the events may be separated and possibly clustered or classified based on the type of event. Low variability is desired so that, again, the events are separable but also so that the effect of the input selected to train the model and for the nominal measure is not significant. Finally, a model that results in a SIRI below a reasonable threshold for all three event types is necessary so

that time periods with events can be identified separately from those with no activity (resulting in a high SIRI).

A series of methods are applied to refine the time series model until acceptable uniqueness and variability are achieved. First, a fifth order AR model is trained on a single car signal and then the trained parameters are used in the predictive model. Figure 13a shows the resulting SIRI when predictions of this model are compared with the nine signals from Figure 12. No two car signatures will ever be the same. Therefore, each of the car signatures are used one-by-one to train the parameters, make a prediction, and compute the SIRI with this same car as the nominal signal. This method is used to determine the variability of the SIRI based on the car selected for training and nominal measurements. Figure 13 illustrates this variability, as the bars show the mean SIRI and the black bars show the standard deviation. Specific values for the SIRI given different inputs are presented in Appendix B. With this first method the mean SIRI or DARHT shots is clearly distinguishable from other events. However, variability is higher than allowed, with the majority of signals having standard deviations greater than the mean SIRI.

The second method seeks to reduce the residual variability. With this method, much like the first, each of the four car signals are independently used to train the five AR model parameters. The difference in this method is that these four sets of parameters are combined to create one parameter vector averaged from the four. These averaged parameters are then implemented for prediction. With this method, variability is only due to the car selected as the nominal signal. Results of this second method are shown in Figure 13b. With this method the variability in all signal types has been reduced. Additionally, DARHT shots are distinguishable from cars and J Divison shots. However, the high values of DARHT shots present an issue for thresholding the SIRI value to detect events, since some undesired activity could be indicated with such a high SIRI value. Additionally, uniqueness of the cars compared to J Division shots is not apparent.

The third method implements a fifth order ARX model rather than the AR model of the previous two, with the goal of having less flexibility in the model by using a car signal as the known exogenous input, Y. Similar to the first method, this method uses a single input and output combination (the first car signal as input and the second car signal as output) to train the ten ARX parameters. These ARX parameters were then used for prediction, and once more the SIRI for each of the nine signals was computed for each of the four cars as the nominal signal. Figure 13c shows the results of this method. This method brings the maximum SIRI value for all events below 500, making a more reasonable region for thresholding to detect events. Variability in these residuals is only slightly reduced for the DARHT shots and is nearly the same for cars and J Division shots.

A fourth, and final, method combines the successful components of the second and third methods. In this method, the twelve possible input and output combinations of car signatures are used to train twelve sets of ARX model parameters. These twelve parameter sets are averaged to determine the ten ARX parameters implemented for prediction. Using this method, we obtained the best results, with the three event types being unique and separable as well as having low variability and falling below a threshold of 400, as shown in Figure 13d. Thus, this fourth method was selected for expanding to a full analysis.
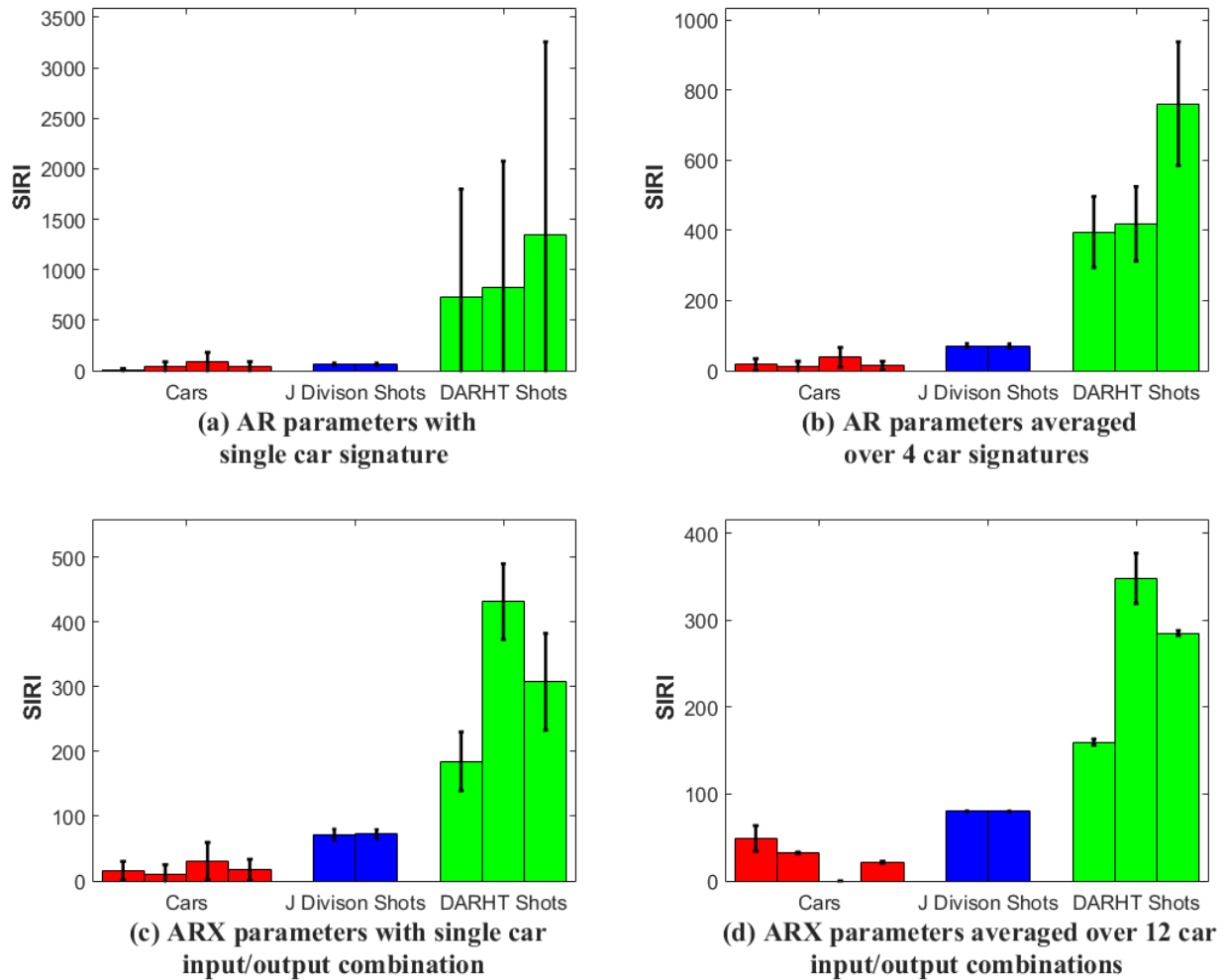
Figure 13. Comparison of SIRI metric for single and averaged AR and ARX model predictions.

Despite the promising results of this autoregressive time series analysis when tested on specific events, complications arise when the analysis is applied to a full day. One issue is the identification of time windows as cars approach and travel past the sensor. These signatures of a car approaching and retreating are similar enough to the nominal car signal to have SIRI values below the detection threshold, thus causing a single car passing the sensor to appear as three or four separate events. To mitigate this issue, consecutive time windows that are identified are grouped together as a single event, as shown in Figure 14. Grouping multiple windows to extend the car signature also helps to distinguish car events from J Division shots, which appear similar in many of their features over the central time frame alone. Finally, grouping consecutive windows avoids the issue of a short duration signal, such as a J Division shot, being split into two events when its center occurs at the edge of a time window.

The second, and more pressing, issue is high variability in the SIRI indicator that is introduced when the ARX analysis is used to continuously evaluate incoming data. Times with "no activity" can have either very high or low SIRI values, in which case the possibility of false positive detection of events is introduced. Thus, although the approach is successful in analyzing times with activity, the increased computational time combined with decreased the effectiveness of the event detection when applied to large time periods with no activity makes the analysis unreasonable for application to the full dataset.
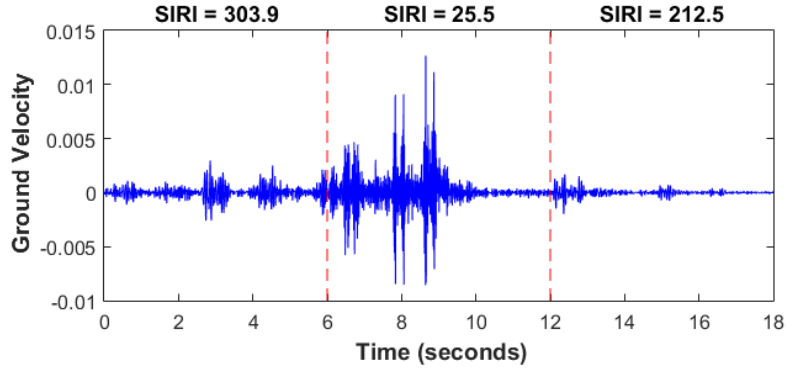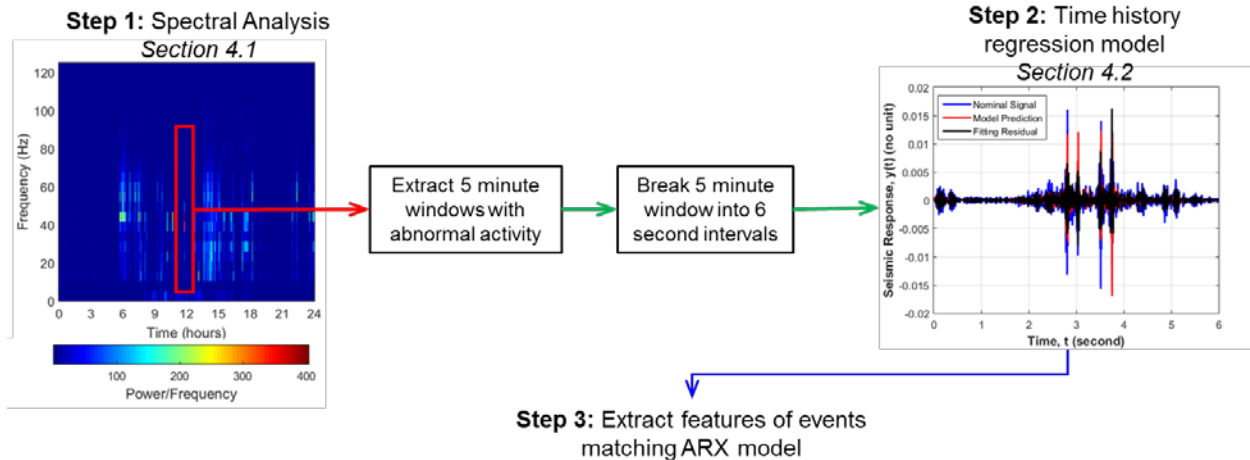
Figure 14. Compilation of three detected time frames to a single event capturing a car passing.

4.3 Two Stage Spectral and Time Series Analysis

Identifying outliers in the spectral content of a day as well as evaluting residuals of an averaged ARX time series model shows promising results for event detection. Neither analysis, however, is sufficient on its own. Rather, the two approaches have different caveats such that a combination of the spectral and time series analysis may complement one another. While the spectral analysis identifies time bins with interesting activity it does not excel at separting events within these time windows, whereas the time series analysis is successful at identifying and separating events when there is activity but gets too variable when times with no activity are evaluated. Building upon the strengths of the two analyses, we develop a new approach in which events are identified using a two-step procedure utilizing information from both the frequency and time domains, illustrated in Figure 15.

