REAL-TIME DATA ANALYTICS FOR CONDITION MONITORING OF COMPLEX INDUSTRIAL SYSTEMS

A Dissertation
Presented to
The Academic Faculty

by

Benjamin Peters

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY in the
H. MILTON STEWART SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

Georgia Institute of Technology
DECEMBER 2021

REAL-TIME DATA ANALYTICS FOR CONDITION MONITORING OF COMPLEX INDUSTRIAL SYSTEMS

Approved by:

Dr. Nagi Gebraeel, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Jianjun Shi
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Kamran Paynabar
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Timothy Lieuwen
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Nicoleta Serban
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: December 03, 2021

To my wife Ashabi, my children Isabella, Abigail, and Uriah, my parents and siblings, and to all the students and faculty that helped me along the way

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Nagi Gebraeel. Dr. Gebraeel took a chance on a student from a small school in south Texas. Despite my early setbacks, he continued to support me and find opportunities for me to grow as a researcher. I would also like to thank Dr. Kamran Paynabar who served as my de facto secondary advisor. Dr. Paynabar contributed his time to aid in several projects and I greatly appreciate all his guidance. Furthermore, I want to thank my other committee members for their contributions during my pursuit of my doctorate's degree. I want to thank Dr. Jianjun Shi for allowing me to use his office to study. I would like to thank Dr. Nicoleta Serban for her guidance on my first publication. Finally, I would like to thank both Dr. Timothy Lieuwen and Dr. Reid Berdanier for their contributions on our project for the Department of Energy.

In addition to faculty members, I also want to show my deepest gratitude to all the students that helped me along the way. First, I want to thank Dr. Murat Yildirim who guided me early in my time at Georgia Tech. I learned so much working alongside him and his contributions to my first publication are greatly appreciated. Second, I want to thank Dr. Xiaolei Fang. Dr. Fang is not only a great esearcher, but a valuable friend who watched over my children during meetings. I greatly appreciate the chance to work with him on several school projects and his direct contribution on the final chapter of this dissertation. Third, I would like to thank Ayush Mohanty for his help in implementing the methodology described in the penultimate chapter of this dissertation. Furthermore, I want to thank all the researchers I worked with on the Department of Energy project: Dr. Nicholas Rock,

Raghul Manosh Kumar, and Eric Deshong. You all did such a great job aiding us with data collection and it was a pleasure working with you all.

Finally, I would like to thank my family. My wife Ashabi encouraged me to continue working toward my goal at achieving my doctorate's degree. She and our children Isabella and Abigail gave me the motivation to continue working hard to accomplish my goals. I also want to thank the newest addition to our family, my son Uriah. I hope my journey in academia encourages him to pursue whatever goals he wishes to accomplish as he matures. I also want to thank my parents and my siblings. I know my move to a new state was difficult on my family. I hope that upon completion of my doctorate's, we can make up for lost time. I want to thank you all for your support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Modern industrial systems are now fitted with several sensors for condition monitoring. This is advantageous because these sensors can provide mass amounts of data that have the potential for aiding in tasks such as fault detection, diagnosis, and prognostics. However, the information valuable for performing these tasks is often clouded in noise and must be mined from high-dimensional data structures. Therefore, this dissertation presents a data analytics framework for performing these condition monitoring tasks using high-dimensional data. To collect this data, the projects discussed utilize either simulated data sets or state-of-the-art, industry-class test-rigs. This enables the validation of data-analytics techniques on real-world data. Chapter 1 elaborates on this idea and outlines the proposed data-analytics framework for condition monitoring. Furthermore, Chapter 1 discusses various challenges with high-dimensional data and introduces the test-rigs used for data acquisition. Finally, Chapter 1 introduces the research topics in the dissertation.

Chapter 2 develops a severity-based diagnosis framework for electric-powered vehicle electric systems with multiple, interacting fault modes. This framework leverages functional data profiles collected from a 4-cylinder automotive engine test-rig. The profiles are collected for various combinations of battery-motor state-of-health. Three regularized multinomial regression models are fitted, each using a unique feature extraction technique, to map profile features to the fault mode combination. Then ensemble methods are proposed that merge the results of these three models to create a more accurate model. The space of fault-mode combinations is partitioned into degrees of overall degradation. Thus,

predictions of the fault-mode combination can be mapped to the overall system degradation.

Chapter 3 discusses two projects related to the gas turbine combustor. These projects utilize industry-class combustors located in the Ben T. Zinn Combustion Laboratory at Georgia Tech. The first project involves utilizing a control chart to detect precursors to lean blowout, an important operational fault that can cause costly powerplant outages and compromise the safety of aircraft passengers. A probability distribution is fitted to these precursors to develop a measure of blowout risk that can be utilized by turbine or aircraft operators. The second project describes a hierarchical feature selection methodology for diagnosing the degradation state of the combustor centerbody, a component that provides stability for the combustor flame and protects the hardware of the combustor from the flame. This methodology utilizes multi-class logistic regression with adaptive group lasso penalty to select an optimal subset of sensors for monitoring. Then multi-class logistic regression with adaptive lasso is used to select features from the remaining sensors to predict centerbody degradation. The results indicate that the proposed methodology is robust to reduction of the size of the training set.

Chapter 4 discusses two projects related to the gas turbine. These projects utilize an industry-class, single-stage gas turbine located at Pennsylvania State University. The first project proposes two approaches for predicting sealing effectiveness, the ability of the combination of turbine geometry and coolant system to prevent ingestion of hot gas into the wheelspace region. The first approach utilizes Linear Regression with Lasso to map features from time-resolved pressure signals recorded either near the rim seal or on the outer casing to the sealing effectiveness. In the second approach, information regarding the

nonlinear relationship between sealing effectiveness and the dominant frequency of the pressure signal was utilized to first make a coarse prediction of the sealing effectiveness. Then, given the coarse prediction, a more granular prediction is made. The result was a much more accurate diagnostic model when utilizing the two-step approach. The second project utilizes IR imaging of the gas turbine blade to diagnose faults in the cooling system. Various feature extraction techniques are analyzed for their predictive ability. Upon analyzing the IR image as a predictor, sparse regions corresponding to hot and cold regions of the turbine blade were found to be highly correlated with the cooling system faults. This demonstrated potential for optimal placement of more cost-effective sensors for real-world implementation.

Chapter 5 discusses a data analytics framework for predicting remaining useful life in systems with multiple failure modes. This framework assumes that the causes of the observed failures are unknown. Therefore, an unsupervised model is proposed. Given that the cause of failure is known, it is assumed that the natural logarithm of the time-to-failure can be modeled as a Gaussian regression problem where the predictors are features extracted from the sensor signals. Thus, the marginal distribution of the natural logarithm of the time-to-failure can be modeled as a finite mixture of Gaussian regressions. To fit the parameters of this model, the log-likelihood is appended with the adaptive group lasso penalty and optimized using a proposed Expectation-Maximization method. This method enables simultaneous clustering of the failures into separate failure modes and optimal sensor selection within each cluster. For prediction of the remaining useful life, a supervised mixture of gaussian model is fitted using the cluster labels and Multivariate Functional Principal Component (MFPC) scores from the selected sensors. Given the

diagnosed failure mode, a Linear Regression with Lasso model is fitted to map MFPC scores from sensors selected for the diagnosed failure mode to the natural logarithm of the time-to-failure. By subtracting the current observation time from the time-to-failure, the remaining useful life is predicted.

Chapter 6 concludes the dissertation by summarizing results and discussing opportunities for future research.

CHAPTER 1. INTRODUCTION

1.1 Background and Motivation

Modern industrial systems are comprised of several interdependent components and are subject to various failure processes. The complex nature of these systems necessitates a systematic framework for detecting anomalies, diagnosing the responsible fault(s), and predicting when the fault(s) will propagate to system failure. Condition monitoring is a process that utilizes sensor data to aid in performing these tasks. It is a key part of any predictive maintenance strategy aimed at reducing cost of downtime from either unnecessary maintenance activities or system failures. Modern systems are instrumented with hundreds, sometimes thousands of sensors. For example, the Siemens gas turbine SGT5-8000H shown in Figure 1.1 is equipped with 3,000 sensors. Several of these sensors are responsible for monitoring the gas turbine's state-of-health (SoH) and performance (Ratcliff et al., 2007).

Figure 1.1 – SGT5-8000H Gas Turbine (Admin, 2011)

Through these sensors, systems generate datasets with rich, highly informative data structures. However, traditional condition monitoring strategies are not equipped to handle the challenges presented by these data structures.

In this dissertation, High-dimensional (HD) data analytics methodologies are developed for real-time condition monitoring of complex industrial systems. By implementing these methodologies, operators can receive real-time updates regarding the system SoH. The foundation of this work is rooted in statistical modeling. Therefore, this work enables the migration of theoretical statistical models to the industrial sector through the demonstration of these models' effectiveness at solving system-relevant problems using real-world data. This data is generated by industrial-class test-rigs located at the Georgia Institute of Technology and Pennsylvania State University. These test-rigs are highly instrumented and thus, can produce data that capture system dynamics under various fault scenarios. Furthermore, this dissertation expands on the literature regarding sensor selection for prognostics by presenting an HD-data analytics framework for optimal selection of sensors for systems with multiple failure modes. The remainder of this chapter is organized as follows: Section 1.2 discusses the data challenges related to various condition monitoring. Section 1.3 describes the industrial systems used to motivate the methodologies proposed in this dissertation along with the test-rigs used for data acquisition. In Section 1.4, the proposed methodologies to address the challenges in Section 1.2 are outlined. The chapter concludes in Section 1.5 with a description of the organization of the dissertation.

1.2 Condition Monitoring and Data Challenges

In condition monitoring, sensor data (e.g. current, voltage, rpm, temperature, pressure, etc.) is collected from a system to assess its SoH. Condition monitoring tasks include anomaly detection, fault diagnosis, and prognostics. In this section, the challenges of using sensor data for these tasks are discussed.

1.2.1 Anomaly Detection

In the context of condition monitoring, anomaly detection consists of identifying when a system's performance deviates from its nominal behavior. Statistical process control (SPC) is a tool for anomaly detection that uses quality characteristics derived from statistics to monitor a process. SPC is usually performed in two phases. In Phase 1, parameters of a statistical distribution are estimated using data recorded while the system operates under nominal operating conditions. In Phase 2, the parameters are monitored in real-time and an alarm is emitted if the observed parameters breach a set of thresholds established in Phase 1.

Traditional SPC methodologies assume the observed data are independent and/or follow a Gaussian distribution with constant mean and variance. However, the sampling frequency of modern sensors is quite fast leading to highly autocorrelated observations. In Shewhart control charts, building a control chart with autocorrelated data can result in narrow control limits, leading to a high false alarm rate. Furthermore, complex processes exhibit nonstationary behavior such as trends and heteroskedasticity. An example is shown in Figure 1.2 where the fuel-to-air ratio in a gas turbine combustor is reduced simulating the process of throttle pullback in a jet engine.

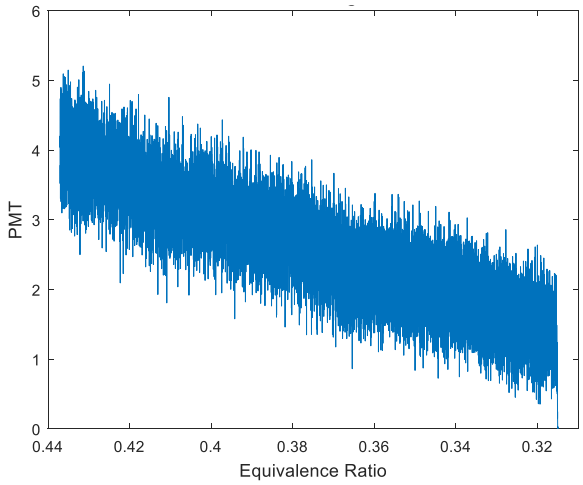


Figure 1.2 – Nonstationary Signal from Jet Engine Throttle Pullback

In Figure 1.2, PMT stands for photomultiplier tube, a sensing technology discussed in Chapter 3.1, whereas Equivalence Ratio is synonymous with fuel-to-air ratio. Obviously, assumptions of independence and Normality with constant mean and variance are violated for this profile. This motivates the need for a robust data curation algorithm that filters nonstationary behavior to make the profile suitable for anomaly detection.

1.2.2 Fault Diagnostics

Fault diagnosis consists of developing a mapping from the feature space to the fault space. These mappings are typically classification models although regression models are also applicable. When monitoring a process over time, the samples recorded by the sensors may resemble a mathematical function. For example, Figure 1.3 displays the electric current and voltage signals recorded during the cranking of a commercial vehicle engine.

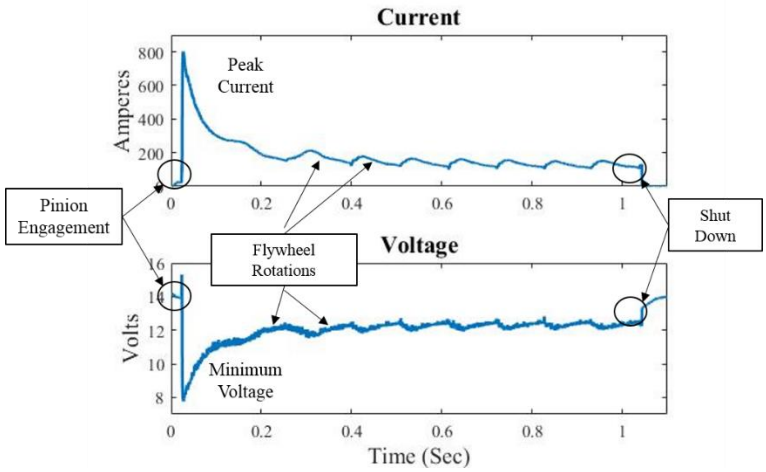


Figure 1.3 – Electric Current and Voltage of Vehicle-engine Start

Both signals capture the various steps of the Vehicle-engine Start (VES) process such as engagement between the starter motor and the engine flywheel, the draw of current by the motor, and the harmonic motion of the engine revolutions. When a fault occurs, it is expected to be captured by a deviation in the observed signal. One of the challenges with functional data is its high dimensionality. Each signal can consist of thousands of samples recorded over a short time span (1-2 seconds). Another challenge is that cycle-to-cycle variation in the starting position of system components can result in a misalignment of signals as shown in Figure 1.4.

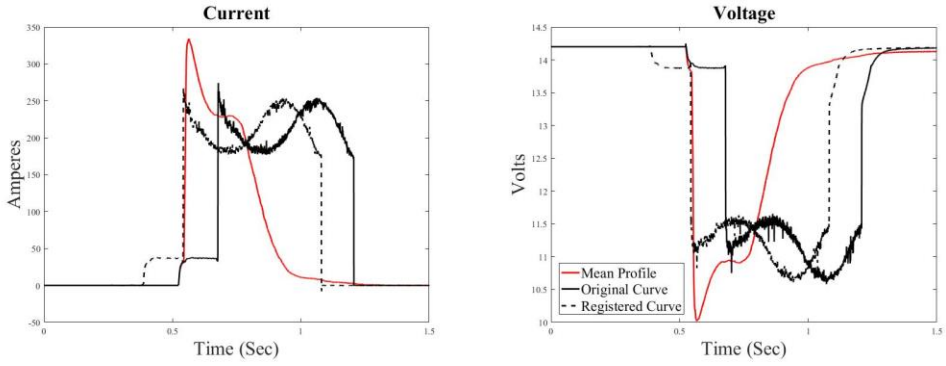


Figure 1.4 – Curve Misalignment

In Figure 1.4, the original signal (solid black line) is offset from the average signal profile (red). Misaligned features compromise diagnostic accuracy because the signal amplitudes at the same time epoch may correspond to different physical phenomena. Therefore, a robust feature engineering algorithm for functional data that accounts for misregistration is needed for accurate diagnosis.

Due to being comprised of various components, complex systems can experience multiple fault modes. For example, the engine can fail to start because of a faulted battery, a faulted starter motor, or due to the combined degradation of both components. The interaction between the component fault states impacts the fidelity of the system, and by extension, the signal profile generated by the system. For example, consider the plots in Figure 1.5 .

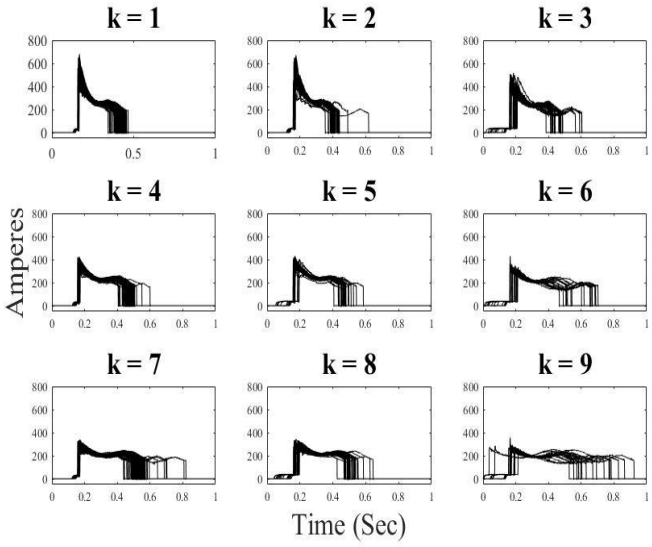


Figure 1.5 – Effect of Multiple Faults on System Performance

Battery degradation is shown from top to bottom whereas starter motor degradation is shown left to right. The ability to distinguish between combinations of fault modes is

crucial for performing diagnosis on complex systems. The challenge in achieving this distinction is compounded by the high dimensionality of the data.

In HD data scenarios, several sensors are utilized to monitor the SoH of a system. Consider the acoustic sensor data recorded during the monitoring of a gas turbine combustor flame in Figure 1.6.

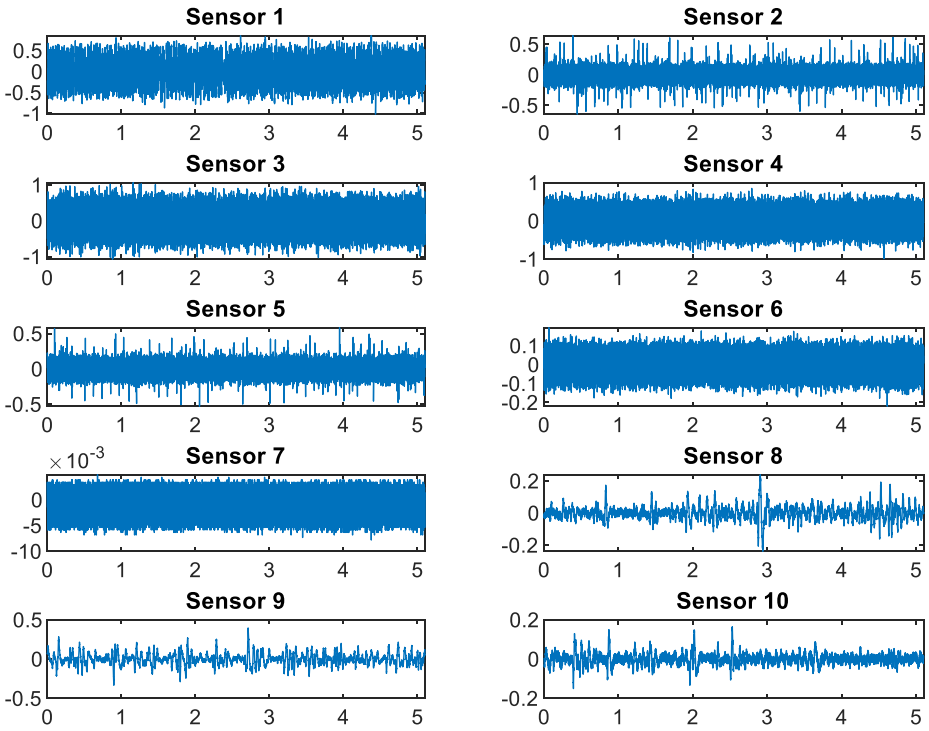


Figure 1.6 – Acoustic Sensor Data for Combustor Flame

Each observation consists of 10 sensors which record over 50,000 data points. The curse of dimensionality states that the number of observations needed to distinguish between fault states increases with the dimensionality of the data. Therefore, a very large number of observations would be needed to separate a dataset into distinct fault states. Furthermore,

the time to train a model would be prohibitively large. Another issue is that sensors may be redundant with each other or completely uncorrelated with degradation, contributing only noise to a prediction. By selecting a subset of informative sensors, it may be possible to improve diagnostic accuracy. Therefore, a framework for selecting an optimal subset of sensors for diagnosis is valuable for reducing the sensing load on large industrial systems.

1.2.3 Prognostics

Prognostics involves predicting the time remaining before the system fails. This time is known as the remaining useful life (RUL) of the system. Prognostics is important because the RUL predictions inform maintenance decisions designed to maximize availability of the system. As discussed previously, complex systems are comprised of multiple components. This means that complex systems suffer from multiple failure modes. However, developing a prognostics model for systems with multiple failure models is non-trivial because multiple failure modes affect the physical processes of the system in different ways. For example, consider the degradation data in Figure 1.7 for a commercial aircraft engine with two failure modes monitored by 21 sensors (Saxena et al., 2008).

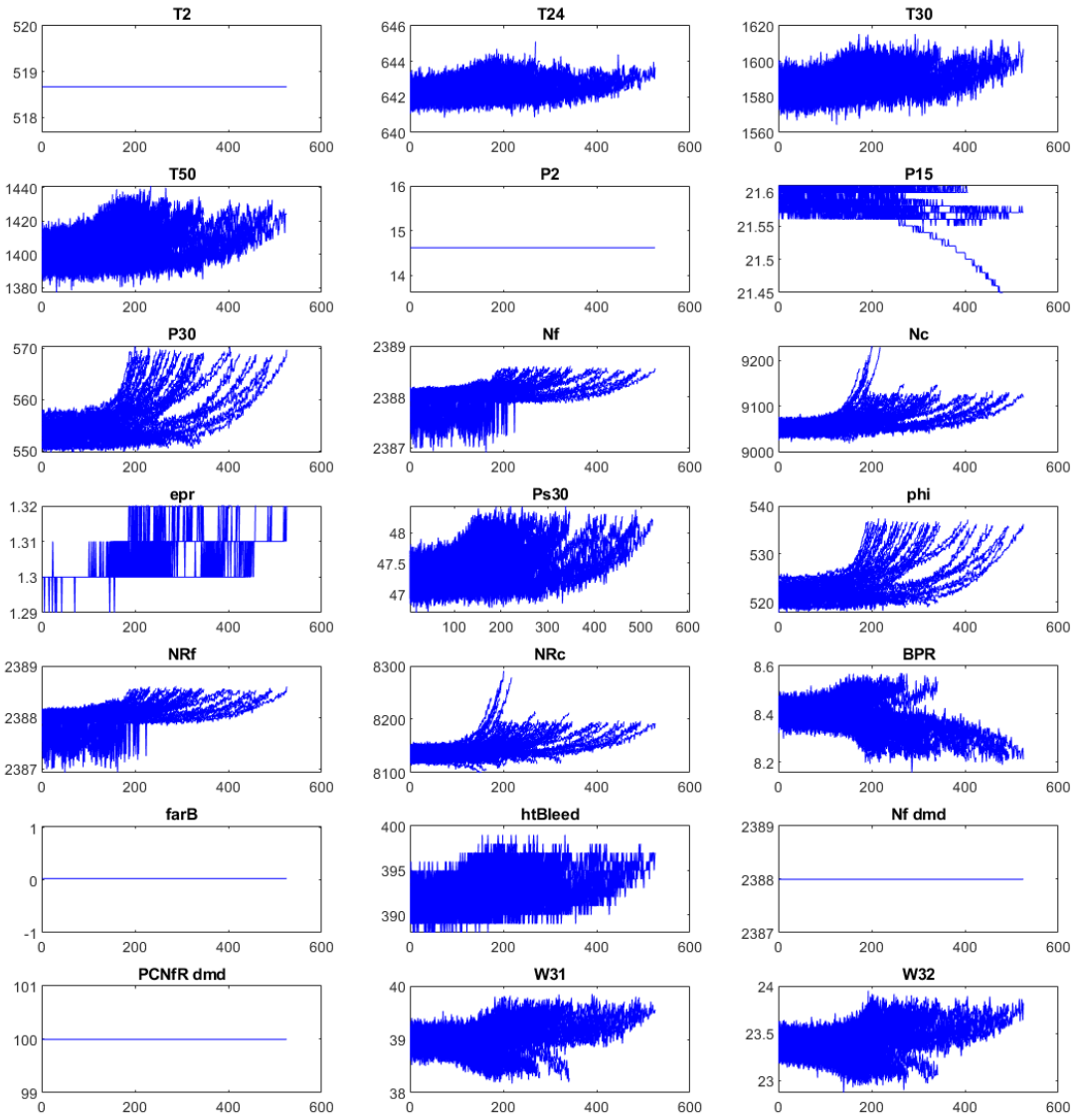


Figure 1.7 – Aircraft Engine Degradation (Saxena et al., 2008)

The effect of various failure modes can be seen by the different trends between signals of the same sensor. When the actual failure modes are not known, a methodology to distinguish between the latent failure modes is necessary. Furthermore, one subset of sensors may be sensitive to one failure mode while another (not mutually exclusive) subset may be sensitive to another failure mode. The ability to identify these subsets can aid in

reducing the sensor load required for accurate prognostics and in reducing RUL prediction error.

1.3 Examples of Complex Systems

The condition monitoring methodologies developed in this work are motivated by failure scenarios common to two industrial systems: the Vehicle-engine Start (VES) system and the gas turbine. In this section, the functionality of these two systems is described. Then, industry-class test-rigs used for generating data are highlighted.

1.3.1 VES System

The first system is the VES system, shown in Figure 1.8.

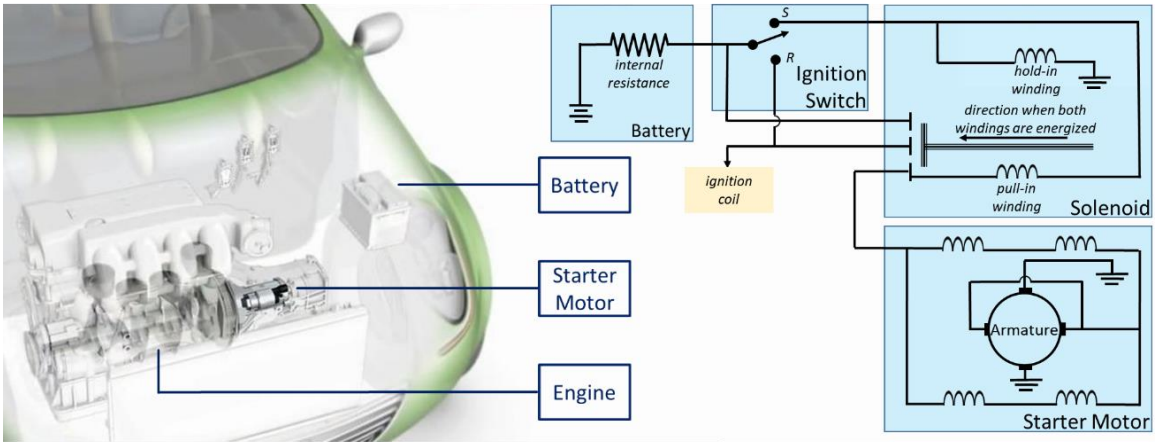


Figure 1.8 – VES System and its Circuit

This image is adapted from a video by Valeo United Kingdom and Republic of Ireland (2014, 1:11). The VES system is comprised of three main components that constitute an engine-start circuit: the battery, starter motor, and engine. When the vehicle key is turned, the ignition switch contacts the starting position (*S*) to close the circuit. This initiates

engine cranking by powering both hold-in and pull-in windings within the solenoid, creating a magnetic field that moves the plunger to connect the three terminals of the solenoid. Once the plunger is in place, hold-in winding remains energized while the pull-in winding becomes inactive. Through this connection, the starter motor receives power from the battery and turns the flywheel, which is connected to the engine. Once the cranking is complete, the ignition switch turns to *R*, thus disconnecting the battery from the motor.

1.3.2 Gas Turbine

The second industrial system utilized for the development of condition monitoring methodologies is the gas turbine. A diagram of the gas turbine is shown in Figure 1.9.

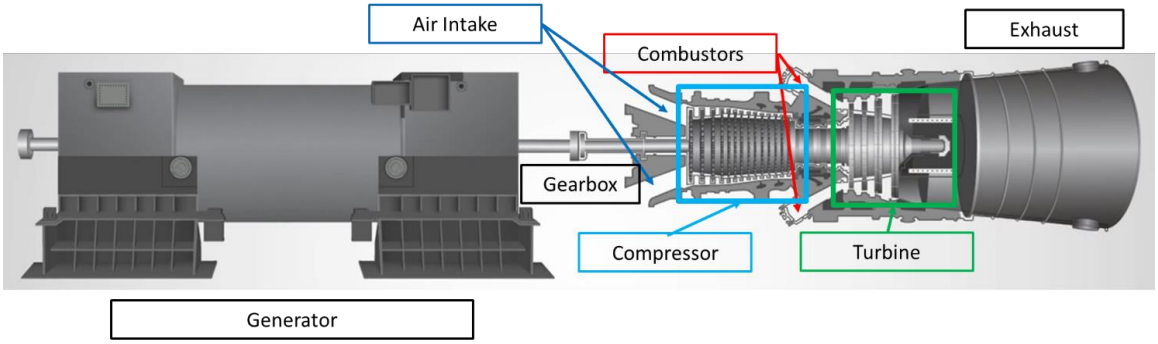


Figure 1.9 – Powerplant Generator Powered by Gas Turbine

The role of the gas turbine is to convert chemical energy from fuel to mechanical energy that is transferred to the generator by a gearbox. The generator uses this energy to create electrical energy that is distributed to cities and neighborhoods via the power grid. Gas turbines are also used to power aircraft engines. The key components of the gas turbine are the compressor, combustor, and turbine, which are labeled in Figure 1.9. Filtered air is

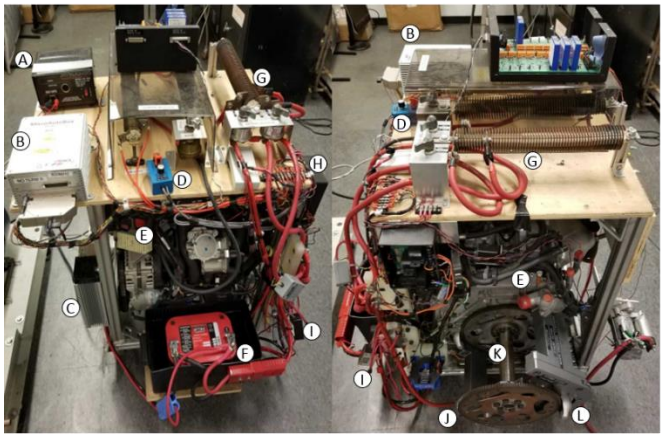
ingested through an air intake, from which the compressor funnels the air through multiple stages of static compressor vanes and rotating compressor blades to create a high-pressure stream. Through this stream, air is transferred to the combustor where it is mixed with fuel and ignited. The hot fuel-air mixture is then used to turn the turbine blades before exiting through the exhaust. The rotation of the turbine blades turns the drive shaft that links the turbine to the generator. For an aircraft engine, rotation of the turbine blades turns the propeller fan of the engine.

1.3.3 Test-rigs

Four industry-class test-rigs that simulate the systems described were utilized to develop condition monitoring methodologies utilizing HD-data analytics. These test-rigs include the VES Test-rig, Combustor Test-rig 1 and Combustor Test-rig 2, and the Turbine Test-rig. The VES and Combustor Test-rigs are located at the Georgia Institute of Technology and the Turbine Test-rig is located at Pennsylvania State University.

1.3.3.1 VES Test-rig

The VES Test-rig is displayed in Figure 1.10.



A. Power Supply
B. Control Unit
C. AC/DC Converter
D. Ignition Switch
E. Engine
F. 12-Volt Battery
G. Resistor
H. PLC Board
I. Voltmeter
J. Flywheel
K. Crankshaft
L. Motor

Figure 1.10 – VES Test-rig

It consists of a 4-cylinder engine equipped with an ammeter, voltmeter, thermocouples, and rpm sensor. The engine is powered by the VES circuit comprised of a 12-volt battery in series with a starter motor. A coiled resistor is placed between this circuit to simulate the increase of internal resistance incurred due to battery degradation. Through repeated cranking of the engine, the test-rig can perform accelerated life testing on the starter motor.

The sensors attached to the test-rig sample data at a rate of ~2 kHz. Thus, the system can generate functional data over the span of an engine cranking cycle. For real vehicles, the starter motor cranks the engine to a certain speed before the fuel source takes over as the primary supplier of energy for the engine. Since the test-rig does not contain a fuel source, the circuit is programmed to disengage the starter motor from the engine once a specified angular speed threshold is reached.

1.3.3.2 Combustor Test-rigs

Two test-rigs are utilized for studying the combustor. These are displayed in Figure 1.11.

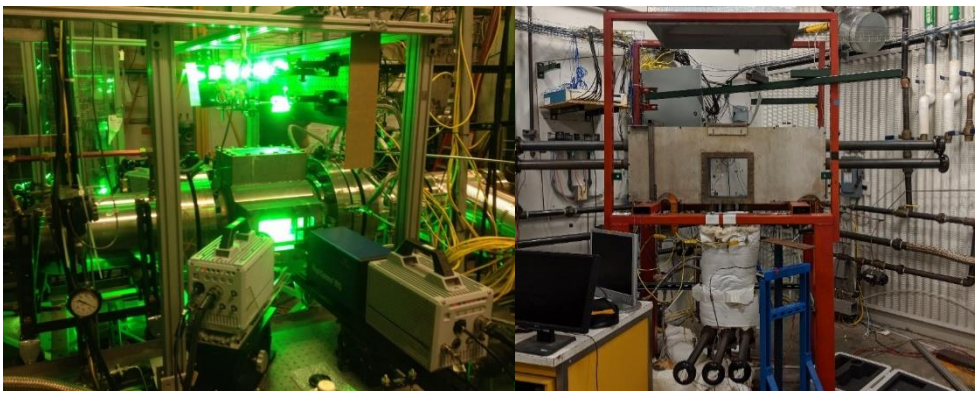


Figure 1.11 – Combustor Test-rigs (Left: Combustor Test-rig 1, Right: Combustor Test-rig 2)

They are located inside the Ben T. Zinn Combustion Laboratory, an 18,000 square foot facility that specializes in combustion experiments. The test-rigs are equipped with compressed and heated air, various fuel types, fresh and recirculating water, and exhaust hoods. They are capable of replicating combustor faults such as lean blowout, a process where the flame is extinguished due to low fuel-to-air ratio. The lab is also capable of outsourcing the manufacture of degraded combustor components such as the centerbody, which serves to stabilize the flame and prevent it from damaging combustor hardware.

To monitor the combustion flame, the lab utilizes a photomultiplier tube and acoustic sensors. The photomultiplier tube aggregates the light intensity in its field of view into a singleton point and has a sampling rate of 10 kHz. The lab also utilizes ten acoustic sensors, each with 40 kHz sampling frequency to monitor combustion. Through this sensing technology, the test-rig can simulate a multiple-sensor environment for monitoring combustor SoH.

1.3.3.3 Turbine Test-rig

The Turbine Test-rig is located in the Steady Thermal Aero Research Turbine (START) Lab at Pennsylvania State University. A photograph of the lab is displayed in Figure 1.12.

Figure 1.12 – Turbine Test-rig in START Lab

The Turbine Test-rig is a single-stage (stator vane/turbine blade) gas turbine consisting of two air compressors that convert ambient air into a high-pressure stream that is either cooled and distributed to various coolant holes in the system or has its temperature raised by the heater to simulate the effect of combustion. The heated air is then used to turn the turbine blade before passing through the exhaust and exiting the system. This test-rig is unique, not only due to its capability to operate at Reynolds and Mach numbers representative of real-world gas turbine engines, but because it is instrumented with several

pressure and temperature sensors for real-time monitoring of the system. In addition to these sensors, infrared imaging is employed to study the effect of upstream faults on the temperature distribution of the blade. These upstream faults are simulated by altering flow rates and temperatures from the compressors and heaters. Due to the various passages for cool air available in the system, the robustness of a condition monitoring methodology to different operating conditions can be tested.