First, the spectral analysis (described in Section 4.1) is completed to identify five minute bins having higher than normal activity at a concentrated frequency band. Within these five minute intervals selected as outliers, specific events are occurring at much shorter windows. These events are separately identified using the time series analysis (described in Section 4.2), in which the five minute window is broken down into six second segments evaluated by computing residual indexes to the trained ARX model. Once an event is detected, features are extracted from this short time window with the goal of classifying the type of event, or at the very least, clustering the events. The features extracted are the same as those of the generic statistical feature extraction for hourly rates (Table 1 and Table 3). Two additional features obtained during the analysis are also included, the first being the SIRI index and the second being the number of frames for which the event occurs (effectively event duration). For events occurring over multiple windows, the residual index feature that is extracted is the minimum SIRI over the event's duration.

19

Figure 15. Two-step procedure for event detection and feature extraction.

4.4 Results of Event Detection

Once refined, this two stage event detection procedure is tested on a single day, August 28, 2014, during which all three event types are known to occur. Figure 16 shows the events detected for this day overlaid with the raw velocity measurements of the SPwest vertical seismic sensor. Note that the blue and green lines in Figure 16 are colored according to logs collected at LANL that indicate when an explosive test has occurred. Figure 17 illustrates specific signals extracted from this day including a car at 05:38 AM, J Division shot at 09:15 AM and DARHT shot at 20:58 PM. A sample of the features extracted for these events are shown in Table 5.
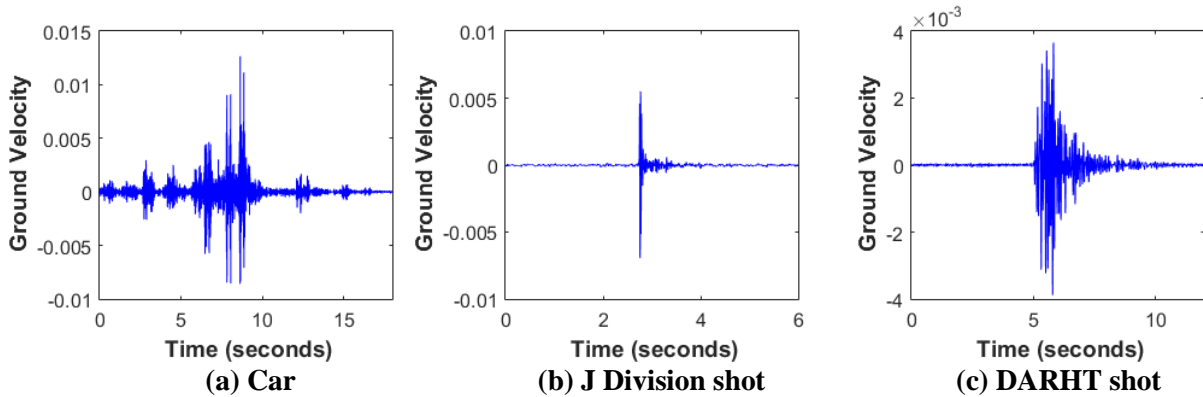

Figure 16. Events detected for August 28, 2014.

20

(a) Car **(b) J Division shot** **(c) DARHT shot**
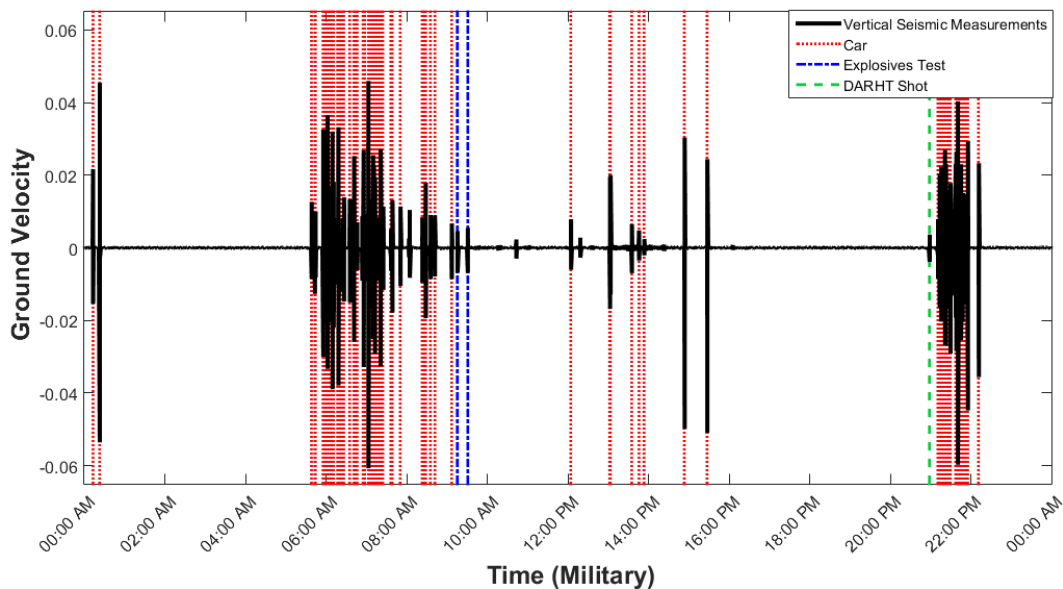
Figure 17. Signals extracted corresponding to features shown in Table 5.

Table 5. Example features extracted for three events (shown in Figure 17).

Event	SIRI	Number of frames	Max velocity	Peak frequency	Zero order spectral moment	First order spectral moment	Second order spectral moment
Car	25.4	3	0.0127	41.0	1.00E-06	4.42E-05	0.0023
J Div. Shot	82.6	1	0.0055	54.7	2.25E-07	1.10E-05	5.78E-04
DARHT	221.27	2	0.0037	10.7	2.77E-07	3.69E-06	5.17E-05

As expected, the majority of events detected can be attributed to car traffic. Therefore, the accuracy of this event detection is evaluated by comparing event counts to car counts manually collected from the video feed at the DARHT entrance gate (Key et al. 2015). Figure 18 shows this comparison of event counts from the seismic data to car counts from the videos per hour. Overall, the two counts compare well, with a few exceptions. The largest discrepancy between the two is during the 21:00 PM – 22:00 PM hour, where the manual car count detects 37 cars and the automated event count detects 24 events. The large discrepancy is attributed to the grouping of cars traveling close together by the automated event detection. This discrepancy could be reduced if the goal of event detection was only to detect cars. If the method were tailored to this specific application the SIRI threshold to 50, thereby eliminating the detection of J Division shots and DARHT shots and likewise allowing cars to be identified by a single window at the time of passing rather than grouping windows.
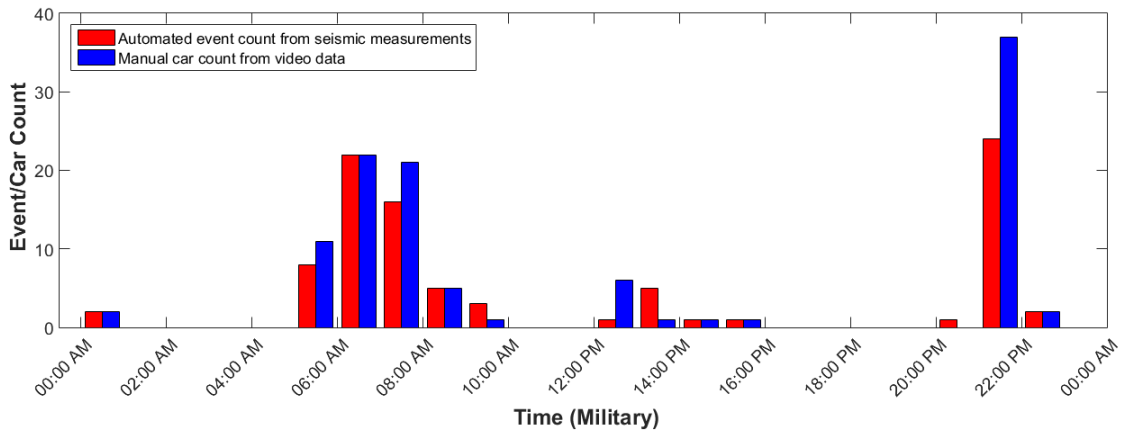

Figure 18. Comparison of automated event extraction and manual car counts for August 28, 2014.

21

Since traffic trends are indicative of shot related operations at DARHT, the number of events detected (also referred to as event density) is compiled into a feature stream, where there are 24 features and each feature is the event count for an hour in the day, shown in Figure 19a. Generation of the feature history in this manner is similar to the hierarchal method used for generating weekly feature histories with generic features, in that one feature per hour should help to identify specific hours that are influential in indicating classes. Similarly, a weekly feature history is constructed using only the Monday to Friday hourly counts with the goal of removing the dominant activity trend of weekdays vs. weekends and instead hone in on trends in the working activity to guide decisions related to DARHT shot preparations.
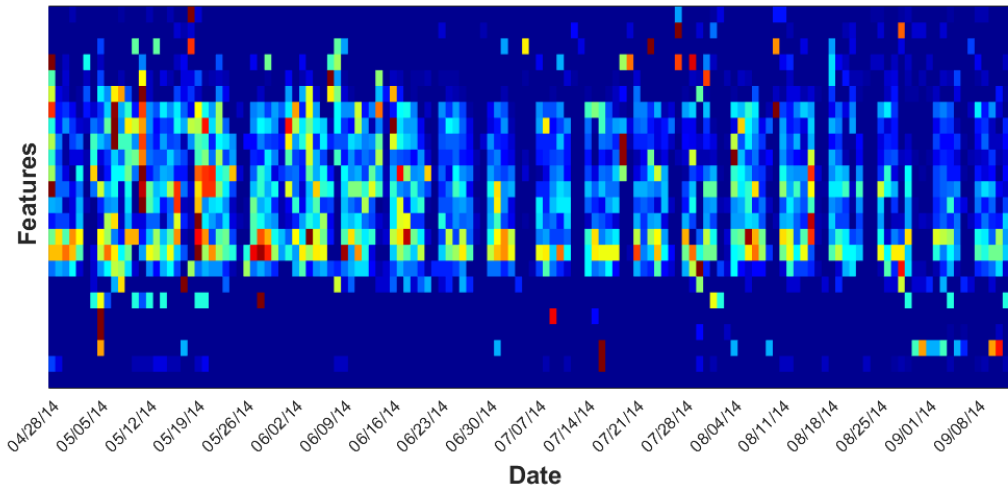
(a) Daily event density feature history composed of 24 features over 140 days
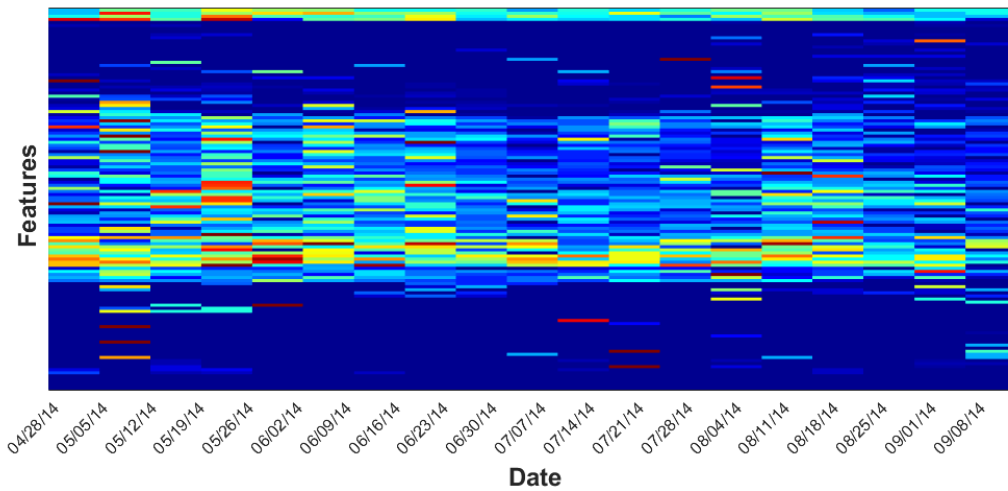
(b) Weekly event density feature history composed of 120 features over 20 weeks

Figure 19. Event density features extracted on (a) daily and (b) weekly time scales.