1.4 Research Overview

In this section, a brief overview of all research presented in this dissertation is provided. This research was motivated by current challenges related to power generation systems.

1.4.1 Condition Monitoring of the VES System

1.4.1.1 Severity-based Diagnosis for Vehicular Electric Systems with Multiple, Interacting Fault Modes

Complex systems are comprised of multiple components that interact in how they degrade and fail. Diagnosing active faults and their severity for these systems is a non-trivial task. Therefore, a data-driven, severity-based diagnosis framework for systems with multiple, interacting failure modes is proposed. The framework is applied to components of the VES system: the battery and the start-stop motor, a modern starter motor used to restart vehicle engines following brief traffic stops. This framework leverages sensor data from several component-fault severity combinations. A functional data analysis approach is utilized to align the features of the signal profiles for these combinations. Then multiple

feature extraction methodologies are employed, and the features are used to fit separate classifiers using Regularized Multinomial Regression. Ensemble methods are then utilized to combine the inferences of the classifiers for higher diagnosis accuracy. Following classification, the component-fault severity combinations are mapped to the system SoH via a lookup table that relates the component fault states to system performance.

1.4.2 Condition Monitoring of the Combustor

1.4.2.1 Data Analytics Method for Detecting Extinction Precursors to Lean Blowout in Spray Flames

Lean blowout (LBO) of the combustion flame results in costly outages for power plants and presents a serious risk to the safety of aircraft passengers. LBO occurs when the fuel-to-air ratio in the combustion mixture becomes too low that the flame is extinguished. Avoiding this by using a high fuel-to-air ratio is not a viable option as doing so wastes fuel and contributes to emissions that harm the environment. Ideally, the engine maintains the combustor flame at a fuel-to-air ratio low enough to minimize emission of toxins, but high enough to avoid blowout.

In the LBO literature, the occurrence of extinction/reignition phenomena has been observed prior to blowout. However, methods of characterizing these phenomena have not been universally accepted. In this work, a data analytics approach to detecting precursors to LBO is developed. This approach consists of three steps: data curation, fault detection, and an adaptive alarm reliability assessment. In the data curation step, a nonstationary signal used to monitor the flame as the fuel-to-air ratio is reduced is filtered using a time series model. The variability of the model residuals is then monitored using an

Exponentially Weighted Root-Mean Squared Error (EWRMS) control chart. As the engine progresses toward blowout, the parameters of the time series model are likely to change, eliciting an alarm if the change is large enough. The goal is to imbue each observed alarm with a value denoting the probability that the alarm is a precursor to blowout. Since the frequency of alarms increases as the flame approaches blowout, the fuel-to-air ratio at the time of an alarm is modeled using a 2-parameter Exponential distribution. After a new alarm is observed, the probability of a true alarm is updated by conditioning on the fuel-to-air ratio measured at the time of the alarm. As the flame approaches blowout, the true alarm probability increases. Therefore, operators can utilize this true alarm probability to gauge the safety of operating the engine at a particular fuel-to-air ratio.

1.4.2.2 Data-driven Fault Detection of Premixer Centerbody Degradation in a Swirl Combustor

In swirl combustors, the centerbody is an important component that provides both stability for the combustor flame and protection of the combustor hardware from flame backpropagation. Due to its proximity to the flame, the centerbody incurs thermal wear. Within the combustion community, the ability to detect degradation of combustion components is limited to visual inspection of the Fourier transform of the raw sensor signal. To encourage an HD data analytics approach to solving this problem, ten acoustic sensors were placed around Combustor Test-rig 2. Four centerbodies were manufactured to various lengths with shorter length representing increased degradation. To diagnose the SoH of the centerbody, features from each sensor signal were extracted using the discrete wavelet transform, which transforms signals in the time domain to signals in the time-frequency domain. Features were engineered by computing the energy within each wavelet

decomposition level. After generating features, a hierarchical approach to selecting features was utilized. The first step involves selecting an optimal subset of acoustic sensors by fitting a Multinomial Regression model with an Adaptive Group Lasso penalty. After removing either the redundant or noninformative sensors, a Multinomial Regression model with Adaptive Lasso penalty is used to select an optimal subset of features from the remaining sensors. This model maps the features selected to one of four degradation classes. The model approach was shown to be robust to decreases in the size of the training data and to additional noise in the acoustic signals.

1.4.3 Condition Monitoring of the Turbine

1.4.3.1 Correlating Time-resolved Pressure Measurements with Sealing Effectiveness for Real-time Turbine Health Monitoring

A hot fuel-air mixture from the combustor is used to rotate the turbine blades and power the system. This is shown in the cross-sectional view of the Turbine Test-rig in Figure 1.13.

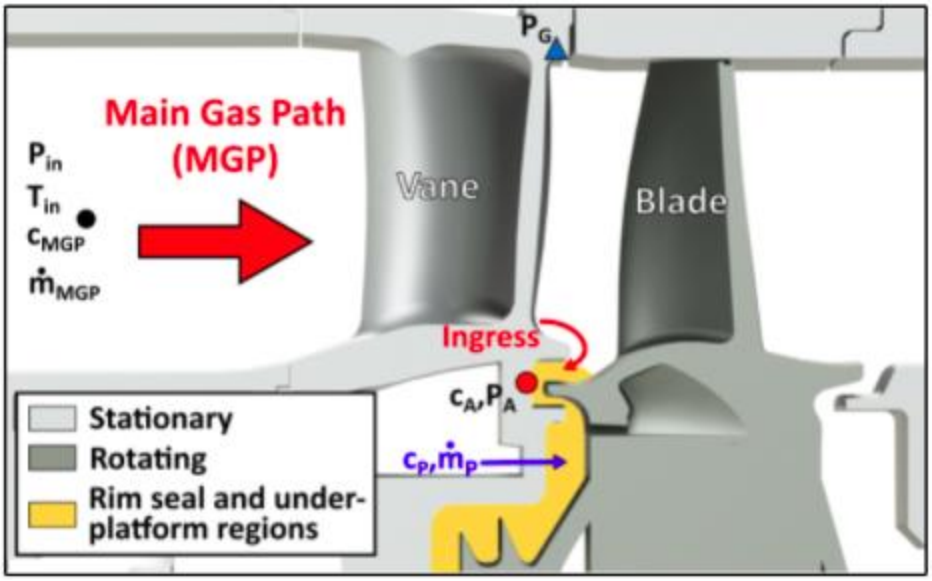


Figure 1.13 – Cross-section Diagram of the Turbine Test-rig

The temperatures for this mixture are very high and can exceed the softening temperature of the turbine components. While the components directly exposed to the heat have cooling mechanisms installed to protect them from the hot mixture, the same is not true for components in the under-platform region. To counter this, the combination of a manufactured rim seal and relatively cool purge flow from the upstream compressor (blue arrow) is utilized to form a seal that protects these components. The ability to measure the effectiveness of the seal from preventing ingress of the hot fuel-air mixture is a challenge in the gas turbine community. The literature regarding sealing effectiveness has focused on the development of physics-based models that attempt to characterize the functional relationship between sealing effectiveness and purge flow rate. However, the applicability of these models is limited since these models do not account for regions of inflection that occur due to the variability in rim seal geometries.

To address the limitations of the physics-based approaches, two data-driven methodologies for monitoring sealing effectiveness are proposed. The first methodology consists of recording the pressure, either near the rim seal or on the outer casing of the turbine, for various sealing effectiveness values. Then, features are extracted from the Discrete Fourier Transform of the pressure signal. These features are mapped to the sealing effectiveness via Linear Regression with Lasso, which selects a subset of the extracted features for predicting sealing effectiveness. The second methodology utilizes domain knowledge about a nonlinear relationship between the sealing effectiveness and the dominant frequency of the pressure. For this methodology, an initial model is fitted to obtain a coarse prediction of the sealing effectiveness. Logistic Regression with Lasso is used to map the extracted features to a binary SoH for the seal (either low or high). Then, a Linear Regression model with Lasso is fitted using only the first five harmonics of the dominant frequency to perform a more granular prediction of sealing effectiveness. This model includes a binary variable corresponding to the coarse SoH of the seal. Through this second methodology, a sealing effectiveness prediction error reduced by 48.1% and 35.4% from the first methodology for the rim seal and the casing respectively.

1.4.3.2 Applying Infrared Thermography as a Method for On-line Monitoring of Turbine Blade Coolant Flow

Turbine blades are more susceptible to failure when subjected to high temperatures. To address this, modern turbine blades are manufactured with cooling holes through which cold air from the compressor passes to regulate the temperature of the blade. Therefore, the flow of cold air from the compressor to the blade is important to monitor. Infrared imaging has recently been implemented for monitoring loss of thermal barrier coating on gas turbine

blades. However, the focus of this work is to use infrared imaging of the blade to infer the SoH of the cooling systems. The challenge is that the effects of changes in operating conditions such as increased main gas path temperature on the image can conflate with the changes due to cooling loss. To address this, infrared images for various combinations of main gas path temperatures and coolant flow rates are sampled. Three methods of feature extraction are utilized: the overall average intensity of the image, column-wise averages of the image, and the image itself. For each of these methodologies, Linear Regression with Lasso is used to model the coolant flow rate as a function of the image features.

The effectiveness of each of these feature extraction techniques in estimating the coolant flow rate is compared for the cases when the main gas path temperature is known and when it is unknown. For the case that the main gas path temperature is unknown, Linear Regression with Lasso is used to model the main gas path temperature as a function of the image features. Since the correlation between the coolant flow rate and the image features is obscured by latent factors that presented themselves during the experiment, models are fitted for the original dataset and an ideal dataset consisting of a subset of observations where the effect of the latent factors is negligible. To account for these latent factors in the original dataset, ancillary measurements are included as predictors in the model. The results demonstrate that inclusion of the ancillary measurements can improve accuracy in predicting coolant flow rate. However, the accuracy of the model using all pixels in the image is only marginally lower than the model using all pixels and the ancillary measurements. Furthermore, only marginal improvement in prediction accuracy is observed for the model using all image pixels exclusively when increasing the number of pixels selected past four. Through analysis of the regression coefficients for the optimal

model, the selected features are shown to aggregate in sparse regions of the image. This indicates potential for achieving accurate prediction of coolant flow rate by placing sensors in these sparse regions.

1.4.4 Predicting Remaining Useful Life in Complex Systems with Multiple Failure Modes

Predicting remaining useful life (RUL) in systems with multiple failure modes is an important, but nontrivial task. The literature pertaining to multiple failure modes is concentrated around developing reliability models for competing risk problems, where competing risks are alternative failures that can precede the primary failure mode of interest. These models tend to be validated with little to no sensor data. The papers that incorporate multiple sensors tend to either not incorporate sensor selection or base their sensor selection on visual inspection of the sensor signals.

This chapter proposes an HD data analytics approach to performing prognostics for systems with multiple failure modes. The premise of this approach is that the functional relationship between the natural logarithm of the time-to-failure ($\ln TTF$) and features from the sensor data is dependent on the failure mode and can be modeled as a Gaussian regression. Therefore, multiple failure modes are modeled using a mixture of Gaussian regressions. An added challenge is that the failure modes may not be known a priori. Therefore, to fit mixture of Gaussian regression model, a novel Expectation-Maximization (EM) algorithm is developed. In the E-step, the observed failures are clustered into their predicted failure modes. In the M-step, the parameters of the mixture of Gaussian regressions model are estimated. For parameter estimation, Adaptive Sparse Group Lasso penalty is incorporated into the optimization problem. Adaptive Sparse Group Lasso is a

regularization strategy that ensures a subset of features from a significant sensor have nonzero regression coefficients, while all features from a non-significant sensor have all zero regression coefficients. The result of the EM algorithm is that the observed failures are clustered by failure mode and subsets of sensors are selected to monitor degradation for each failure mode. The signals for the selected sensors are then fused using Multivariate Functional Principal Component Analysis, a functional data analysis technique that transforms the sensor signals into a low-dimensional, but highly informative set of predictors. MFPCA is first used on all selected sensors to train a supervised mixture of Gaussians model. After monitoring a currently degrading system up to time t^*, the selected sensors' signals are transformed to MFPC scores and used to diagnose the system fault. Given this diagnosis, Linear Regression with Lasso penalty is used to model $\ln TTF$ as a linear function of the MPFC scores extracted from the signals corresponding to sensors selected for the diagnosed failure mode. Since not all MFPC scores may be correlated with the $\ln TTF$, the Lasso penalty can remove uninformative MFPC scores from the model. This model is then used to predict RUL by calculating the difference between the predicted time-to-failure $(\exp(\ln TTF))$ and the current observed time t^*.

This methodology is applied to a commercial aircraft engine simulation dataset with two failure modes: high pressure compressor and the fan. A comparison of the proposed model with alternative models demonstrates that clustering and sensor selection are important steps to ensure an accurate RUL prediction. Furthermore, the methodology performs comparably to an existing statistics-based modeling approach for multiple failure modes in the literature.

1.5 Organization of Dissertation

The remainder of the dissertation is organized as follows. In Chapter 2, the development of a severity-based multiple-fault diagnosis algorithm for estimating the SoH of the VES system is described. In Chapter 3, two methodologies are presented for condition monitoring of the combustor. These methodologies are designed to solve problems related to detecting early onset of LBO and selecting sensors for diagnosing centerbody degradation. In Chapter 4, two methodologies are presented for condition monitoring of the gas turbine. These methodologies revolve around developing statistical models to estimate sealing effectiveness (Section 4.1) and to estimate SoH of the blade cooling system. In Chapter 5, the methodology for predicting RUL in systems with multiple failure modes is proposed. Like in Section 3.2, this methodology employs a sensor selection strategy to aid in condition monitoring. The dissertation is concluded in Chapter 6 with a summary of the dissertation and a discussion of some best practices.

CHAPTER 2. SEVERITY-BASED DIAGNOSIS FOR VEHICULAR ELECTRIC SYSTEMS WITH MULTIPLE, INTERACTING FAULT MODES

2.1 Introduction

The reliability of engineering systems depends on the health states of their constituent components. For example, automotive electric power generation and storage (EPGS) systems, which are used for starting the automobile engine, depend on the health states of the alternator, battery, and starter motor. It is customary to represent the component states of the system as either healthy or faulted. This binary representation is sometimes not sufficient because it does not capture the severity of the component faults. Due to complex interactions between various system components, system failure will most likely depend on the combinations of component-fault severities. Thus, a more precise characterization of system failure can be obtained by modeling component health states and their interactions. Identifying fault severity and its impact on system performance also enables life extension and limits unnecessary part replacements.

This chapter presents methodology for performing severity-based diagnosis of systems with multiple interacting fault modes. One of the main contributions of this methodology is the ability to diagnose the presence and severity of multiple, interacting fault modes in applications involving high-dimensional data. The proposed modeling framework identifies the state-of-health (SoH) of system components and how interactions between their fault severity (degradation) levels impact overall system performance. This

approach is applicable to conventional time series data and high-dimensional data structures, e.g. spectral and profile data, hereafter referred to as functional data. Functional data analysis tools such as curve registration and dimensionality reduction are used for automatic synthesis and extraction of degradation features from the data. Degradation in high-dimensional functional data is unique because it can manifest itself in different ways ranging from rising frequency amplitudes to complex phase shifts. In this chapter, the proposed methodology is applied to high-dimensional data generated by a Vehicular Engine Start (VES) test-rig. This test-rig enables the acquisition of information-rich data profiles for a vast array of component-fault states. Through the proposed methodology, the utility of these profiles for diagnosis is demonstrated. The importance of developing diagnostic frameworks for such systems stems from the fact that approximately 88% of Americans own an automobile (Poushter, 2015). Thus, driver dissatisfaction, walk-home incidents, and potential accidents that occur due to EPGS faults are serious concerns for a vast majority of people.

The remainder of the chapter is outlined as follows. Section 2.2 provides a review of the literature on multiple fault diagnosis. In Section 2.3, the diagnosis methodology is described in detail. In Section 2.4, the capabilities of the VES Test-rig are discussed. In Section 2.5, the proposed methodology is demonstrated using the VES Test-rig as a case study. The chapter concludes in Section 2.6 with a summary of the findings, a description of the challenges involved in implementing the methodology, and a discussion regarding potential for future work.

2.2 Literature Review

Approaches to fault diagnosis are categorized as either model-based or data-driven. Model-based approaches, which rely on physical models of the system, characterize faults as deviations of observed system outputs from the outputs predicted by the physical model. Stringer et al (2012) provide a review of model-based diagnosis approaches with a focus on aerospace transmission systems. The effectiveness of model-based approaches depends on the integrity of the physical model. However, developing this model becomes challenging as systems become more complex. In contrast, data-driven approaches characterize faults via their relationship with sensor data. Jardine et al (2006) classify data-driven approaches into two categories, artificial intelligence (AI) and statistical learning models. Modern AI approaches for diagnosis revolve around deep learning architectures such as convolutional neural networks (Chen et al, 2017, Hoang et al, 2017, Wu et al, 2019), deep belief networks (Tamilselvan and Wang, 2013), and deep autoencoders (Liu et al, 2018). Alternatively, statistical learning methods are comprised of tools such as clustering techniques (Ghimire et al, 2011) and support vector machines (SVM) (Islam and Kim, 2019). Whether to use a model-based or data driven approach is a subjective choice that can depend heavily on the application. What truly distinguishes diagnosis approaches from one another is the level of complexity of the system they are designed to diagnose. In general, diagnosis approaches are designed for the following applications: single fault, multiple mutually exclusive faults, and multiple simultaneous faults.

The literature is rich with diagnosis approaches for single fault applications. For example, Zhang, Y. et al (2010) develop a model-based approach to detect belt slip in the automotive EPGS system. They establish that for a given set-point voltage, the battery

voltage and field duty cycle are linearly related. They detect belt slip by monitoring the distance of battery voltage and field duty cycle from the hypothetical line defining their relationship. From the data driven perspective, Liu et al (2018) use deep autoencoders to fuse time, frequency, and time-frequency features used for classifying gear crack length under various operating conditions. As the complexity of industrial systems grows, so does the need for diagnosis methodologies that model multiple faults. Multiple Fault Diagnosis (MFD) is a popular research area geared toward these complex systems. One framework for MFD assumes that faults are mutually exclusive. Zhang, F. et al (2014) diagnose faults in rotating machinery by selecting the fault state with minimum Kullback-Leibler divergence between the distribution of features within that fault state and the distribution of those features combined with the test observation. Li et al (2008) and Scacchioli et al (2014) both develop physics-based models for diagnosing mutually exclusive faults in an EPGS system. Li et al (2008) utilize an alternator simulation model to diagnose belt-slip, rectifier diode, and voltage regulator faults, while Scacchioli et al (2014) develop physics models for the alternator, battery, and voltage regulator to create a set of residuals used to diagnose belt-slip, diode short, and regulator faults. Kazemi et al (2019) incorporate a hybrid model-based and data-driven approach for diagnosing multiple faults in a gas pressure regulating station. Principal component analysis (PCA) is applied to data acquired from a real station to define a low-dimensional subspace for the normal operating condition. Then, fault data acquired from both the real system and their model is projected onto this subspace and SVM is used to classify the various fault states. In a data-driven approach, Zhang, X. et al (2010) use principal component analysis to develop a set of residuals used to diagnose the belt-slip, diode short, and regulator faults.

The main limitation of the mutually exclusive assumption is that it assumes the presence of only one fault at a time, which can lead to missed detections if multiple faults are present simultaneously. Methodologies developed for diagnosing the presence of multiple faults include work by Islam and Kim (2019), Wu et al (2019), Jaramillo et al (2017), and Kodali et al (2013). Islam and Kim diagnose the presence of multiple faults on bearings using multi-class SVM, whereas Wu et al (2019) use a convolutional neural network to diagnose various faults in helical and planetary gearboxes. Jaramillo et al (2017) use a two-stage data fusion technique via Bayesian inference to estimate the most likely fault condition present in the system given the states of diagnostic tests. Kodali et al (2013) uses the Dynamic Multiple Fault Diagnosis (DMFD) algorithm, which employs a Factorial Hidden Markov Model to estimate the most likely fault sequence given the state of observed diagnostic tests, to diagnose multiple faults in an EPGS system. Another application of DMFD is for diagnosing multiple faults in the regenerative braking system of a hybrid electric vehicle (Sankavaram et al, 2014). In the application to the EPGS, Kodali et al (2013) list eleven faults related to the alternator, battery, and control units regulating the system outputs. They develop twelve diagnostic tests and demonstrate their relationship with the faults in a table referred to as the D-matrix. The derived D-matrix serves as an input to the DMFD algorithm. While the D-matrix is often associated with the DMFD and its various extensions, Yan et al (2014) integrated the D-matrix with PCA to diagnose multiple faults in air handling units of Heating Ventilation and Air Conditioning (HVAC) systems. The work in this chapter differs from these methodologies in that it incorporates a granular representation of the fault state instead of a binary representation. It also accounts for how various combinations of component-fault states interact to impact system

performance. This is possible due to the ability to collect data for a dense mesh of component-fault severity combinations via the VES Test-rig. While the objective of many MFD methodologies is to determine the optimal set of component-fault states, the proposed methodology expands on this by relating the inferred fault states to the system operability. Thus, the importance of each fault state is comparable. In the next section, a detailed explanation of the diagnosis methodology is provided.

2.3 Methodology for Diagnosing State-of-health of Systems with Multiple, Interacting Fault Modes

In this section, a methodology for diagnosing the SoH of systems with multiple, interacting fault modes is developed. This methodology considers complex systems where the severity of component faults has a direct impact on system performance. When multiple faults are present, each with their own respective severity level, their combined effect on system performance manifests itself in profiles acquired from sensors monitoring the system. This approach revolves around using these profiles to estimate the severity level of each component fault and mapping the estimated component-fault severity combination to the system SoH.

The proposed methodology is comprised of an offline stage and an online stage. The offline stage has two principal operations: model fitting and system health-state definition. In model fitting, multi-dimensional profile data is acquired for several component-fault severity combinations and class labels are assigned to these combinations. The data is partitioned into two sets: a training set and a validation set. The training data is used to train a classifier that maps features extracted from the profiles to the set of class labels. To

extract features, a two-step approach is employed. First, curve registration is used to align the salient characteristics of the profiles. In this step, a target function is defined to register all other profiles. Then, the following transforms are applied to extract fault-based features from the registered profiles: Functional Principal Component Analysis (FPCA), Short-time Fourier Transform (STFT), and Discrete Wavelet Transform (DWT). For each of these statistical transformations, an individual Regularized Multinomial Regression (RMR) model is trained. These RMR models are used to perform classification on the observations from the validation set. Several ensemble methods are developed using the classification accuracy of the RMR models on the validation set to improve the classification accuracy of the individual RMR models.

The second operation in the offline stage is to define the health states for the overall system. First, a system health indicator is identified. Next, hierarchical clustering is employed to cluster the component-fault severity combinations according to the system health indicator. To define the system health states, the clusters are ranked by their centroid value. In the online stage, new observations are sampled from the present state of the system. After curve registration and feature extraction, the features of the observations are passed through the three RMR models. Then, ensemble methods are used to combine the outputs of the three models. To classify the observations, an aggregate form of the traditional maximum a posteriori (MAP) criterion is employed. Since the classes refer to combinations of component-fault severity states, performing classification is the same as estimating the health state of the individual components simultaneously. Once the component-fault severity combination is estimated, the system SoH is inferred by referencing the ranking of the cluster to which the combination belongs.

2.3.1 Offline Stage – Model Fitting

Consider a system with C interdependent components. Let $k_c \in \{0, 1, \dots, K_c\}$ denote the fault state of component c, ordered in increasing severity where $k_c = 0$ indicates no fault. Let $\boldsymbol{k} = [k_1, \dots, k_C]'$ denote the component-fault severity combination. This vector characterizes the system SoH in terms of the SoH of each component in the system. The number of possible states that $\boldsymbol{k}$ can assume is $K = \prod_{c=1}^{C}(K_c + 1)$. Thus, an integer-valued class label $\mathcal{k} \in \{1,2, \dots, K\}$ is assigned to each state. For example, $\mathcal{k} = 1$ when $\boldsymbol{k} = [0, \dots,0]'$ and $\mathcal{k} = K$ when $\boldsymbol{k} = [K_1, \dots, K_C]'$. A supervised learning approach is used to train the models. First, a training set comprised of multi-dimensional profiles from each class is acquired. Suppose the system is monitored by P sensors. For the discrete time domain $\mathcal{T} = \{0,1, \dots, T-1\}$, the nth profile observation at time $t \in \mathcal{T}$ is denoted by $\boldsymbol{s}_n[t] = [s_n^1[t], \dots, \boldsymbol{s}_n^P[t]], n = 1, \dots, N$. Let $y_n \in \{1,2, \dots, K\}$ denote the class label for the nth observation. The goal is to train a model that receives the multi-dimensional profiles as input and outputs a class label. However, profile data often suffers from misregistration and high dimensionality. To address these issues, a two-step approach for feature extraction is proposed.

2.3.1.1 Feature Extraction – Curve Registration

As faults propagate, profiles used for monitoring the system grow out of phase. Misregistered profiles are not comparable because their features at a specified time represent different system dynamics. Furthermore, misregistration obscures the correlation between profile features and faults rendering the extraction of informative features difficult. Curve registration is proposed to correct misregistration. Curve registration aligns

the salient features of profiles through transformations of time t called the warping functions. To perform curve registration for sensor $p, p = 1, \dots, P$, a target function s_*^p is defined and all other profiles are transformed so that they are aligned to this target function. For profile s_n^p, the warping function h_n^p is a monotonically increasing function of time t that is equal to t at the boundaries of $\mathcal{T}$. Registration of profile s_n^p is performed by evaluating $s_n^p(t)$ at $t = h_n^p(t)$.

The warping function is defined by the following general formulation by Ramsay and Silverman (2005).

$$h_n^p(t) = D_0^{n,p} + D_1^{n,p} \int_0^t \exp W^{n,p}(u)\, du\,, n = 1, \dots, N \tag{2.1}$$

where $D_0^{n,p} = 0$ and $D_1^{n,p} = \frac{T-1}{\int_0^{T-1} \exp W^{n,p}(u) du}$ because of the properties of h_n^p. $W^{n,p}$ is an unconstrained function that specifies the warping function. Since $W^{n,p}$ is unconstrained, it is represented using the following linear expansion $W^{n,p} = \sum_{r=1}^{R} w_r^{n,p} B_r$, where $B_r, r = 1, \dots, R$ are basis functions. To register s_n^p to the target s_*^p the value of $\boldsymbol{w}^{n,p} = (w_1^n, \dots, w_R^n)$ that minimizes the misregistration between the two profiles is determined. According to Ramsay and Silverman (2005), the misregistration is characterized using Principal Component Analysis (PCA). Since two profiles perfectly in phase form a straight line when plotted against each other, the result of PCA on these two profiles yields a minimum eigenvalue of zero. Therefore, the degree of misregistration can be characterized as the departure from unidimensionality, which is signified by a nonzero minimum eigenvalue. To perform curve registration, define the target function for sensor $p, p = 1, \dots, P$ as the

average over all signals, $s_*^p(t) \coloneqq N^{-1}\sum_{n=1}^{N} s_n^p(t)$. For observation n from sensor p, the curve registration algorithm is initialized by setting $\boldsymbol{w}^{n,p} = \mathbf{0}$ and computing the cross-product matrix:

$$\mathbb{T}_n^p\left(h_n^p\right) = \begin{bmatrix} \int \left\{s_*^p(t)\right\}^2 dt & \int \left\{s_*^p(t) s_n^p\left(h_n^p(t)\right)\right\} dt \\ \int \left\{s_*^p(t) s_n^p\left(h_n^p(t)\right)\right\} dt & \int \left\{s_n^p\left(h_n^p(t)\right)\right\}^2 dt \end{bmatrix} \tag{2.2}$$

Misregistration for sensor $p, p = 1, \ldots, P$ is computed as the minimum eigenvalue of $\mathbb{T}_n^p\left(h_n^p(t)\right)$, $MINEIG\left(\mathbb{T}_n^p\left(h_n^p(t)\right)\right)$. Next, the value of $\boldsymbol{w}^{n,p}$ that minimizes $MINEIG\left(\mathbb{T}_n^p(h_n)\right)$ is found using a method such as Newton's method or stochastic gradient descent. Given the optimal solution $\widehat{\boldsymbol{w}}^{n,p}$, the warping function is defined by Equation 2.1 where $W^{n,p} = \sum_{r=1}^{R} \widehat{w}_r^{n,p} B_r$. A demonstration of curve registration is provided in Figure 2.1.

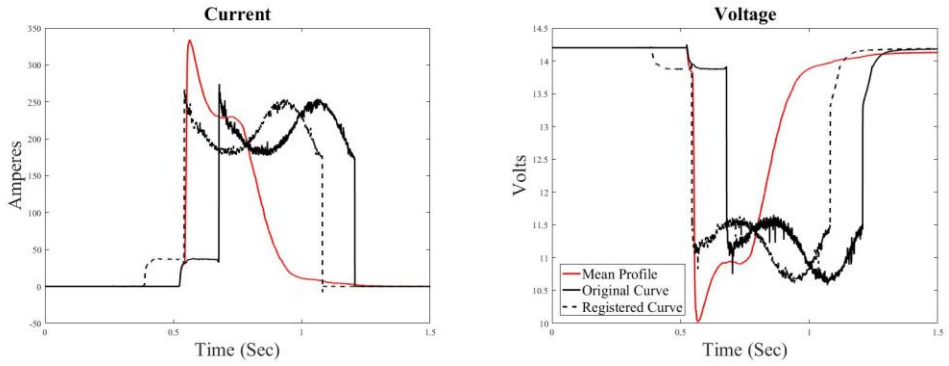


Figure 2.1 – Curve Registration

Figure 2.1 shows curve registration for a current profile (left) and a voltage profile (right). The red curves represent the target functions for each sensor. The solid black curve is the

unregistered profile. Due to a component fault, there is a noticeable delay before current is drawn. As demonstrated by the dashed black curve, curve registration shifts the profile to the left to align it with the target function.

For $n = 1, \dots, N$, the curve registration algorithm starts by setting $\boldsymbol{w}^{n,p} = \mathbf{0}$. Therefore prior to curve registration, $h_n^p(t) = t$ and the feature $MINEIG\left(\mathbb{T}_n^p(h_n)\right), p = 1, \dots, P$ represents the amount of misregistration between an initially unregistered profile s_n^p and the target profile s_*^p. Let δ_n^p denote the amount of misregistration for profile $n, n = 1, \dots, N$ and sensor $p, p = 1, \dots, P$. Since misregistration can occur due to system faults, this feature is combined with the features extracted using statistical transforms on the profiles. This feature is denoted by the vector $\boldsymbol{\delta}_n = [\delta_n^1, \dots, \delta_n^P]'$.

2.3.1.2 Statistical Transformation of Profile Data

Functional data introduces modeling difficulties due to its complexity and high dimensionality. Therefore, statistical transformations are used to extract potential fault-based features from the profiles. Computation of the transformations involves projecting the profiles onto a selected set of basis functions. For the proposed methodology, the eigenfunction, Fourier, and wavelet bases are selected to deploy FPCA, STFT, and DWT respectively. These transformations are applied to the set of registered training profiles $\left\{\tilde{\boldsymbol{s}}_p^n\right\}_{n=1}^N, p = 1, \dots, P$ and combine the extracted features with the misregistration feature $\Delta = \begin{bmatrix} -\boldsymbol{\delta}_1' - \\ \vdots \\ -\boldsymbol{\delta}_N' - \end{bmatrix}$.

Functional Principal Component Analysis (FPCA)

Functional Principal Component Analysis (FPCA) is a technique for capturing dominant modes of variation in functional data. A discussion of FPCA and its application to condition monitoring is found in Fang et al (2017). The eigenfunction basis for sensor $p, p = 1, \dots, P$ is found by first estimating the covariance function as:

$$\widehat{Cov}_p[t, t'] = \frac{1}{N-1}\sum_{n=1}^{N}\left(\tilde{s}_p^n[t] - \hat{\mu}_p[t]\right)\left(\tilde{s}_p^n[t'] - \hat{\mu}_p[t']\right), p = 1, \dots, P \qquad (2.3)$$

where $\hat{\mu}_p[t] = \frac{1}{N}\sum_{n=1}^{N}\tilde{s}_p^n[t]$ and $t, t' \in \mathcal{T}$. Using the Karhunen-Loeve decomposition, the covariance function is expanded as follows:

$$\widehat{Cov}_p[t, t'] = \sum_{m=1}^{\infty}\gamma_{p,m}\eta_{p,m}[t]\eta_{p,m}[t'], p = 1, \dots, P \qquad (2.4)$$

where $\gamma_{p,1} \geq \gamma_{p,2} \geq \cdots$ are ordered eigenvalues for sensor p and $\eta_{p,m}, m = 1,2,$ ... are the corresponding eigenfunctions for sensor p. To obtain the FPC scores, the de-meaned profiles are projected onto the eigenfunctions. The FPC score for observation n of eigenfunction m from sensor p is $\xi_{p,m}^n = \sum_{t=0}^{T-1}\left(\tilde{s}_p^n[t] - \hat{\mu}_p[t]\right)\eta_{p,m}[t]$. Since $\lambda_{p,m} \to 0$ as $m \to \infty$, it is possible to represent the profiles using a finite weighted sum. Given N observations, the maximum number of eigenfunctions is $M_p = \max(N-1, T)$. Let $\boldsymbol{\xi}_p^n = \left[\xi_{p,1}^n, \xi_{p,2}^n, \dots, \xi_{p,M_p}^n\right]'$ denote the FPC score feature vector for observation n from sensor p. The FPCA feature matrix for sensor p is the $N \times M_p$ matrix:

$$FPCA_p = \begin{bmatrix} - \xi_p^{1\prime} - \\ - \xi_p^{2\prime} - \\ \vdots \\ - \xi_p^{N\prime} - \end{bmatrix}, p = 1, \dots, P \tag{2.5}$$

The P FPCA matrices are concatenated with the registration feature vector to form the $N \times \left(\sum_{p=1}^{P} M_p + P\right)$ matrix $X_{FPCA} = [FPCA_1 | \dots | FPCA_P | \Delta]$.

Short-time Fourier Transform (STFT)

It is common for fault information to be visible in the frequency spectrum of a profile. The Short-time Fourier Transform (STFT) is used to evaluate this frequency spectrum as the profile changes in time. Thus, it is preferable over the Fourier transform for nonstationary profiles. For a detailed discussion on STFT, the reader is referred to Gao and Yan (2011). The STFT utilizes a sliding window function centered at specified time epochs and computes the Fourier transform on the segment of the profile within this window. The STFT of $\tilde{s}_p^n$ evaluated at time realization τ and frequency ω is:

$$STFT_{\tilde{s}_p^n}^{W}(\tau, \omega) = \sum_{t=0}^{T-1} \left(\tilde{s}_p^n[t] \mathcal{W}(t - \tau)\right) e^{-i\omega t} \tag{2.6}$$

where $\mathcal{W}(t) = 0.54 - 0.46 \cos\left(2\pi\left(\frac{t}{T-1}\right)\right)$ is the Hamming window function and i denotes an imaginary number. For each profile, the STFT is evaluated at 7 time epochs equally spaced between 0 and $T - 1$ and 257 frequencies equally spaced between 0 and π. Then, the STFT matrix is reshaped to a column vector and the modulus of the vector

elements is computed. Let $STFT_p^n$ denote the feature vector for observation n from sensor p. The STFT feature matrix for sensor p is:

$$STFT_p = \begin{bmatrix} - STFT_p^{1\prime} - \\ - STFT_p^{2\prime} - \\ \vdots \\ - STFT_p^{N\prime} - \end{bmatrix}, p = 1, \dots, P \tag{2.7}$$

The P STFT matrices are concatenated with the registration feature vector to form the $N \times (1799 + 1)P$ matrix $X_{STFT} = [STFT_1| \dots |STFT_P|\Delta]$.

Discrete Wavelet Transform (DWT)

Like the STFT, the Discrete Wavelet Transform (DWT) yields a time-frequency representation of a profile. Unlike the STFT however, the DWT yields a transformation with varying frequency resolution. This is achieved through dilations and translations of wavelet functions. A full discussion of wavelets can be found in Gao and Yan (2011). Given the father wavelet φ, the mother wavelet function is defined as $\psi(t) = \varphi(2t) - \varphi(2t-1)$. A basis can be constructed in the space of square integrable functions by scaling and translating the wavelet functions using the equations: $\varphi_{j,ŧ}(t) = 2^{\frac{j}{2}}\varphi\left(2^j t - ŧ\right)$ and $\psi_{j,ŧ}(t) = 2^{\frac{j}{2}}\psi\left(2^j t - ŧ\right)$ where j and ŧ correspond to the dilation and position parameters respectively. Assuming the profiles are sampled from the space of square-integrable functions, $\tilde{s}_p^n(t)$ can be approximated as:

$$\tilde{s}_p^n(t) = \frac{1}{\sqrt{T}} \sum_{ŧ \in Z_{j_0}} c_{j,ŧ}^{n,p} \varphi_{j,ŧ}[t] + \frac{1}{\sqrt{T}} \sum_{j=j_0}^{J} \sum_{ŧ \in Z_j} d_{j,ŧ}^{n,p} \psi_{j,ŧ}[t] \tag{2.8}$$

where $c_{j,ł}^{n,p} = \sum_{t=0}^{T-1} \tilde{s}_p^n[t]\varphi_{j,ł}[t]$ are the approximation coefficients and $d_{j,ł}^{n,p} = \sum_{t=0}^{T-1} \tilde{s}_p^n[t]\psi_{j,ł}[t]$ are the detail coefficients. $J = \log_2 T$ is the maximum number of decomposition levels permitted, $J - (j_0 - 1)$ is the number of decomposition levels computed, and Z_j is the set of indices for decomposition level j. To extract features using the DWT, a full wavelet decomposition ($j_0 = 1$) is performed on each profile. The order 4 symlet is selected as the father wavelet. The advantages of this wavelet are its near symmetry and the ability to perform a smoother reconstruction of the profile than the Haar wavelet. For each level of the decomposition, the energy in that level is quantified by computing the sum of squared coefficients: $DWT_p^n = \left[\sum_{ł \in Z_J}[c_{J,ł}^{n,p}]^2, \sum_{ł \in Z_J}\left[d_{J,ł}^{n,p}\right]^2, \sum_{ł \in Z_{J-1}}\left[d_{J-1,ł}^{n,p}\right]^2, \ldots, \sum_{ł \in Z_1}\left[d_{1,ł}^{n,p}\right]^2\right]'$. Thus, the DWT feature matrix for sensor p is:

$$DWT_p = \begin{bmatrix} - DWT_p^{1\prime} - \\ - DWT_p^{2\prime} - \\ \vdots \\ - DWT_p^{N\prime} - \end{bmatrix}, p = 1, \ldots, P \tag{2.9}$$

The P DWT matrices are concatenated with the registration feature vector to form the $N \times (J+2)P$ matrix: $X_{DWT} = [DWT_1| \ldots |DWT_P|\Delta]$.

2.3.1.3 Regularized Multinomial Regression (RMR) Model

The matrices X_{FPCA}, X_{STFT}, and X_{DWT} serve as design matrices used to fit the RMR models. Prior to fitting the model, the columns in each matrix are standardized to zero mean and unit variance. A separate model is fitted for each statistical transformation. The purpose of these models is two-fold: to select the most informative features via

regularization and to perform classification. Recall that $y_n, n = 1, \dots, N$ denotes the class label for observation n. For statistical transformation t, $t \in \{FPCA, STFT, DWT\}$, define $\boldsymbol{x}_n^t$ as the nth row of design matrix X_t. Then, $\boldsymbol{x}_n^t, n = 1, \dots, N$ are a set of observed predictors used to train the RMR model for statistical transformation t.

Multinomial regression is an example of a general linear model. Thus, it satisfies the following conditions: the distribution of the response is a member of the exponential family and some function of the expected response is linear in a set of predictors. The response variable $\boldsymbol{Y}_n$ for observation n can take on one of K mutually exclusive classes, corresponding to the component-fault severity combinations. It is represented using a "one-hot vector" where $\boldsymbol{Y}_n[k] = 1$ if $y_n = k$ and $\boldsymbol{Y}_n[k] = 0$ otherwise. Suppose $\{\boldsymbol{Y}_n\}_{n=1}^N = \left\{\left[y_{n,1}, \dots, y_{n,K}\right]'\right\}_{n\in\{1,\dots,N\}}$ is a set of independent multinomial distributed random response vectors. The distribution of $\boldsymbol{Y}_n$ is $\mathbb{p}(\boldsymbol{Y}_n|\boldsymbol{\rho}) = \prod_{k=1}^K \rho_k^{y_{n,k}}$, where $\boldsymbol{\rho}$ is the mean vector whose elements $\rho_k \coloneqq \mathbb{P}\{y_{n,k} = 1\}$ are nonnegative and sum to unity. Given $\boldsymbol{x}_n^t$, the conditional mean for the multinomial distribution $\rho_k^t = \mathbb{P}\{y_{n,k} = 1|\boldsymbol{x}_n^t\}$ can be rewritten using Bayes' Theorem:

$$\rho_k^t = \frac{\mathbb{p}(\boldsymbol{x}_n^t|y_{n,k} = 1)\rho_k}{\sum_{l=1}^K \mathbb{p}(\boldsymbol{x}_n^t|y_{n,l} = 1)\rho_l}, k = 1, \dots, K \tag{2.10}$$

For all k elements of the conditional mean, the natural logarithm of the numerator is a function of the kth element of mean parameter $\boldsymbol{\rho}$. According to the general linear modeling framework, the log of the numerator of ρ_k^t is modeled as a linear combination of the predictors for all k classes: $\ln[\mathbb{p}(\boldsymbol{x}_n^t|Y_{n,k} = 1)\rho_k] = \beta_{0,k}^t + (\boldsymbol{x}_n^t)'\boldsymbol{\beta}_k^t, k = 1, \dots, \mathcal{K}$.

Taking the exponential of both sides of this equation and substituting into Equation 3.10 yields:

$$\rho_{\hbar}^{\mathrm{t}} = \frac{e^{\beta_{0,\hbar}^{t} + (\boldsymbol{x}_n^{t})' \boldsymbol{\beta}_{\hbar}^{t}}}{\sum_{l=1}^{K} e^{\beta_{0,l}^{t} + (\boldsymbol{x}_n^{t})' \boldsymbol{\beta}_{l}^{t}}}, \hbar = 1, \dots, K \tag{2.11}$$

To solve for parameters $\left\{\beta_{0,\hbar}^{t}, \boldsymbol{\beta}_{\hbar}^{t}\right\}_{\hbar=1}^{K}$, the penalized negative log-likelihood of the data $\{(\boldsymbol{x}_n^{t}, y_n)\}_{n\in\{1,\dots,N\}}$ shown below is minimized.

$$\begin{aligned} \ell\left(\left\{\beta_{0,\hbar}^{t}, \boldsymbol{\beta}_{\hbar}^{t}\right\}_{\hbar=1}^{K}\right) = & -\left[\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{\hbar=1}^{K} y_{n,\hbar}\left(\beta_{0,\hbar}^{t} + (\boldsymbol{x}_n^{t})'\boldsymbol{\beta}_{\hbar}^{t}\right) - \log\sum_{l=1}^{K} e^{\beta_{0,\hbar}^{t} + (\boldsymbol{x}_n^{t})'\boldsymbol{\beta}_{l}^{t}}\right)\right] \\ & + \lambda\left[\frac{(1-\alpha)\|\mathrm{B}_t\|_F^2}{2} + \alpha\sum_{d=1}^{D}\|\boldsymbol{\beta}_d^{t}\|_2\right] \end{aligned} \tag{2.12}$$

The penalty term is referred to as the group-lasso term (Simon et al., 2012) as it selects the same group of features for each of the K classes. The lasso parameter λ imposes a sparse selection of features for modeling and the elastic net parameter $\alpha \in [0,1]$ accounts for correlated features by causing them to either be selected or removed together. $\|\mathrm{B}_t\|_F^2$ is the Frobenius norm of the coefficient matrix $\mathrm{B}_t \in \mathbb{R}^{D_t \times K}$ and $\boldsymbol{\beta}_d^{t}$ is the dth row vector of B_t where the dimension D_t is the number of features extracted for statistical transformation t. Let $\widehat{\mathrm{B}}_t$ denote the estimated coefficient matrix for statistical transformation $t, t \in \{FPCA, STFT, DWT\}$. These matrices are used to map features from newly observed profiles to conditional means: $\boldsymbol{\rho}_{FPCA}^{\mathrm{t}}, \boldsymbol{\rho}_{STFT}^{\mathrm{t}}$, and $\boldsymbol{\rho}_{DWT}^{\mathrm{t}}$. Next, ensemble methods that combine the conditional means from each statistical transformation with the aim of developing more accurate classifiers than the individual RMR models are proposed.

2.3.1.4 Ensemble Methods

When a set of classifiers are accurate and diverse, an ensemble method can produce a classifier with higher accuracy than the individual classifiers (Dietterich, 2000). Accuracy refers to the ability of a classifier to make the correct estimate at a rate greater than 50% whereas diversity is the characteristic where multiple classifiers make different errors on a set of new observations. To develop a set of ensemble methods, a validation set of profiles is acquired in a similar manner as the training set. Let $\{\boldsymbol{s}_v\}_{v=1}^{V}$ denote the set of validation profiles and let $\{y_v\}_{v=1}^{V}$ denote their class assignments. These profiles are registered to the previously defined target function and their features are extracted using the three statistical transformations. Let $\{\boldsymbol{x}_v^t\}_{v=1}^{V}$ denote the standardized features from the validation set using statistical transformation t. The estimates $\widehat{\mathrm{B}}_t, t \in \{FPCA, STFT, DWT\}$, obtained by solving the optimization problem in Equation 2.12, are used to map the feature vector $\boldsymbol{x}_v^t$ to the conditional mean $\boldsymbol{\rho}_{v,t}^{\mathrm{t}}$. The maximum-a-posteriori (MAP) estimate is utilized to perform classification. The classification accuracy of the validation set using statistical transformation t is $acc_t = \frac{1}{V}\sum_{v=1}^{V} 1\{y_v^t = y_v\}$.

The accuracies of the RMR models on the validation set are utilized to develop the following four ensemble methods:

Ensemble Method 1: Most accurate classifier - In this method, the statistical transformation t that yielded the highest classification accuracy for the validation set is selected.

$$\boldsymbol{\rho}_{Ens1}^{\mathrm{t}} = \boldsymbol{\rho}_{t^*}^{\mathrm{t}};\ t^* = \arg\max_{t \in \{FPCA, STFT, DWT\}} acc_t \tag{2.13}$$

Ensemble Method 2: Average conditional mean over t

$$\boldsymbol{\rho}_{Ens2}^{\mathrm{t}} = \frac{1}{3}\left(\boldsymbol{\rho}_{FPCA}^{\mathrm{t}} + \boldsymbol{\rho}_{STFT}^{\mathrm{t}} + \boldsymbol{\rho}_{DWT}^{\mathrm{t}}\right) \tag{2.14}$$

Ensemble Method 3: Weighted average conditional mean – In this method, weights are created proportional to the classification accuracy. First, min-max normalization is used to scale the accuracies between 0 and 1. Then, the scaled accuracies are normalized to sum to unity. Let w_t denote the weight for transform t.

$$\boldsymbol{\rho}_{Ens3}^{\mathrm{t}} = w_{FPCA}\boldsymbol{\rho}_{FPCA}^{\mathrm{t}} + w_{STFT}\boldsymbol{\rho}_{STFT}^{\mathrm{t}} + w_{DWT}\boldsymbol{\rho}_{DWT}^{\mathrm{t}} \tag{2.15}$$

Ensemble Method 4: Most confident classifier – In this method, the conditional mean that contains the highest confidence amongst the transformations is selected. For statistical transformation t, let k_t^* denote the class with the highest confidence.

$$\boldsymbol{\rho}_{Ens4}^{\mathrm{t}} = \boldsymbol{\rho}_{t^8}^{\mathrm{t}};\ t^* = \arg\max_{t \in \{FPCA, STFT, DWT\}} \boldsymbol{\rho}_t^{\mathrm{t}}[k_t^*] \tag{2.16}$$

The development of the ensemble methods concludes the model training operation. Given this model, a newly observed profile is mapped to an ensemble method conditional mean. Since the classes refer to component-fault severity combinations, this model estimates the fault severity of the system components simultaneously. To ascertain the system SoH, it is useful to analyze the impact of faults on system performance.

2.3.2 Offline Stage – Definition of System Health States

For the second operation in the offline stage, a function that maps the component-fault severity combination to the system SoH is defined. First, the space of system health states is defined according to the impact of component-fault severity combinations on the system performance. To define this space, a system health indicator is identified. This indicator should show a decline in performance as the component faults become more severe. Bottom-up hierarchical clustering is employed to group the K component-fault severity combinations. Bottom-up hierarchical clustering starts with an initial set of clusters, often the individual data points or a prespecified set of cluster centroids. The distance between each cluster is computed and the two closest clusters are merged. The merged clusters are removed and replaced with a new cluster representing the merged clusters. This is repeated until only one cluster remains. The number of clusters can be chosen by placing a threshold on the value of the distance metric. After clustering the component-fault severity combinations, the system states are defined by ranking the clusters in order of their centroid values. Suppose the component-fault severity combinations are represented by a C-dimensional array. The clustering partitions the array into ranked system states. Thus, the array can be seen as a tabular function that maps the component-fault severity state to the system SoH.

2.3.3 Online Stage – Diagnosis of Newly Observed Profiles

In the online stage, newly observed profiles are used for system diagnosis. Let $\{\boldsymbol{s}_r\}_{r=1}^{R}$ denote a set of R consecutively observed profiles. Feature extraction is performed by first registering the profiles to the predefined target functions and then extracting the

FPCA, STFT, and DWT features. These features are standardized using the means and standard deviations from the training design matrices. After passing the standardized features through the RMR models and combining the outputs using an ensemble method, the posterior means for the R observations $\boldsymbol{\rho}_1^{\mathrm{t}}, \ldots, \boldsymbol{\rho}_R^{\mathrm{t}}$ are obtained. Define $\boldsymbol{\rho}^{(*)}$ as the vector with elements:

$$\boldsymbol{\rho}^{(*)}[k] = \frac{\prod_{r=1}^{R} \boldsymbol{\rho}_{\boldsymbol{r}}^{\mathrm{t}}[k]}{\sum_{l=1}^{K} \prod_{r=1}^{R} \boldsymbol{\rho}_{\boldsymbol{r}}^{\mathrm{t}}[l]}, k \in \{1, \ldots, K\} \tag{2.17}$$

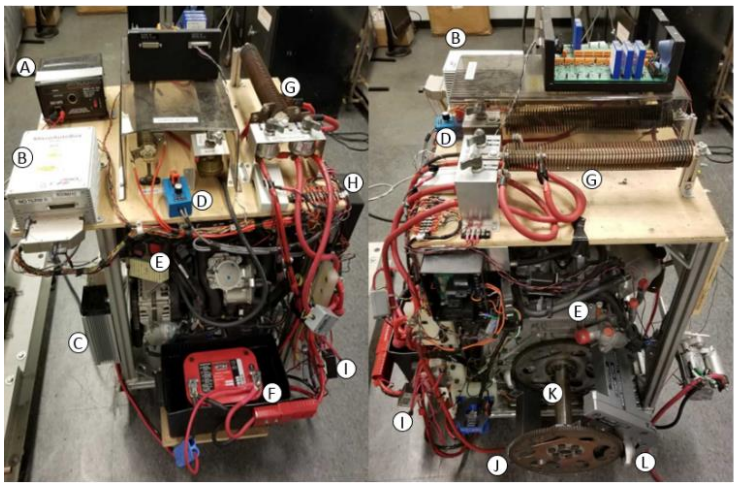
Once $\boldsymbol{\rho}^{(*)}$ is calculated, the maximum a posteriori (MAP) estimate is used to assign the group of observations to a component-fault severity combination. This estimate is motivated by the possibility that a single observation may be classified to a class with low confidence. By taking multiple observations, the uncertainty of one observation can be overcome and a more confident estimate can be obtained. Since this estimate aggregates the results from multiple observations, it is referred to as Aggregate MAP. Note that if $R = 1$, $\boldsymbol{\rho}^{(*)}[k] = \boldsymbol{\rho}_1^{\mathrm{t}}[k]$ and Aggregate MAP reduces to the traditional MAP estimate. Once the component-fault severity combination is estimated, the system SoH is inferred by the tabular function defined in Section 2.3.2. In Section 2.5, the methodology is applied to the Vehicle-Engine Start (VES) system. Prior to demonstrating the proposed methodology, the VES system and the capabilities of the test-rig are briefly discussed.

2.4 Vehicle-engine Start (VES) System

The circuit for the VES system was described in Chapter 1. For the remainder of this section, the VES Test-rig and its capabilities are described in further detail.

2.4.1 VES Test-rig

The VES Test-rig, shown in Figure 2.2, was utilized to simulate the engine-cranking process. The test-rig consists of a 4-cylinder engine with a 12-volt Li-ion battery and three slots for testing multiple motors simultaneously. A coiled resistor is placed between the battery and the motor. The resistor simulates battery degradation and allows for testing of the motors under various battery-fault severity levels. The time and duration of cranking is controlled by the MicroAutobox® prototyping unit. This test-rig is displayed in Figure 2.2.



A. Power Supply
B. Control Unit
C. AC/DC Converter
D. Ignition Switch
E. Engine
F. 12-Volt Battery
G. Resistor
H. PLC Board
I. Voltmeter
J. Flywheel
K. Crankshaft
L. Motor

Figure 2.2 – VES Test-rig

The test-rig is fitted with thermocouples, an rpm sensor, an ammeter, and a voltmeter that measure temperature, engine rpm, current, and voltage respectively. The thermocouples are attached to the outer shell of the motors to monitor their temperature as a safety precaution. The engine rpm is a measure of overall system performance and system failure is defined as failure of the engine to reach an angular velocity of 250 rpm within one second of cranking. The speed threshold was determined empirically and was validated using the criterion provided by Halderman (1988). The current and voltage profiles are acquired from

the ammeter and voltmeter respectively. These profiles capture the electro-mechanical activity of the system during engine cranking and are used for modeling.

2.4.2 VES System Components

This study focuses on two critical and interdependent components of the VES system: the Li-ion battery and the start-stop motor. In contrast to traditional starter motors, start-stop motors restart the engine following short idle periods such as stopping at a traffic light or during heavy traffic. The start-stop motor has a significant impact on driver experience as even a small delay in restarting the engine is considered a very serious fault. The fidelity of the motor relies on the thickness of its brushes, which are responsible for passing current from the motor wiring to the rotating armature. As the thickness decreases, contact between the brush and the armature weakens, compromising the ability of the motor to crank the engine. The battery fault level also impacts system performance since a faulted battery provides insufficient current to crank the engine. Through the VES Test-rig, faults in the motor and the battery are induced. By cranking the engine for a significant amount of time, the brushes are naturally degraded to increasingly higher levels of fault severity. Faulty battery conditions are simulated using the coiled resistor. The resistor simulates increased internal resistance, which serves as a measure of battery health (Saxena et al, 2012). Using the sensors mounted on the test-rig, current, voltage, and engine rpm signals are acquired during the cranking process. The current and voltage profiles when the battery and motor are both in their healthy state are displayed in Figure 2.3.

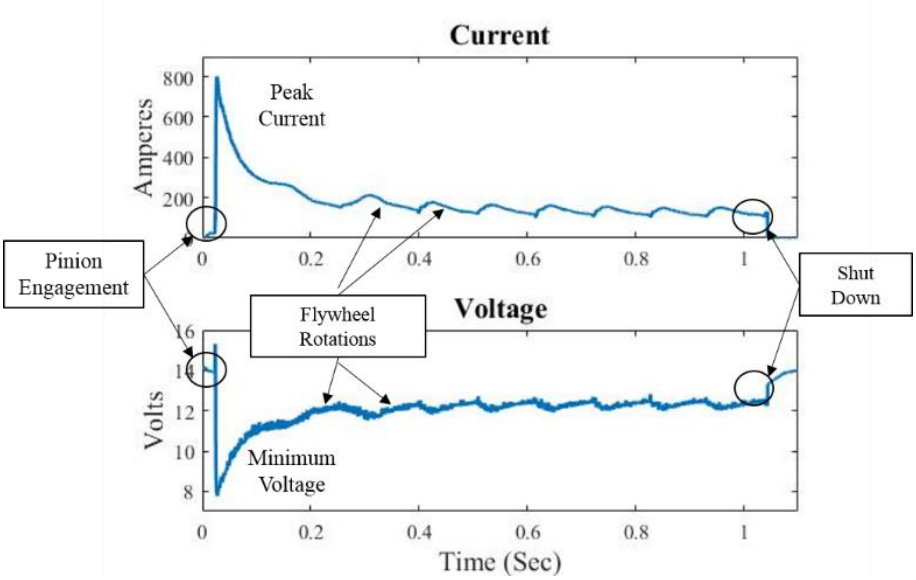


Figure 2.3 – Current and Voltage Profiles

The profiles in Figure 2.3 were recorded during the first second of engine cranking. The initial portions of both profiles show the pinion engaging with the flywheel. Once engaged, the motor quickly draws a large amount of current from the battery, hence the large peak in the current and the subsequent, sharp drop in voltage. The oscillatory portion of the profiles represent when the motor is drawing a steady amount of current to crank the engine. The oscillations in the profiles are due to the changes in the position of the crankshaft as the flywheel turns. Once the motor stops cranking the engine, the current drops to zero and the voltage returns to its initial potential. In a real-life scenario, the duration of cranking is dependent on how quickly the VES system can crank the engine to its target rpm. In this experiment, the target angular speed threshold is set to 250 rpm. Natural variability in the process and degradation results in variability in the stopping time as shown in Figure 2.4.

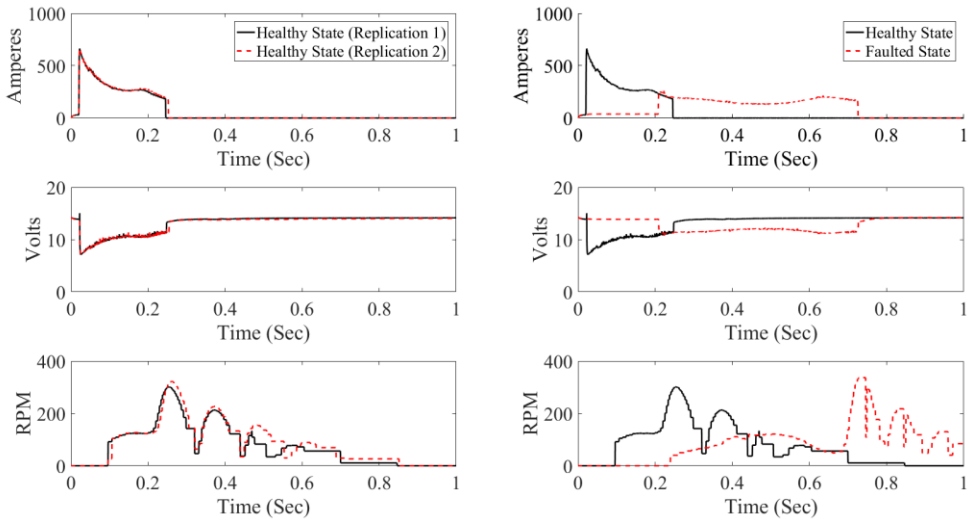


Figure 2.4 – Variability in Cranking Time

The left hand column of Figure 2.4 compares two different profiles when the system is in a healthy condition. As shown in the bottom left, two rpm signals reach the threshold at different times resulting in two different stopping times. The right hand column shows the large difference in performance of the system when the components are in their most severe fault state. Due to the severe battery condition, the amount of current drawn is significantly lower and the amount of time to reach the threshold is much larger. Furthermore, the low thickness of the motor brushes causes a long delay from when the pinion engages to when the motor draws the current. Therefore, profiles acquired when the motor is faulted are likely to be out of phase with profiles collected when the system components are healthy. The next section demonstrates how the proposed methodology utilizes these characteristics to train the diagnosis models.

2.5 Case Study: Vehicle-engine Start (VES) System Diagnosis

This case study considers delayed engine start as the failure mode of interest. As shown in Section 2.4, a delayed start is linked to the health of the battery and the motor. Therefore, the VES system is modeled as being comprised of these two components. Implementation of the proposed methodology requires data snapshots for the various component-fault severity combinations. Through a designed experiment, snapshots are acquired over a fine mesh of battery and motor levels. These snapshots consist of current, voltage, and engine rpm profiles recorded simultaneously. To define component-fault states, the battery and motor levels are clustered using features extracted from the unprocessed profiles. Using the defined fault states and the profile data, the diagnosis methodology is demonstrated for the VES system.

2.5.1 Data Acquisition Via Design of Experiments

For the proposed experiment, faults are seeded into the components of the test-rig and data is recorded under various component-fault severity combinations. The resistor (labeled G in Figure 2.2) is used to simulate various battery health levels by adding resistance to the cranking circuit. Meanwhile, brush degradation in the motor is induced by repeated engine cranking. The component levels for the battery and the motor are displayed in Table 2.1.

Table 2.1 – Battery Levels

Battery Level	Added Resistance (Ω)	Battery Level	Added Resistance (Ω)	Battery Level	Added Resistance (Ω)	Motor Level	Brush Thickness (in.)	Motor Level	Brush Thickness (in.)
1	0.000	**6**	0.013	**11**	0.027	**1**	0.401	**6**	0.221
2	0.001	**7**	0.015	**12**	0.029	**2**	0.340	**7**	0.184
3	0.004	**8**	0.018	**13**	0.032	**3**	0.291	**8**	0.163
4	0.007	**9**	0.021	**14**	0.035	**4**	0.255	**9**	0.136
5	0.010	**10**	0.024			**5**	0.231	**10**	0.126

The data acquisition process consists of conducting a two-way factorial experiment where factor levels are equivalent to those displayed in Table 2.1. Since the motor level cannot be efficiently randomized, the motor level is held constant while the data acquisition for the battery level follows a randomized order. For a fixed motor level, current, voltage, and engine rpm profiles are sampled for each of the 14 battery levels. This procedure is replicated six times before proceeding to degrade the motor to its next fault level. This is repeated until motor brush failure, which occurred at the tenth level.

2.5.2 Hierarchical Clustering for Defining Component States

As the number of battery/motor fault severity combinations is large (140), bottom-up hierarchical clustering is employed to group the combinations into homogeneous clusters. In all further applications of hierarchical clustering, the distance metric is defined as the euclidean distance between cluster centroids where the centroid is the arithmetic mean of the points in the cluster. The feature used for clustering is the wavelet feature described in Subsection 2.3.1.2, but for the unregistered current and voltage profiles. Since battery level has a more prominent effect on the behavior of the profiles than the motor levels, the battery levels are clustered first. Then for each battery cluster, the motor levels are clustered. For battery level clustering, the number of initial clusters is set to 14.

Similarly, the number of initial motor levels is set to 10. The results of hierarchical clustering on the battery and motor level are displayed in Figure 2.5.

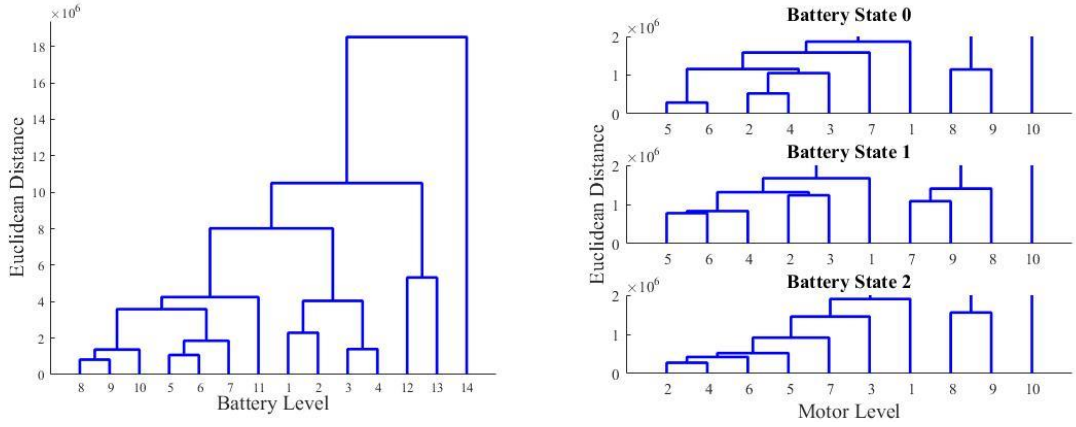


Figure 2.5 – Hierarchical Clustering (Left: Battery, Right: Motor)

Figure 2.5 displays the results as a dendrogram. The heights of the vertical lines represent distances between the clusters. For example, the distance between levels 8 and 9 is small compared to the distance between levels 12 and 13. Horizontal lines represent linkages between clusters. Using the dendrogram in the left side of Figure 2.5, three battery states are defined: State 0 refers to when the battery level is between 1 and 4 (0Ω – 0.007Ω) , State 1 refers to when the battery level is between 5 and 7 (0.007Ω – 0.015Ω), and State 2 refers to when the battery level is between 8 and 10 (0.015Ω – 0.024Ω). Levels 11 through 14 are discarded because instances of system failure occur at these levels even when the motor is healthy. While the elements of cluster (1, 2, 3, 4) are further apart than those of the other two clusters, better clustering performance for the motor levels is observed with this configuration than if two separate clusters were defined. Motor clustering is shown on the right side of Figure 2.5. For States 0 and 2 of the battery, motor levels 8 and 9 cluster together, whereas levels 7, 8, and 9 cluster together for State 1 of the battery. Since level 7

is closer to the first six motor levels in two of the three battery states, this level is assigned to the healthy motor state. The motor states are defined as follows: State 0 refers to when the motor is between levels 0 and 7 (greater than 0.184 in.), State 1 refers to when the motor level is between 8 and 9 (0.136 in. – 0.184 in.), and State 2 refers to when the motor level is at 10 (less than 0.136 in.). Since the number of component states is three for each component, the number of component-fault severity combinations is nine. The class labels for these combinations are displayed in Table 2.2.

Table 2.2 – Class Labels for Component-fault Severity Combinations

Battery	**Motor**		
	0	**1**	**2**
0	1	2	3
1	4	5	6
2	7	8	9

Using the results of hierarchical clustering, the VES system is framed as being comprised of $C = 2$ interdependent components. Let $c = 1$ denote the index for the battery component and $c = 2$ denote the index for the motor component. The highest degree of fault severity for both components is $K_c = 2, c = 1,2$. Thus, $k_c \in \{0,1,2\}$ for $c = 1,2$. Table 5.2 shows how class labels are assigned to each component-fault severity combination. For example, when $\boldsymbol{k} = [1,1]'$, the class label is $\mathcal{k} = 5$. During one cranking cycle, the test-rig generates voltage, current, and engine RPM profiles. The current and voltage profiles are used for modeling while the engine rpm profiles are used for defining the system health indicator. For the purposes of modeling, one observation is a 2-dimensional profile where s_n^1 refers to the nth current profile and s_n^2 refers to the nth voltage profile. Each profile consists of $T = 2048$ points sampled at a 2 kHz rate.

In the experiment, six observations are acquired for each combination of factor levels. After discarding profiles for battery levels 11-14, the number of observations is 600. To demonstrate the proposed methodology, the 600 observations are partitioned into a training, validation, and testing set. For the training set, half the observations of each class are sampled at random without replacement. Due to the clustering, there is an unequal number of samples in each class. Therefore, 100 observations are sampled with replacement from each of the nine classes to obtain a training set with $N = 900$ observations. The remaining 300 observations are split into a validation and testing set. The validation set has 152 observations and the testing set has 148 observations. The slight discrepancy is due to some classes containing an odd number of observations. In the offline stage of the proposed methodology, a classifer is trained using the training and validation observations and the system health states are defined using the system health indicator. In the online stage, diagnosis is achieved by assigning a new observation to one of the classes. Then, the estimated class is mapped to the system SoH. The testing set is used to assess the online performance of the proposed methodology. In this demonstration, the methodology is repeated ten times for random partitions of the data and a graphical summary of the performance is reported.

2.5.3 Offline Stage – Model Fitting

2.5.3.1 Feature Extraction – Curve Registration

Curve registration consists of defining a target function and aligning all profiles to this target function. The target function s_*^p for sensor $p, p = 1, \dots, P$ is the mean of all profiles in the training set. Recall that to align a profile s_n^p to s_*^p, a numerical iterative method is used to find the warping function that minimizes the misregistration between the two profiles. The warping function is specified by $\boldsymbol{w}^{n,p}$, the vector of weights from the linear expansion of $W^{n,p}$. Initially, $(\boldsymbol{w}^{n,p})^0 = \mathbf{0}$. Thus, $\left(h_n^p\right)^0(t) = t$ and the initial registered profile $\left(\tilde{s}_n^p\right)^0$ is equal to s_n^p. The misregistration criterion is intiailized as $\delta_n^p = MINEIG\left(\mathbb{T}_n^p((h_n)^0)\right)$. The set of minimum eigenvlaues $\boldsymbol{\delta}_n = [\delta_n^1, \dots, \delta_n^P]'$ serves as the first feature extracted for diagnosis. On the first iteration, the weight vector is updated to $(\boldsymbol{w}^{n,p})^1$ and the warping function is updated to $\left(h_n^p\right)^1$. Given this warping function, estimating the registered function consists of two steps. First, the inverse warping function $\left(\left(h_n^p\right)^{-1}\right)^1(t)$ is estimated by interpolating the relationship t vs. $\left(h_n^p\right)^1(t)$. Second, the registered function $\left(\tilde{s}_{\boldsymbol{n}}^p\right)^1$ is estimated by interpolating the relationship $\left(\tilde{s}_{\boldsymbol{n}}^p\right)^0(t)$ vs. $\left(\left(h_n^p\right)^{-1}\right)^1(t)$. Given the target function and the estimated registered function, the misregistration criteria is computed as $MINEIG\left(\mathbb{T}_n^p((h_n)^1)\right)$. This procedure continues iteratively until convergence. The results of curve registration for the current and voltage profiles are shown in Figure 2.6.

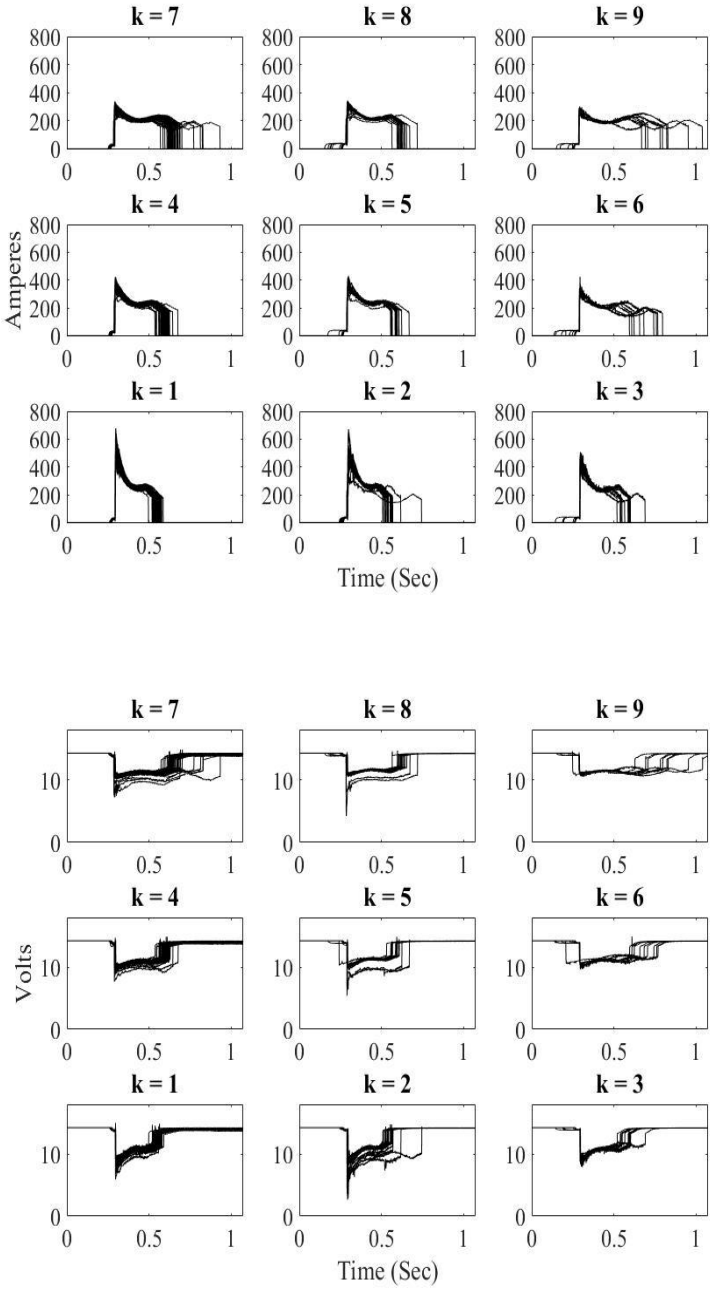


Figure 2.6 – Registered Profiles for All Classes (Top: Current, Bottom: Voltage)

2.5.3.2 Statistical Transformations of Profile Data

Following curve registration, features are extracted by applying the three statistical transforms to the registered profiles. When computing FPC scores, there were originally

$N = 600$ observations. Therefore, the number of FPC scores for each sensor is $N - 1 = 599$. The FPC scores are computed prior to the partitioning into training, validation, and testing. When combined with the misregistration feature, the FPCA design matrix X_{FPCA} is a 900×1200 matrix. To compute the STFT of a profile, the STFT is evaluated at the following epochs: $\tau = [256, 512, 768, 1024, 1279, 1535, 1791]$ for 257 equally spaced frequency values between 0 and π. Thus, the STFT matrix is 257×7. After taking the modulus of the STFT matrix, reshaping it into a vector, and concatenating it with the misregistration feature, the STFT design matrix X_{STFT}, which has dimension 900×3600, is formed. For the DWT, the maximum number of decomposiiton levels is $J = 11$. Therefore, twelve features are extracted from each current and voltage profile. Thus, the DWT design matrix X_{DWT} is a 900×26 matrix.