5. Classifying State of Operation

Ultimately, features extracted from the data are used to classify the mode of operation of the DARHT facility. Classifying the state of the facility requires two components. The first component is a model of the facility when it is operating in the classes of interest. This model is developed by training a statistical classifier using training data for which the class is known. The second component is test data for which the class is unknown and the decision maker wishes to predict the state. There are a variety of classifications that may be applied for this analysis depending on the level of detail desired. However, the accuracy of the classification will depend on the quality and availability of data for training. Additionally, the type of feature history needed for training and prediction (i.e. daily or weekly) will depend on the classification desired. In this section, we present several classification options and select one for a more detailed analysis. Classification will be completed using generic statistical features as well as the more specific event density features, with the goal of comparing the two methods to determine if one approach is more successful. In total, eight different classification possibilities were evaluated. Each of these classifications represents a different level of complexity as well as time frame.

Four classifications are evaluated for analysis at a weekly resolution, progressing from the most simplistic classification (two class option of shot week/not shot week) to the most complex (twelve class option of number of weeks leading to shot). It should be noted that only twenty weeks of data are available for the time period during which seismic and acoustic measurements were collected. Among these twenty weeks there are only three examples of shot weeks. This lack of data makes training a classifier for testing its accuracy on hold out data difficult, if not unreasonable. Therefore, all of the weeks are used as both training and test data. A comparison of the accuracy of these classifiers is shown in Table 6. As expected, due to the fact that the same data is used for both training and prediction, the classifiers perform well with over 90% accuracy for both generic statistical and event density features for the first three classifications. The fourth classification, number of weeks leading to shot, is not possible to conduct since the limited data results in insufficient examples (12 possible classes and only 20 weeks of data).

Similarly, four classifications are evaluated for analysis at a daily resolution, again starting simple (two class option of weekday/weekend) and progressing to most complex (82 class option of number of days leading to shot). Daily resolution reduces the limited data issue encountered at the weekly scale, with 140 days of data available including 78 normal work days, 40 weekend days, 9 operating (B) Fridays, 10 non-operating (A) Fridays, 3 holidays, and 3 shot days. For this reason, training and test data can be separated at the daily scale, with 121 days being kept for training and 19 days being held out for prediction. The 19 days selected for prediction are taken from the middle of the summer, falling from June 27 to July 15. In the first three classifications event density features consistently out performs generic statistical features, with the accuracy of both cases deteriorating as the classification becomes more advanced. Though the availability of data for classification at a daily resolution is better than that available for a weekly resolution, the fourth classification, lead up to shot day, is still not possible due to the data limitations (82 possible classes and only 140 days of data).

For the sake of simplification, we will focus on only one classification for the remainder of this report. The daily resolution three class problem of working days, DARHT Fridays off, and weekends or holidays will be implemented from here on out. This classification has been selected for a few reasons. Mainly, there is increased availability of data for testing and training at the daily resolution. Additionally, the three class problem presents a more interesting classification than the two class problem without having a lack of information reducing the accuracy too much, as seen in the five class problem. While the remainder of the report focuses on this one classification, results of the other possible classifications may be found in Appendix A.

Table 6. Success rates of classifiers with different classification and weekly features.

Classification	Weekly (Mon. – Fri.) Feature History	
	Generic Statistical Features	*Event Density Features*
Shot week / Not shot week (2 classes)	95.0% (19 of 20)	100.0% (20 of 20)
Shot Tuesday / Not shot Tuesday (2 classes)	95.0% (19 of 20)	95.0% (19 of 20)
Shot week / Prep week / Normal week (3 classes)	90.0% (18 of 20)	95.0% (19 of 20)
Number of weeks leading to shot (12 classes)	Not enough data to separate classes	Not enough data to separate classes

Table 7. Success rates of classifiers with different classification and daily features.

Classification	Daily Feature History	
	Generic Statistical Features	*Event Density Features*
Weekday / Weekend (2 classes)	84.2% (16 of 19)	100.0% (19 of 19)
Working / Friday off / Weekend or Holiday (3 classes)	73.7% (14 of 19)	94.7% (18 of 19)
Working / Friday on / Friday off / Weekend / Holiday (5 classes)	57.9% (11 of 19)	84.2% 16 of 19
Number of days leading to shot (82 classes)	Not enough data to separate classes	Not enough data to separate classes

5.1 Classifier Development

In order to classify the state of operations on a given day, a statistical model must first be developed that can compare features of test data to features of training data for each of the classes. Recall the large dimensionality of the feature histories to be used, with 108 features per day for the generic statistical features and 24 features per day. Thus, we first use Principal Component Analysis (PCA) to reduce the dimensionality of the data. The model developed herein is based on the modes of the number of principal components necessary to capture approximately 55% of the variability in the data (in this case, the first 3 principal components).

After the data has been reduced to these principal components a model is trained for every class of interest (in this case, 3 classes). Hyperparameters of this model include a mean vector, μ_k, and covariance matrix, Σ_k, as defined in Eq. 8 and Eq. 9 where k represents the number of principal components. Given the mean vector for a new set of features for prediction, y, the Mahalanobis distance is calculated for every class, M_k, as defined by Eq. 10. Once these distances have been determined, the class with the minimum distance is selected as the predicted class. Classification using PCA and Mahalanobis distance is a simple, well established method for classification and we recognize more advanced statistical classifiers, such as the elastic net, which could also be implemented and used for feature selection (Zuo and Hastie 2005).

$$\mu_k = \begin{Bmatrix} \mu_{(1)k} \\ \mu_{(2)k} \\ \vdots \\ \mu_{(n)k} \end{Bmatrix} \qquad\qquad\text{(Eq. 8)}$$

$$\Sigma_k = \begin{bmatrix} \sigma^2_{(1)k} & \sigma_{(1,2)k} & \cdots & \sigma_{(1,n)k} \\ \sigma_{(2,1)k} & \sigma^2_{(2)k} & \cdots & \sigma_{(2,n)k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{(n,1)k} & \sigma_{(n,2)k} & \cdots & \sigma^2_{(n)k} \end{bmatrix} \qquad\qquad\text{(Eq. 9)}$$

$$M_k = (\mu_k - y)^T \Sigma_k^{-1} (\mu_k - y) \qquad\qquad\text{(Eq. 10)}$$

5.2 Results of Classification

Recall from Sections 3 and 4 that two distinct methods were used for generating feature histories: generic statistical feature extraction and event detection. In this section, the Mahalanobis distance classification is implemented for both feature histories using the daily rates and the three classes defined in Table 8.

Table 8. Daily modes of operation for DARHT in three class model.

Classes of Days
Working
DARHT Friday off (A Friday)
Weekend or Holiday

First, the first three principal components are determined for the feature histories. Figure 20 shows a scatter plot of the first three principal components for all 140 days during which seismic and acoustic measurements were collected. Separation between working days (red squares) and weekends and holidays (blue circles) is noticeable for both generic statistical features (Figure 20a) and event density features (Figure 20b). However, separation between working days and Fridays off (green triangles) is not as clear, particularly for the generic statistical features. Friday off is a class posing an interesting problem, as the facility is non-operational but activity is still present, such as maintenance crews or staff working at nearby facilities.

After dimensionality of the data is reduced by PCA, the mean vector, μ_k, and covariance matrix, Σ_k, are trained for each of the three classes. Next, the mean vector for each training day, y, is used to compute the Mahalanobis distance for every class, where the class with the minimum distance is selected as the class prediction for the prediction day. Figure 21 shows the Mahalanobis distance values for the 19 prediction days.

Figure 22 shows results of the classification, where the squares represent the correct class for the test day and the circles represent the class predicted by the Mahalanobis distance model. In this figure, days that are correctly classified can be identified by squares containing circles, where misclassified days are shown as empty squares. Table 9 provides details on the accuracy of the classifier trained with generic statistical features, which has a total accuracy of 73.7%. Table 10 shows the same statistics for the classifier trained with even density features, which has a total accuracy of 94.7%.
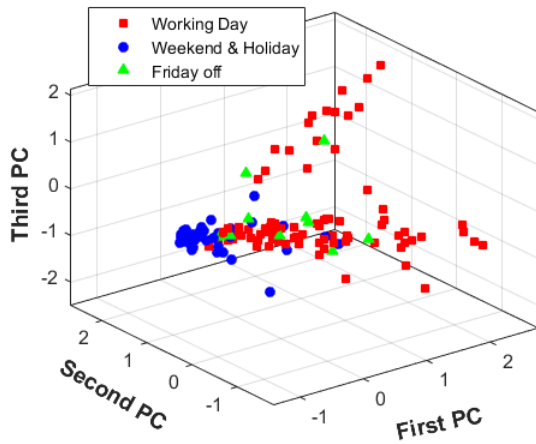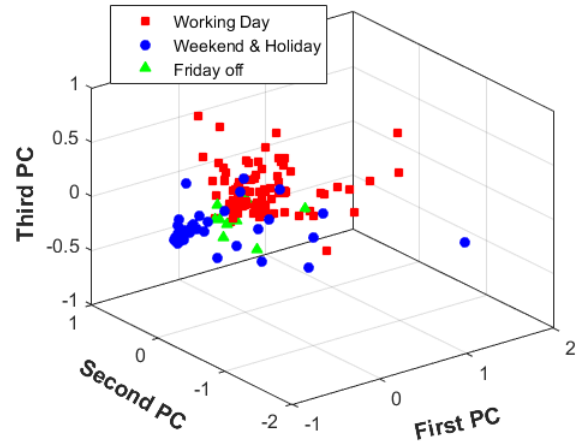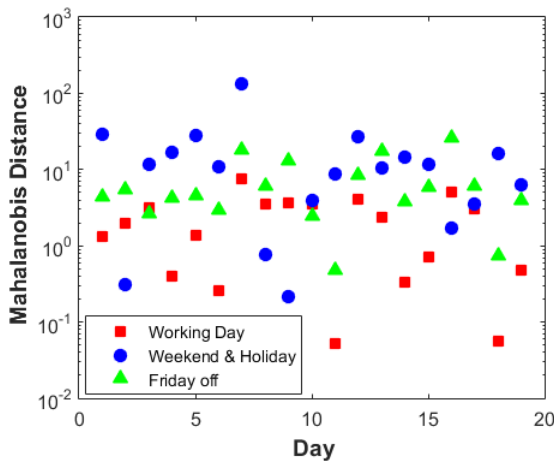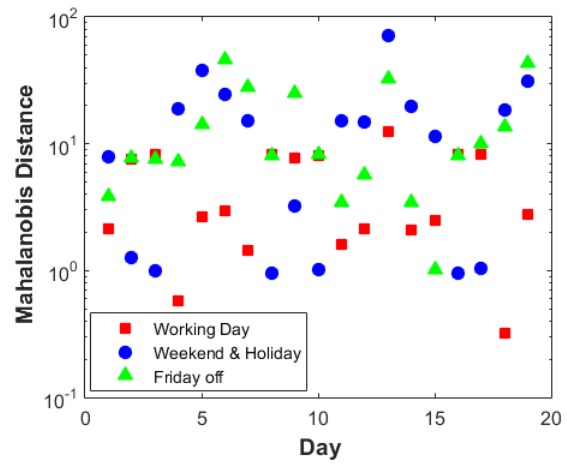
(a) Generic Statistical Features　　　　　**(b) Event Density Features**
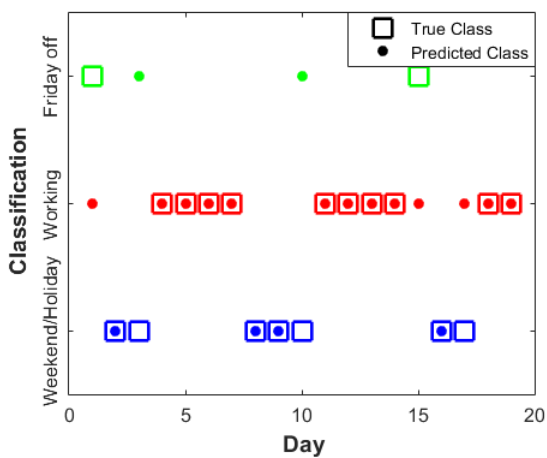**Figure 20. Separation of data in first three principal components.**
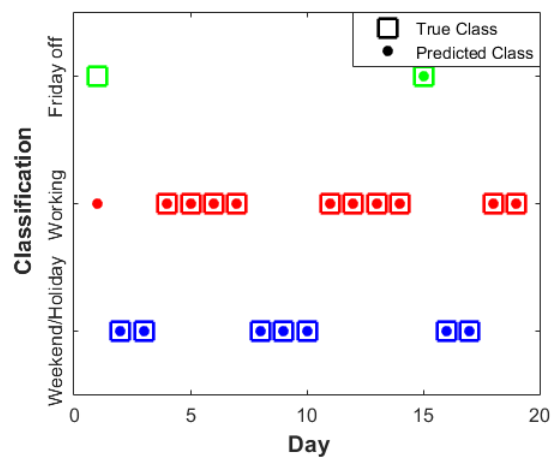

(a) Generic Statistical Features　　　　　**(b) Event Density Features**
Figure 21. Mahalanobis distances for 19 test days.

(a) Generic Statistical Features　　　　　**(b) Event Density Features**

Figure 22. Predicted class compared to true class for 19 test days.

26

Table 9. Success rate of classifier using generic statistical features.

	Weekend & Holiday	Working Day	Friday Off
Correctly Classified	4	10	0
Incorrectly Classified	3	0	2
Percent Correct	**57%**	**100%**	**0%**

Table 10. Success rate of classifier using event density features.

	Weekend & Holiday	Working Day	Friday Off
Correctly Classified	7	10	1
Incorrectly Classified	0	0	1
Percent Correct	**100%**	**100%**	**50%**

Results of these classifications provide a few key insights that should be noted. First, working days can be classified very well with 100% accuracy in both the generic statistical features and the event density features. Weekends and holidays are accurately classified with the event density features, having a success rate of 100%. Generic statistical features, however, only correctly predict weekends and holidays 57% of the time. The success of event density features over generic statistical features for classifying weekends and holidays indicates that traffic activity is more clearly indicative of operational activities at DARHT than other trends in the data. Finally, Friday off is the most difficult classification for both feature types. This result could be explained by the fact that though the staff at DARHT related to operations are off on A Fridays there is still activity, such as maintenance crews and staff at neighboring facilities, generating traffic activities on these days. Further analysis combining other data sets, such as cyber or video, are expected to provide additional information needed to improve the distinction between Fridays off and normal working days.

6. Robustness to Uncertainty

Classification, as demonstrated in Section 5, with the nominal signal can be provided to a decision maker who needs to determine a course of action based on the state of operations. However, accuracy of this nominal prediction made by the statistical classifier can only be assured if there is no uncertainty in the classification. The assumption of no uncertainty, however, in this type of classification is far from reasonable. Uncertainty may be introduced to the classification at several levels including the measurements, features, parameters of the statistical classifier, or model form of the statistical classifier. For example, well controlled data collection may be completed for collection of training data resulting in data with very low levels of noise from external activities. However, measurements deployed in the field where data is collected for prediction are not likely to be as well controlled. Therefore, uncertainty may be introduced by our measurements in the form of noise due to uncontrollable activities around the sensors such as strong winds, thunder and lightning during a storm, distant sirens, or animals nearby. In this case we are now extracting features from noisy measurements to be used in a statistical classifier that has been trained on clean, well-controlled measurements. Robustness analysis is a tool that can be used to determine how the classification may change as the level of uncertainty in the measurements increases. By quantifying the robustness in addition to the classification given a nominal measurement set, a decision maker can be provided with information regarding the expected state of operation as well as an understanding of how uncertainty will affect this state. Essentially, the more robust a prediction is to uncertainty, the more reliability the decision-maker can place on this prediction.

For the seismic and acoustic data presented herein, the only source of uncertainty considered is noise in the raw measurements. This single source of uncertainty is used to demonstrate the concept of how a classification can change given uncertainty in the system. Figure 23 illustrates the classification with respect to robustness as it is completed in this application. First, noise is added to the measurements at varying horizons of uncertainty, α. Within each horizon of uncertainty, perturbations are made to the level of noise introduced in each of the nine sensor measurements, where the level of noise can vary anywhere from zero up to the uncertainty bound. Several perturbations are completed within each horizon of uncertainty and each run results in a set of distances related to the three predictions, which then drives the classification. Among the set of runs within each level there will be a minimum and maximum distance for each class that defines the range of distances possible in the classification given that level of uncertainty.
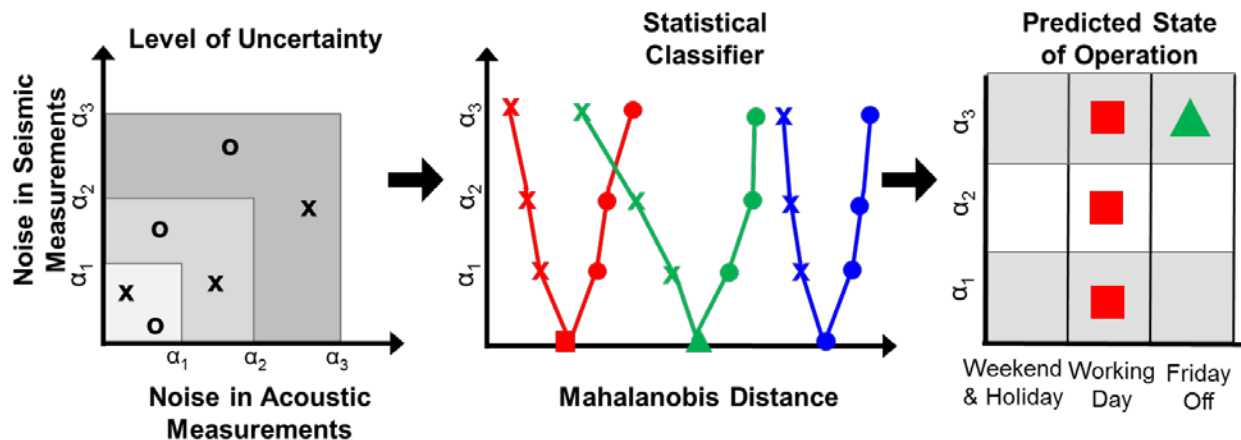

Figure 23. Illustration of robustness analysis of classification prediction for a single day.

There are two possibilities for how the classification can be affected as uncertainty is added to the measurements. One option is that the class with the minimum distance, and thus the predicted state of operation, could change (Figure 24a). Additionally, the classification may not change, but the minimum value for one class could cross the threshold for the maximum value of another class resulting in a set of

28

possible predictions considering the uncertainty, resulting in a set of possible states of operation rather than a nominal prediction (Figure 24b). A non-unique set of classifications exists when the minimum distance from a class falls within the bounds of the distance for the predicted class. For example, in Figure 24b the minimum distance for the 'Friday off' class is less than the maximum possible distance for the 'Working day' class at horizons of uncertainty greater than α_2. Accordingly, it is not possible to make a distinction between which of these two classes a decision maker should use.
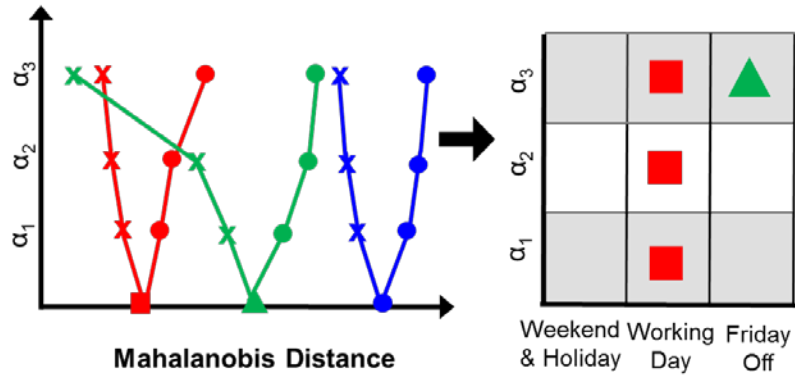
(a) Change in classification from working day to Friday off at horizon of uncertainty α_3.
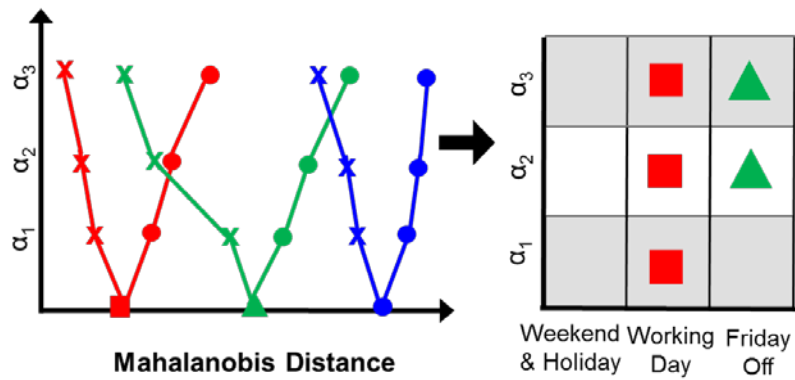
(b) Non-unique classification of working day and Friday off beginning at horizon of uncertainty α_2.
Figure 24. Changes resulting in classification from introduction of uncertainty.