2.5.3.3 Regularized Multinomial Regression (RMR)

Following feature extraction, the three design matrices are used to train three regularized multinomial regression (RMR) models. The elastic net parameter α is set to 0.99 to promote a sparse model. The lasso parameter λ determines how many parameters are retained in the model. This parameter is selected using 10-fold cross-validation. The results of cross-validation are shown in Figure 2.7.

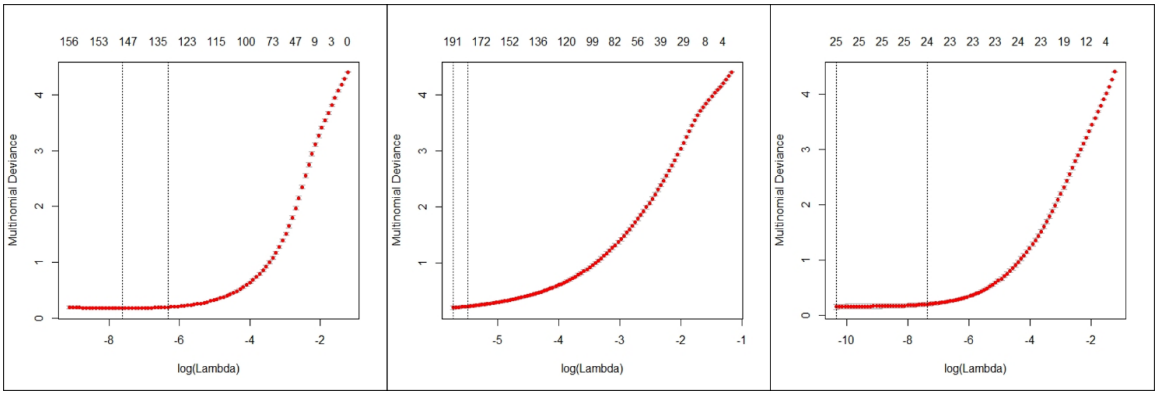


Figure 2.7 – Cross-validation Plots (Left to Right: FPCA, STFT, DWT)

For the plots in Figure 2.7, the x-axis contains the $\ln \lambda$ values. The dotted line on the left-hand side denotes the choice of λ that minimizes the multinomial deviance and the top horizontal axis refers to the number of features retained by the model for the corresponding λ values. The plots indicate that the number of features selected are 149, 191, and 25 for the FPCA, STFT, and DWT classifiers respectively.

2.5.3.4 Ensemble Methods

After the RMR classifiers are trained, their performance on the validation set is analyzed to develop the ensemble methods. First, curve registration is used to register the validation profiles to the target profile defined using the training set. Next, the three statistical transforms are applied. Note that for FPCA, the validation profiles are projected onto the eigenfunction basis computed using the training profiles. Each observation is standardized according to the mean and standard deviation of the design matrices from the training set. Finally, the standardized features are passed through their respective RMR classifier. Over the ten iterations, the average accuracy for the three classifiers is as follows: FPCA – 0.738, STFT – 0.774, DWT – 0.844. Thus, the DWT outperforms the other

statistical transformations on average. As is shown in Figure 2.8, it outperforms the other two over all but one iteration.

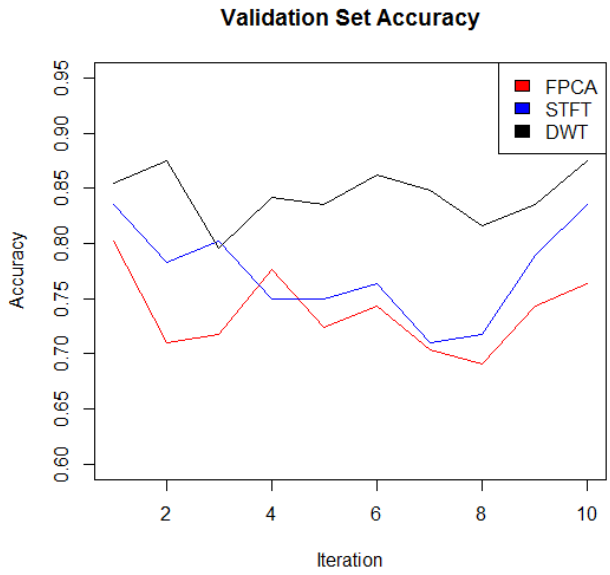


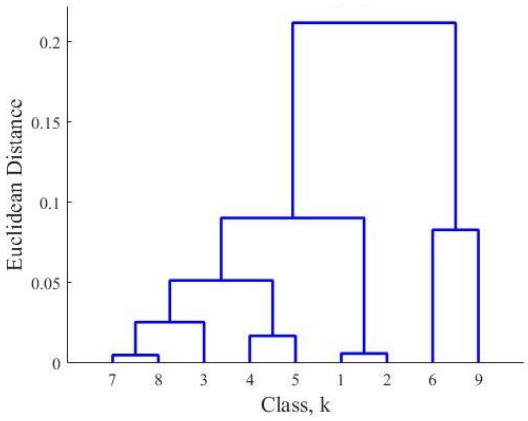
Figure 2.8 – Overall Accuracy for Validation Set

Therefore, Ensemble Method 1 selects the DWT classifier in all but the third iteration. When creating weights for Ensemble Method 3, the DWT classifier receives the largest weight for all but the third iteration. Conversely, the FPCA classifier has the worst performance for all but the fourth iteration. Other than for the fourth iteration, a weight of zero is assigned to the FPCA classifier for Ensemble Method 3. The STFT classifier is assigned an intermediate weight for eight out of the ten iterations. Before proceeding to the online stage, the definition of the system health states is discussed.

2.5.4 Offline Stage – Definition of System Health States

To define the system health states, the system health indicator is defined as the time (in seconds) for the VES system to crank the engine to its target angular velocity of 250

rpm. This feature is extracted from the engine rpm profile. Next, hierarchical clustering is performed to group the nine classes defined in Table 2.2 according to the system health indicator. The initial centroids are chosen to be the average system health indicator for each class. The centroids for each class and the dendrogram are displayed in Figure 2.9.



Battery	Motor		
	0	1	2
0	1) 0.256	2) 0.262	3) 0.353
1	4) 0.310	5) 0.327	6) 0.494
2	7) 0.376	8) 0.381	9) 0.577

Figure 2.9 – Dendrogram and Cluster Centroids for Component-fault Severity Combinations

Hierarchical clustering results in the following clusters: (1, 2), (4, 5), (3, 7, 8), (6, 9) color coded in Figure 2.9. The clusters are ranked in descending order according to their average cranking time. For example, State 1 (white) is the healthy state and consists of classes 1 and 2. State 2 (green) is the mildly degraded state and consists of classes 4 and 5. State 3 (yellow) is the highly degraded state and consists of classes 3, 7, and 8. Finally, State 4 (red) is the system failure state and consists of classes 6 and 9. An interesting result is when $\boldsymbol{k} = [0,2]'$, the system SoH is not in the system failure state. Despite the low brush thickness, this combination resides is in the highly degraded cluster because the spring-loaded nature of the brushes allows enough current to be drawn to rotate the armature and crank the engine if supplied with enough current. At more severe battery levels, the amount of current is not enough to overcome the brush fault and the system fails. In practice, the

table in Figure 2.9 is used to infer the system SoH once the component-fault severity combination has been estimated. The benefit of this approach is that the combinations are imbued with physical meaning, and they can be compared based on the direness of their impact on system health.

2.5.5 Online Stage – Diagnosis Using Newly Observed Profiles

In the online stage, diagnosis is performed by acquiring data for the current state of the system. This can be either one observation or a set of consecutive observations. If one observation is sampled, the traditional maximum a posteriori (Trad. MAP) criterion is used for classification. Otherwise, the Aggregate MAP (Agg. MAP) criterion is used. The performance of the proposed modeling framework is analyzed using the testing set. The procedure for performing classification on the testing set is the same as with the profiles from the validation set, except the ensemble methods are deployed prior to estimating the component-fault severity combination. Since these observations are sampled at random, Aggregate MAP ($R = 2$) is demonstrated on a single observation in the following way. For an observation from class $\mathscr{k} \in \{1, \dots, 9\}$, Equation 2.17 is applied to combine this observation with each observation from the same class separately. The results of applying the criteria are averaged and the combined observations are classified according to the MAP criterion. Thus, an expected class assignment of an observation is obtained if it is assumed that it was sampled consecutively with another observation. Recall that the proposed modeling framework was performed ten times for various partitions of the data set. Figure 2.10 displays box and whisker plots for the classification accuracy of eight models using both Traditional MAP and Aggregate MAP with $R = 2$.

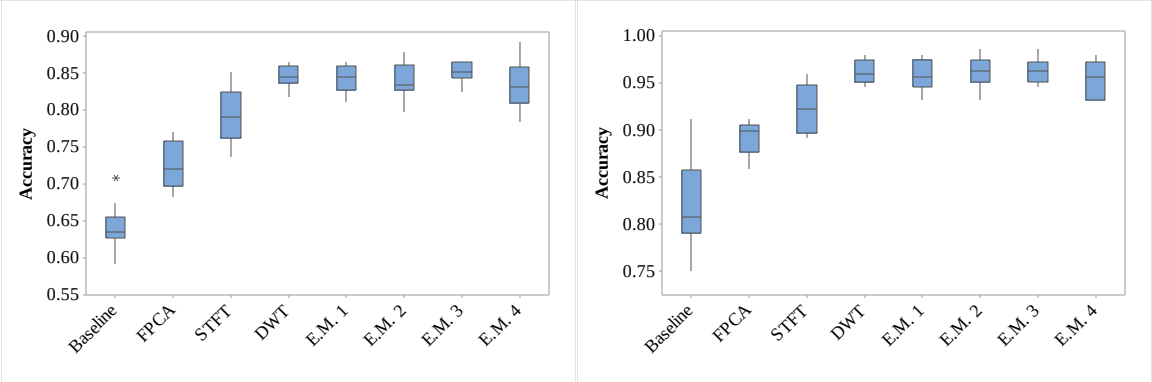


Figure 2.10 – Box and Whisker Plots for Classification Accuracy (Left: Trad. MAP, Right: Agg. MAP)

The eight models displayed in Figure 2.10 include a baseline model, where the raw current and voltage profiles are used as feature vectors to train the RMR model, the three individual models (FPCA, STFT, DWT), and the four ensemble methods denoted as E.M. #. The baseline model is shown to perform the worst for both criteria. The DWT performs the best for all individual models, which is consistent with the results in Figure 2.10. However, the one instance when the STFT model performed better than the DWT for the validation set resulted in Ensemble Method 1 performing slightly worse than the DWT model. Outside of that one instance, Ensemble Method 1 performs the same as the DWT model. For Ensemble Methods 2 and 4, the performance is quite varied. For example, the maximum of both methods is higher than that of all other models. However, their median and minimum values fall below that of the DWT model. Ensemble Method 2 averages the conditional mean vectors from the three individual models whereas Ensemble Method 4 selects the conditional mean vector whose element contains the highest confidence. For both of these methods, the influence of the worst performing classifier, FPCA, may adversely impact the classification accuracy of the ensemble methods. Ensemble Method 3 uses a weighted average of the conditional mean vectors from the three individual

models. Since the min-max normalization removes the impact of the lowest performing classifier of the validation set, this results in Ensemble Method 3 having the highest median classification accuracy of the ensemble methods. Furthermore, its median value is greater than all the individual models.

As shown on the right-hand side of Figure 2.10, the Aggregate MAP criterion improves the classification accuracy of all the models. In general, the conclusions from the Traditional MAP criterion carry over to the Aggregate MAP criterion. The primary difference is that the median value of Ensemble Method 2 is greater than that of the DWT. However, the minimum of Ensemble Method 2 is lower than that of the DWT. The median value of Ensemble Method 3 is equal to that of Ensemble Method 2, but its minimum is higher than that of Ensemble Method 2 and is equal to that of DWT. DWT has a slightly higher third quartile than Ensemble Method 3, but its maximum is less than that of Ensemble Method 3.

Both the classification accuracy of the models and the performance of the classifiers at assigning observations to the correct system state are analyzed. In Figure 2.9, the classes are partitioned into system states-of-health. An observation is deemed correctly assigned if its estimated class falls into the correct ranked cluster. For example, if an observation is from class 3, it is deemed correctly assigned if the estimated class is 3, 7, or 8 since these classes all encompass the highly degraded state, State 3. Since the system states are ranked, incorrectly assigned observations are penalized by squaring the deviation between the assigned state and the actual state. For all models, the mean squared error (MSE) and average over all iterations are computed. Table 2.3 displays the average MSE for all models over the ten iterations.

Table 2.3 – Mean Squared Error for Estimating System SoH

Model	State 1		State 2		State 3		State 4		Total	
	Trad. MAP	**Agg. MAP**	**Trad. MAP**	**Agg. MAP**	**Trad. MAP**	**Agg. MAP**	**Trad. MAP**	**Agg. MAP**	**Trad. MAP**	**Agg. MAP**
Benchmark	0.067	0.008	0.090	0.033	0.151	0.109	0.046	0.028	0.353	0.178
FPCA	0.051	0.003	0.060	0.015	0.130	0.083	0.043	0.032	0.284	0.133
STFT	0.042	**0.000**	0.045	0.009	0.093	0.048	0.039	0.037	0.218	0.094
DWT	0.053	0.001	0.052	0.004	0.058	**0.008**	**0.030**	**0.014**	0.193	**0.027**
E. M. 1	0.055	0.001	0.050	0.005	0.061	0.009	0.035	0.016	0.201	0.030
E. M. 2	**0.020**	**0.000**	**0.032**	**0.002**	0.074	0.016	0.034	0.016	**0.159**	0.034
E. M. 3	0.047	**0.000**	0.049	0.003	**0.056**	0.011	0.032	0.016	0.184	0.030
E. M. 4	0.028	**0.000**	0.035	0.006	0.066	0.014	0.040	**0.014**	0.170	0.034

On average, the total MSE is lowest for Ensemble Method 2 when using Traditional MAP and for DWT when using Aggregate MAP. For Traditional MAP, the STFT has a lower MSE than the DWT for States 1 and 2. Since the Ensemble Methods 2-4 incorporate the STFT, they also have lower MSE than the DWT for States 1 and 2. The DWT has lower MSE than the other individual models for States 3 and 4. In Ensemble Method 3, the weight for the DWT is highest for all but one repeated trial. Therefore, it has higher MSE than Ensemble Methods 2 and 4 for States 1 and 2, but lower MSE than the other Ensemble Methods for States 3 and 4. The effect of Aggregate MAP removes much of the error in the DWT for States 1 and 2 resulting in the differences in total MSE to be dependent on States 3 and 4 where the DWT performs strongest. Of the Ensemble Methods, Ensemble Method 3 has the lowest total MSE, which is very close to that of DWT.

2.6 Conclusion

In this chapter, a methodology for diagnosing the presence and severity of multiple, interacting faults in complex industrial systems was proposed. This methodology consisted of training multiple models and combining their performance using ensemble methods.

The methodology was demonstrated using VES Test-rig. Using the traditional maximum-a-posteriori (MAP) criterion, the most accurate individual model was the DWT. Ensemble Method 1, which selects the most accurate individual model for the validation set, selected the DWT for all but one iteration. Thus, Ensemble Method 1 performed equal to DWT in all but one instance. Ensemble Method 3 outperformed the DWT in terms of median overall classification accuracy. When inferring the system SoH, Ensemble Methods 2-4 had less error than the individual models. The application of Aggregate MAP greatly improved the overall classification accuracy for all models. While Ensemble Methods 2 and 3 had slightly higher median overall classification accuracy, their error in inferring the system SoH was slightly higher than the DWT.

While the proposed methodology is designed to be applicable to common industrial systems, it does require some domain knowledge to identify both important and diagnosable faults and a system health indicator. Tools such as Failure Mode and Effect Analysis (FMEA), the Pareto chart, and the Ishikawa diagram are useful for identifying these faults, but knowledge of how to induce seeded faults is also necessary. For applications such as bearings, techniques to induce seeded faults are widely known. For other applications, domain expertise on the degradation process of the system components is likely needed. In the automotive industry however, it is common to record data snapshots of the system for various levels of component-health states during the preproduction stage. Therefore, there exists the ability to record data following a designed experiment like ours in the case study. Thus, the proposed methodology can be highly impactful for the automotive industry.

In future work, validation of the proposed methodology for the case of more than two fault modes is planned. Furthermore, high-dimensional sensors such as infrared cameras are being used for monitoring complex industrial systems. These sensors produce images and video streams which may contain more fault information but are of higher dimension than signal profiles. Therefore, extensions of the proposed methodology to account for these advanced data structures are planned.

CHAPTER 3. CONDITION MONITORING OF THE COMBUSTOR

Traditional methodologies for turbine condition monitoring attempt to map changes in the properties of the main gas path to component degradation. However, these methodologies are ineffective for monitoring combustor degradation because this type of degradation does not significantly affect the gas path properties (Tahan et al., 2017). Instead, various sensors (acoustic, vibration, optical, etc.) are recommended for condition monitoring of the combustor. In this chapter, two methodologies for combustor condition monitoring are proposed. The first methodology involves detecting early precursors to lean blowout, a serious operational fault for both terrestrial gas turbines and turbines for aviation. For terrestrial turbines, lean blowout can result in costly outages. In aviation, lean blowout poses a significant risk to the health of the passengers. Since lean blowout is related to the equivalence (fuel-to-air) ratio, the proposed methodology leverages the relationship between the alarm frequency and the equivalence ratio to determine a true alarm probability. This probability can aid the operator in controlling the equivalence ratio to maintain a steady flame. The second methodology utilizes acoustic sensors for monitoring centerbody degradation. The centerbody is a component of the combustor that provides stability for the flame while protecting the combustor hardware from the flame. For the proposed methodology, features are mapped from the acoustic signatures to the degradation level of the centerbody. Furthermore, regularization techniques that address the problem of model overfitting and aid in the selection of a subset of informative sensors for monitoring are introduced.

To develop these condition monitoring methodologies, experiments were conducted at the Ben T. Zinn Combustion Laboratory located in the Georgia Institute of Technology. An arial view of this laboratory is displayed in Figure 3.1.

Figure 3.1 – Arial View of Ben T. Zinn Combustion Laboratory

These experiments utilized two combustor test-rigs. The first test-rig, displayed in Figure 3.2, operates similarly to an aviation combustor.

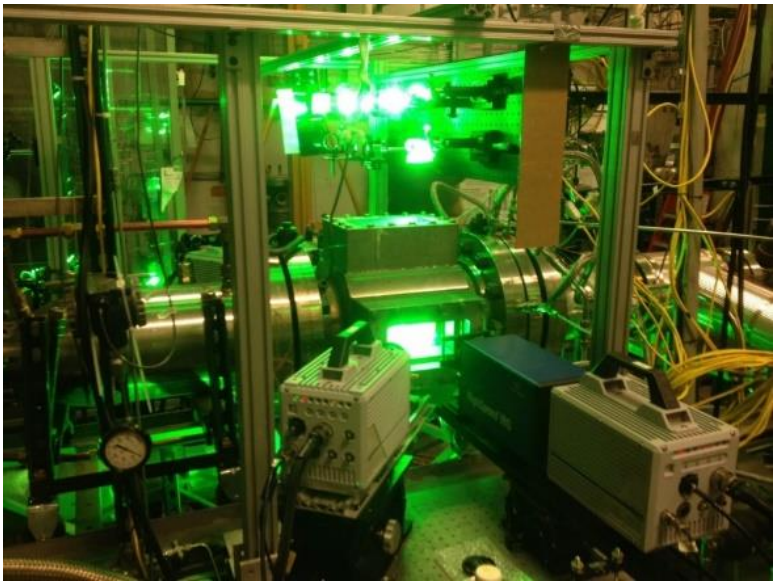


Figure 3.2 – Combustor Test-rig 1

Diagrams of this test-rig published by Rock et al. (2019) are displayed in Figure 3.3.

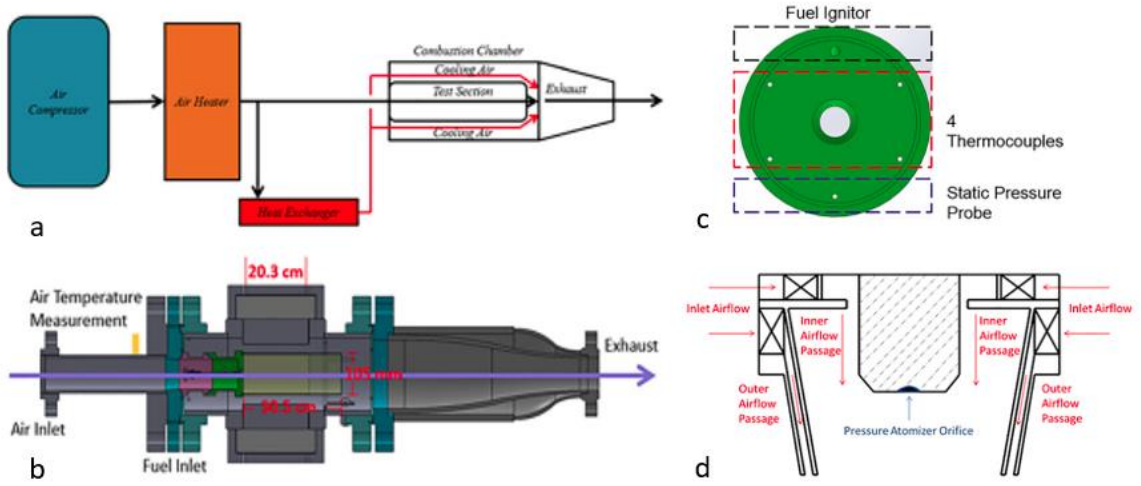


Figure 3.3 – Diagrams of Combustor Test-rig 1 (Rock et al., 2019): a) Air Preparation and Routing Diagram, b) Schematic of the Experimental Combustor, c) Bulkhead Instrumentation Placement Schematic, d) Schematic of the Swirler and Fuel Injector Configuration

Figure 3.3a displays the air flow system. For this system, the air compressor provides an air stream with pressure up to 20 atm (2,026 kPa). The air heater raises the temperature of the air to a value between 350-750 K. This temperature is measured at the location labeled in Figure 3.3b and is continuously monitored. A portion of the air is cooled to 320 K by the heat exchanger, which then distributes the cold air around the liner to keep it cool. The combustion products and cool air then mix downstream and exit the system through the exhaust. To maintain pressure in the combustion chamber, an orifice plug is installed at the exhaust exit.

The liner, displayed in Figure 3.3b, is comprised of a 30.5 cm long, 105 mm inner diameter quartz section with a stainless-steel wall called the “bulkhead” located on its back end. The bulkhead, displayed in Figure 3.3c, contains both a fuel ignitor and thermocouples

for measuring the temperature. The liner is housed in the combustion chamber, which consists of four quartz windows with the following dimensions: 20.3 × 10.8 × 5.1 cm. These windows are located on the front, back, top, and bottom surfaces of the chamber. The front and top window are visible in Figure 3.2. These windows enable optical access to the combustor flame. Details of the sensing technology are discussed in Section 3.1. A mixture of fuel and air enters the liner through the nozzle, a configuration comprised of both the swirler and the fuel injector. The nozzle is displayed in Figure 3.3d, where it is rotated 270°. In the nozzle, air travels through two swirler passages: the inner and outer passage. The fuel is injected through the Pressure Atomizer Orifice. The air and fuel are then mixed so they can be used for combustion.

The second test-rig along with its schematic are shown in Figure 3.4.

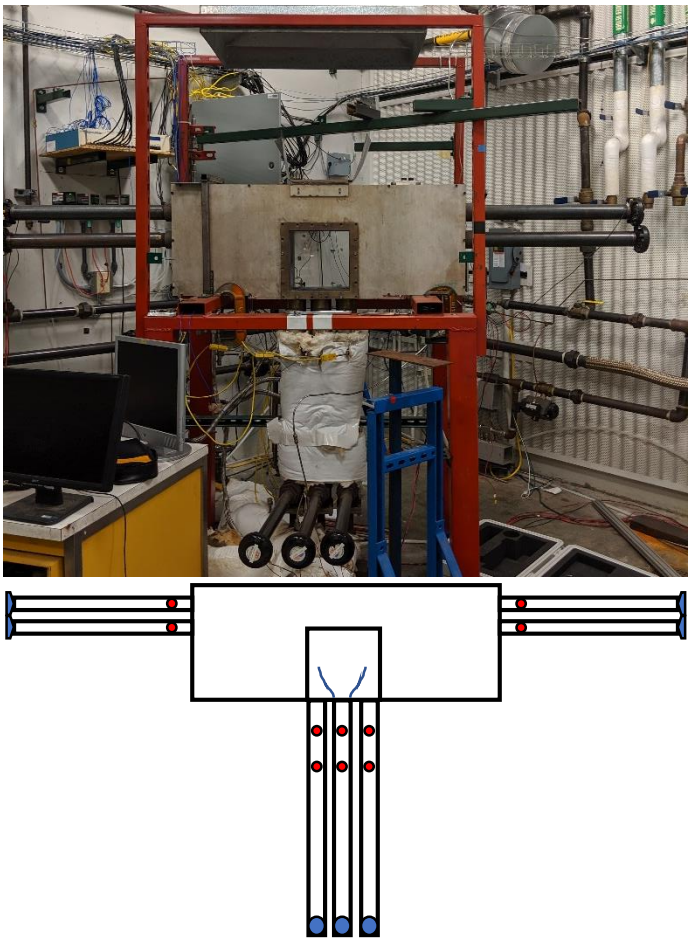


Figure 3.4 – Combustor Test-rig 2 (Kumar et al., 2020)

Combustor Test-rig 2 consists of a premixed combustor that simulates combustion for terrestrial gas turbines used in power plants. Premixing refers to the fuel and air being mixed prior to combustion instead of at the point of combustion to reduce emissions of toxins into the environment. The test-rig is a laboratory scale, unwrapped annular combustor and it consists of a combustion chamber with dimensions 1.16 m × 0.4 m × 0.15 m. An air compressor provides pressurized air at 862 kPa, which is then heated by a flow heater. The combustor burns natural gas that is sourced at 172 kPa at atmospheric conditions. Quartz windows enable optical access of the combustor flame. The test-rig is

fitted with 10 acoustic sensors. Details related to these sensors are provided in Section 3.2 when discussing the experiment.

The benefit of these test-rigs is they enable the acquisition of turbine-representative data sets for the development of condition monitoring methodologies for the combustor. The remainder of this chapter describes these methodologies. In Section 3.1, control charts are used to detect precursors to lean blowout and the distribution of the alarms is used to imbue future alarms with a true alarm probability. In Section 3.2, general linear modeling is utilized to map features extracted from the acoustic sensors to the degradation level of the centerbody. Section 3.3 contains the references from the previous two sections.

3.1 Data Analytics Method for Detecting Extinction Precursors to Lean Blowout in Spray Flames

3.1.1 Introduction and Literature Review

Lean blowout (LBO) is a critical operational fault for both terrestrial and aviation gas turbines. For terrestrial gas turbines, the need for lean operation to prevent toxic emissions of carbon monoxide (CO) and nitrogen oxide (NOx) increases the risk of LBO, which results in costly outages. For aviation, the throttle pullback of a jet engine can result in LBO if too much fuel is reduced. It is imperative that this is avoided as LBO threatens the safety of the aircraft passengers. For both applications, the risk of LBO is inversely proportional with the fuel-to-air ratio, or equivalence ratio (ϕ). Therefore, the equivalence ratio at LBO (ϕ_{LBO}) sets the lean operation limit of a combustor. However, there is high variability in ϕ_{LBO} even when the operational conditions are static. A preset threshold on ϕ could result in the turbine burning too rich in some instances, while experiencing LBO

in other instances. Therefore, a methodology that can detect when LBO is imminent in real-time aids turbine operators in adaptively setting ϕ to a level that is sufficiently lean while accounting for the risk of LBO.

Shanbhogue et al. (2009) described LBO as the two-stage process illustrated in Figure 3.5.

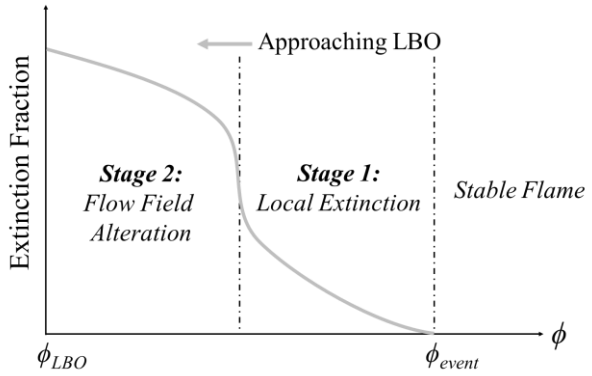


Figure 3.5 – Conceptual Schematic Blowoff Model (Shanbhogue et al., 2009)

At high ϕ, the flame is stable. However, it is using up excessive amounts of fuel and emitting a high amount of toxins. As ϕ reduces, the flame enters Stage 1 where intermittent local extinction/reignition events are observed. However, the shape and structure of the flame is consistent with the stable flame. As the equivalence ratio is reduced, the flame enters Stage 2, where fundamental changes in the flow field occur. These changes were also observed by Nair and Lieuwen (2007). For an explanation of the physical basis of this phenomenon, the reader is referred to Erickson and Soteriou (2011) and Emerson et al. (2012). To identify precursors to LBO, several researchers have correlated features of OH* chemiluminescence emissions to the described phenomena in the combustion flame prior to blowout. Nair and Lieuwen (2005) placed thresholds on wavelet coefficients of the OH*

data. At high ϕ, the wavelet coefficients do not cross the thresholds. However, as ϕ is reduced, the frequency of threshold crossings increases. In their dissertation, Thiruchengode (2006) presented a double threshold for detecting local extinction/reignition events, defined as a crossing of the lower threshold and a return crossing of the upper threshold. This is designed to filter out false alarms that occur from crossings of the upper threshold. Thiruchengode (2006) showed that the frequency of extinction/reignition events increased as LBO approached. This double threshold methodology was also utilized by Prakash et al. (2006), Zhang et al. (2010), and Rock et al. (2020). Other monitoring methods include the work of Yi and Gutmark (2007) who monitored the normalized root mean squared error of the filtered OH* chemiluminescence signal and Unni and Sujith (2016) who monitored the recurrence plot of pressure signal with length N. The recurrence plot is an $N \times N$ matrix whose (i, j)th element equals one if the amplitudes of the time series at times i and j are within a threshold and zero otherwise. These plots are shown to exhibit unique patterns as blowout approaches.

The challenge in implementing these monitoring strategies is the arbitrariness in determining appropriate thresholds. Furthermore, the thresholds do not account for the effect of autocorrelation and nonstationarity on the false alarm rate. Therefore, this section proposes a Statistical Process Control (SPC) approach for identifying LBO precursors. First, the data curation step utilizes time series modeling to transform raw OH* chemiluminescence signals to zero mean signals with constant variance. This model is fitted while the flame is in Phase 1. As the flame approaches LBO, the time series model parameters change. To detect these changes, an Exponentially Weighted Root-Mean Squared Error (EWRMS) control chart is utilized. The control limits (thresholds) of this

monitoring scheme are determined by a false alarm rate specified prior to implementation. Precursors to LBO are identified as instances when the EWRMS test statistic breaches the control limits. As a novel contribution, each precursor is imbued with a probability that the observed alarm is a true indicator of LBO. This probability can be used by turbine operators as an indicator of LBO risk. The remainder of this section is organized as follows. In Subsection 3.1.2, the experiment to collect the OH* chemiluminescence data is discussed. Subsection 3.1.3 discusses the data analytics methodology. This includes the steps of data curation, LBO detection, and adaptive alarm reliability. Subsection 3.1.4 concludes the LBO portion of the chapter.

3.1.2 Data Acquisition

The OH* chemiluminescence measurements analyzed in this study were acquired in the pressurized spray combustor described in the Chapter 3 introduction. The photomultiplier tube (PMT) was used to monitor the OH* chemiluminescence emission as the flame transitioned from Phase 1 to LBO. The PMT integrates all light intensity in its field of view and outputs a singleton point value. This data was sampled at 10 kHz over a 50 second interval. As previously mentioned, studies have shown that extinction/re-ignition physics are known to change significantly based on differences in the air inlet temperature (Rock et al., 2020). Therefore, the blowout detection methodology was applied to data recorded at two different air inlet temperatures: 300 K and 450 K. This enables the quantification of the physical differences between these two air temperatures. Ten PMT samples were acquired at air inlet temperatures of 300 K and 450 K. The Hamamatsu H5784-04 PMT was stored inside of a box during data acquisition to prevent background light from interfering with the chemiluminescence measurements. A Newport spectral filter with a

center frequency of 310 nm and a half width of 10 nm was used to reject emission that was not associated with the OH* radical. Further details about the PMT measurements can be found in Rock et al. (2020).

3.1.3 Methodology

This subsection proposes a data analytics framework for detecting precursors to LBO. This framework consists of three major activities. First, an automated data curation framework is used to transform the PMT signal to a white noise process, i.e., an uncorrelated process with zero mean and constant variance. This step is important because it filters nonstationary behavior and serial correlation from the PMT signal that would otherwise cause a high frequency of false alarms to be detected. Next, a detection algorithm designed to detect changes in the behavior of the PMT signal as the equivalence ratio is reduced toward the blowout limit is employed. Finally, alarms emitted from the detection algorithm are imbued with a probability of denoting a true LBO event. This probability is calculated in an adaptive fashion to incorporate a knowledge of previously observed alarms.

3.1.3.1 Data Curation

The data curation framework consists of an offline phase and an online phase. In the offline phase, five of the ten experimental realizations were set aside for model training. Using these realizations, a time series model is trained. This model captures the dynamics of the PMT signal while the flame is in Phase 1 operation. In the online phase, the fitted model is used to forecast future values of the PMT signal. A simple transformation of the forecast errors is assumed to follow a white-noise process.