Uncertainty is incorporated in the classification as white noise added to the seismic and acoustic velocity measurements where the horizon of uncertainty is defined as a percentage of the standard deviation of the velocity measurement. For a simple demonstration of the concept, robustness curves are developed varying the horizon of uncertainty from 10% to 50%, with twelve simulations at each level. It should be noted that this number of simulations is insufficient for fully exploring the uncertainty space. This is evidenced by the trend in the robustness curves, which is not monotonically increasing or decreasing. As such, these results should simply be taken as a conceptual demonstration and more runs (on the order of hundreds or thousands at every uncertainty level) need to be evaluated. This aspect of the robustness analysis is particularly meaningful in this application due to the randomness introduced to the uncertainty perturbations that is characteristic of the white noise being added. The maximum and minimum values for the Mahalanobis distance of the three classes at each level are used to populate robustness curves and determine the predicted classes. Figure 25 shows the robustness curves for Wednesday, June 25, a day that is robust to uncertainty. The robustness of this day is apparent as the optimal distances determined for each of the three classes never intersect each other and the prediction remains at the correct classification of working day, as shown in Figure 26. In contrast, Figure 27 shows the classification of Saturday, June 28, a day that is not robust to noise in the measurements. In this extreme case the distances for all of the classes vary wildly and intersect

29

one another at varying levels of uncertainty. This point is further demonstrated in Figure 28, where none of the classes can be distinguished from one another given 20% or 30% noise in the measurements.
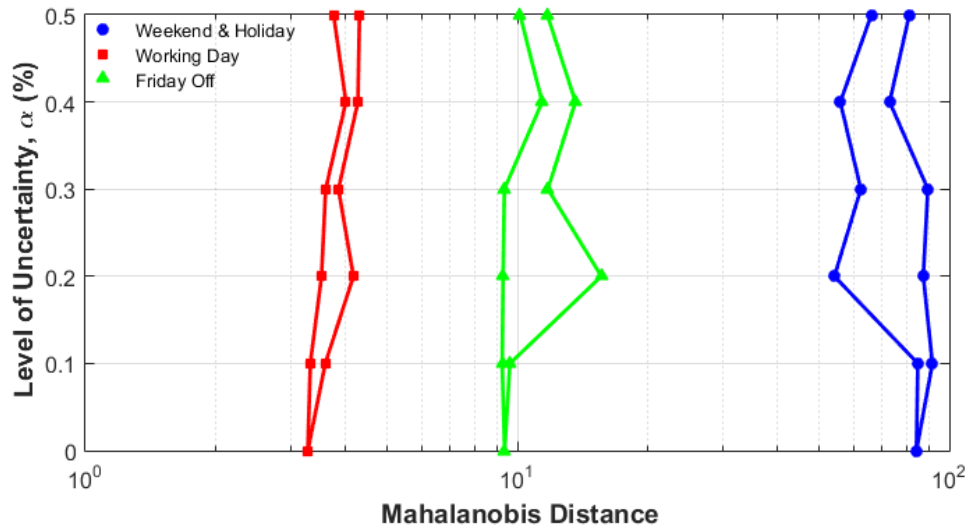

Figure 25. Shift in Mahalanobis distance calculations given increasing uncertainty levels for a day with robust data: Wednesday, June 25, 2014.
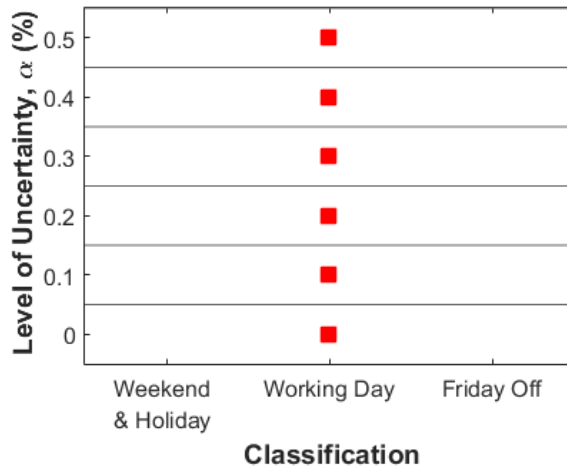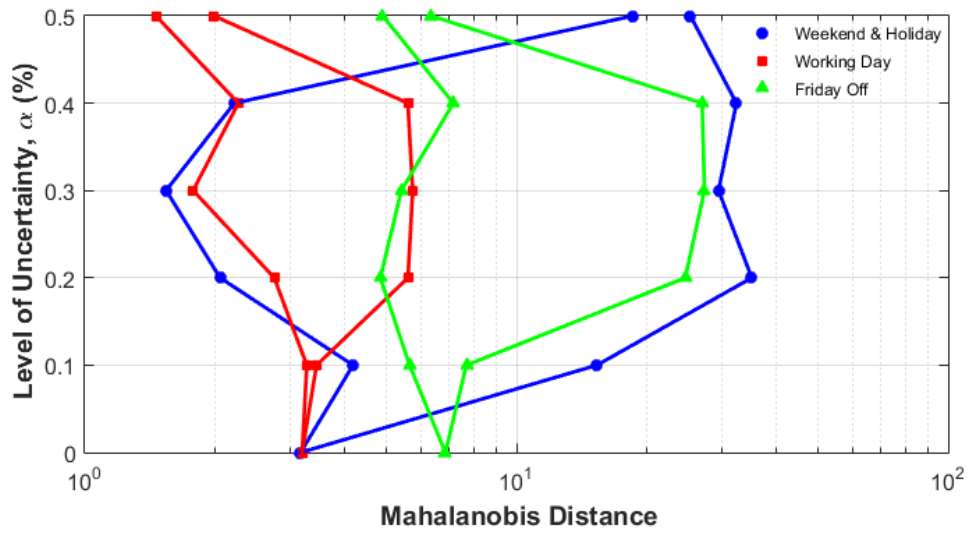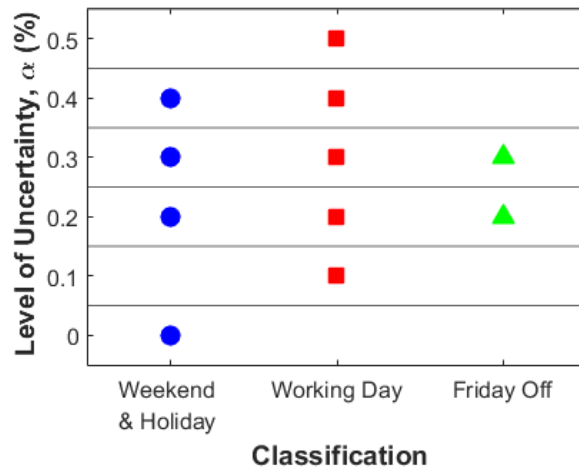

Figure 26. Demonstration of robustness, where the classification does not change with up to 50% uncertainty, for Wednesday, June 25, 2014.

Figure 27. Shift in Mahalanobis distance calculations given increasing uncertainty levels for a day with non-robust data: Saturday, June 28, 2014.

Figure 28. Demonstration of lack of robustness, where the decision can change and classification cannot determine a single state of operation given uncertainty, for Saturday, June 28, 2014.

7. Conclusions and Future Work

The state of operations of the Dual-Axis Radiographic Hydrodynamic Test (DARHT) facility at Los Alamos National Laboratory, given nine data streams of seismic and acoustic measurements, has been assessed through a generalized, automated framework for (1) feature extraction, (2) data fusion, (3) classification and (4) robustness analysis. Development and testing of this framework makes strides towards the ultimate goal of establishing a tool for rapidly providing decision makers with useful information in classifying the operations of a facility. Given data at known operation states for training and data for new, unknown states, the framework provides a nominal prediction of the state of DARHT with some indication of the robustness of this prediction to sources of uncertainties and lack of knowledge.

The framework has been evaluated using two unique feature sets, one composed of generic statistical features, and one capturing the density of events throughout the day. Extraction of events is completed through a new two-stage process developed by combining the strengths of well-established detection methods, including spectral analysis, outlier detection, and residuals of autoregressive models. The benefit of this approach is the capability to detect events of different types with only the knowledge of a car signature being needed, and then further extract features of these events. We found that classification using event densities extracted from this two-stage process, which can be directly related to traffic activity levels at the facility, performs better than classification using generic features for a three class model classifying working days, DARHT Fridays off, and weekends.

Finally, the concept of robustness of the decision is demonstrated through a preliminary analysis where uncertainty is added to the system through noise in the measurements. In this sense, robustness assesses the level of noise a measurement can tolerate before the class prediction changes. It should be pointed out that the robustness analysis presented herein is only a demonstration of concept and does not entail a full quantification of robustness. Furthermore, the robustness analysis herein is only completed with the generic statistical feature histories. A comprehensive comparison of the generic statistical features to the event density features calls for robustness of the event extraction to also be evaluated.

Our analysis concludes with some suggestions for future work. The first is an expansion of the robustness analysis presented herein to include a more complete assessment, which may involve uncertainties such as the classifier model parameters or perhaps the model form itself. Furthermore, the present analysis only evaluates the change in a decision for one given day. Continued exploration of robustness could include an assessment of the classifier itself and how robust a particular classifier is to uncertainty. Second, completion of the analysis using a more extensive data set would provide better quantification of the classification ability with the given features, specifically in the context of shot related activities. The seismic and acoustic data available as part of the LANL multi-intelligence grand challenge only covers 20 weeks and 3 shot days. Accordingly, the lack of data presents a challenge for classification on a weekly scale as well as classification specific to pre-shot and post-shot activities. Finally, the multi-intelligence component of this analysis should be extended by incorporating all of the available data (network, radiofrequency, and video) to combine different types of information for informing the classification as well as the value of information gained from each data type. Van Buren 2016 provides further investigation into this larger dataset.

References

Basseville, M., Mevel, L., Goursat, M., "Statistical Model-based Damage Detection and Localization: Subspace-based Residuals and Damage-to-noise sensitivity ratios," *Journal of Sound and Vibration*, vol. 275, pp. 769-794, 2004.

Casleton, E., Skurikhin, M., Van Buren, K., Wendelberger, J., "Imputation and Clustering for the DARHT Multi-Intelligence Grand Challenge," *Technical Report LA-UR-16-20772*, Los Alamos National Laboratory, Los Alamos, New Mexico, 2016.

Doebling, S.W., Farrar, C.R., Prime, M.B., Shevitz, D.W., "Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review," *Technical Report LA-13070-MS*, Los Alamos National Laboratory, Los Alamos, New Mexico, 1996.

Key, E., Van Buren, K.L., Warren, W., Hemez, F.M., "Video Analysis in Multi-Intelligence," Technical *Report LA-UR-16-xxxxx*, Los Alamos National Laboratory, New Mexico, 2016.

Farrar, C.R. and Worden, K., "An Introduction to Structural Health Monitoring," *Philosophical Transactions of the Royal Society A*, Vol. 365, pp. 303-315, 2007.

Scarpetti, R.D., "DARHT – Dual Axis Radiographic Hydrotest facility brochure," *Technical Report LALP-08-031*, Los Alamos National Laboratory, Los Alamos, New Mexico, 2008.

Sohn, H., Farrar, C.R., Hunter, N.F., Worden, K., "Structural Health Monitoring Using Statistical Pattern Recognition Techniques," ASME Journal of Dynamic Systems, Measurement, and Control, Vol. 123, No. 4, pp. 706-711, 2001a.

Sohn, H., Farrar, C.R., Hunter, N.F., Worden, K., "Applying the LANL Statistical Pattern Recognition Paradigm for Structural Health Monitoring to Data from a Surface-Effect Fast Patrol Boat," *Technical Report LA-13761-MS*, 2001b.

Van Buren, K.L., "Answering the Multi-Intelligence Challenge through Robust Decision-Making," *Technical Report LA-UR-16-22078*, Los Alamos National Laboratory, Los Alamos, New Mexico, 2016.

Zuo, H., Hastie, T., "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society, Series B*, Vol. 67, No, 2, pp. 301-320, 2005.

Wellman, M.C., Srour, N., Hillis, D.B., "Feature Extraction and Fusion of Acoustic and Seismic Sensors for Target Identification," *AeroSense'97*, International Society for Optics and Photonics, pp. 139-145, 1997.

Worden, K., Manson, G., Fieller, N.R.J., "Damage Detection using Outlier Analysis," *Journal of Sound and Vibration*, Vol. 229, No. 3, pp. 647 – 667, 2000.

APPENDIX A

Table A1. Success rates of classifiers with different classification with daily and weekly features.