The time series model used in this methodology is the ARIMA(2,1,1)-IGARCH(1,1). Its formulation is shown below:

$$\nabla X_i = \mu + \phi_1 \nabla X_{i-1} + \phi_2 \nabla X_{i-2} + \theta_1 a_{i-1} + a_i$$

$$a_i \sim N(0, \sigma_i^2)$$

$$\sigma_i^2 = Var(a_i | a_{i-1}) = \alpha_0 + \alpha_1 a_{i-1}^2 + \eta_1 \sigma_{i-1}^2, \alpha_0 > 0, \alpha_1, \eta_1 \geq 0, \alpha_1 + \eta_1 = 1 \quad (3.1.1)$$

The first line in Equation 3.1.1 denotes the Autoregressive Integrated Moving Average (ARIMA) portion of the model. In this equation, $\nabla X_i \coloneqq X_i - X_{i-1}$ is the difference in the PMT signal at consecutive epochs. Since a single differencing operation is performed on the raw data, the I term is 1. The AR term refers to the dependence of ∇X_i on its previous two epochs and the MA term refers to the dependence of ∇X_i on the model residual at epoch $i - 1$. The order of the model was selected via the Hyndman-Khandakar algorithm (Hyndman and Khandakar, 2008). To account for the increasing signal-to-noise ratio in the PMT signal, the conditional variance of the model residuals $a_i, i = 1,2, \ldots$ is modeled using an Integrated Generalized Autoregressive Conditional Heteroscedasticity (IGARCH) process of order (1,1). Normally, the GARCH model is used to model processes with short bursts of increased variation. By imposing the constraint that the coefficients sum to one, the IGARCH can model processes with a trend in the variability. In this framework, an ARIMA-IGARCH model is fitted to the portions of the training PMT signals when the flame is in Phase 1. A special case of this model, the ARMA-GARCH, was used by Pham and Yang (2010) for estimating the wear and fault state of a methane compressor in a

petrochemical plant. The parameters of the ARIMA-IGARCH are averaged over all training realizations to construct a global time series model. Using this model, one step ahead forecasts on the differenced PMT signal and the conditional variance are performed. This is shown in , where the dark gray portion reflects the part of the signal used for modeling and the light gray portion reflects the remainder of the signal.

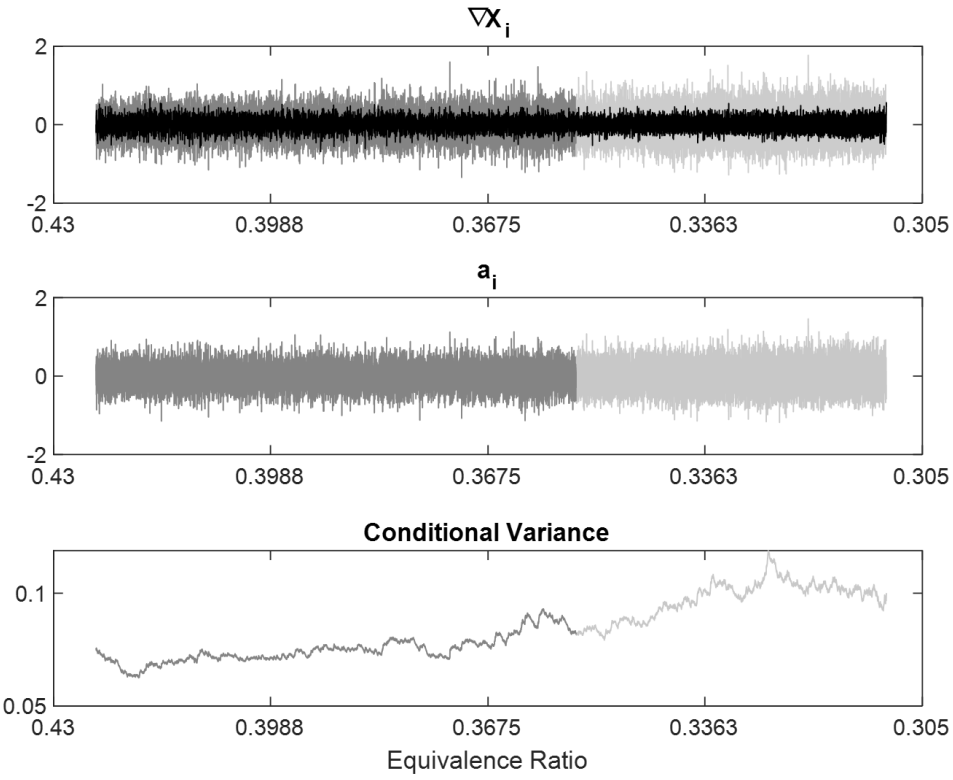


Figure 3.6 – ARIMA-IGARCH Fit

The black curve on the top plot demonstrates the one-step ahead forecast of ∇X_i, whereas the bottom plot displays the conditional variance estimation. Finally, the middle plot displays the residuals from predicting the differenced PMT signal. Now, suppose σ_i^2 is a known quantity. Then $Z_i = \frac{a_i}{\sigma_i} \sim N(0,1)$. Therefore, by estimating the conditional variance

of a_i, the model residuals are transformed into a zero mean, uncorrelated process with constant variance. This is shown in Figure 3.7.

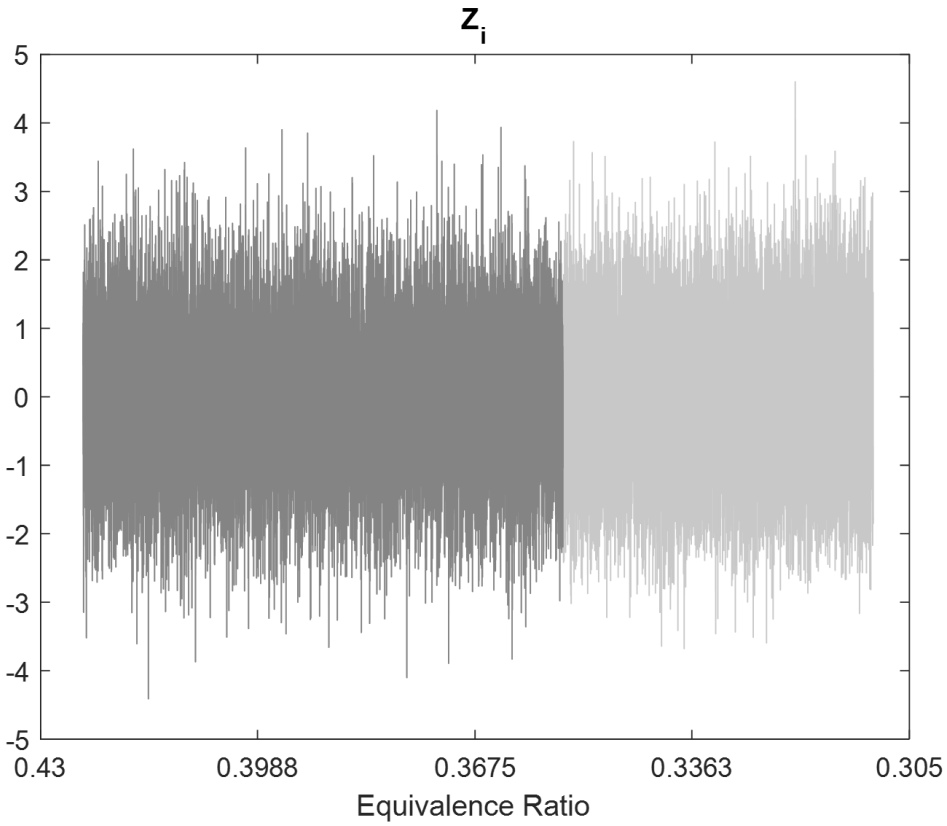


Figure 3.7 – Standard Normal Process

Following the transformation from the differenced PMT signals to $Z_i, i = 1,2, ...$, the data has been properly curated and can be used to develop an LBO detection algorithm.

3.1.3.2 LBO Detection Algorithm

In Figure 3.7, the variability appears to increase as the flame nears blowout. Therefore, a control chart for detecting changes in the process variance is used for detecting LBO precursors. Two commonly used control charts for monitoring variance are the R chart, which monitors the range of samples, and the S chart, which monitors the sample

standard deviation of samples. A detailed discussion of these control charts is provided by Montgomery (2009). One drawback of these control charts is they do not incorporate any memory of previous observations, i.e. state of the process is completely specified by the most recently observed sample. This makes the R chart and S chart insensitive to small changes in the process that can be caused by incipient faults. Therefore, this section proposes the use of the Exponentially Weighted Root Mean Squared-error (EWRMS) control chart developed by MacGregor and Harris (1993). Even after fitting a time series model, the residuals may still be slightly autocorrelated. Therefore, rational subgroups of size b are collected and the average of each subgroup is recorded to compute the quality characteristic. Let $\bar{Z}_k = b^{-1}\sum_{i=b(k-1)}^{bk} Z_i$ denote the average of the kth rational subgroup of size b. The quality characteristic at batched index k is computed as follows:

$$S_k = \sqrt{(1-\gamma)S_{k-1}^2 + \gamma\bar{Z}_k^2}, k = 1,2,\dots,\lceil b^{-1}T_n \rceil \tag{3.1.2}$$

where γ is a constant $(0 < \gamma \leq 1)$ that determines the amount of weight placed on the memory of the quality characteristic and $\lceil\cdot\rceil$ is the ceiling operator that rounds the input to the smallest integer larger than the input. By assumption, $Z_i, i = 1,2,\dots,b$ are normally and independently distributed with zero mean and unit variance. Using the property that $Var\left(b^{-1}\sum_{i=1}^{b} Z_i\right) = b^{-2}\sum_{i=1}^{b} Var(Z_i) = b^{-1}$, the quality characteristic is initialized to $S_0 = b^{-1}$. To determine the control limits, the quality characteristic is computed when the flame is in Phase 1. For a predetermined false alarm rate α, the upper and lower control limit are equal to the $100(1-\alpha)$th and 100αth percentiles of the quality characteristic, respectively. The control chart parameters are: $\gamma = 0.2$, $b = 10$, and $\alpha = 0.0027$, where

$\gamma = 0.2$ was chosen by inspecting the sensitivity of the control charts to changes in the variability. As γ approaches one, the detectability of the control chart decreases, whereas the control chart becomes more sensitive as γ approaches zero. The choice of $\gamma = 0.2$ protects against a high frequency of alarms when the equivalence ratio is high but also enables detectability of significant changes in the flame dynamics. The rational subgroup size is selected to be the smallest that removes the remaining autocorrelation. Finally, the false alarm rate $\alpha = 0.0027$ is selected to be equivalent to that of Shewhart control charts. The control charts monitoring the training realizations for air temperatures 300 K and 450 K are displayed in Figure 3.8 and Figure 3.9, respectively.

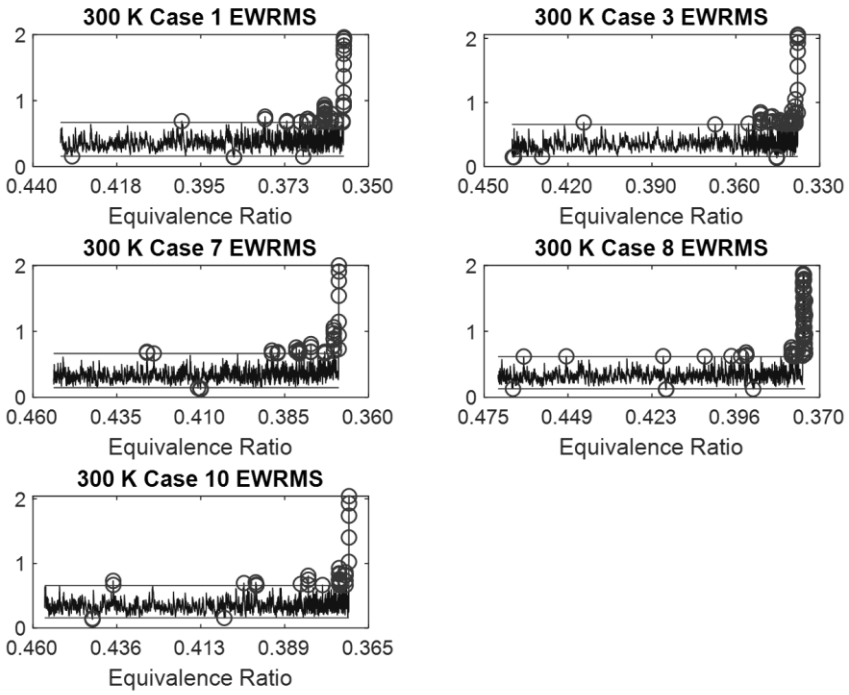


Figure 3.8 – EWRMS Control Charts of Training Realizations for 300 K Air Temperature

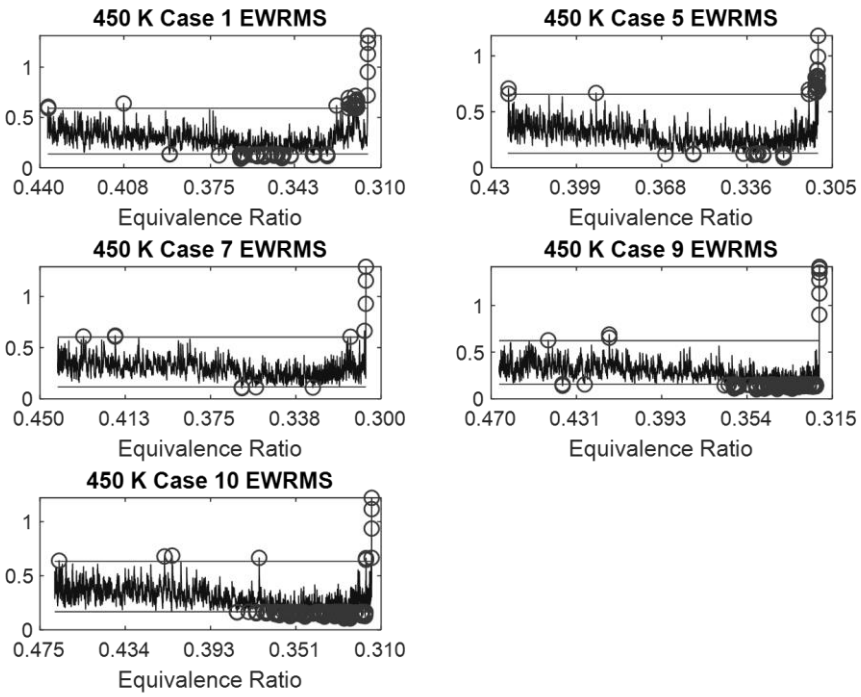


Figure 3.9 – EWRMS Control Charts of Training Realizations for 450 K Air Temperature

The control chart alarms are highlighted by open circles. In the 450 K case, the quality statistic decreases prior to increasing near blowout. This is due to overestimation of the conditional variance of error process a_i, which results from a decrease in the rate of growth of the variability in a_i following the transition from Phase 1 to Phase 2. For the 300 K case, the rate of growth stays constant. Thus, the control charts do not exhibit the same behavior. As blowout is approached, the ARIMA portion of the time series model begins to forecast new observations poorly, leading to large residuals. This is reflected by the large quality characteristic values near blowout.

3.1.3.3 Adaptive Alarm Reliability

The final step of the LBO detection framework is to imbue alarms from the control chart with a measure of confidence that the alarms are true indicators of blowout. For both air temperatures, the equivalence ratios of the alarms from the training cases are used to fit a probability distribution. There are two challenges to fitting this distribution. The first is that there is variability in the equivalence ratio at blowout between cases of the same operating conditions. If the distribution is fitted for too wide a domain, then the ability to detect blowout instances which occur at equivalence ratios higher than the average equivalence ratio at blowout is compromised. To address this issue, a lower bound is set as the median blowout equivalence ratio of the training cases. Thus, the equivalence ratios of alarms occurring below this threshold are raised to the threshold. The second challenge is that the flame stability for some cases is more sensitive to LBO precursors than other cases, causing the less sensitive cases to have little impact on the fitted distribution. This challenge is addressed as follows. First, the domain of the equivalence ratio is partitioned into a set of bins. For each training case, the proportion of alarms in each bin relative to the total number of alarms for the case is computed. By using proportions, the less sensitive cases have equal influence on the fitted distribution. These proportions are then averaged over all training cases. For $m = 1,2, \dots, M$, let p_m denote the average proportion of alarms that occur within bin m. To fit a distribution, a random sample of size L is generated by sampling Lp_m random points from bin m for all M. Since the rate of alarms is expected to grow as the equivalence ratio approaches its lower bound, a 2-parameter exponential distribution is fitted to the random sample. The probability density function (PDF) for the 2-parameter distribution is:

$$f(\phi;\nu,\delta)=\frac{1}{\nu}\exp\left(-\frac{1}{\nu}(\phi-\delta)\right),\nu>0,\phi\geq\delta \tag{3.1.3}$$

where ν is the scale parameter, which reflects the dispersion in the distribution, and δ is the threshold parameter that places a lower bound on the domain of the equivalence ratio ϕ. Model fitting consists of estimating the scale and threshold parameters using maximum likelihood estimation. For the 2-parameter exponential distribution, the maximum likelihood estimates are $\hat{\delta}=\min_{m=1,\ldots,M}\phi_m$ and $\hat{\nu}=M^{-1}\sum_{m=1}^{M}\phi_m-\hat{\delta}$. The fitted distributions for both air temperatures are compared in Figure 3.10.

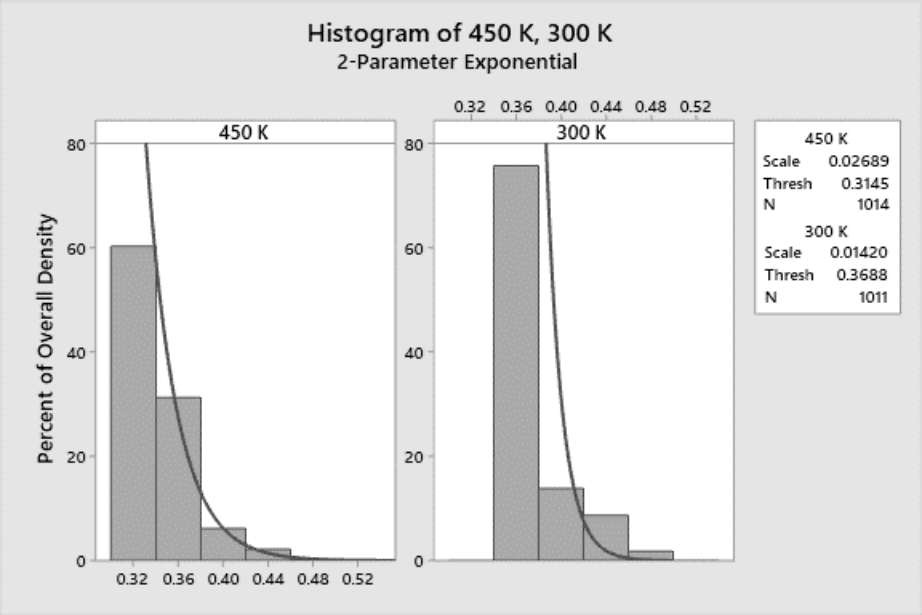


Figure 3.10 – Fitted Distributions for Air Temperatures 300 K and 450 K

In these histograms, observed equivalence ratios are partitioned into six bins. The height of each bar represents the percentage of data accounted for by each bin. The two distributions demonstrate that the lower threshold on equivalence ratio is reduced for increased air temperature.

The LBO detection framework is applied by first using the global time series model to perform a one-step ahead forecast of the PMT signal. After observing the true value, the residual and the conditional variance are estimated and used to obtain Z_i. After a rational subgroup is collected, the EWRMS control chart is constructed and used to monitor the blowout process. When one of the control limits is breached, an alarm is emitted. Each alarm is imbued with a value denoting the probability of a true LBO event. To compute this value, the fitted PDF is discretized into 50 equally spaced bins spanning the equivalence ratio range of 0.3 to 0.5. For bin $\mathscr{b}$, $\mathscr{b} = 1,2, \dots ,50$, the amplitude is calculated as follows:

$$\mathscr{f}'_{\mathscr{b}} = \frac{1}{0.004}\int_{0.5-0.004\mathscr{b}}^{0.5-0.004(\mathscr{b}-1)} f\left(\phi; \hat{v}, \hat{\delta}\right) d\phi \tag{3.1.4}$$

The amplitudes are then normalized to sum to unity.

$$\mathscr{f}_{\mathscr{b}} = \frac{\mathscr{f}'_{\mathscr{b}}}{\sum_{\mathbb{b}=1}^{50} \mathscr{f}'_{\mathbb{b}}} \tag{3.1.5}$$

Now suppose an alarm is observed at $\phi = 0.41$. This equivalence ratio falls into bin 23. Since new alarms can only occur at equivalence ratios less than 0.41, only the density in bins 23-50 is considered. Therefore, the updated probability of an alarm in bin 23 being caused by a true LBO event is $\mathscr{f}_{23} = \frac{\mathscr{f}'_{23}}{\sum_{\mathbb{b}=23}^{50} \mathscr{f}'_{\mathbb{b}}}$. Through this adaptive method, the probability of a true LBO event is assured to approach 1 as the equivalence ratio approaches the fitted threshold $\hat{\delta}$. To demonstrate this method, the control charts of the testing cases and the probability of a true LBO event are displayed in Figure 3.11.

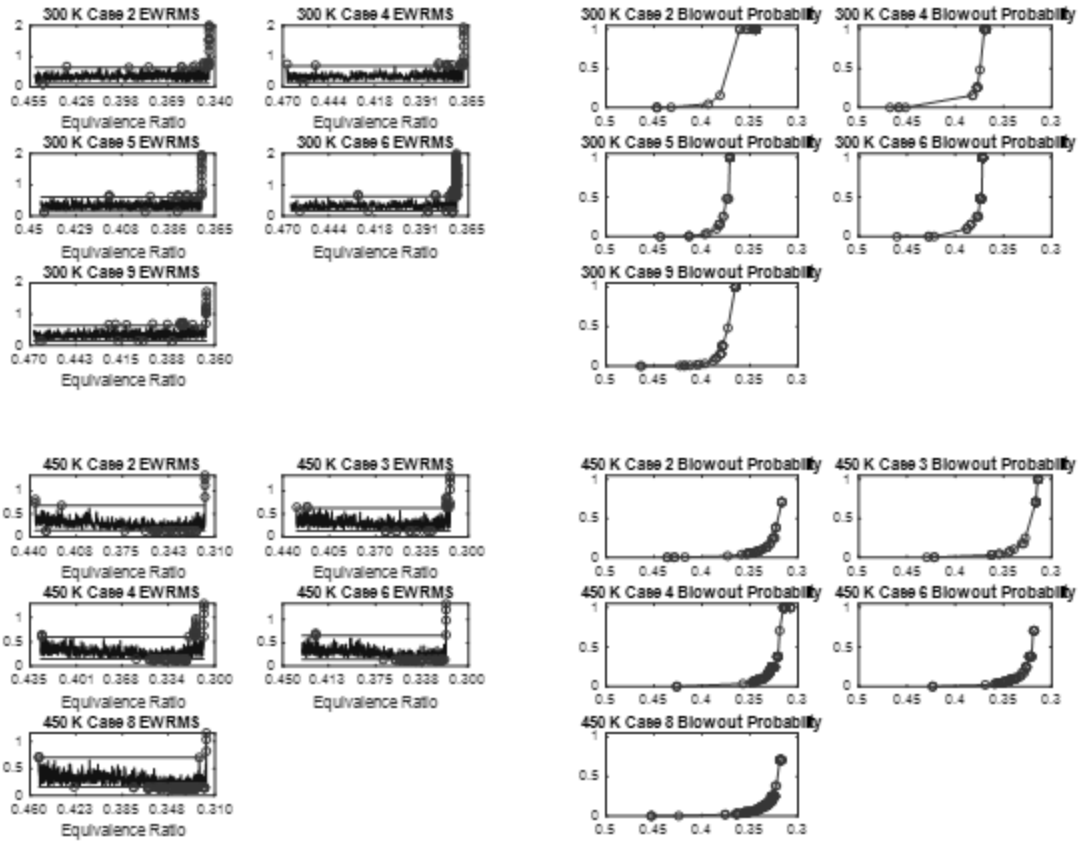


Figure 3.11 – Control Charts for 300 K (Top) and 450 K (Bottom) Test Cases

The blowout probability plots on the right of Figure 3.11 indicate that there is a smooth increase in true LBO alarm probability as the equivalence ratio is increased. A threshold designed to balance the risk of LBO with emissions standards can be incorporated to alert the operator of impending blowout. If incorporated as a controller, this could aid aviation operators in avoiding a blowout while landing.

3.1.4 Conclusion

Since many of the existing methodologies for forecasting lean blowout (LBO) involve a high level of arbitrariness, this section presents a data analytics framework for predicting LBO. It applies statistical modeling to chemiluminescence measurements that were taken during the lean blowout process by a photomultiplier tube (PMT). The key component of this methodology is that the false alarm rate is prespecified and thresholds are intrinsically linked to the false alarm rate. The novel contribution of this methodology is the imbuement of alarms emitted by the control chart with a probability of denoting a true LBO event. This probability is updated adaptively with every observed alarm. While this methodology was applied using the photomultiplier tube, it is not restricted to this sensing technology. Time series modeling can be applied to time series data from any source (e.g. acoustic sensor data). The key is to identify trends such as drift, periodicity, and heteroskedasticity and adjust the time series model accordingly.

The methodology developed in this section has value for both the research and applied combustion communities. There has been considerable interest recently in understanding the lean blowout process of alternative fuels with various physical and chemical properties. Thus, there are plans to extend this methodology to multiple fuels with the goal of studying the effects of fuel properties on LBO performance. Furthermore, the focus of future work is to develop a control system that allows combustor operating conditions to be mapped to an LBO boundary without bringing the combustor to complete blowout. This can be accomplished by leveraging this methodology to determine the probability that an alarm is indicative of LBO and adjusting the combustor when flame stability becomes threatened.

3.2 Data-driven Fault Detection of Premixer Centerbody Degradation in a Swirl Combustor

3.2.1 Introduction and Literature Review

The implementation of new standards to reduce the emission of NOx and CO has resulted in a transition from non-premixed combustion to lean premixed combustion. While this transition has the impact of reducing emissions, it also leaves the gas turbine more susceptible to operational faults such as lean blowout and combustion instability. The former was discussed in Section 3.1. However, the latter also bears attention because it can result in degraded combustor components, which negatively impacts combustor performance. Furthermore, this degradation can result in debris that propagates downstream and damages more expensive turbine components. Therefore, the ability to monitor combustor hardware is valuable to the maintainability of the gas turbine. In this section, a condition monitoring strategy is proposed for detecting degradation of combustor components.

The literature for condition monitoring for gas turbines has primarily been focused on gas turbines in general. Tahan et al. (2017) presented a review of condition maintenance methodologies for gas turbines. These methodologies can be categorized as performance-based and non-performance based. Performance-based methods utilize gas-path analysis to identify causes of reduced engine efficiency such as debris in axial compressors, erosion, and blade corrosion. Tahan et al. (2017) note that this type of analysis is not appropriate for combustion components as their degradation is not a primary source of performance deterioration and therefore, would not be detected through gas path analysis. An alternative

class of methodologies are non-performance-based methods, which utilize analysis of oil samples, vibration sensors, and acoustic sensors to monitor mechanical faults such as misalignment, bearing faults, and loss of lubrication. These methodologies hold more promise since models can be trained to learn relationships between changes in component degradation and changes in sensor features. For example, Madhavan et al. (2014) used blade-tip timing to measure blade vibration and related vibration anomalies to blade damage.

Compared to the turbine components, methodologies for condition monitoring of combustor components are very sparse. These methodologies typically revolve around analyzing changes from a baseline Fourier spectrum that occur due to combustor faults. For example, the Calpine Turbine Maintenance Group (Sewell et al., 2005) and Power Technology (Goy et al., 2005) reported success in detecting faults using the discrete Fourier transform for a series of case studies. Sewell et al. (2005) identified three types of combustion instabilities: low-, midrange-, and high-frequency dynamics. Low-frequency dynamics (LFD) refer to the 10-50 Hz frequency range. Oscillations in this range are called cold tones as their amplitudes are inversely proportional to flame temperature. An example of a fault resulting in a cold tone shift is debris being lodged in the swirler which elicited an alarm at 23.75 Hz. Conversely, midrange-frequency dynamics (MFD) refer to the 100-250 Hz frequency range and are called hot tones because their amplitude rises with flame temperature. Examples of faults eliciting hot tone alarms are pilot-nozzle cracks (156 Hz) and transition-piece failures (225 Hz). High-frequency dynamics (HFD), called screech, are rare but cause rapid deterioration. Examples of HFD-induced failures include cracks in combustor cans and liners (Goy et al., 2005).

An assumption of the case study methodologies is that the correlation between the spectral signatures and the component faults are invariant to temporal effects such as changes in ambient conditions, seasonal effects, or degradation. However, to maintain emissions standards, turbine operators tune modern turbines to account for these temporal effects. As Emerson et al. (2018) mention, these tuning processes can mask fault signatures. Noble et al. (2019) presented a methodology to improve robustness of detection algorithms to these temporal effects via a historical database that can be queried to find signatures in line with the current operating conditions. This methodology, patented by Angello et al. (2017), was used to detect combustor faults in the combustor such as damage to the combustor liner and the pilot nozzle. Furthermore, it could detect instrumentation faults such as those highlighted by Emerson et al. (2018).

The proposed methodology differs from these studies in two fundamental ways. First, data is transformed using a discrete wavelet transform (DWT) instead of a discrete Fourier transform. The DWT decomposes the raw signal into frequency ranges of various bandwidths. This enables a systematic method for extracting features by computing the energy content within each of these frequency bins. Second, instead of simply detecting that a fault has occurred, this algorithm detects the severity of degradation of the fault. To demonstrate the methodology, the case study of the combustor centerbody is used. The centerbody is a vital component that provides stability for the combustor flame and protects combustion hardware from the flame. Due to its proximity to the flame, the centerbody is exposed to thermal stresses. To simulate degradation, four centerbodies were manufactured to various lengths as shown in Figure 3.12.

Figure 3.12 – Manufactured Centerbodies of Different Lengths

To detect centerbody degradation, a real-time machine learning-based classification framework is utilized using acoustic pressure sensor signals. The framework utilizes Multi-class Logistic Regression to train and test a classification model using this signal data. The study aims to identify the state of degradation in the centerbody amongst 4 possible degradation levels.

The remainder of this section is organized as follows. Subsection 3.2.2 discusses the experimental setup. Subsection 3.2.3 introduces the data analytics-based framework for fault diagnosis. It revolves around a feature engineering framework that processes the high-dimensional sensor data into a manageable set of features for training the diagnosis model. Finally, the results of applying the proposed methodology to the problem of centerbody diagnosis are evaluated in Subsection 3.2.4. The section concludes in Subsection 3.2.5 with a summary of the results and a description of the scope for future work.

3.2.2 Experimental Setup

This study utilized Combustor Test-rig 2 displayed in Figure 3.4. For this test-rig, acoustic sensors were mounted at the ten locations marked by red circles. The sensors are broken into two groups: transverse sensors that are located ~2.54 cm from the combustor side wall and longitudinal sensors, which are mounted on the premixer tubes ~23.5 cm upstream of the dump plane – 2 sensors per tube. While the transverse sensors are more sensitive to degradation, they have a lower operational temperature range than the longitudinal sensors.

Centerbody degradation was simulated by manufacturing centerbodies of four different lengths – 87.4 mm, 76.2 mm, 63.5 mm, and 50.8 mm as shown in Figure 3.12. During this study, only the center nozzle was operated. The nozzle settings were $\phi = 0.8$ and $u_{bulk} = 35$ m/s, with a preheat temperature of $T_{ph} = 500$ K. For each observation, a centerbody was installed and the flame was monitored for 5.1 seconds while the acoustic sensors recorded data at a 40 kHz sampling frequency. Ten observations were recorded for each of the centerbody lengths to provide enough samples for training the machine learning model.

3.2.3 Data-driven Diagnosis Methodology

3.2.3.1 Data Curation

Sensors mounted on commercial systems are exposed to extreme conditions and can sometimes fail. Similarly, the data acquisition system and read-write operations can sometimes generate unintended errors that result in outliers and/or missing data. Data

curation ensures that a reasonable level of data quality is maintained. This is achieved by studying the characteristics of historical sensor data (e.g., standard deviation, mean, range of outputs, etc). Based on the historical sensor characteristics of the 10 mounted sensors, 2 sensors were found to be malfunctioning and were removed from the analysis.

3.2.3.2 Feature Engineering

Raw sensor data, especially high-dimensional data such as spectral data require statistical transformations to help extract unique signal features. In this study, this is accomplished using the Discrete Wavelet Transform (DWT). The DWT has been used in previous studies to perform condition monitoring of wind turbines (Yang et al., 2010, Watson et al., 2010, García Márquez et al., 2012) and mechanical tools (Li et al., 2000, Li et al., 1999, Velayudham et al., 2005, Zhu et al., 2009). In addition to providing a systematic method of aggregating the spectral information of the signal, the DWT provides a more informative representation of nonstationary signals than the Fourier transform. It works by decomposing a signal into distinct frequency bands, thus enabling a robust partitioning of the frequency domain. This allows for the representation of signals by their energy content in each frequency band (Tobon-Mejia et al., 2012, Zarei and Poshtan, 2007). The following subsection provides mathematical details of how to construct the features that are used for modeling.

Feature Extraction

Feature extraction is a prerequisite for any diagnosis methodology. It serves to represent a sensor signal using a small number of features relative to the length of the signal. In this framework, the sensor signals are first decomposed using the DWT with the

Daubechies 3-type wavelet (Daubechies, 1992). The DWT decomposes the standardized acoustic pressure signal into a weighted sum of wavelet basis functions as follows:

$$a_g(t) = \sum_{k \in \mathcal{K}_{j_0}} c^g_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{\mathcal{J}-1} \sum_{k \in \mathcal{K}_j} d^g_{j,k} \psi_{j,k}(t) \tag{3.2.1}$$

where $t = 0,1,\dots,T-1$. Each signal consists of $T = 204{,}000$ points. Therefore, a full wavelet decomposition ($j_0 = 0$) applied to one of the signals allows for $\mathcal{J} = \lfloor \log_2 T \rfloor = 17$ decomposition levels. Each level corresponds to a particular frequency range. For example, if a signal is sampled at 40 kHz, the Nyquist frequency is 20 kHz. When $j = 16$, the vector of detail coefficients $\boldsymbol{d}^g_{16} = \left(d^g_{16,0}, d^g_{16,1}, \dots\right)^T$ represents the signal information in the frequency range of 10-20 kHz. When $j = 15$, the vector $\boldsymbol{d}^g_{15}$ represents the signal information in the frequency range of 5-10 kHz. Thus, the frequency range is repeatedly halved for every level of the decomposition. After exhausting all levels of the decomposition, the vector of approximate coefficients $\boldsymbol{c}^g_0 = \left(c^g_{0,0}, c^g_{0,1}, \dots\right)^T$ represents the signal information in the frequency range of 0-0.152 Hz.

The wavelet coefficients for sensor g are represented by vector $\boldsymbol{q}^g = \left(\boldsymbol{q}^{gT}_1, \boldsymbol{q}^{gT}_2, \dots, \boldsymbol{q}^g_{\mathcal{J}+1}{}^T\right)^T$ where $\boldsymbol{q}^g_1 = \boldsymbol{c}^g_0$ and $\boldsymbol{q}^g_{j+2} = \boldsymbol{d}^g_j$ for $j = 0,1,\dots,\mathcal{J}-1$. To construct features, the energy associated within the m^{th} frequency band is calculated as the mean squared value of the coefficients in $\boldsymbol{q}^g_m, m = 1,2,\dots,\mathcal{J}+1$. The vector of these energies is denoted as the *component* feature vector $\boldsymbol{z}_g = \left(z^g_1, z^g_2, \dots, z^g_{\mathcal{J}+1}\right)^T$ corresponding to signal $\boldsymbol{a}_g$ where

$$z_m^g = \frac{\|\boldsymbol{q}_m^{\boldsymbol{g}}\|_2^2}{length(\boldsymbol{q}_m^{\boldsymbol{g}})} \tag{3.1.2}$$

By repeating this process for all working sensors, the *composite* feature vector $= (\boldsymbol{z}_1^T, \boldsymbol{z}_2^T, \dots, \boldsymbol{z}_G^T)^T$ can be constructed. In Figure 3.13, the amplitudes of the extracted features for each of the eight active sensors in the combustor test rig are displayed. Since there are 18 features per sensor signal, the total number of features is $p = 144$.

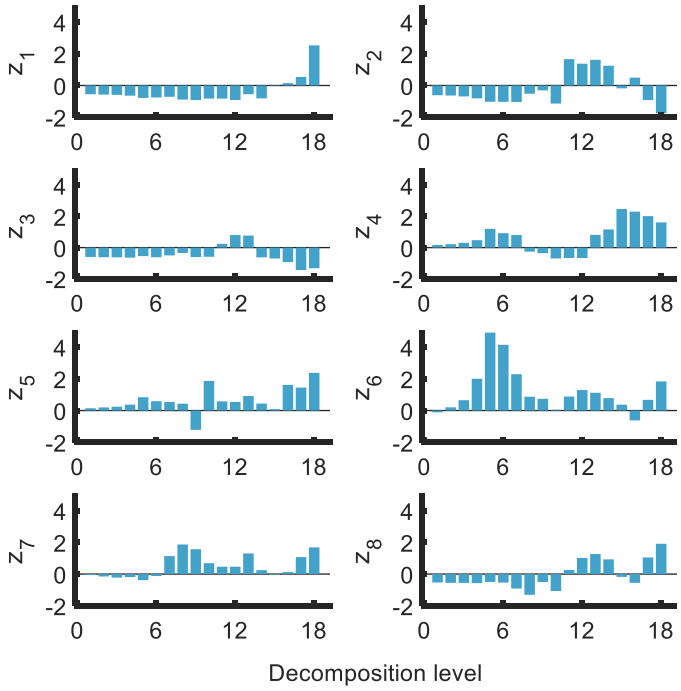


Figure 3.13 – Component Feature Vectors

In the following subsections, a demonstration on how to reduce the number of model feature further is provided. However, this is preceded by an introduction to the diagnosis model.

3.2.3.3 Multiclass Logistic Regression for Diagnosis

For diagnosis, a ML-based classifier, namely multiclass logistic regression (MLR), is used to identify the severity of the degradation state of the centerbody. Multiclass logistic regression is a generalization of traditional linear regression to multinomial response data (discrete number of possible outcomes). It has been used for diagnosing degradation states in automotive electric power generation and storage systems (Peters et al., 2020). Consider a training set consisting of the following: $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$, where $\boldsymbol{x}_i$ is the i^{th} observed composite feature vector of length p and $\boldsymbol{y}_i$ is a vector whose elements $y_{i,k}$ equal one if observation i belongs to class k and zero otherwise. In MLR, the conditional probability that an observation belongs to class k given an observed composite feature vector $\boldsymbol{x}_i$ is modeled as follows:

$$P(Y_i = k|\boldsymbol{X}_i = \boldsymbol{x}_i) = \frac{\prod_{k=1}^{K}\left(e^{\beta_{0k}+\boldsymbol{\beta}_k^T\boldsymbol{x}_i}\right)^{y_{i,k}}}{\sum_{l=1}^{K} e^{\beta_{0l}+\boldsymbol{\beta}_l^T\boldsymbol{x}_i}} \tag{3.2.3}$$

where β_{0k} is the kth intercept coefficient and $\boldsymbol{\beta}_k$ is the kth vector consisting of p regression coefficients.

The model is trained using maximum likelihood estimation, where the goal is to find the parameters (i.e. $\{\beta_{0k}, \boldsymbol{\beta}_k\}_{k=1}^{K}$) that maximize the conditional distribution of Y_i given $\boldsymbol{X}_i$ using the training data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$. Assuming the observations are independently and identically distributed, the likelihood function is formulated as

$$\mathcal{L}(\{\beta_{0k}, \boldsymbol{\beta}_k\}_{k=1}^{K}) = \prod_{i=1}^{N} P(Y_i = k | \boldsymbol{X}_i = \boldsymbol{x}_i) \tag{3.2.4}$$

Maximizing the natural logarithm of the likelihood function, or log-likelihood, is often more tractable and results in the same optimal solution (Bishop, 2006). Regularization is employed to select features. Regularization is a Machine Learning tool that penalizes the magnitudes of the parameters to reduce model complexity especially when there are too many predictor variables and when some predictors are highly correlated.

Let $\mathcal{P}(\{\boldsymbol{\beta}_k\}_{k=1}^{K})$ denote the penalty. Subtracting this penalty from the log-likelihood yields the following optimization problem:

$$\max_{\{\beta_{0k}, \boldsymbol{\beta}_k\}_{k=1}^{K}} \frac{1}{N}\left[\sum_{i=1}^{N}\left(\sum_{k=1}^{K} y_{i,k}(\beta_{0k} + \boldsymbol{\beta}_k^T \boldsymbol{x}_i) - \log\left(\sum_{k=1}^{K} e^{\beta_{0k} + \boldsymbol{\beta}_k^T \boldsymbol{x}_i}\right)\right)\right] - \lambda \mathcal{P}(\{\boldsymbol{\beta}_k\}_{k=1}^{K}) \tag{3.2.5}$$

Fitting the MLR model is equivalent to solving this optimization problem for $\{\beta_{0k}, \boldsymbol{\beta}_k\}_{k=1}^{K}$. An algorithm for solving Equation 3.2.5 is provided by Friedman et al. (2010). K-fold cross-validation is employed to select the tuning parameter λ. Observations are partitioned into K groups of equal size called folds. Setting one of the folds aside, the remaining data is used to train the model and its accuracy is tested on the remaining fold. This is repeated for each fold and the average over all folds is computed. For MLR, the performance metric is the Multinomial Deviance (*MD*) computed as follows:

$$MD = -2\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log\left(\frac{P(Y_i = k|\boldsymbol{X}_i = \boldsymbol{x}_i)}{y_{i,k}}\right) \quad (3.2.6)$$

The optimal λ is the value of λ that minimizes the average cross-validated Multinomial Deviance. In the next subsection, the methodology of utilizing this general MLR model within a hierarchical framework for feature selection is discussed.

3.2.3.4 Hierarchical Feature Selection

Recall that the signal energy content within each frequency represents the signal features extracted from the raw sensor data. These signal features are still large in number and therefore present challenges when training conventional ML algorithms, namely model overfitting. Model overfitting occurs when a model is trained too closely to a dataset that it cannot be generalized to other datasets. This happens when uninformative sources of variation (e.g. noise) are treated as primary elements of the model structure. To address this challenge, a hierarchical feature selection approach that utilizes regularization is proposed. The hierarchical feature extraction approach consists of two steps. The first step selects the most informative sensors that should be used for diagnosis. A byproduct of this step can be viewed as optimal sensor placement, especially in scenarios involving redundant sensors. Once the most informative sensors are selected, the second step focuses on extracting the key signal features generated by the DWT. The signal features retained from the selected sensors are then used to train the classifier.

Prior to sensor selection, initial parameter estimates $\{\beta_{0k}^{ridge}, \boldsymbol{\beta}_k^{ridge}\}_{k=1}^{K}$ are obtained by solving the problem in Equation 3.2.5 with $\mathcal{P}(\{\boldsymbol{\beta}_k\}_{k=1}^{K}) = \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{K}\beta_{k,j}^2$. This is

known as the ridge penalty and the parameter estimates are denoted as the ridge estimates. The ridge penalty is used to avoid problems with optimizing the nonpenalized log-likelihood function such as trivial solutions (all parameters are zero) or nonunique maximizers (when some parameters are infinite) (Bishop, 2006). Next, the Adaptive Group Lasso (Least Absolute Shrinkage and Selection Operator) (Friedman et al., 2010, Tibshirani, 1996, Pencer, 2016, Zou, 2006) is used to perform sensor selection. This technique leverages the group structure of the composite feature vector $\boldsymbol{x_i}$ where $\mathcal{J}+1$ features are extracted from each sensor signal. Adaptive Group Lasso penalizes the parameters of all features within a group together. Thus, if a group is noninformative, all parameters within this group are reduced to zero and the group is removed from the analysis. In the context of this problem, a group refers to the signal features associated with a single sensor. That is, reducing the group to zero implies removing the respective sensor from the analysis. To implement Adaptive Group Lasso, a weight for each of the G groups is calculated as follows:

$$w_g = \frac{1}{\sum_{j \in group\ g} \sqrt{\sum_{k=1}^{K} \left(\beta_{k,j}^{ridge}\right)^2}}, g = 1,2,\dots,G \tag{3.2.7}$$

Using these weights, Adaptive Group Lasso is performed by solving Equation 3.2.5 with the penalty term defined as follows:

$$\mathcal{P}(\{\boldsymbol{\beta}_k\}_{k=1}^{K}) = \sum_{g=1}^{G} w_g \sum_{j \in group\ g} \sqrt{\sum_{k=1}^{K} \beta_{k,j}^2} \tag{3.2.8}$$

After removing the noninformative sensors, Adaptive Lasso (Bishop, 2006, Zou, 2006) is used to select a subset of features for fitting the diagnostic model. Let $\mathcal{G} = \{1,2,\dots,G\}$ denote the set of sensors and $\mathcal{G}^r$ denote the subset of $\mathcal{G}$ consisting of sensors remaining after sensor selection. Furthermore, let $\boldsymbol{x}^r$ denote the composite feature vector containing only features from the $\mathcal{G}^r$ sensors and $\{\boldsymbol{\beta}_k^r\}_{k=1}^K$ denote the vectors of parameters associated with the sensor features. Replacing $\boldsymbol{x}_i$ and $\{\boldsymbol{\beta}_k\}_{k=1}^K$ in Equation 3.2.5 with $\boldsymbol{x}_i^r$ and $\{\boldsymbol{\beta}_k^r\}_{k=1}^K$ respectively, the parameters are initialized by solving Equation 3.2.5 with the ridge penalty. Using estimates $\left\{\hat{\beta}_{0k}^r, \widehat{\boldsymbol{\beta}}_{\boldsymbol{k}}^r\right\}_{k=1}^K$, the weights are computed for each individual feature as follows:

$$w_j = \frac{1}{\sqrt{\sum_{k=1}^K \left(\hat{\beta}_{k,j}^r\right)^2}}, j = 1,2,\dots,p' \tag{3.2.9}$$

where p' is the number of remaining features after sensor selection. Adaptive Lasso penalizes the parameters for each feature individually, resulting in parameters for noninformative features being reduced to zero. To perform adaptive Lasso, Equation 3.2.5 is solved with the penalty term defined as follows:

$$\mathcal{P}(\{\boldsymbol{\beta}_k^r\}_{k=1}^K) = \sum_{j=1}^{p'} w_j \sqrt{\sum_{k=1}^K \left(\beta_{k,j}^r\right)^2} \tag{3.2.10}$$

Figure 3.14 and Figure 3.15 display the results of sensor selection and individual feature selection respectively. In both figures, the green bars denote features that were removed from the model.

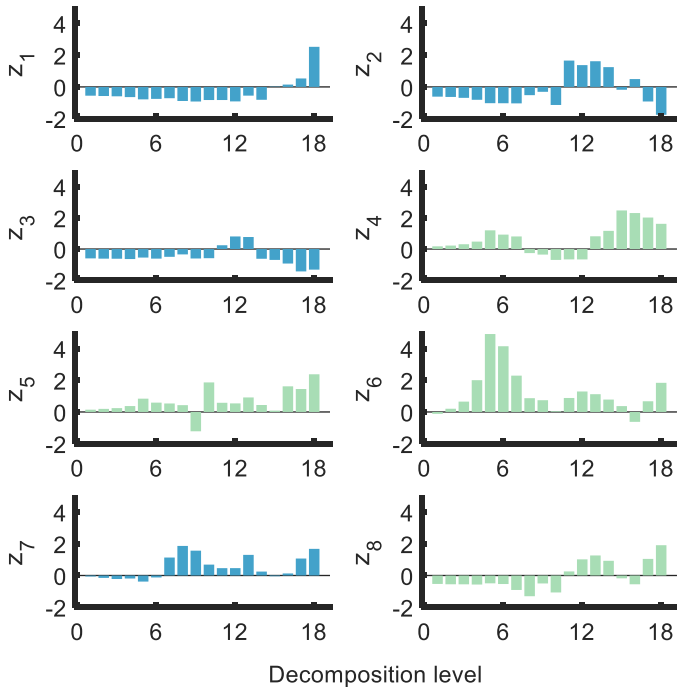


Figure 3.14 – Component Feature Vectors Showing the Selected Sensors (Blue) and Rejected Sensors (Green)

Figure 3.14 demonstrates that Adaptive Group Lasso resulted in sensors 1, 2, 3, and 7 being selected and the remaining sensors being removed. Of these, only one (upper right) transverse sensor was selected. This may be explained by the proximity of the longitudinal sensors from the combustor flame. Figure 3.15 demonstrates the results of Adaptive Lasso. This penalty removes most of the remaining features from the model. Thus, diagnosis of the centerbody requires the monitoring of only a small subset of features.

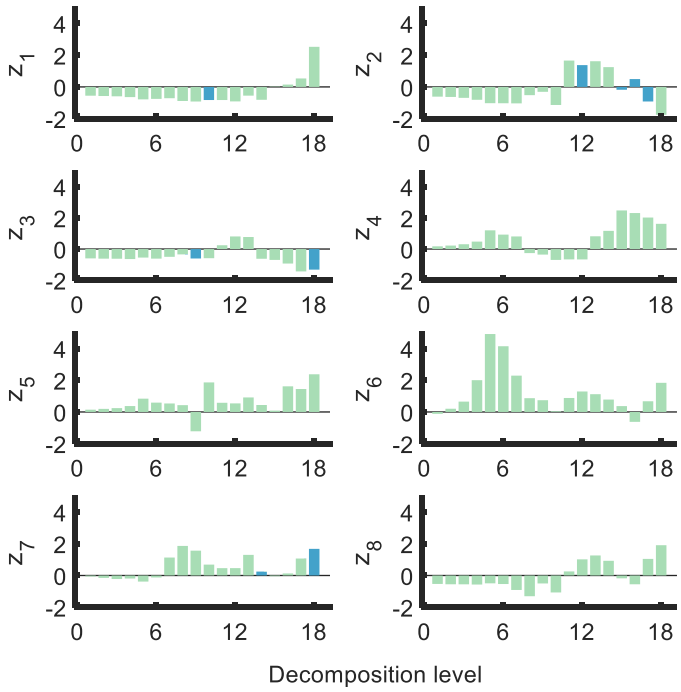


Figure 3.15 – Component Feature Vectors, z_i, Showing the Selected Features (Blue) and the Rejected Features (Green) Corresponding to Each Sensor

Let $\{\beta_{0k}^*, \boldsymbol{\beta}_k^*\}_{k=1}^K$ denote the solution from performing Adaptive Lasso. To perform diagnosis, Equation 3.2.4 with $\{\beta_{0k}, \boldsymbol{\beta}_k\}_{k=1}^K = \{\beta_{0k}^*, \boldsymbol{\beta}_k^*\}_{k=1}^K$ is used to estimate the probability that a new observation belongs to class k for all K classes. Diagnosis is performed by assigning the observation to the class with the largest probability. The following subsection presents the results of the diagnosis step and several studies to determine the robustness and accuracy of the proposed methodology.

3.2.4 Results

To evaluate the proposed methodology, an experiment was conducted where the combustor was monitored under four centerbody degradation states ($K = 4$). One observation consists of $G = 8$ signals of length $T = 204{,}000$. As previously discussed, the feature extraction methodology results in 18 features per signal for a total of $p = 144$

features. Ten observations were recorded for each centerbody length. The dataset was divided into two halves: training and testing. For each degradation state, five observations were sampled without replacement for training and the remaining observations are used for testing. Thus, the number of observations in the training set is $N = 20$. To estimate the average performance of the proposed methodology, 100 iterations of the model with different partitions of the data were performed.

To demonstrate the effectiveness of the proposed hierarchical feature selection approach, its performance is compared to a baseline MLR model using only the ridge penalty, which does not remove any features from the model. The performances of the baseline model and the model with hierarchical feature selection are summarized in Table 3.1 and Table 3.2 respectively.

Table 3.1 – Confusion Matrix (Baseline)

		Actual State (mm)			
		87.4	**76.2**	**63.5**	**50.8**
Predicted State (mm)	**87.4**	1	0.01	0	0
	76.2	0	0.98	0.008	0.032
	63.5	0	0.01	0.992	0.002
	50.8	0	0	0	0.966

Table 3.2 – Confusion Matrix (Hierarchical Feature Selection)

		Actual State (mm)			
		87.4	**76.2**	**63.5**	**50.8**
Predicted State (mm)	**87.4**	1	0	0	0
	76.2	0	1	0	0
	63.5	0	0	1	0
	50.8	0	0	0	1

The performance of each classifier is presented as a confusion matrix, a table used to describe how a classifier performs on test data with known class labels. The values on the diagonal represent the rate of true positives for each class. For example, in Table 3.1, the

rate of true positives for a centerbody of length 76.2 mm is 0.98. Similarly, the true negative, false positive, and false negative rates for each class can be defined. Using the 76.2 mm class as an example, the true negative rate for this class is the rate at which an observation with length other than 76.2 mm is classified as having length other than 76.2 mm. The false positive rate for this class is the rate at which an observation with length other than 76.2 mm is classified as having length 76.2 mm. Finally, the false negative rate for the 76.2 mm class is the rate at which an observation with length 76.2 mm is classified as having a length different from 76.2 mm. A summary of the true positive rate (TPR), true negative rate (TNR), and false positive rate (FPR) for all classes for both models are displayed in Table 3.3.

Table 3.3 – Classifier Metrics for Baseline Model (BL) and Hierarchical Feature Selection (HFS)

Class	**87.4**		**76.2**		**63.5**		**50.8**	
Model	**BL**	**HFS**	**BL**	**HFS**	**BL**	**HFS**	**BL**	**HFS**
TPR	1	1	0.980	1	0.992	1	0.966	1
TNR	0.997	1	0.987	1	0.996	1	1	1
FPR	0.003	0	0.013	0	0.004	0	0	0

3.2.4.1 Reduced Training Size Study

Table 3.1,

Table 3.2, and Table 3.3 show that the model employing the hierarchical feature selection approach achieves a perfect classification. This is likely due to training the model on 50% of the data. To test the robustness of the proposed methodology, the accuracy of the proposed approach is evaluated for varying sample sizes. In Table 3.4, the accuracy of the baseline model is compared with the model using hierarchical feature selection when using 50% ($N = 20$), 40% ($N = 16$), 30% ($N = 12$), and 20% ($N = 8$) of the recorded

data for training. An equal number of observations were sampled from each degradation state to ensure all model classes were equally represented. The number of observations per class are shown in parentheses in Table 3.4.

Table 3.4 – Classification Accuracy for Models Using N Total Observations for Training

N (# Obs/Class)	Model	
	Baseline model	**Hierarchical Feature Selection**
20 (5)	0.985	1.000
16 (4)	0.974	0.999
12 (3)	0.961	0.994
8 (2)	0.882	0.956

The results in Table 3.4 demonstrate that for all choices of N, the hierarchical feature selection approach improves on the baseline model. Furthermore, while the baseline model accuracy decreases steadily as the size of the training data decreases, the accuracy for the hierarchical model stays consistently high. While both models show a sharp decline when reducing the sample size from $N = 12$ to $N = 8$, the accuracy of the hierarchical model remains above 95%.

To supplement the above results, Table 3.5, Table 3.6, and Table 3.7 display the confusion matrices of the hierarchical feature selection model for each sample size.

Table 3.5 – Confusion Matrix for $N = 16$ (Hierarchical Feature Selection)

Predicted State (mm)	**Actual State (mm)**			
	87.4	**76.2**	**63.5**	**50.8**
87.4	1	0	0	0
76.2	0	1	0	0.003
63.5	0	0	1	0
50.8	0	0	0	0.997

Table 3.6 – Confusion Matrix for $N = 12$ (Hierarchical Feature Selection)

Predicted State (mm)	Actual State (mm)			
	87.4	**76.2**	**63.5**	**50.8**
87.4	1	0.001	0.001	0
76.2	0	0.994	0.003	0.014
63.5	0	0.004	0.996	0
50.8	0	0	0	0.986

Table 3.7 – Confusion Matrix for $N = 8$ (Hierarchical Feature Selection)

Predicted State (mm)	Actual State (mm)			
	87.4	**76.2**	**63.5**	**50.8**
87.4	0.990	0.013	0.009	0.015
76.2	0.010	0.960	0.036	0.039
63.5	0	0.015	0.951	0.023
50.8	0	0.013	0.004	0.924

Table 3.4 and the confusion matrices in Table 3.5, Table 3.6, and Table 3.7 show that the hierarchical model is robust to changes in the sample size until 20% of the data is used for training. While the classification accuracy is still high for this case, several observations are misclassified to degradation states far from the actual value. Thus, training with 30% of the data appears to be the bottom threshold for achieving reliable diagnosis of the centerbody. To further demonstrate this, the TPR, TNR, and FPR are reported in Table 3.8 for the confusion matrices in Table 3.5, Table 3.6, and Table 3.7.

Table 3.8 – Classifier Metrics for Hierarchical Feature Selection for Different Sample Sizes

Class	N	TPR	TNR	FPR
87.4 mm	**16**	1	1	0
	12	1	0.999	0.001
	8	0.990	0.988	0.012
76.2 mm	**16**	1	0.999	0.001
	12	0.994	0.994	0.006
	8	0.960	0.972	0.028
63.5 mm	**16**	1	1	0
	12	0.996	0.999	0.001
	8	0.951	0.988	0.013
50.8 mm	**16**	0.997	0.986	0.924
	12	1	1	0.995
	8	0	0	0.005

3.2.4.2 Reduced Sensor Study

As a final test of the proposed methodology, the analysis was carried out by only considering the longitudinal and transverse acoustic sensors. The longitudinal group consisted of five working sensors and the transverse group consisted of three working sensors. As demonstrated previously, it was found that the hierarchical model had significantly better performance than the baseline model. Therefore, subsequent results are only shown for the former.

Longitudinal Sensors

Table 3.9 shows the confusion matrix for the hierarchical feature selection approach using only the longitudinal sensors, using 50% ($N = 20$) of the recorded data for training. With an overall accuracy of 99.75%, the model with only longitudinal sensors is marginally less accurate than the model with all sensors (100%).

Table 3.9 – Confusion Matrix for $N = 20$ (Longitudinal Sensors – Hierarchical Feature Selection)

Predicted State (mm)	Actual State (mm)			
	87.4	**76.2**	**63.5**	**50.8**
87.4	1	0	0	0
76.2	0	0.996	0.006	0
63.5	0	0.004	0.994	0
50.8	0	0	0	1

Once again, the high accuracy can be attributed to the large training size used. A similar investigation with decreasing training size was performed using 40% ($N = 16$) and 30% ($N = 12$) of the recorded data for training. These results are shown in Table 3.10 and Table 3.11.

Table 3.10 – Confusion Matrix for $N = 16$ (Longitudinal Sensors)

Predicted State (mm)	Actual State (mm)			
	87.4	**76.2**	**63.5**	**50.8**
87.4	1	0	0	0
76.2	0	0.992	0	0
63.5	0	0.008	1	0
50.8	0	0	0	1

Table 3.11 – Confusion Matrix for $N = 12$ (Longitudinal Sensors – Hierarchical Feature Selection)

Predicted State (mm)	Actual State (mm)			
	87.4	**76.2**	**63.5**	**50.8**
87.4	1	0	0	0
76.2	0	0.991	0.005	0.001
63.5	0	0.009	0.995	0.006
50.8	0	0	0	0.993

The method appears to be robust to the decreasing training size with the overall accuracy never falling below 99%.

Transverse Sensors

In a similar fashion, Table 3.12 shows the confusion matrix for the hierarchical feature selection approach using only the transverse sensors, using 50% ($N = 20$) of the recorded data for training. While all 8 sensors and the purely longitudinal sensors had overall accuracies of 100% and 99.75% respectively, the purely transverse sensor study performed poorly with an overall accuracy of 85.6%. This suggests that the longitudinal sensors are possibly better suited to the detection of faults presented in this work, which is supported by 75% of the selected sensors coming from the longitudinal group.

Table 3.12 – Confusion Matrix for $N = 20$ (Transverse Sensors – Hierarchical Feature Selection)

		Actual State (mm)			
		87.4	**76.2**	**63.5**	**50.8**
Predicted State (mm)	**87.4**	1	0	0	0
	76.2	0	0.826	0.280	0.068
	63.5	0	0.170	0.714	0.048
	50.8	0	0.004	0.006	0.884

Once again, the model was subjected to the decreasing training size study to evaluate the limits of its performance. These results are shown in Table 3.13 and Table 3.14.

Table 3.13 – Confusion Matrix for $N = 16$ (Transverse Sensors – Hierarchical Feature Selection)

		Actual State (mm)			
		87.4	**76.2**	**63.5**	**50.8**
Predicted State (mm)	**87.4**	0.992	0.002	0.007	0.003
	76.2	0.007	0.782	0.358	0.052
	63.5	0.001	0.213	0.623	0.042
	50.8	0	0.003	0.012	0.903

Table 3.14 – Confusion Matrix for $N = 12$ (Transverse Sensors – Hierarchical Feature Selection)

Predicted State (mm)	Actual State (mm)			
	87.4	**76.2**	**63.5**	**50.8**
87.4	0.98	0	0	0.007
76.2	0.02	0.739	0.401	0.084
63.5	0	0.254	0.584	0.049
50.8	0	0.007	0.015	0.86

As expected, the performance gets worse as the training size is decreased with the overall accuracies for the $N = 16$ and $N = 12$ cases dropping to 82.5% and 79% respectively.

3.2.5 Conclusion

This section introduced a data-driven diagnosis methodology to estimate premixer centerbody degradation in a swirl stabilized combustor. The methodology is capable of processing high-resolution signals recorded by multiple sensors simultaneously. It consists of a feature engineering framework where the DWT is used to construct features for modeling. Through a hierarchical approach, a subset of features correlated with the degradation of the centerbody are selected. These features are used to fit an ML-based classifier to estimate centerbody degradation. The degradation of the centerbody was modeled by manufacturing centerbodies of different lengths and data was collected on a laboratory scale combustor test-rig.

The hierarchical model successfully estimated the level of centerbody degradation in the combustor as it yielded a 0.015-0.074% improvement in accuracy from the baseline model. Furthermore, classification accuracy stayed consistent for the hierarchical model when reducing the size of the training set down to 30% of the recorded observations. However, the accuracy of the baseline model reduced by an average of 0.012 points per 4

total observations when reducing the sample size to 30%. It was found that the transverse sensors on their own were not strong predictors for centerbody degradation as the longitudinal sensors outperformed them in terms of classification accuracy. This observation was consistent across both the baseline model and the hierarchical feature selection model – the latter however performed better than the former in all tested cases. The longitudinal sensors performed very well in terms of the fault detection and were also robust to the decreasing size of the training set. Overall, the proposed methodology has been shown to perform at high accuracy rates by monitoring just a few features. This suggests a high potential for direct integration into existing condition monitoring systems for easy and efficient monitoring.

Practical challenges associated with the manufacturing and assembly of the centerbody in the test rig limited the number of degradation levels that could be tested. Plans for future work include testing the proposed methodology with more granular changes in the centerbody length and applying the methodology to other systems (e.g. spray combustion rigs and bluff body rigs) and to other types of combustion faults. Other future work includes studying the changes in the flame dynamics due to degradation and its relationship with the acoustic pressure.

CHAPTER 4. CONDITION MONITORING OF GAS TURBINE

Following combustion, a hot fuel-air mixture enters the turbine through the main gas path (MGP) where it turns the turbine blades before being expelled through an exhaust. Rotation of the turbine blades causes the connected power shaft to turn, generating mechanical power for the system attached (e.g. generator or propeller fan). While a hot fuel-air mixture is important for ensuring efficient turbine operation, the temperature of the fuel-air mixture can exceed the material softening point of the turbine components (Bunker, 2017). To avoid degradation of the turbine components, cooling mechanisms have been incorporated into the design of modern gas turbines. Examples of these cooling mechanisms are visible in Figure 4.1, which displays a diagram of a turbine stage consisting of a stator vane and rotating blade (Owen, 2010a).

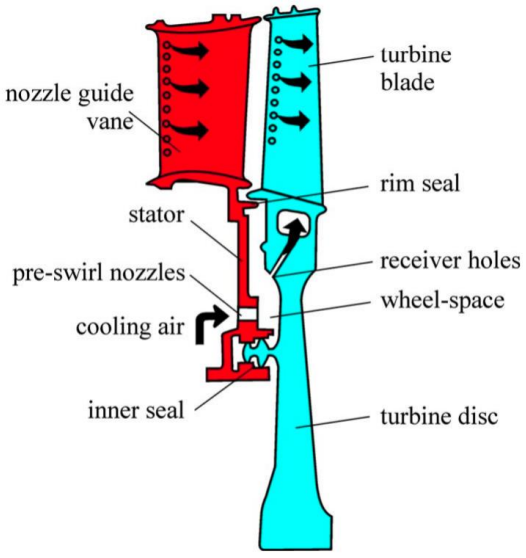


Figure 4.1 – Turbine Stage (Owen, 2010a)

In Figure 4.1, the MGP passes through the nozzle guide vane and turbine blade. Cooling holes and thermal barrier coatings are used to prevent damage to these components. However, these techniques are only applicable to the components in the MGP. The under-platform components such as the stator, wheel-space, and turbine disk are vulnerable to ingress from the MGP. To protect these components, sophisticated rim seal geometries and the pumping of cooling air from the compressor (called purge flow) are incorporated.

The literature on condition monitoring of these cooling mechanisms is quite limited. For the rim seal, the literature has primarily focused on understanding the flow physics surrounding the rim seal to improve the design of rim seal/purge flow configurations. Physics-based models have been proposed for monitoring the effectiveness of the seal to prevent ingress. However, the applicability of these models is limited because they cannot account for the variability in flow physics between different rim seal geometries. For the components in the MGP, the literature has focused primarily on using sensors to monitor degradation of the turbine blade. More recently, infrared imaging has been incorporated for condition monitoring. However, the health of the components in the MGP is intrinsically linked to the rate at which coolant air can be pumped through these components. Therefore, a methodology of inferring coolant flow rate from the temperature distribution of the blade is valuable.

In this chapter, data-driven methodologies are proposed for condition monitoring of these cooling systems. In Section 4.1, two models for predicting sealing effectiveness are proposed. Both models rely on sensor data for inference. The primary difference is that one of the models incorporates domain knowledge into its formulation, whereas the other is purely data driven. In Section 4.2, infrared imaging of the turbine blade is used to predict

the coolant flow rate for a variable MGP temperature. The accuracy of various feature extraction techniques in predicting coolant flow rate are analyzed and the tradeoff between accuracy and sensing capability is evaluated.

The models presented in the chapter were fitted and validated using data generated by the Turbine Test-rig housed in the Steady Thermal Aero Research Turbine (START) facility at Pennsylvania State University. A diagram of the facility is displayed in Figure 4.2.

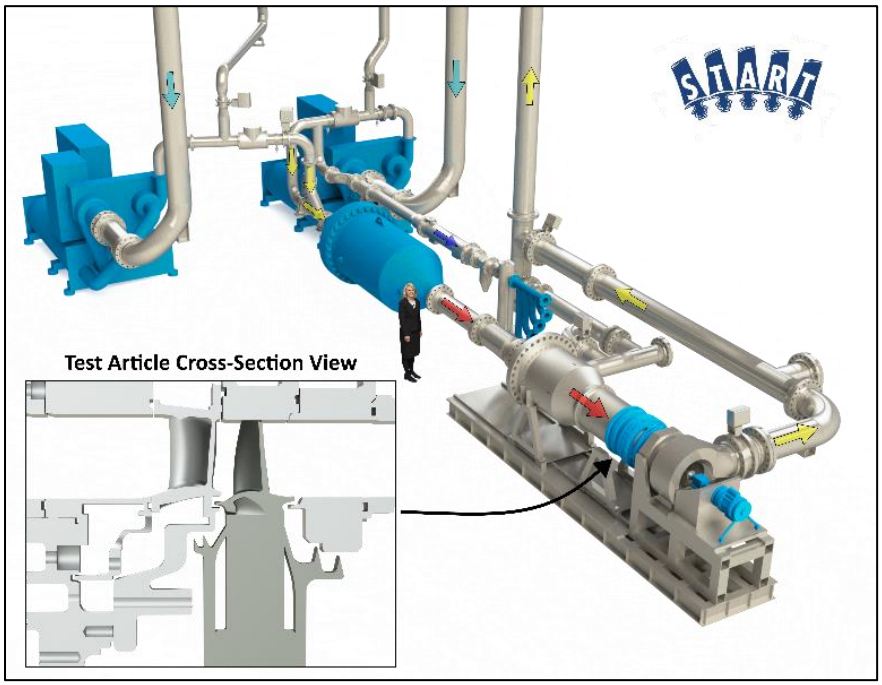


Figure 4.2 – START Lab

The START facility operates continuously to emulate operation of a real-world gas turbine engine. While the facility can induce transient operations, the work in this chapter is conducted under steady operating conditions. The facility utilizes an open-loop flow path that begins with two 1.1 MW compressors that intake air with a maximum flow capacity

of 11.3 kg/s at 480 kPa. The compressors raise the air temperature to approximately 380 K upon exiting the compressor. From the compressor, the air splits into two paths. Most of the air proceeds to a natural gas heater chamber that provides high-temperature MGP flow to the turbine (675 K at nominal flow rate). The remainder of the air is diverted to a heat exchanger that cools the air to about 273 K. This air is then distributed through various cooling paths to the turbine, providing coolant to the blades and purge flow to the rim seal. The high-temperature MGP flow then enters the test section comprised of a single-stage turbine. After passing through the test section, the hot fuel-air mixture exits the system through the exhaust. There are several sensors placed within the test-rig to monitor temperatures, pressures, and flow rates. These sensors are described later in the chapter as they become relevant to the proposed condition monitoring methodologies.

4.1 Correlating Time-Resolved Pressure Measurements with Sealing Effectiveness for Real-time Turbine Health Monitoring

4.1.1 Introduction and Literature Review

The reliability of turbine components is important because they impact engine performance and are costly to replace (Tahan, 2017). While several sensing technologies exist to monitor turbine components in the MGP, the same is not true for the under-platform region. While it is possible to place sensors on the static components in the under-platform region, this is infeasible for the rotating components which also experience the effects of MGP ingress (Bohn et al., 1995) Since direct sensing of under-platform component temperatures is not feasible, the development of a methodology for quantifying sealing effectiveness is vital for maintaining the health of critical turbine components in the under-platform region. Most of the research regarding sealing effectiveness has focused on understanding the physical dynamics of flows into and out of the under-platform region to aid in the design of more effective rim seal/purge flow configurations. For example, Johnson et al. (1994) provided a summary of various physical process that drive rim seal ingress. In their summary, Scobie et al. (2016) categorized factors for ingress as externally induced (EI), rotationally induced (RI), or a combination of external and rotational (CI). While rim seal/purge flow configuration design is important for efficient turbine operation, there remains a demand for a methodology to perform real-time condition monitoring on the rim seal effectiveness once a design is implemented.

Quantification of the sealing effectiveness has been primarily accomplished using the gas concentration method (Clark et al., 2006), which uses a tracer gas (such as carbon

dioxide) to distinguish between purge flow and MGP flow. While this methodology can be implemented in experimental rigs, it is not practical for fielded turbines. Fortunately, a functional relationship between sealing effectiveness and purge flow rate has been observed. Therefore, research efforts have focused on formalizing this functional relationship. In a two-part paper, Owen modeled the rim seal as an orifice with passages for ingress and egress. In these papers, the author derives a set of "orifice equations" for RI ingress (Owen, 2010a) and both EI and CI ingress (Owen, 2010b). These orifice equations demonstrate how the sealing effectiveness depends primarily on the purge flow rate. Sangan et al. utilized these equations to derive "effectiveness equations" for EI ingress (Sangan et al., 2012a) and for RI ingress (Sangan et al., 2012b). These equations include parameters that can be estimated through statistical methods applied to experimental data. Given these parameter estimates, the sealing effectiveness can be extrapolated by inputting known engine conditions into the effectiveness equations. A comparison of theoretical results with experimental results from applying the effectiveness equations was provided by Owen (2012).

Using the link between the test-rig and engine conditions, Owen et al. (2014) developed a methodology for predicting sealing effectiveness using pressure measurements. This methodology is very useful as it only requires a few pressure measurements. However, challenges in implementing this methodology include extensive experimentation and simulation. Furthermore, this methodology along with the other physics-based models assume that the sealing effectiveness increases smoothly with increasing purge flow rate. However, for some rim seal geometries, inflection points in the

sealing effectiveness have been observed. For example, consider the relationship between sealing effectiveness and purge flow rate graphed in Figure 4.3.