Classification	Weekly (Mon. – Fri.) Feature History	
	Generic Statistical Features	*Event Density Features*
Shot week / Not shot week (2 classes)	95.0% (19 of 20)	100.0% (20 of 20)
Shot Tuesday / Not shot Tuesday (2 classes)	95.0% (19 of 20)	95.0% (19 of 20)
Shot week / Prep week / Normal week (3 classes)	90.0% (18 of 20)	95.0% (19 of 20)
Number of weeks leading to shot (12 classes)	Not enough data to separate classes	Not enough data to separate classes
Classification	**Daily Feature History**	
	Generic Statistical Features	*Event Density Features*
Weekday / Weekend (2 classes)	84.2% (16 of 19)	100.0% (19 of 19)
Working / Friday off / Weekend or Holiday (3 classes)	73.7% (14 of 19)	94.7% (18 of 19)
Working / Friday on / Friday off / Weekend / Holiday (5 classes)	57.9% (11 of 19)	84.2% 16 of 19
Number of days leading to shot (82 classes)	Not enough data to separate classes	Not enough data to separate classes

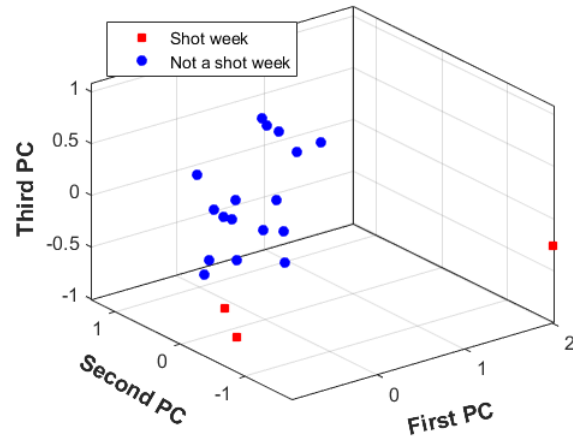*A.1 Classification of Shot Week vs. Not a Shot Week*

Table A2. Weekly modes of operation for DARHT in two class model.

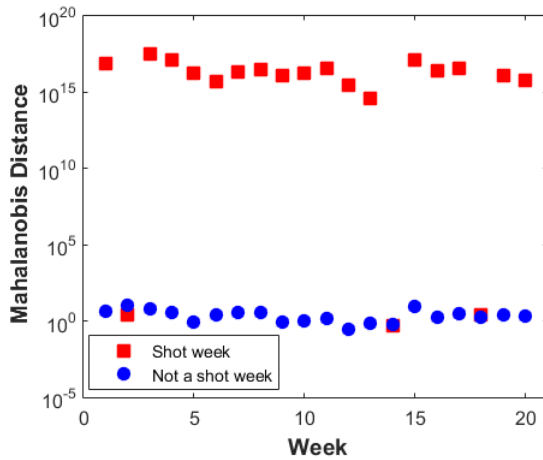| Classes of Weeks |
| --- |
| Shot week |
| Not shot week |
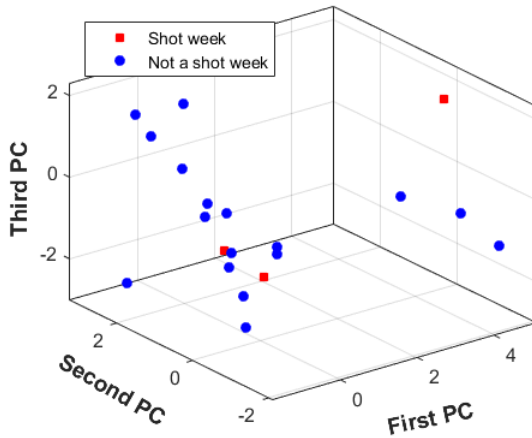
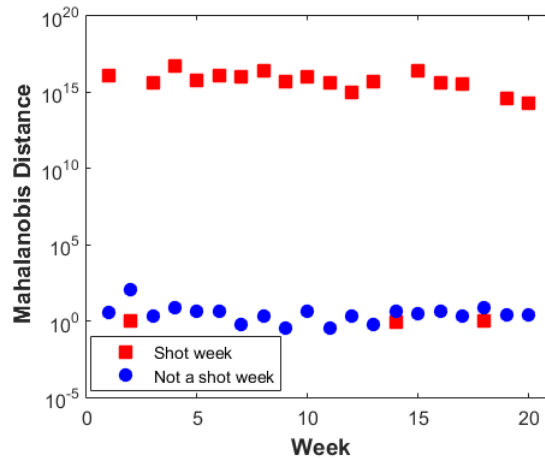**(a) Generic Statistical Features**   **(b) Event Density Features**
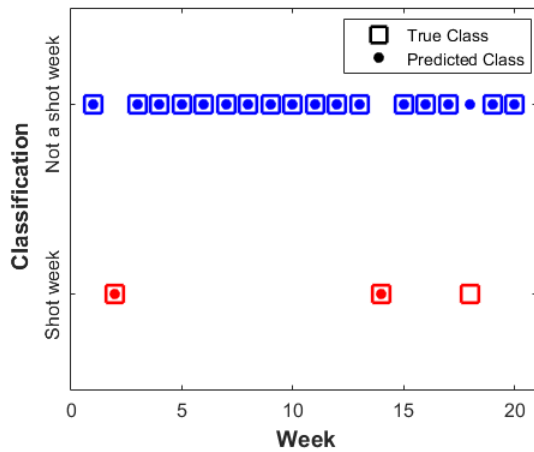**Figure A1. Separation of data in first three principal components.**
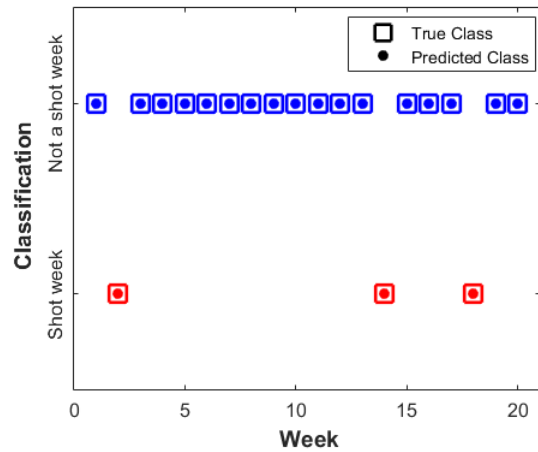

(a) Generic Statistical Features   **(b) Event Density Features**
Figure A2. Mahalanobis distances for 20 weeks.

(a) Generic Statistical Features   **(b) Event Density Features**

Figure A3. Predicted class compared to true class for 20 weeks.

Table 11. Success rate of classifier using generic statistical features.

	Shot Week	Not a Shot Week
Correctly Classified	2	17
Incorrectly Classified	1	0
Percent Correct	**33%**	**100%**

Table 12. Success rate of classifier using event density features.

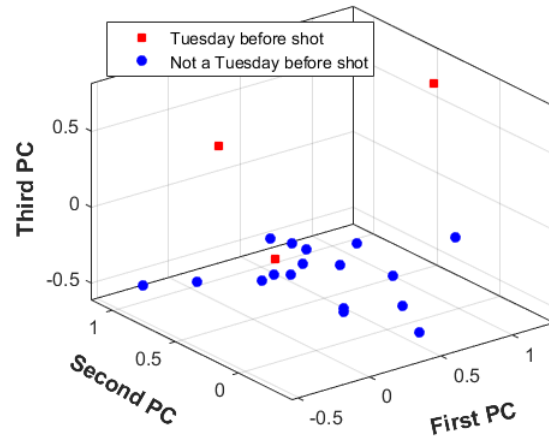|	Shot Week	Not a Shot Week
Correctly Classified	3	17
Incorrectly Classified	0	0
Percent Correct	**100%**	**100%**

A.2 Classification of Shot Tuesday vs. Not a Shot Tuesday

Table A3. Weekly modes of operation for DARHT in two class model.

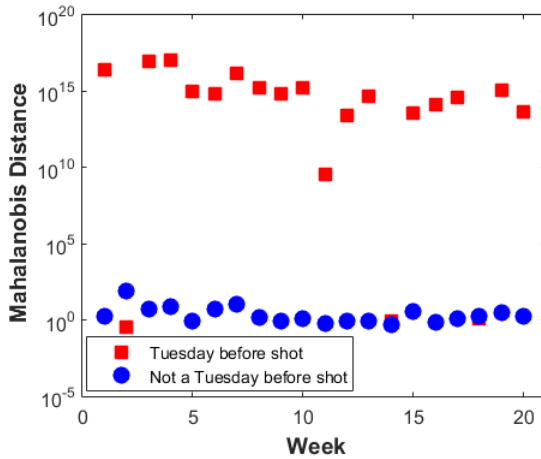| Classes of Weeks |
| --- |
| Tuesday before shot |
| Not a Tuesday before shot |
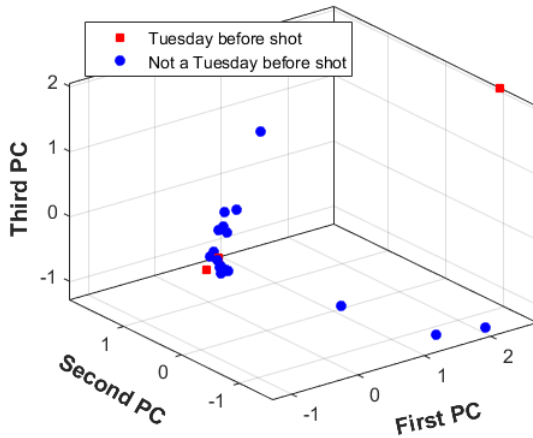
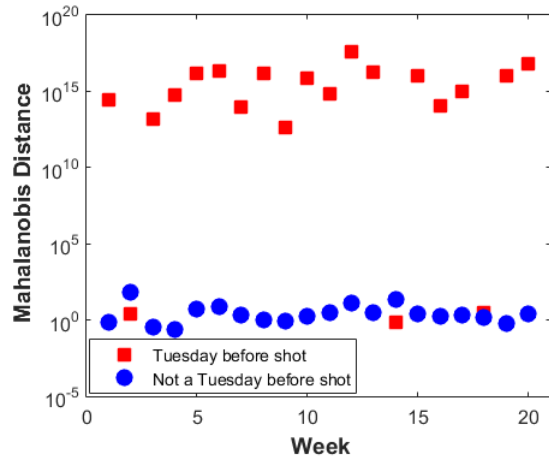(a) Generic Statistical Features   (b) Event Density Features
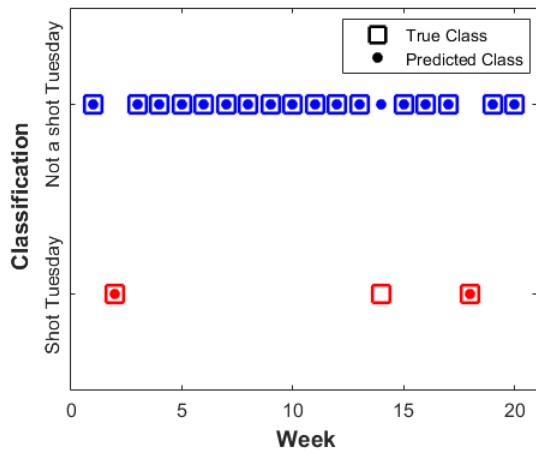**Figure A4. Separation of data in first three principal components.**
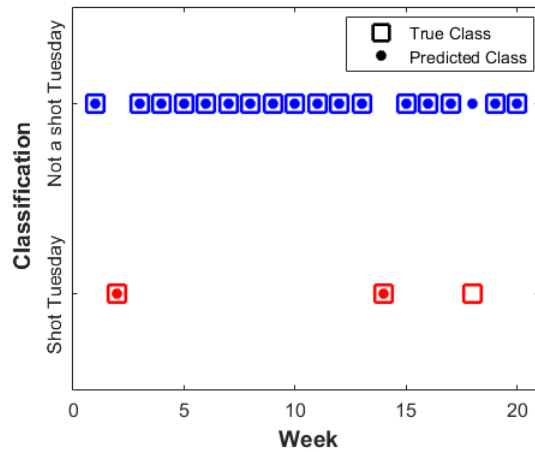

(a) Generic Statistical Features   (b) Event Density Features
Figure A5. Mahalanobis distances for 20 weeks.