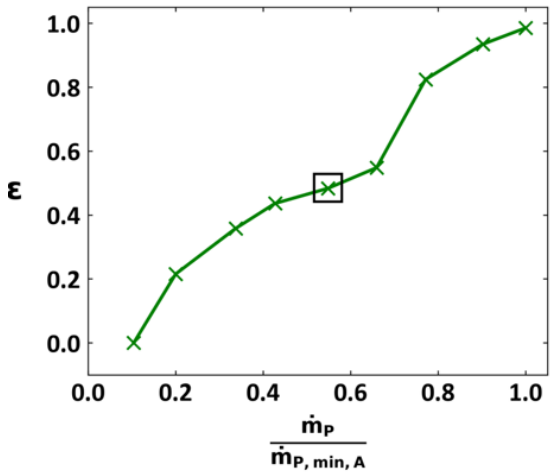


Figure 4.3 – Sealing Effectiveness vs. Normalized Purge Flow Rate

In Figure 4.3, ε is the sealing effectiveness whereas $\frac{\dot{m}_p}{\dot{m}_{p,min,A}}$ is the purge flow rate normalized by the minimum purge flow rate required to prevent ingress into the under-platform region. The rate at which ε increases starts to slow until the normalized purge flow rate reaches 0.6 where the rate increases again. While Scobie et al. (2016) demonstrated accurate predictions of sealing effectiveness using the model by Owen et al. (2014), Clark et al. (2018) showed that the orifice model was not sufficient for capturing the inflection point since it does not account for the complex flow physics that occur in some rim seal geometries. Therefore, application of the model by Owen et al. (2014) for gas turbines with certain rim seal geometries is likely to yield large prediction errors.

Due to the variety of rim seal/purge flow configurations, it is very challenging to develop a generalizable physics-based model. Therefore, this section presents two data-driven methodologies for predicting sealing effectiveness. The advantage of these

methodologies is that they are robust to different rim seal/purge flow configurations. Furthermore, they can be implemented with one pressure sensor located outside rim seal cavity. This contrasts with the physics-based model by Owen et al. (2014) that required multiple sensors, including one in the wheel-space region. While a critique of data-driven models is their lack of interpretability, this critique is addressed using regularization, a machine learning technique that can select a subset of informative features for modeling. Thus, the proposed modeling strategies can quantify the importance of features from the pressure signal for predicting sealing effectiveness. The methodologies presented are both inherently data driven. However, the second methodology incorporates domain knowledge into the formulation of its models. Since it requires the fitting of two models, the second methodology is labeled two-step (2S) whereas the first methodology is labeled data-driven (DD), The data used to develop these methodologies was acquired through an experiment using the Turbine Test-rig. This experiment is described in Subsection 4.1.2.

4.1.2 Experimental Methodology

The Turbine Test-rig is used to generate a dataset for studying the relationship between sealing effectiveness and pressure sensor data. A cross-section view of the Turbine Test-rig is displayed in Figure 4.4.

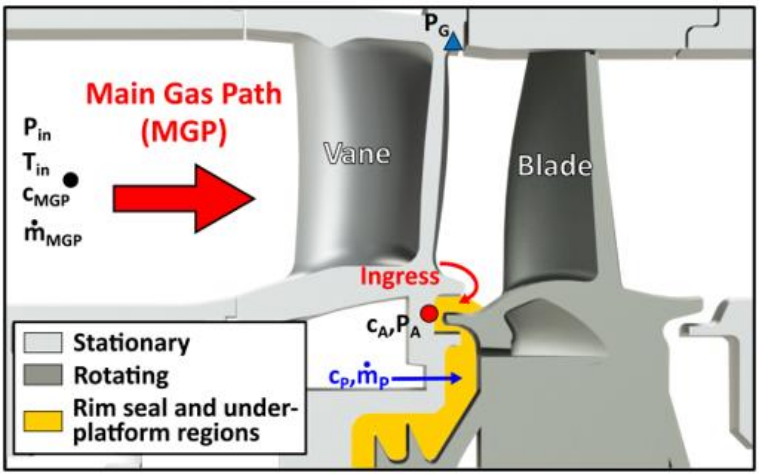


Figure 4.4 – Cross-section View of Single-stage Turbine Test-rig

For this experiment, data was collected under two turbine operating points, where an operating point is defined by a unique combination of turbine settings. These settings include the inlet absolute total pressure (P_{in}), total pressure ratio (P_{in}/P_{out}), mass flow rate of the fuel-air mixture ($\dot{m}_{MGP}$), rotating speed of the blade (RPM), and the inlet temperature (T_{in}). While the exact values for these settings cannot be disclosed for proprietary reasons, the relative difference between the operating points can be seen in Table 4.1.

Table 4.1 – Operating Conditions for OP2

Parameters	OP2
Inlet Absolute Total Pressure, P_{in}	0.9×OP1
Total Pressure Ratio, (P_{in}/P_{out})	0.82×OP1
Mass Flow Rate, $\dot{m}_{MGP}$	0.88×OP1
Rotating Speed, RPM	0.88×OP1
Inlet Temperature, T_{in}	1×OP1

For each operating point, the purge flow rate $(\dot{m}_P)$ was set to a value less than or equal to the minimum purge flow rate required to fully seal the rim cavity at sensor P_A $\left(\dot{m}_{P,min,A}\right)$. At each of these set points, pressure signals were sampled at a non-dimensional frequency of $f_s/f_D \approx 600$ for 500 revolutions. The signals were sampled from two locations: the rim seal (P_A) and the outer casing (P_G).

To determine the sealing effectiveness, the gas concentration method utilizing a carbon dioxide tracer gas was used to determine the hot fuel-air mixture concentration at the inlet (c_{MGP}), the rim seal (c_A), and at the purge flow location (c_P). The sealing effectiveness is thus calculated using Equation 4.1.1

$$\varepsilon = \frac{c_A - c_{MGP}}{c_P - c_{MGP}} \tag{4.1.1}$$

In total, nine different purge flow rates are utilized. The relationship between the purge flow rate and the sealing effectiveness is displayed in Figure 4.3, where the purge flow rate is normalized by $\dot{m}_{P,min,A}$. 9 samples of pressure signals were recorded for each purge flow rate for OP1 and 18 samples were recorded for each purge flow rate for OP2. There are two exceptions for OP2, where the first purge flow rate has 14 samples and the final purge

flow rate has 16 samples. In this study, two methodologies for modeling sealing effectiveness using the pressure signals are proposed. These methodologies are first developed for Operating Point 1 (OP1). After comparison of the two methodologies, the sensitivity of the more accurate methodology to the settings of OP2 is analyzed.

4.1.3 Data-driven Methodology

The data-driven methodology starts with performing initial preparation of the raw pressure signal for analysis. A low pass filter is applied to remove frequency content above 30 f/f_D because most of the salient information lies below this frequency and because the blade pass frequency needed to be removed for proprietary reasons. To extract features from the pressure signal, a fast Fourier transform (FFT) is performed and the energy content within consecutive frequency intervals centered at integer values of the nondimensional frequency (f/f_D) is aggregated. Following feature extraction, the 81 observations are randomly partitioned into training and testing sets. For the training set, Linear Regression with Lasso is used to select informative frequency bands and to fit a predictive model. Then, the sealing effectiveness is predicted for the test observations. To assess the general performance of this methodology, the partitioning of the data into training and testing was repeated 250 times and the median and range of the prediction error are reported.

4.1.3.1 Feature Extraction Using Discrete Fourier Transform

Let $\tilde{P}[t], t = 0,1,, \dots, N-1$ denote the time-resolved pressure signal following application of the low-pass filter to the raw sensor signal, where N is the length of the signal. The filtered pressure signal is displayed in Figure 4.5.

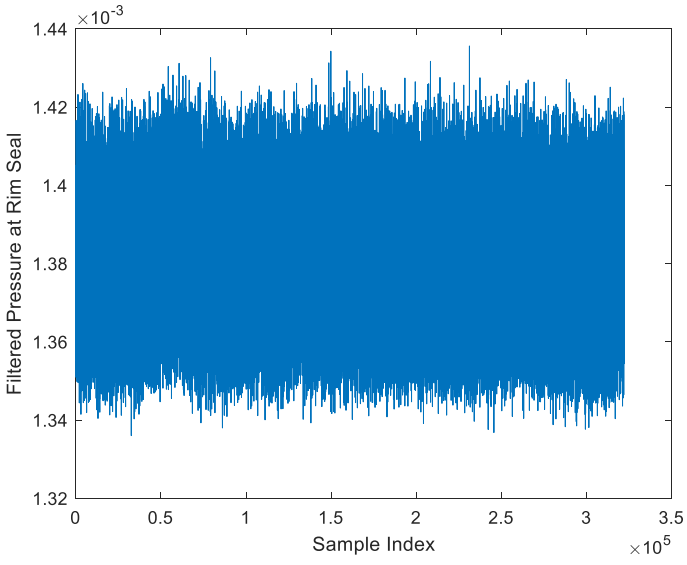


Figure 4.5 – Filtered Pressure Signal Recorded at Rim Seal

To extract features, the data in Figure 4.5 is transformed using the FFT. The FFT is a computationally efficient algorithm that performs a discrete Fourier transform of a time-resolved signal, transforming the signal into the frequency domain. This is an appropriate strategy because the signal does not exhibit any apparent nonstationary behavior, which is expected as the pressure sensor is monitoring rotational movement under static operating conditions. The equation for the discrete Fourier transform is shown below:

$$y\left[\frac{F_s}{N}k\right] = \sum_{t=0}^{N-1} \tilde{P}[t] \exp\left(-j\frac{2\pi}{N}kt\right), k = 0,1,\dots,N-1 \qquad (4.1.2)$$

The discrete Fourier transform for the filtered pressure signal is displayed in Figure 4.6 where amplitude equals the absolute value of $y\left[\frac{F_s}{N}k\right]$.

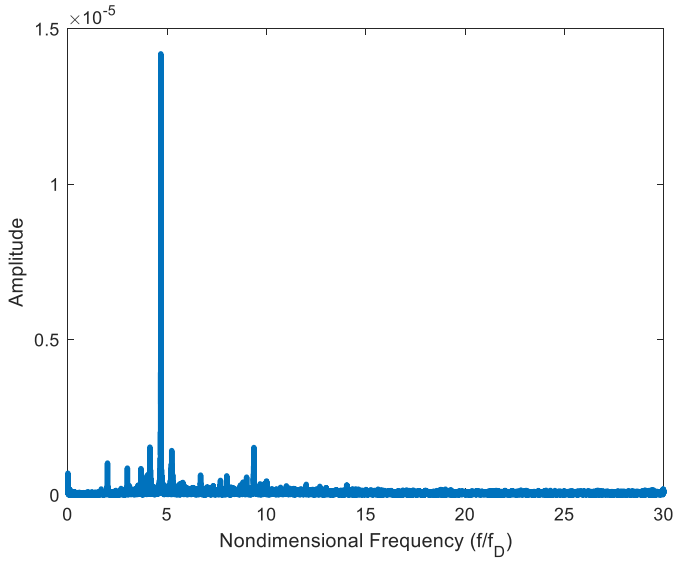


Figure 4.6 – Discrete Fourier Transform of Filtered Pressure Signal Recorded at Rim Seal

Due to the filtering in the Data Preparation step, the nondimensional frequency range is truncated at $f/f_D = 30$. Features are extracted by creating frequency bands of length $\Delta(f/f_D) = 1$ centered at discrete nondimensional frequency values. For example, the frequency band for feature 1 consists of frequencies between 0.5 and 1.5. Within this frequency band, the energy of the discrete Fourier transform amplitudes is calculated as the sum of squared amplitudes. This is formalized as follows:

$$x_l = \sum_{\frac{F_s}{N}k=l-0.5}^{l+0.5} \left| y\left[\frac{F_s}{N}k\right] \right|^2, l = 1,2,\dots,29 \tag{4.1.3}$$

Given these features and the observed sealing effectiveness values, the diagnostics model is trained.

4.1.3.2 Train Linear Regression with Lasso Model

Following feature extraction, the data is randomly partitioned into a training and test set. A 75-25 percent split where 75% of the data is selected for training and the remaining 25% is used for testing. This corresponds to 7 samples per purge flow rate setting used for training and the remaining 2 samples used for testing.

The diagnostics model used is a Linear Regression with Lasso model. Lasso is a regularization strategy used to improve prediction accuracy by shrinking the regression coefficients, reducing the variance of the model at the cost of incurred bias (Tibshirani, 1996). Since the regression coefficients of some predictors are reduced to zero, Lasso can perform variable selection. Prior to fitting the model, the predictors are standardized. Since this ensures that all predictors are of equal scale, the amplitude of the regression coefficients reflect the relative importance of the features to predicting sealing effectiveness. Let M denote the total number of samples. Standardization is performed as follows:

$$\tilde{x}_{i,l} = \frac{x_{i,l} - \bar{x}_l}{s_l}, l = 1,2, \dots ,29 \tag{4.1.4}$$

where $\bar{x}_l = \frac{1}{M}\sum_{i=1}^{M} x_{i,l}$ and $s_l = \frac{1}{M-1}\sum_{i=1}^{M}\left(x_{i,l} - \bar{x}_l\right)$ are the sample mean and sample standard deviation respectively. Let ε_i denote the ith observed sealing effectiveness and $\tilde{\boldsymbol{x}}_i = \left(\tilde{x}_{i,1}, \tilde{x}_{i,2}, \dots , \tilde{x}_{i,29}\right)^T$ denote the ith vector of predictors. To fit Linear Regression with Lasso, the following optimization problem is solved.

$$MD = -2\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log\left(\frac{P(Y_i = k|\boldsymbol{X}_i = \boldsymbol{x}_i)}{y_{i,k}}\right) \tag{4.1.5}$$

In Equation 4.1.5, $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^{29}$ are the bias term and the vector of regression coefficients respectively. Solving the optimization problem in Equation 4.5 involves finding $(\hat{\beta}_0, \widehat{\boldsymbol{\beta}})$ that minimizes the objective function given the tuning parameter $\lambda \in \mathbb{R}_{\geq 0}$. In practice, λ is selected via 10-fold cross-validation. The problem in Equation 4.1.5 is convex and therefore can be solved via coordinate descent.

4.1.3.3 Predicting Sealing Effectiveness Using Linear Regression with Lasso

Once the model is fitted, the sealing effectiveness can be monitored by sampling the pressure signal, extracting the frequency band features, and using the features as inputs into the Linear Regression with Lasso model. The sealing effectiveness is predicted as follows:

$$\hat{\varepsilon} = \hat{\beta}_0 + \tilde{\boldsymbol{x}}_*\widehat{\boldsymbol{\beta}} \tag{4.1.6}$$

where $\hat{\varepsilon}$ is the predicted sealing effectiveness for standardized input predictors $\tilde{\boldsymbol{x}}_*$ with estimated regression coefficients $\hat{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$**.** To evaluate the performance of the proposed modeling approach, the root mean squared error (RMSE) between the actual sealing effectiveness and the values predicted by the model is computed. Since this error can depend on the random partitioning of the data into training and testing, the random partitioning and model fitting is repeated 250 times and descriptive statistics on the errors are reported. In Figure 4.7, the median RMSE over all 250 iterations along with the range of the RMSE are displayed for both sensor locations.

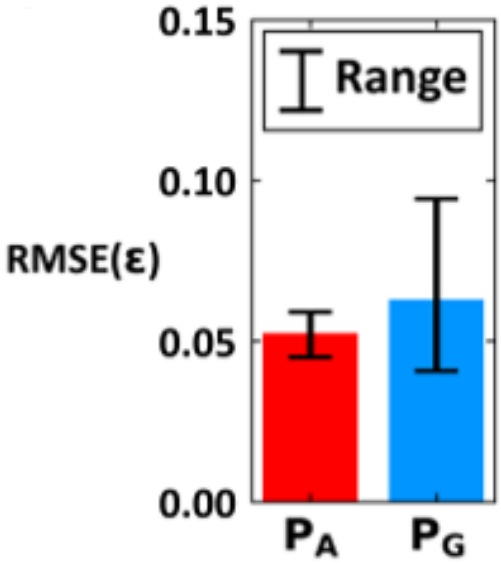


Figure 4.7 – Distribution of RMSE for Two Sensor Locations

The first observation from Figure 4.7 is that the rim seal sensor provides higher accuracy in predicting sealing effectiveness than the outer casing. The errors also have less variability for the rim seal sensor than for the outer casing sensor. This is expected since the rim seal sensor is closer to the source of purge flow than the outer casing sensor. The rim seal Linear Regression model is capable of distinguishing between the following intervals of sealing effectiveness: $[1.2 \times 10^{-4}, 0.22], 0.36, [0.44,0.48], 0.55,0.82,$ and $[0.93,0.99]$. This can be seen in Figure 4.8 by analyzing the boxplots of predictions at each sealing effectiveness setpoint.

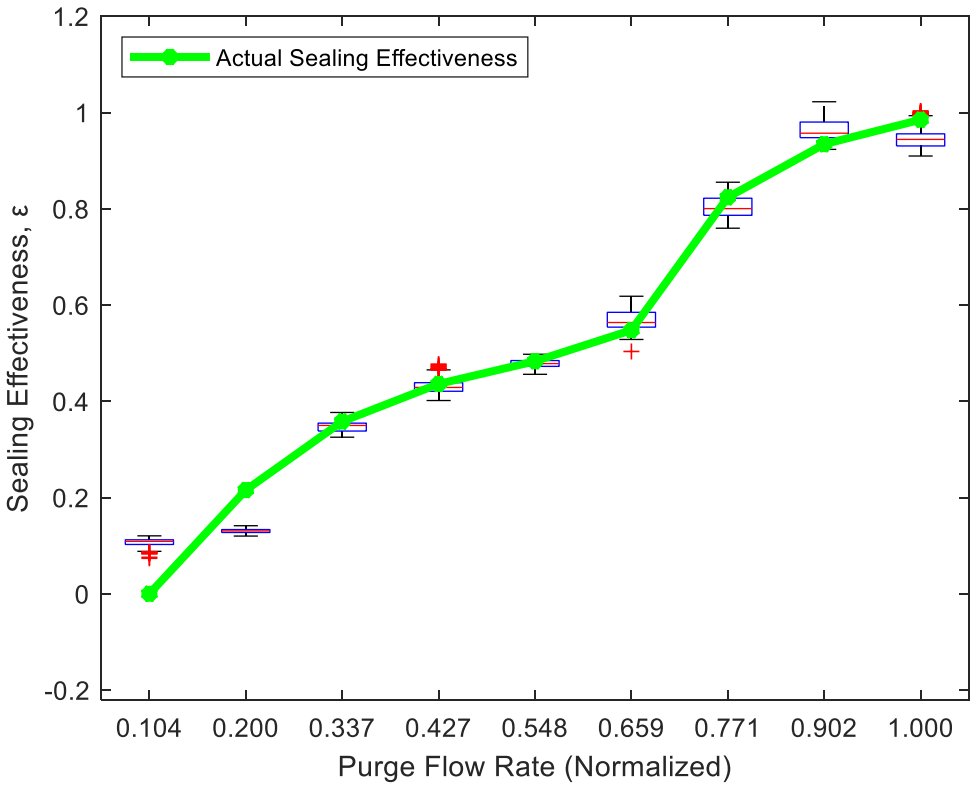


Figure 4.8 – Predicted ε at Rim Seal for DD Methodology

From Figure 4.8, it can also be seen that the locations with the lowest accuracy are near the boundaries of the sealing effectiveness. For comparison, Figure 4.9 displays the sealing effectiveness predictions made using sensor data from the outer casing. The predictions for the outer casing sensor are more accurate than the rim seal predictions for the low sealing effectiveness values. However, the predictions are not very precise as there is strong overlap between boxplots for different sealing effectiveness values. Therefore, stratifying the sealing effectiveness into distinct intervals is much more difficult with the outer casing predictions than the rim seal predictions.

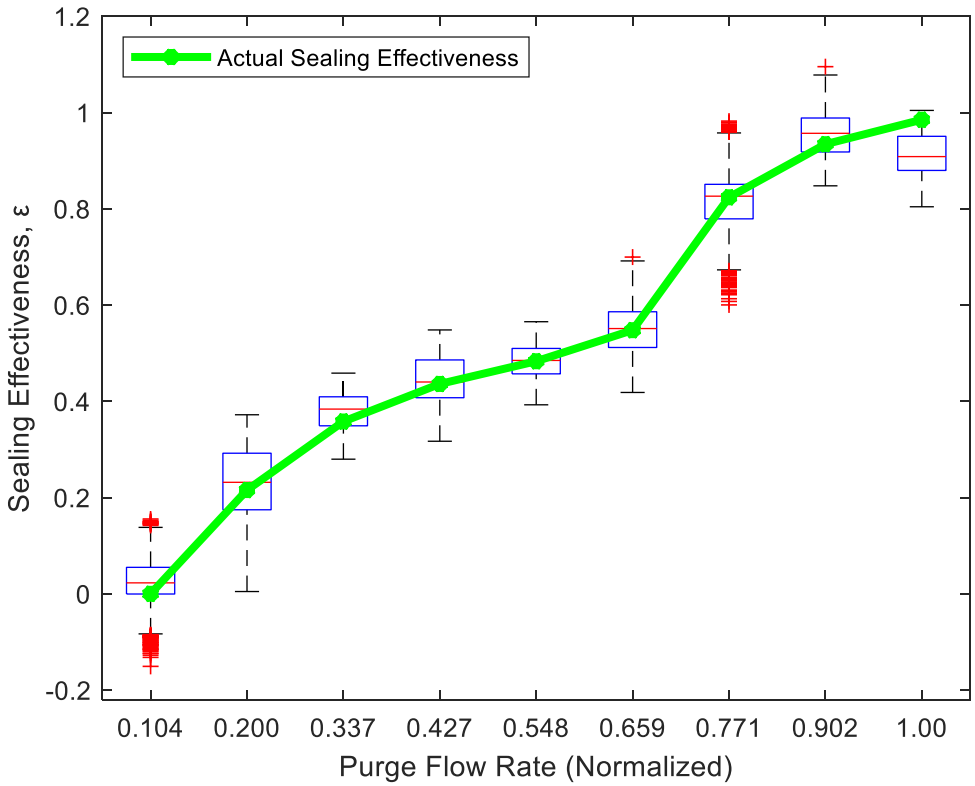


Figure 4.9 – Predicted ε at Outer Casing for DD Methodology

From analyzing Figure 4.8 and Figure 4.9, the data-driven model can capture the inflection in the relationship between the sealing effectiveness and the purge flow rate that the physics-based models were not capable of doing.

In Figure 4.10, the average regression coefficients over all 250 partitions are displayed.

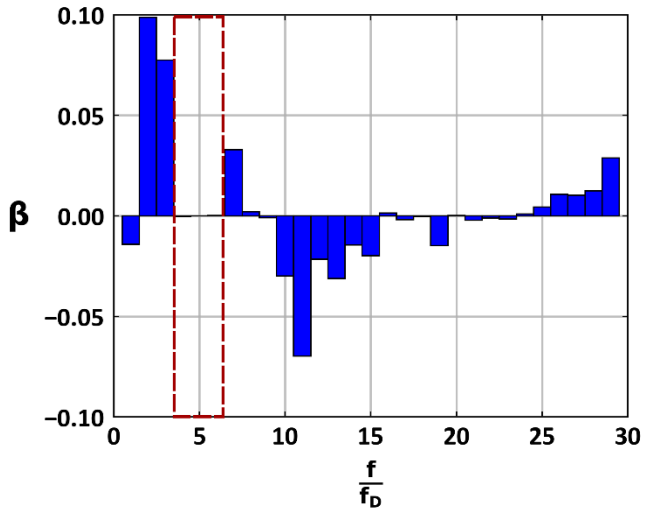


Figure 4.10 – Average Regression Coefficients Over All 250 Partitions

By analyzing the regression coefficients, it can be seen that $\bar{x}_5$ has a zero coefficient. The lack of significance of this feature is interesting given that $f/f_D = 5$ is the dominant frequency in Figure 4.6. By plotting the sealing effectiveness vs. $\bar{x}_5$ in Figure 4.11, the lack of significance is shown to be a result of the nonlinear relationship between the two variables.

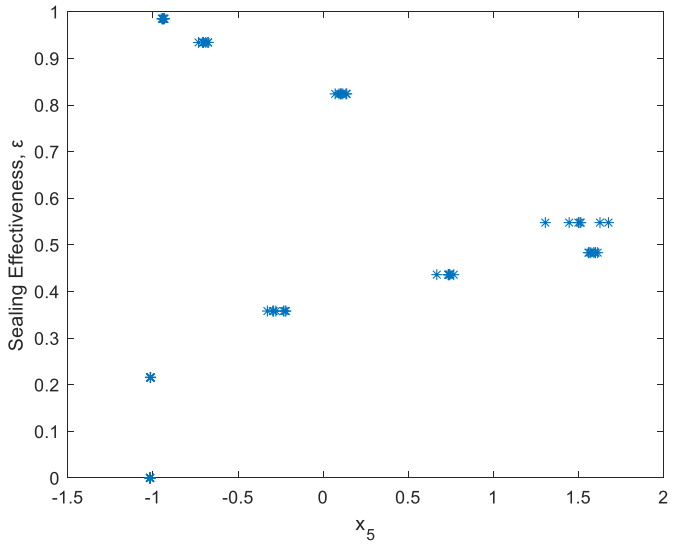


Figure 4.11 – Nonlinear Relationship Between Sealing Effectiveness and Frequency Band 5

The dominant frequency is caused by rotating flow structures that increase in strength as the purge flow increases. However, these rotating flow structures begin to weaken once the purge flow reaches a normalized value of about 0.55. Therefore, while the feature $\tilde{x}_5$ is not linearly correlated with the sealing effectiveness, there are distinct regions of strong correlation. In the two-step methodology this domain knowledge of the physics regarding this feature is leveraged to improve prediction accuracy.

4.1.4 Two-step Methodology

Through the feature extraction step, the two-step methodology is identical to the data-driven methodology. After partitioning the data into a training and test set, two models are trained. The first model is designed to obtain a coarse prediction of the sealing effectiveness. More specifically, Logistic Regression with Lasso is fitted. This model maps

the 29 frequency band features to one of two states: $\varepsilon \geq 0.5$ and $\varepsilon < 0.5$. To fit this model, the following optimization problem is solved:

$$\min_{(\theta_0, \boldsymbol{\theta}) \in \mathbb{R}^{29+1}} -\left[\frac{1}{M}\sum_{i=1}^{M} z_i(\theta_0 + \tilde{\boldsymbol{x}}_i^T\boldsymbol{\theta}) - \ln\left(1 + e^{\theta_0 + \tilde{\boldsymbol{x}}_i^T\boldsymbol{\theta}}\right)\right] + \eta\|\boldsymbol{\theta}\|_1 \qquad (4.1.7)$$

where $z_i = 1$ if $\varepsilon_i \geq 0.5$ and $z_i = 0$ otherwise. Solving this problem involves finding the parameters $\left(\hat{\theta}_0, \hat{\boldsymbol{\theta}}\right)$ that minimize the objective function in Equation 4.1.7 given the tuning parameter $\eta \in \mathbb{R}_{\geq 0}$. Like with λ in the Linear Regression with Lasso model, η is selected via 10-fold cross-validation. The objective function in Equation 4.1.7 is convex and so the problem can be solved via coordinate descent.

In the second model, a model is fitted to obtain a more granular prediction of the sealing effectiveness given that whether $\varepsilon \geq 0.5$ or $\varepsilon < 0.5$ is known. The model fitted is a Linear Regression with Lasso using $\tilde{x}_5, \tilde{x}_{10}, \tilde{x}_{15}, \tilde{x}_{20}, \tilde{x}_{25}$ and the natural logarithm of these features as predictors. These features represent the first five harmonics of the dominant frequency. The use of the natural logarithm is to account for the curvature seen in Figure 4.10. Included in the Linear Regression with Lasso model is the binary variable z, which equals 1 if the sealing effectiveness is greater than or equal to 0.5 and equals 0 otherwise. For this problem, following optimization problem is solved:

$$\min_{\gamma_0,\omega_0,\{\gamma_k,\omega_k,\phi_k,\psi_k\}_{k=1}^{5}\in\mathbb{R}^{22}} \sum_{i=1}^{M}\left(\varepsilon_i - \gamma_0 - \omega_0 z_i - \sum_{k=1}^{5}(\gamma_k + z_i\omega_k)\tilde{x}_{i,5k} - \sum_{k=1}^{5}(\phi_k + z_i\psi_k)\ln\tilde{x}_{i,5k}\right)^2 + \rho\left(|\omega_0| + \sum_{k=1}^{5}(|\gamma_k| + |\omega_k| + |\phi_k| + |\psi_k|)\right) \tag{4.1.8}$$

Like with the previous two models, solving Equation 4.1.8 involves finding the parameters $\hat{\gamma}_0, \hat{\omega}_0, \{\hat{\gamma}_k, \hat{\omega}_k, \hat{\phi}_k, \hat{\psi}_k\}_{k=1}^{5}$ that minimize the objective function in Equation 4.1.8 given the tuning parameter $\rho \in \mathbb{R}_{\geq 0}$. Like with λ and η, ρ is determined by 10-fold cross-validation. The objective function is convex and can be solved using coordinate descent.

Once the two models are fitted, the sealing effectiveness is predicted by first sampling the pressure signal and extracting the 29 frequency band features. These features are used to infer whether $\hat{\varepsilon} \geq 0.5$ ($\hat{z} = 1$) or $\hat{\varepsilon} < 0.5$ ($\hat{z} = 0$) using the following decision rule: If $\frac{e^{\hat{\theta}_0 + \tilde{x}_i^T\hat{\boldsymbol{\theta}}}}{1+e^{\hat{\theta}_0 + \tilde{x}_i^T\hat{\boldsymbol{\theta}}}} > 0.5, \hat{z} = 1$, otherwise $\hat{z} = 0$. Next, only $\tilde{x}_5, \tilde{x}_{10}, \tilde{x}_{15}, \tilde{x}_{20}, \tilde{x}_{25}$ and the natural logarithm of these features from the sampled pressure signal are utilized along with the inferred sealing effectiveness state $\hat{z}$ to predict the sealing effectiveness. If $\hat{z} = 0$, the sealing effectiveness is predicted using the following equation:

$$\hat{\varepsilon}_i = \hat{\gamma}_0 + \sum_{k=1}^{5}\hat{\gamma}_k\tilde{x}_{i,5k} + \sum_{k=1}^{5}\hat{\phi}_k\ln\tilde{x}_{i,5k} \tag{4.1.9a}$$

Conversely, if $\hat{z} = 1$, the sealing effectiveness is predicted using Equation 4.9b.

$$\hat{\varepsilon}_i = (\hat{\gamma}_0 + \hat{\omega}_0) + \sum_{k=1}^{5} (\hat{\gamma}_k + \hat{\omega}_k)\tilde{x}_{i,5k} + \sum_{k=1}^{5} \left(\hat{\phi}_k + \hat{\psi}_k\right) \ln \tilde{x}_{i,5k} \qquad (4.1.9b)$$

As with the data-driven methodology, the two-step methodology is analyzed by performing 250 random partitions of the data into training and testing, fitting the two models with the training set, and predicting sealing effectiveness with the test set. The descriptive statistics of the RMSE for the two-step methodology are displayed alongside that of the data-driven methodology for both the rim seal and outer casing in Figure 4.12.

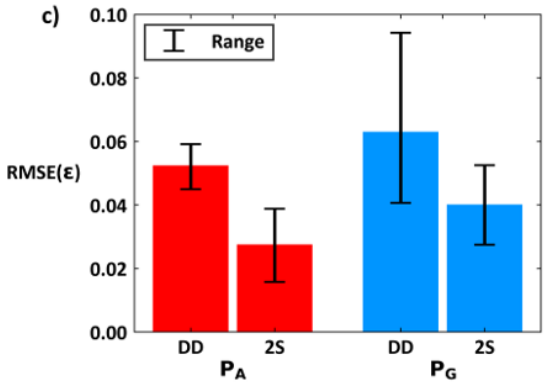


Figure 4.12 – Comparison of Distribution of RMSE between DD and 2S Methodologies

From Figure 4.12, the two-step methodology greatly improves the model accuracy. For the rim seal, the two-step model results in a 48.7% reduction in median RMSE whereas the casing model experiences a 35.1% reduction in median RMSE using the two-step model. In addition to analyzing the overall accuracy, the distribution of predictions at each purge flow rate are shown in Figure 4.13 and Figure 4.14 for the rim seal and outer casing respectively.

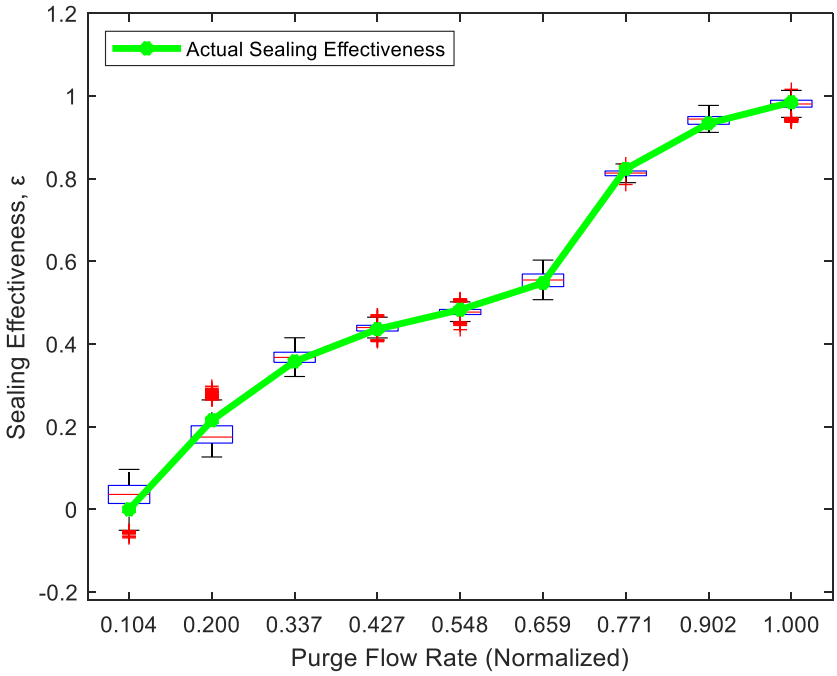


Figure 4.13 – Predicted ε at Rim Seal for 2S Methodology

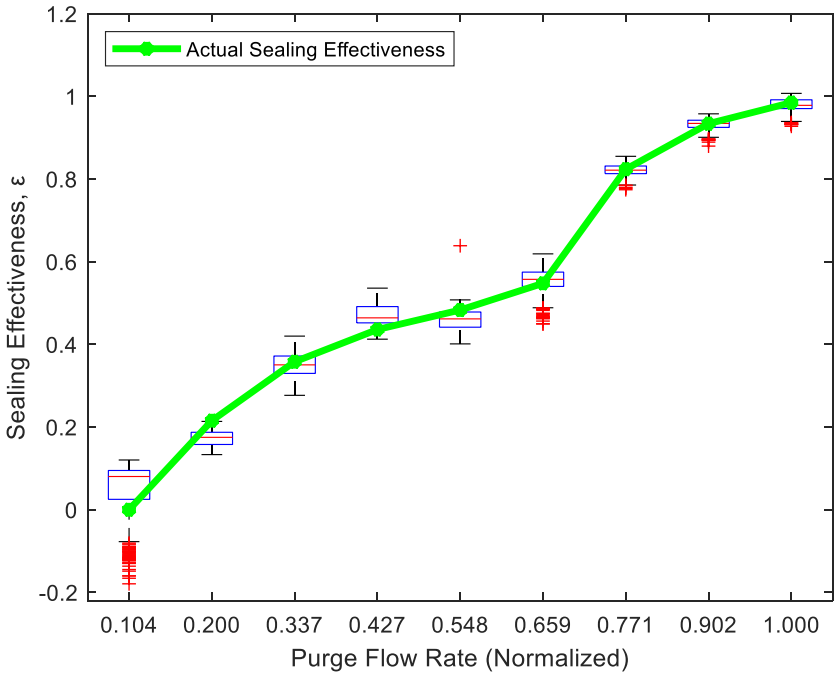


Figure 4.14 – Predicted ε at Outer Casing for 2S Methodology

For predictions made at the rim seal, the deviations at the boundaries in the data-driven methodology have been corrected. Furthermore, for each purge flow rate, there are no 75^{th} percentile values of predicted sealing effectiveness that overlap with the 25^{th} percentile of the succeeding purge flow rate. This indicates that there is strong distinction between predictions made at different purge flow rates. For predictions made at the outer casing, only the predictions made at normalized purge flow rate of 0.427 show significant deviation from the actual sealing effectiveness. Otherwise, the predicted sealing effectiveness at the outer casing are distinctive as well.