(a) Generic Statistical Features   (b) Event Density Features

Figure A6. Predicted class compared to true class for 20 weeks.

Table 13. Success rate of classifier using generic statistical features.

	Tuesday before a shot	Not a Tuesday before a shot
Correctly Classified	2	17
Incorrectly Classified	1	0
Percent Correct	**33%**	**100%**

**Table 14. Success rate of classifier using event density features.**

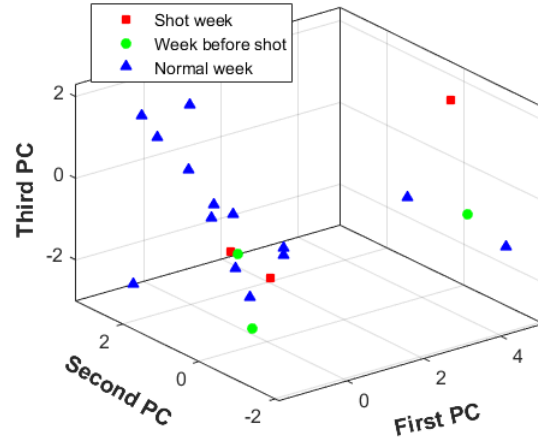|	Tuesday before a shot	Not a Tuesday before a shot
Correctly Classified	2	17
Incorrectly Classified	1	0
Percent Correct	**33%**	**100%**

A.3 Classification of Shot week vs. Prep week vs. Normal week

Table A4. Weekly modes of operation for DARHT in three class model.

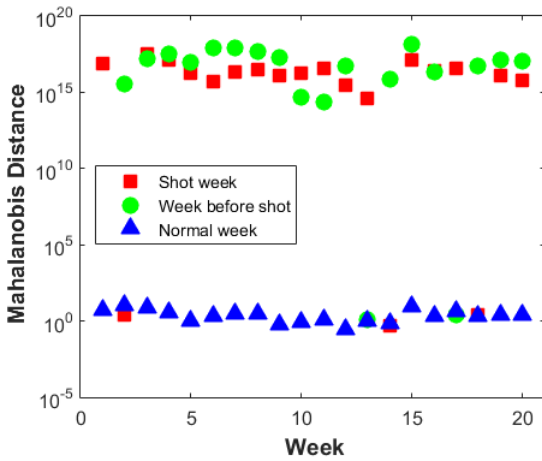| Classes of Weeks |
| --- |
| Shot week |
| Week before shot (prep) |
| Normal week |
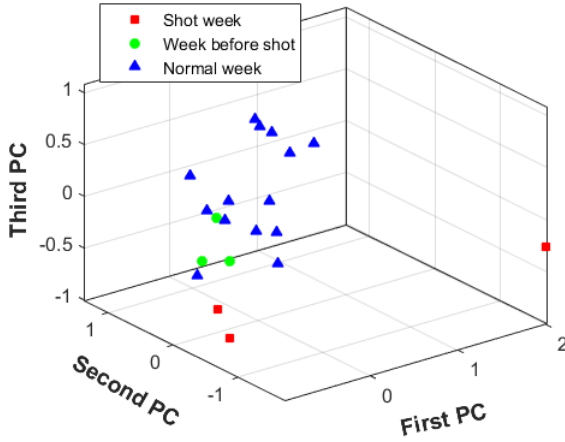
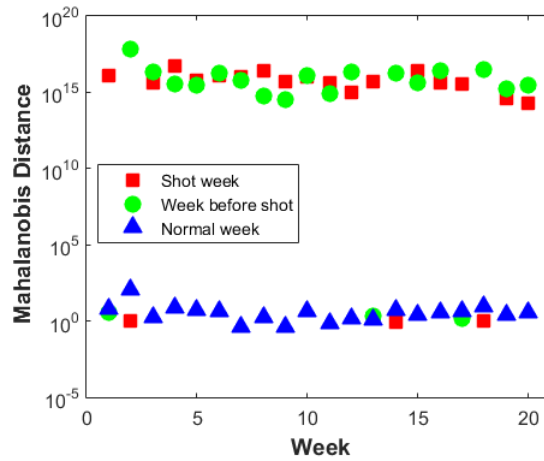**(a) Generic Statistical Features**   **(b) Event Density Features**
Figure A7. Separation of data in first three principal components.
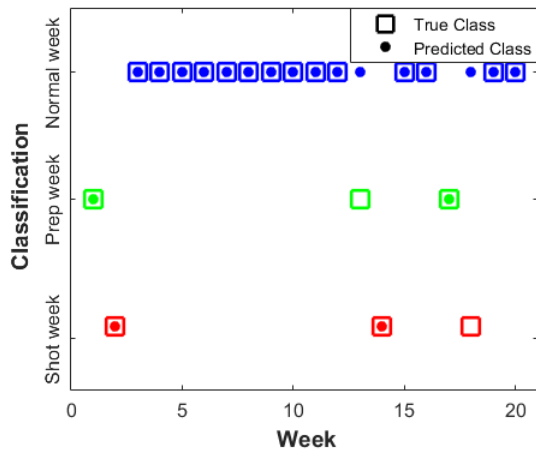

(a) Generic Statistical Features   **(b) Event Density Features**
Figure A8. Mahalanobis distances for 20 weeks.

(a) Generic Statistical Features   (b) Event Density Features

Figure A9. Predicted class compared to true class for 20 weeks.

Table 15. Success rate of classifier using generic statistical features.

	Shot week	Prep week	Normal week
Correctly Classified	2	2	14
Incorrectly Classified	1	1	0
Percent Correct	**33%**	**33%**	**100%**

Table 16. Success rate of classifier using event density features.

	Shot week	Prep week	Normal week
Correctly Classified	3	2	14
Incorrectly Classified	0	1	0
Percent Correct	**100%**	**33%**	**100%**

Table A5. Daily modes of operation for DARHT in two class model.

Classes of Days
Weekday
Weekend & Holiday
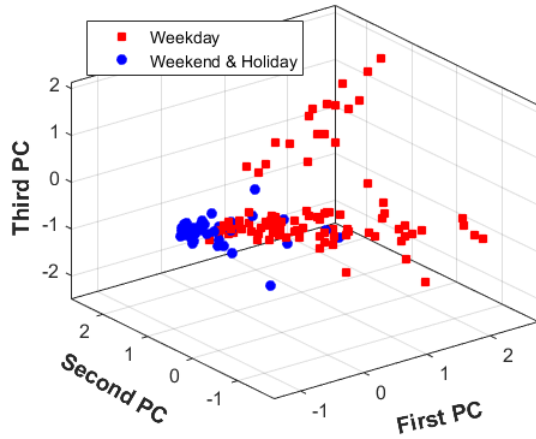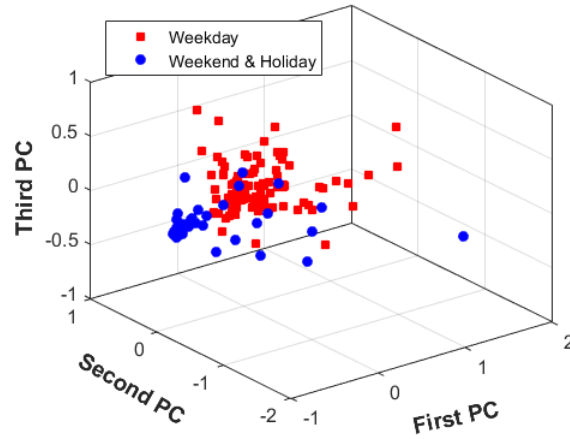
(a) Generic Statistical Features   (b) Event Density Features

Figure A10. Separation of data in first three principal components.
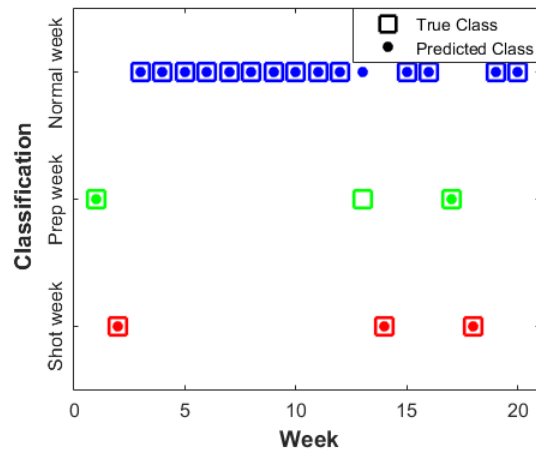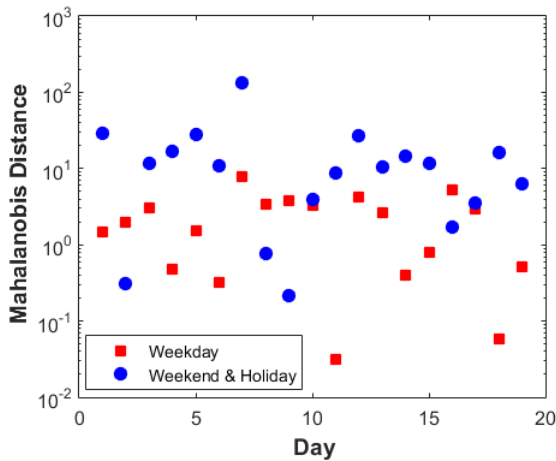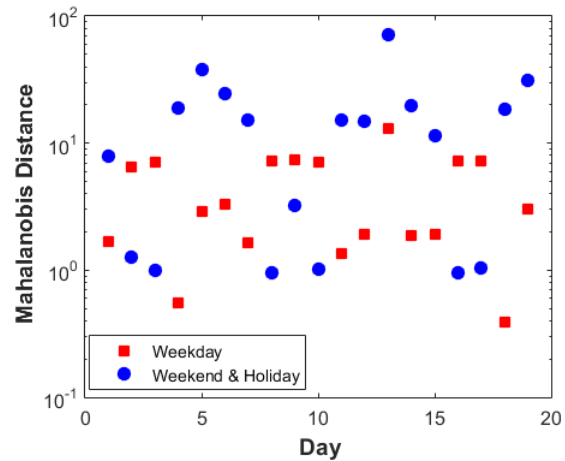

(a) Generic Statistical Features   (b) Event Density Features

Figure A11. Mahalanobis distances for 19 test days.

(a) Generic Statistical Features   **(b) Event Density Features**

Figure A12. Predicted class compared to true class for 19 test days.

Table 17. Success rate of classifier using generic statistical features.

	Weekday	Weekend & Holiday
Correctly Classified	12	4
Incorrectly Classified	0	3
Percent Correct	**100%**	**100%**

Table 18. Success rate of classifier using event density features.

	Weekday	Weekend & Holiday
Correctly Classified	12	7
Incorrectly Classified	0	0
Percent Correct	**100%**	**100%**

Table A6. Daily modes of operation for DARHT in five class model.

Classes of Days
Working day
Friday on (B Friday)
Friday off (A Friday)
Weekend
Holiday
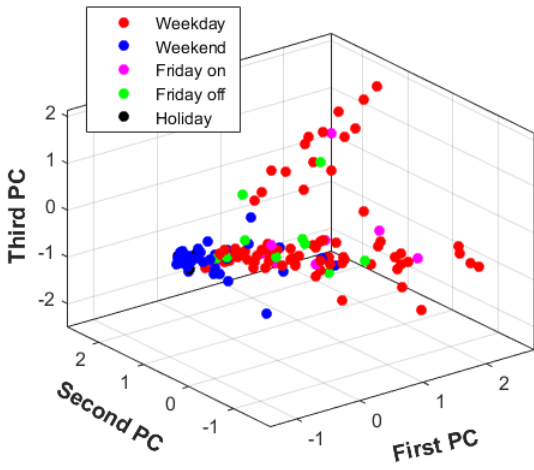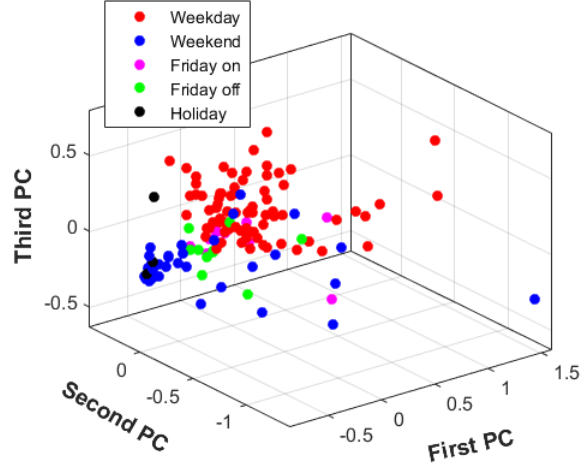
(a) Generic Statistical Features	(b) Event Density Features

Figure A13. Separation of data in first three principal components.
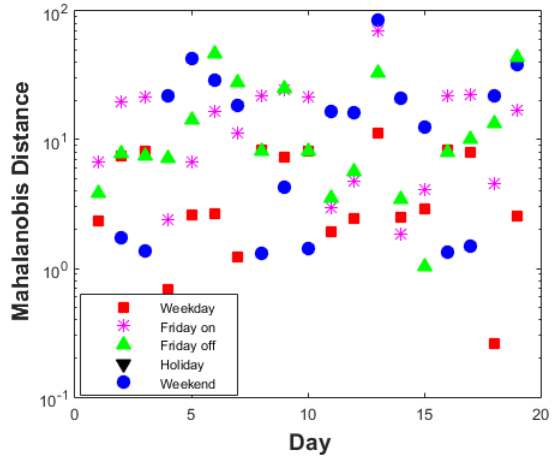
(a) Generic Statistical Features	(b) Event Density Features

Figure A14. Mahalanobis distances for 19 test days.

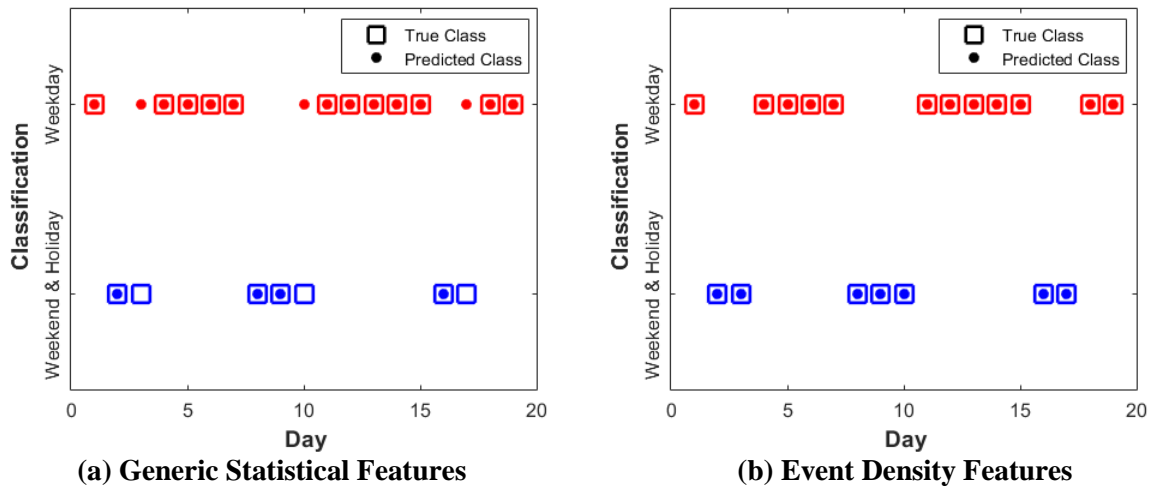**(a) Generic Statistical Features**   **(b) Event Density Features**
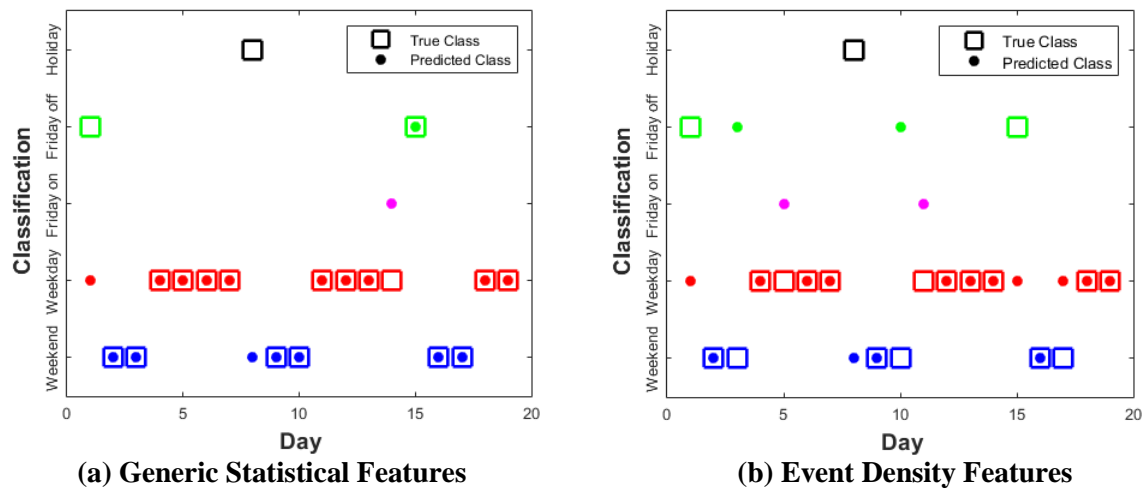
Figure A15. Predicted class compared to true class for 19 test days.

Table 19. Success rate of classifier using generic statistical features.

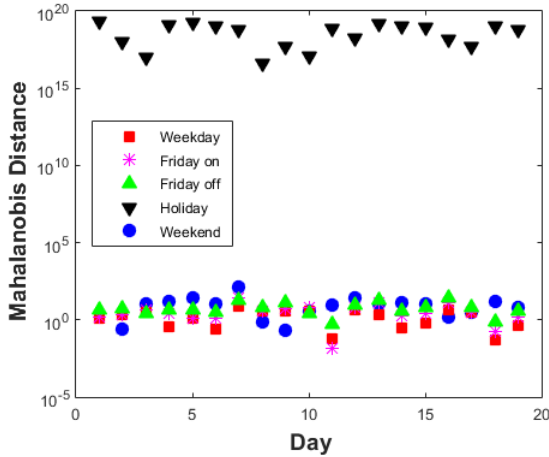|	Weekday	Friday on	Friday off	Weekend	Holiday
Correctly Classified	9	0	1	6	0
Incorrectly Classified	1	0	1	6	1
Percent Correct	**90%**	**N/A**	**50%**	**100%**	**0%**

Table 20. Success rate of classifier using event density features.

	Weekday	Friday on	Friday off	Weekend	Holiday
Correctly Classified	8	0	0	3	0
Incorrectly Classified	2	0	2	6	1
Percent Correct	**80%**	**N/A**	**0%**	**50%**	**0%**

APPENDIX B

Table A1. Resulting prediction residuals for AR model trained on a single car input.

Nominal Signal	Shift in Residual Index								
	Car 1	Car 2	Car 3	Car 4	J Div. 1	J Div. 2	DARHT 1	DARHT 2	DARHT 3
Car 1	0.0	111.2	223.9	110.9	64.2	49.2	2327.8	2692.9	4203.6
Car 2	4.9	0.0	61.0	31.1	74.2	72.7	195.0	195.4	374.2
Car 3	24.6	26.0	0.0	21.9	72.1	73.2	86.7	112.9	202.3
Car 4	11.6	16.7	50.6	0.0	47.7	51.2	308.7	318.8	593.3
Mean	10.3	38.5	83.9	40.9	64.6	61.6	729.5	830.0	1,343.4
Std. Dev.	10.7	49.7	97.1	48.4	12.0	13.2	1,069.3	1,244.8	1,191.3

Table A2. Resulting prediction residuals for AR model using averaged parameters from the four available car inputs.

Nominal Signal	Shift in Residual Index								
	Car 1	Car 2	Car 3	Car 4	J Div. 1	J Div. 2	DARHT 1	DARHT 2	DARHT 3
Car 1	0.0	13.9	67.6	21.4	65.6	64.1	501.1	529.3	945.2
Car 2	12.3	0.0	47.13	6.5	69.5	68.5	427.4	452.1	817.1
Car 3	40.3	32.0	0.0	27.6	79.5	78.6	258.4	275.2	523.3
Car 4	17.6	6.1	38.1	0.0	71.6	70.5	395.1	418.3	760.9
Mean	17.6	13.0	38.2	13.9	71.6	70.4	395.5	418.7	761.6
Std. Dev.	16.9	13.9	28.3	12.8	5.9	6.1	101.6	106.4	176.6

Table A3. Resulting prediction residuals for ARX model trained on a single car input/output combination (Car 1 as input and Car 2 as output).

Nominal Signal	Shift in Residual Index								
	Car 1	Car 2	Car 3	Car 4	J Div. 1	J Div. 2	DARHT 1	DARHT 2	DARHT 3
Car 1	-	-	-	-	-	-	-	-	-
Car 2	4.2	0.0	57.0	32.6	76.2	75.2	196.6	405.5	338.4
Car 3	32.0	27.6	0.0	18.4	76.5	77.4	134.4	390.6	222.5
Car 4	10.5	0.5	35.0	0.0	61.9	64.55	223.0	498.5	362.3
Mean	15.6	9.4	30.7	17.0	71.5	72.4	184.7	431.5	307.7
Std. Dev.	14.6	15.8	28.7	16.3	8.3	6.9	45.5	58.5	74.8

Table A4. Resulting prediction residuals for ARX model using averaged parameters from the twelve available car input/output combinations.

Nominal Signal	Shift in Residual Index								
	Car 1	Car 2	Car 3	Car 4	J Div. 1	J Div. 2	DARHT 1	DARHT 2	DARHT 3
Car 1	43.3	33.5	0.0	23.1	80.6	80.5	164.4	390.9	283.2
Car 2	41.1	32.4	0.0	21.3	79.9	79.9	156.1	339.6	282.8
Car 3	40.6	31.5	0.0	20.7	80.1	79.7	159.2	326.3	288.8
Car 4	70.9	31.8	0.0	21.2	80.1	79.8	159.3	336.2	286.5
Mean	49.0	32.3	0.0	21.6	80.2	80.0	159.8	348.3	285.3
Std. Dev.	14.7	0.9	0.0	1.1	0.3	0.4	3.4	29.0	2.8