4.1.5 Application of Methodology to Other Operating Points

The results observed thus far have demonstrated that the two-step methodology is capable of accurately predicting the sealing effectiveness. Thus far, the model has only been assessed at one operating point. However, the applicability of this methodology to other operating points is also analyzed. Table 4.1 displayed the relative difference between OP1 and OP2. The two-step methodology was performed using the data from OP2 and the results are compared to those from OP1. The comparison utilized the predictions made using the outer casing sensor since this is more likely to be feasible in a real-world scenario. Furthermore, both datasets were combined into one and a combined model is trained. Since the frequencies are nondimensionalized by the disk rotating frequency, the features are comparable between operating points despite OP2 operating at a slower rpm. Figure 4.15 displays the comparison between these three models.

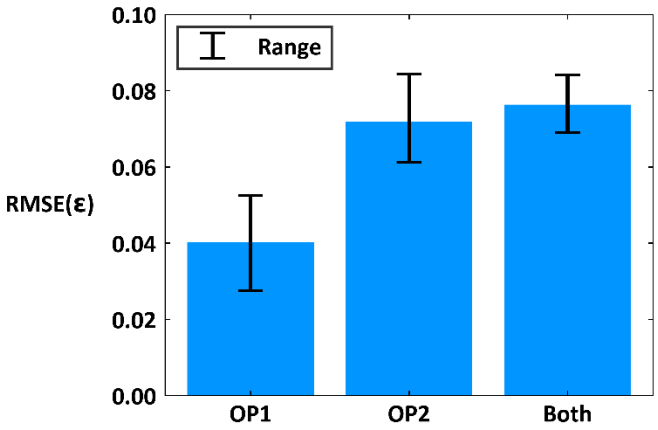


Figure 4.15 – Comparison of 2S Model for Different Operating Points

Figure 4.15 shows that the median error for OP2 is nearly double that of OP1. Furthermore, the combined model has a similar median RMSE as OP2. The actual sealing effectiveness vs. predicted sealing effectiveness plot in Figure 4.16 shows that the functional relationship between the sealing effectiveness and the purge flow rate for OP2 is different from that of OP1. Furthermore, the increase in median RMSE for the outer casing is due to large errors when predicting low sealing effectiveness, the region where the functional relationship is most different. However, the predictions at the other purge flow rates are accurate.

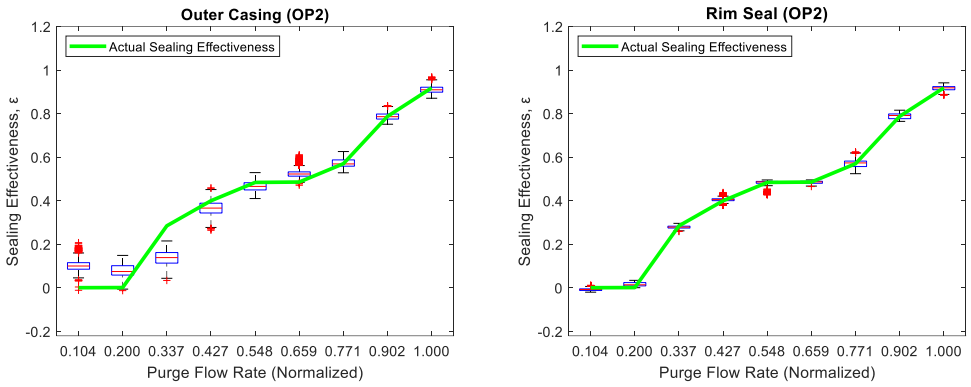


Figure 4.16 – Comparison of 2S Model at OP2 for Outer Casing (Left) and Rim Seal (Right)

The right-hand side of Figure 4.16 shows that modeling with the rim seal yields very accurate results. In fact, the median RMSE for OP2 using the rim seal is 0.013, about half the median RMSE for OP1 using the rim seal. This indicates that the increase in error for the outer casing is a result of attenuation of the dominant frequency from the rim seal to the outer casing obscuring the correlation between the sealing effectiveness and the dominant frequency. This can be seen from the scatterplot of the sealing effectiveness vs. the dominant frequency $\tilde{x}_5$ displayed in Figure 4.17.

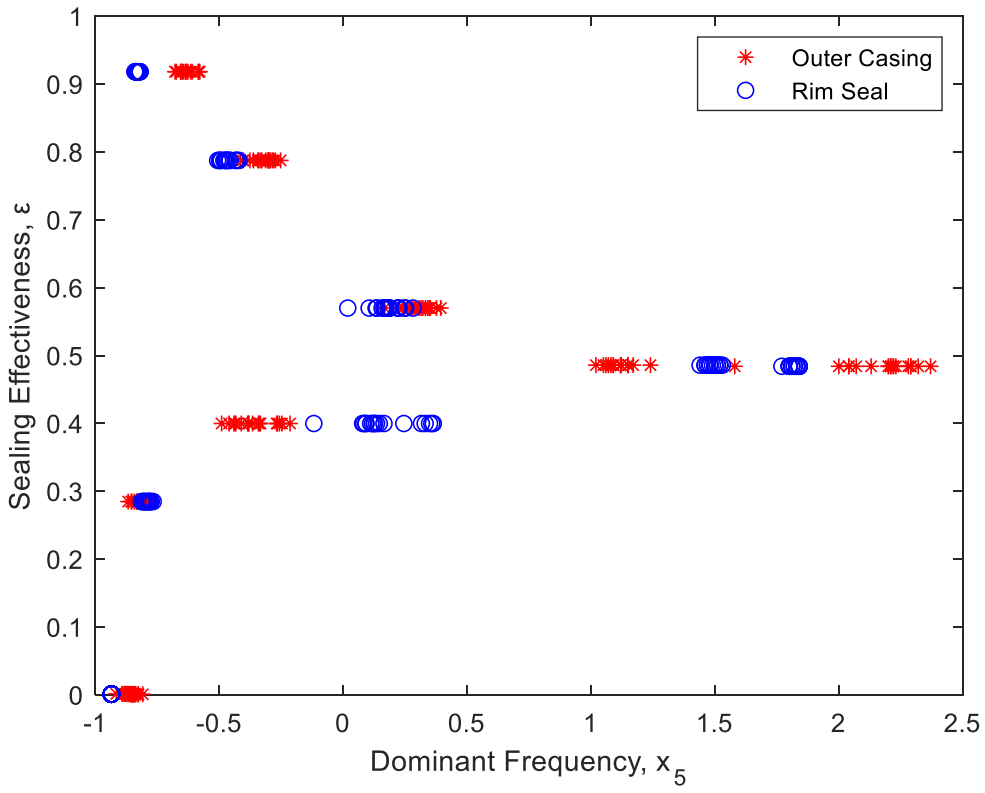


Figure 4.17 – Scatterplot of Sealing Effectiveness vs. Dominant Frequency (OP2)

In Figure 4.17, the rate of growth in the sealing effectiveness is higher for the outer casing than the rim seal. However, in the experiment, the purge flow is being altered and the pressure signal changes in response to the corresponding change in sealing effectiveness. Therefore, the steep rise in sealing effectiveness indicates that the outer casing is not capable of detecting the change in sealing effectiveness due to the purge flow rate changing from 0.104 to 0.200. This compromises model adequacy and results in high prediction errors. Due to its proximity to the purge flow, the rim seal sensor is more responsive to this change and thus, achieves very accurate results. This indicates that the two-step methodology is sound and is only limited by the outer casing sensor's lack of responsiveness to the impact of a different operating point on the functional relationship

between the purge flow rate and the sealing effectiveness. While application to other operating points is needed for a full study on the robustness of the methodology, these results are positive indicators that the two-step methodology can be applied for different operating points.

4.1.6 Conclusion

In this study, two methodologies were developed for performing real-time monitoring of sealing effectiveness in an industry-class gas turbine. These methodologies were trained using time-resolved pressure sensors generated by the Turbine Test-rig at engine-relevant conditions and with engine-relevant hardware. Two pressure sensors, one located at the rim seal near the purge flow and the other located on the outer casing were installed to monitor the system under various purge flow rates. The signals from these pressure sensors were utilized to develop two methodologies for predicting sealing effectiveness.

The first methodology (DD) utilized a Linear Regression with Lasso model to directly model the sealing effectiveness as a linear function of frequency features extracted from the pressure signal. The second methodology consists of a two-step approach (2S) that incorporates domain knowledge. First. Logistic Regression with Lasso with all features is used to obtain a coarse prediction of the sealing effectiveness and then Linear Regression with Lasso is performed using only the dominant frequency and its harmonics as predictors to perform a more granular prediction. Both methodologies demonstrated the capability of accurately predicting the sealing effectiveness. However, the two-step approach was

capable of reducing the median RMSE of the predictions by 48.7% and 35.1% for the rim seal and outer casing sensor respectively.

Due to its higher accuracy, the two-step methodology was tested on another operating point. The conditions of this operating point altered the functional relationship between the sealing effectiveness and the purge flow rate, compromising the outer casing's ability to detect sealing effectiveness changes when it is low. This resulted in the median RMSE almost double that of OP1. However, the outer casing is capable of distinguishing between higher sealing effectiveness values. Implementing the two-step methodology for OP2 using the rim seal resulted in higher accuracy than for OP1. Therefore, the two-step methodology using the rim seal sensor can perform real-time health monitoring the turbine sealing effectiveness.

4.2 Applying Infrared Thermography as a Method for Online Monitoring of Turbine Blade Coolant Flow

4.2.1 Introduction and Literature Review

To ensure efficient turbine operation, modern gas turbines operate with inlet temperatures surpassing 1673 K (Bogard and Thole, 2006). This is greater than the 1588.7 K incipient melting temperature for nickel-based turbine blades (Koff, 2004). The high temperatures render the turbine blades susceptible to faults such as: high cycle fatigue, oxidation, sulphidation, hot corrosion, creep, and erosion (Meher-Homji and Gabriles, 1998). To mitigate the risk of these faults, gas turbines are equipped with a combination of external and internal cooling mechanisms. Externally, cold air from the compressor is routed to the turbine section to cool the turbine components. The blades themselves are designed with internal cooling passages that deliver coolant to the blade surface via film cooling holes. Koff (2004) demonstrated that turbine blades designed with these cooling mechanisms enable operation at inlet temperatures over 588 K greater than the incipient melting temperature. Since inlet temperature is intrinsically linked to turbine efficiency, the integrity of turbine cooling mechanisms is important for ensuring high turbine efficiency.

Despite the importance of the cooling flow to turbine blade health, gas turbine health monitoring methodologies in the literature do not focus on monitoring blade coolant. Instead, condition monitoring efforts have focused on degradation of the turbine blade. For example, vibration sensors have been used to detect both geometric and structural turbine blade faults. Cox et al. (2015) utilized offsets in the peaks of a sinusoidal vibration signal

to detect bending in the turbine blade. Meanwhile, Lim and Leong (2010) used wavelets to obtain a time-frequency representation of the vibration spectrum, enabling discrimination of creep rub and eccentricity rub in the turbine blade. Blade tip clearance data has also been utilized for monitoring turbine blade health. For example, Goel et al. (2008) proposed percentile ranking of the sample mean of blade tip clearance data to detect blade tip degradation. While these techniques can detect turbine faults, their suitability for detecting coolant flow faults is limited since the techniques would only capture the resulting effect of a coolant flow loss. Conversely, infrared imaging has been shown to detect losses in thermal barrier coating, which is used in conjunction with the cooling mechanisms (LeMieux, 2005, Markham et al., 2014). Therefore, this section studies the efficacy of infrared imaging as a tool for monitoring coolant flow.

To assess the efficacy of infrared imaging, various models are proposed that map features extracted from the infrared image of the blade to the Coolant Flow Rate. To test the robustness of the models to changes in the Main Gas Path (MGP Temperature, an experiment is conducted where infrared images are recorded for various combinations of MGP Temperature and Coolant Flow Rate. During the experiment, the correlation between the MGP Temperature and the Coolant Flow Rate was affected by latent factors. To control for these factors, additional measurements were recorded and included in the model and the difference in accuracy of predicting the Coolant Flow Rate with and without the additional measurements is analyzed. To conclude, the potential for thermal imaging to inform placement of more cost-effective thermal sensors is assessed. The remainder of this section is organized as follows. Section 4.2.2 describes the experimental approach and data collection while Section 4.2.3 describes how features are extracted from the thermal

images. In Section 4.2.4, the regression model that forms the foundation for all models proposed is formulated. Section 4.2.5 details the results of the various modeling approaches. The findings are summarized in the conclusion in Section 4.2.6.

4.2.2 Experiment

For this project, data acquisition was conducted as follows. An infrared camera was used to record images of the gas turbine blade. A diagram of this can be seen in Figure 4.18.

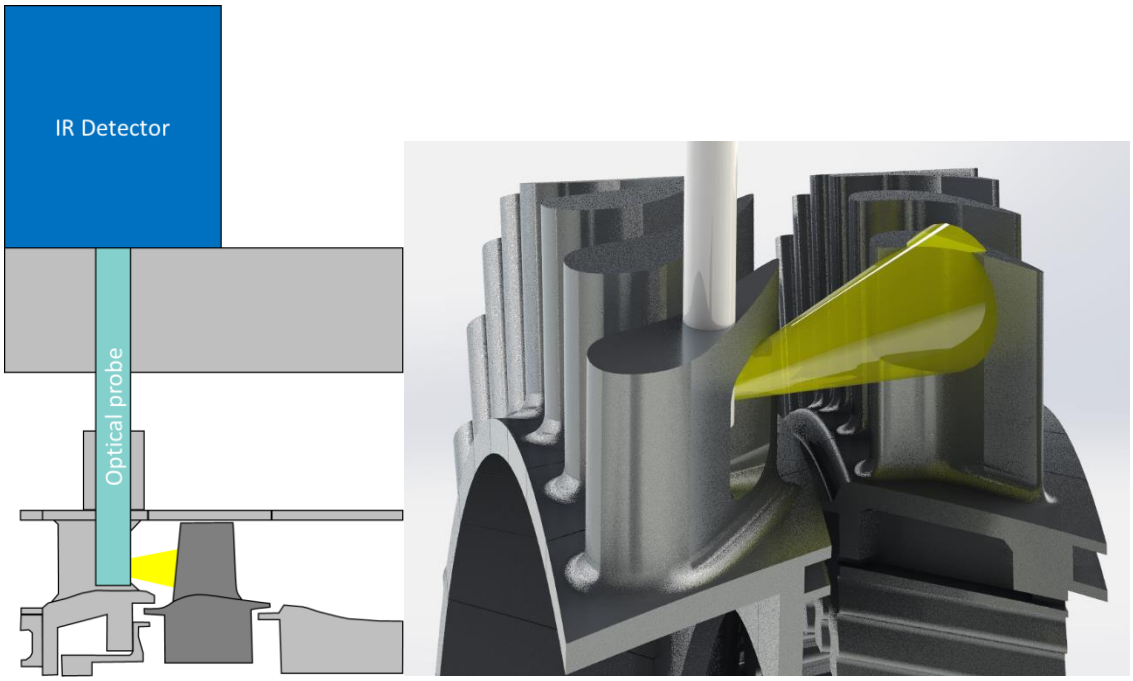


Figure 4.18 – Cross-section View of Turbine with Camera (Left), 3D Diagram of Camera and Blade (Right)

Images were recorded for various combinations of MGP Temperature (T_{MGP}) and Coolant Flow Rate ($\dot{m}_{cool}$). In total, four values of T_{MGP} and five values of $\dot{m}_{cool}$ were considered. The number of samples for each combination are displayed in Table 4.2.

Table 4.2 – Number of Samples for Each $T_{MGP}/\dot{m}_{cool}$ Combination

Coolant Flow Rate	Main Gas Path Temperature				
	Lv 1. 0.966	**Lv. 2. 1.395**	**Lv 3. 1.778**	**Lv 4. 2.114**	**Lv 5. 2.314**
Lv 1. 1.324	4	4	23	8	8
Lv 2. 1.391	0	0	16	8	8
Lv 3. 1.428	4	4	4	4	4
Lv 4. 1.445	4	8	16	15	11

Measurements were not all recorded on the same day. Therefore, the effect of lurking variables was observed in the data. To account for this, three ancillary sensor measurements were included with the thermal images. These sensors were the Coolant Temperature (T_{cool}) and the rate of change at the time of measurement of the temperature measured at the inner diameter near the rim seal $\left(\frac{dT_{ID}}{dt}\right)$ and at the outer diameter of the turbine $\left(\frac{dT_{OD}}{dt}\right)$. Ideally, the coolant temperature would be constant, and the rates of change would be zero. However, regulating these parameters is not a trivial task. For example, a positive rate of change in the temperature indicates that the turbine is heating due to an increase in T_{MGP} or a decrease in $\dot{m}_{Cool}$. While an operator may assume that the temperatures in the turbine have reached a steady state, minor differences in the rate of change of the temperature may still have significant effects on the images. Therefore, these additional sensors can aid in estimating the Coolant Flow Rate in the presence of minor transients.

4.2.3 Feature Engineering

In this study, infrared images are used to predict Coolant Flow Rate. The sampled images for the extreme combinations of T_{MGP} and $\dot{m}_{Cool}$ are provided in Figure 4.19.

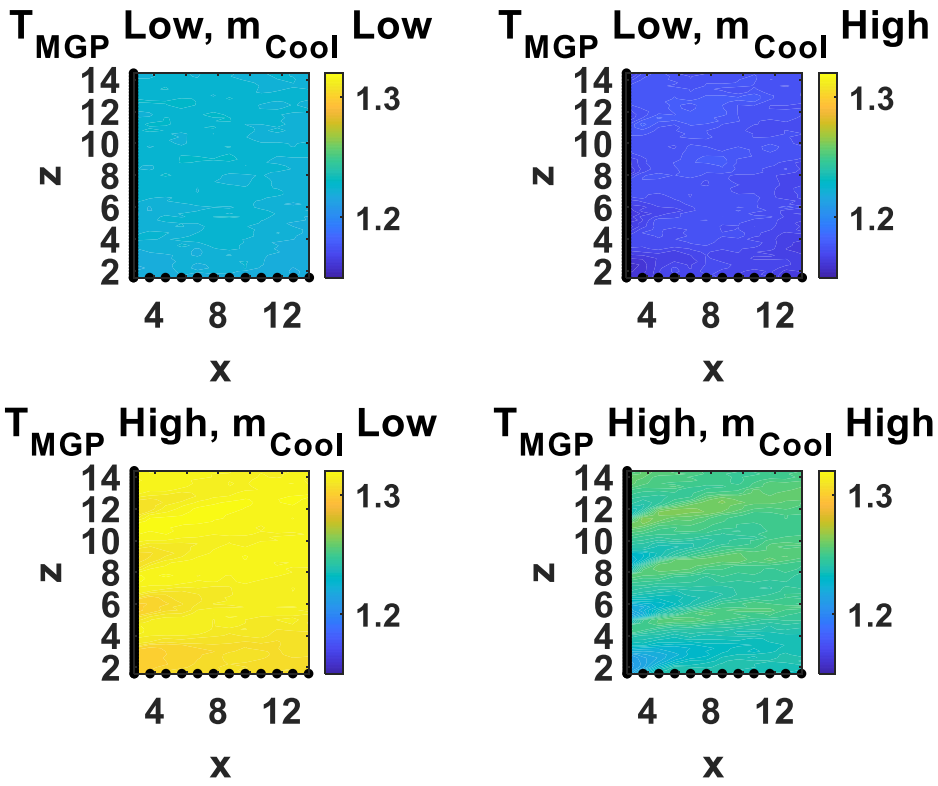


Figure 4.19 – Sample Infrared Image

It should be noted that these are not the true images. Instead, these are contour plots formed by concatenating normalized image columns. The true images are not sharable for proprietary purposes. The color maps are scaled equally across all images. While the scaling obscures the contours of the individual images, it allows for contrasts to be visible between the different operating modes. For example, when both T_{MGP} and $\dot{m}_{Cool}$ are at their lowest values, the image is a neutral light blue. However, an increase in $\dot{m}_{Cool}$ to its highest value reduces the temperature sharply, resulting in a dark blue image with somewhat visible contours. Conversely, an increase in T_{MGP} sharply increases the temperature resulting in a bright yellow image with dark yellow accents. These accents are

due to the cooling holes in the turbine blade. They become more visible when both operating conditions are at their highest values.

To predict Coolant Flow Rate using these images, the following three feature extraction techniques are employed.

- 2D Avg: Overall average pixel intensity, x_{2D}
- ID Avg: Column-wise pixel averages, $\boldsymbol{x}_{1D} = \left(x_{1D,1}, \ldots, x_{1D,30}\right)^T$
- Image: Individual pixels, $\boldsymbol{x}_{Image} = \left(x_{1,1}, x_{2,1}, \ldots, x_{63,1}, \ldots, x_{63,63}\right)^T$

The overall average pixel intensity is representative of a pyrometer with a focal area equivalent to that of the infrared camera. The column-wise pixel averages have been used for studying the functional relationship between cooling and downstream distance. Therefore, it is worth investigating as a potential predictor for predicting Coolant Flow Rate. The final technique utilizes the entire image, treating each pixel as an individual feature. In this study, the ability of each of these techniques along with the ancillary sensors $\left(T_{Cool}, \frac{dT_{ID}}{dt}, \frac{dT_{OD}}{dt}\right)$ in predicting Coolant Flow Rate is evaluated.

4.2.4 Linear Regression with Lasso

All models evaluated in this chapter were fitted using Linear Regression with Lasso. As explained in Section 4.1, Lasso can perform variable selection because it causes some regression coefficients to shrink to zero. To fit this model, the following optimization problem is solved:

$$\min_{(\beta_0,\boldsymbol{\beta})\in\mathbb{R}^{P+1}} \frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \lambda\|\boldsymbol{\beta}\|_1 \tag{4.2.1}$$

Here, $y_i \in \mathbb{R}$ is the ith observed Coolant Flow Rate and $\boldsymbol{x}_i \in \mathbb{R}^P$ is the corresponding vector of P standardized predictors. Since the objective function is convex, this can be solved using coordinate descent given the tuning parameter $\lambda \geq 0$, which is determined by 10-fold cross-validation. To train the models presented in this chapter, half of the data for each $T_{MGP}/\dot{m}_{Cool}$ combination was selected for training and the other half was used for testing the model. Linear Regression with Lasso was fitted using the training set and the parameters $(\hat{\beta}_0, \widehat{\boldsymbol{\beta}})$ were estimated. For observations in the test set, the Coolant Flow Rate is predicted using the equation:

$$\hat{y}_i = \hat{\beta}_0 + \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} \tag{4.2.2}$$

The accuracy of the model is determined by computing the Root Mean Squared Error (RMSE) for test observations. Like the sealing effectiveness study, the data was partitioned into training and testing 250 times and the median and range of the RMSE values were reported. Therefore, all models fitted are analyzed according to the distribution of the RMSE.

4.2.5 Analysis of Coolant Flow Rate Predictions Using Linear Regression with Lasso

Two datasets are utilized for this analysis. The first dataset consists of all samples recorded in the experiment described in Subsection 4.2.2. The second dataset is a downsampled version of the first dataset. This downsampling is designed to remove the

influence of lurking variables. This is accomplished by selecting samples where the ancillary sensor measurements $\left(T_{Cool}, \frac{dT_{ID}}{dt}, \frac{dT_{OD}}{dt}\right)$ have little variability. The purpose of this is to test the Linear Regression with Lasso model on an "ideal" dataset sampled when no lurking variables are present. The effect of downsampling on the distribution of the ancillary features is shown in Figure 4.20.

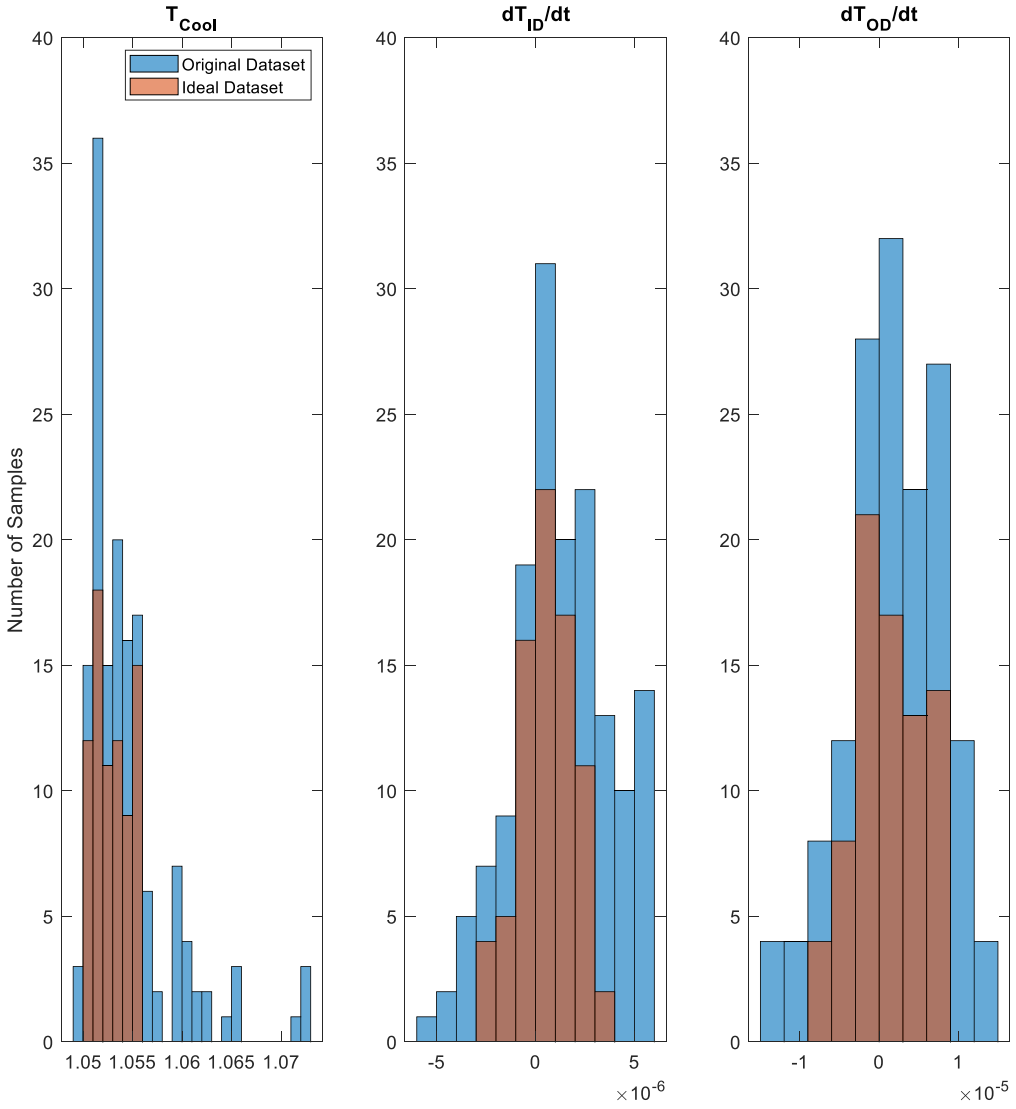


Figure 4.20 – Histograms of Ancillary Factors

Figure 4.20 draws a comparison between the distributions of the ancillary sensor measurements for the original dataset and the ideal dataset. The variability in the ancillary sensor measurements is reduced through the downsampling. In Subsection 4.2.5.1, the accuracies of the proposed models are analyzed for the ideal dataset.

4.2.5.1 Comparison of $\dot{m}$ Estimation Using Ideal Dataset with or without T_{MGP}

As can be seen in Figure 4.19, changes in the MGP Temperature significantly affect the temperature distribution on the surface of the blade. This study compares the performance of each feature extraction technique when including T_{MGP} as a known measurable quantity with the performance of each feature extraction technique on their own. The predictor variables and response variable for each model are shown in Table 4.3

Table 4.3 – Summary of Modeling Approaches with and without T_{MGP} as a Predictor

Model	Training Predictors	Response, y
2D Avg without T_{MGP}	x_{2D}	$\dot{m}_{Cool}$
1D Avg without T_{MGP}	$\boldsymbol{x}_{1D}^{T}$	$\dot{m}_{Cool}$
Image without T_{MGP}	$\boldsymbol{x}_{Image}^{T}$	$\dot{m}_{Cool}$
2D Avg with T_{MGP}	$x_{2D}, T_{MGP}, T_{MGP}x_{2D}$	$\dot{m}_{Cool}$
1D Avg with T_{MGP}	$\boldsymbol{x}_{1D}^{T}, T_{MGP}, T_{MGP}\boldsymbol{x}_{1D}^{T}$	$\dot{m}_{Cool}$
Image with T_{MGP}	$\boldsymbol{x}_{Image}^{T}, T_{MGP}, T_{MGP}\boldsymbol{x}_{Image}^{T}$	$\dot{m}_{Cool}$

The comparison in RMSE between the models with and without T_{MGP} as a predictor is shown in Figure 4.21.

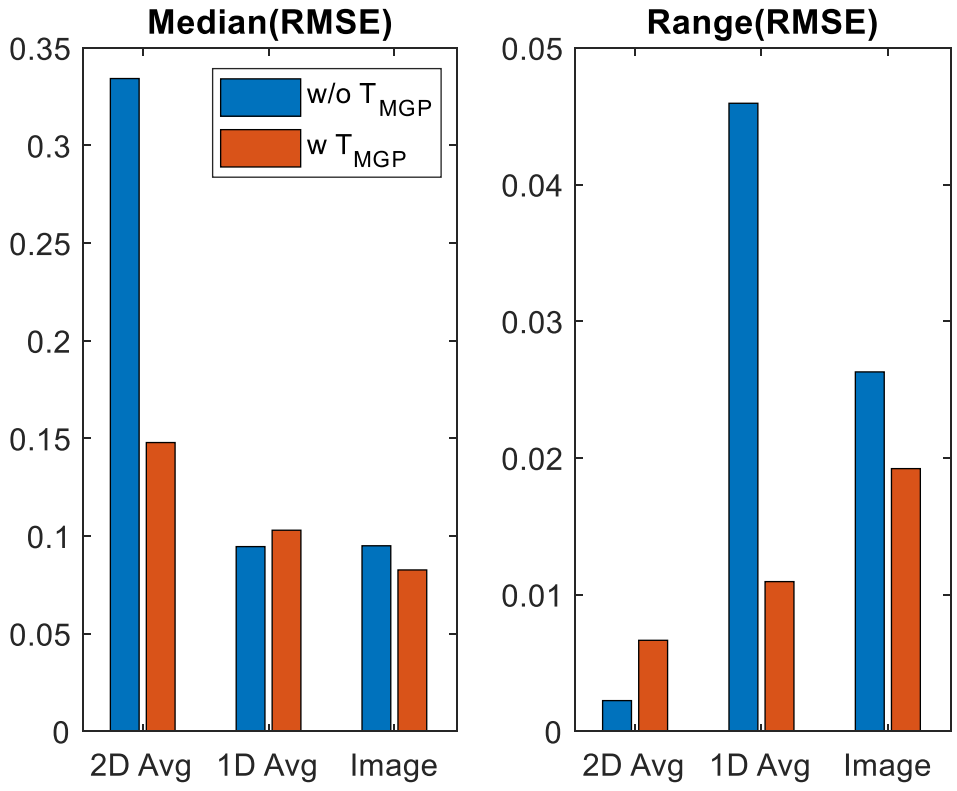


Figure 4.21 – Median and Range of RMSE for Estimating $\dot{m}$ with and w/o T_{MGP}

From Figure 4.21, the RMSE is much lower for the 1D Avg and the Image features than the 2D Avg feature. This is expected because the 1D Avg and Image are more informative than the 2D Avg. When compared with each other, the 1D Avg and Image have comparable accuracy. While the accuracy for the 1D Avg and the Image features does not significantly improve when including T_{MGP} as a predictor, the same is not true for the 2D Avg. Including T_{MGP} into the model greatly improved the ability of the 2D Avg to predict Coolant Flow Rate. However, this inclusion assumes that T_{MGP} is a measurable quantity, which may not be realistic. Therefore, whether similar improvement in accuracy is possible if T_{MGP} is inferred from the infrared image features is tested.

Linear Regression with Lasso is used to generate three models where T_{MGP} is modeled as a function of the infrared image features. Table 4.4 displays the MGP Temperature models.

Table 4.4 – Summary of MGP Temperature Models

Model	Training Predictors	Response
2D Avg	x_{2D}	T_{MGP}
1D Avg	$\boldsymbol{x}_{1D}^{T}$	T_{MGP}
Image	$\boldsymbol{x}_{Image}^{T}$	T_{MGP}

The distributions of the RMSE for these models are shown in Figure 4.22.

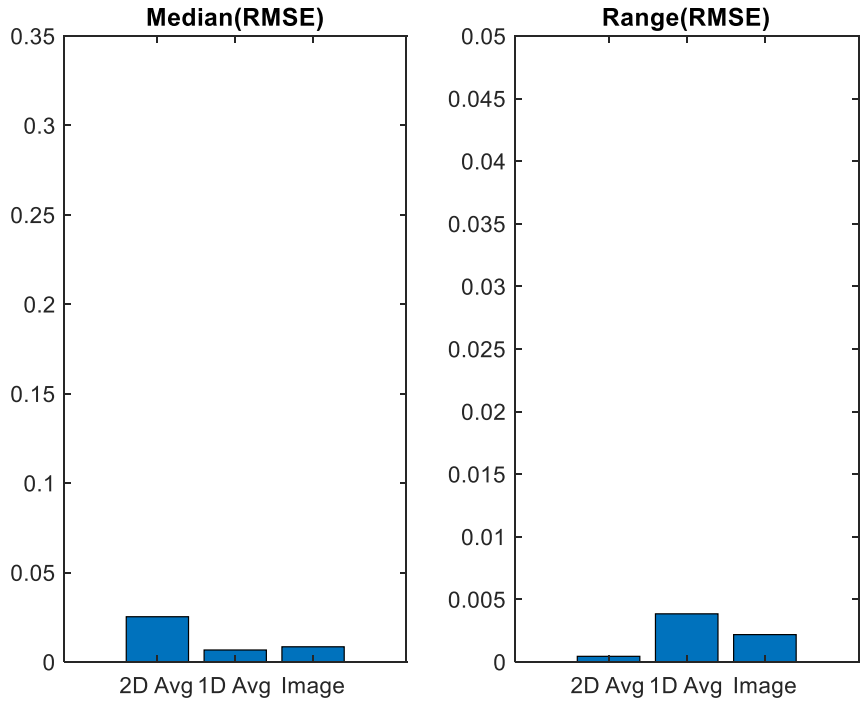


Figure 4.22 – Modeling T_{MGP} Using Infrared Image Features

The RMSE when predicting T_{MGP} is low, demonstrating that the features are strong predictors for predicting MGP Temperature. This is not surprising given that the MGP is acting directly on the blade. However, it is important to assess the utility of the inferred T_{MGP} as a predictor for Coolant Flow Rate. Since significant improvement in Coolant Flow Rate prediction only occurred for the 2D Avg, the analysis is limited to this case.

The methodology consists of modeling the Coolant Flow Rate using the 2D Avg along with the measured T_{MGP}. The Coolant Flow Rate is modeled as follows:

$$y = \beta_0 + \beta x + \gamma T_{MGP} + \theta x T_{MGP} \tag{4.2.3}$$

where y is the Coolant Flow Rate, x is the 2D average, and $\{\beta_0, \beta, \gamma, \theta\}$ are model parameters. Linear Regression with Lasso is used to fit the Coolant Flow Rate model. Then, T_{MGP} is modeled using individual Linear Regression with Lasso models for each feature extraction technique. T_{MGP} is inferred for the test data for each of these regression models. The inferences for T_{MGP} are used along with the infrared image features to predict Coolant Flow Rate. The RMSE for the Coolant Flow Rate models with T_{MGP} as an inferred value are shown in Figure 4.23. For comparison, the result if T_{MGP} is known is shown.

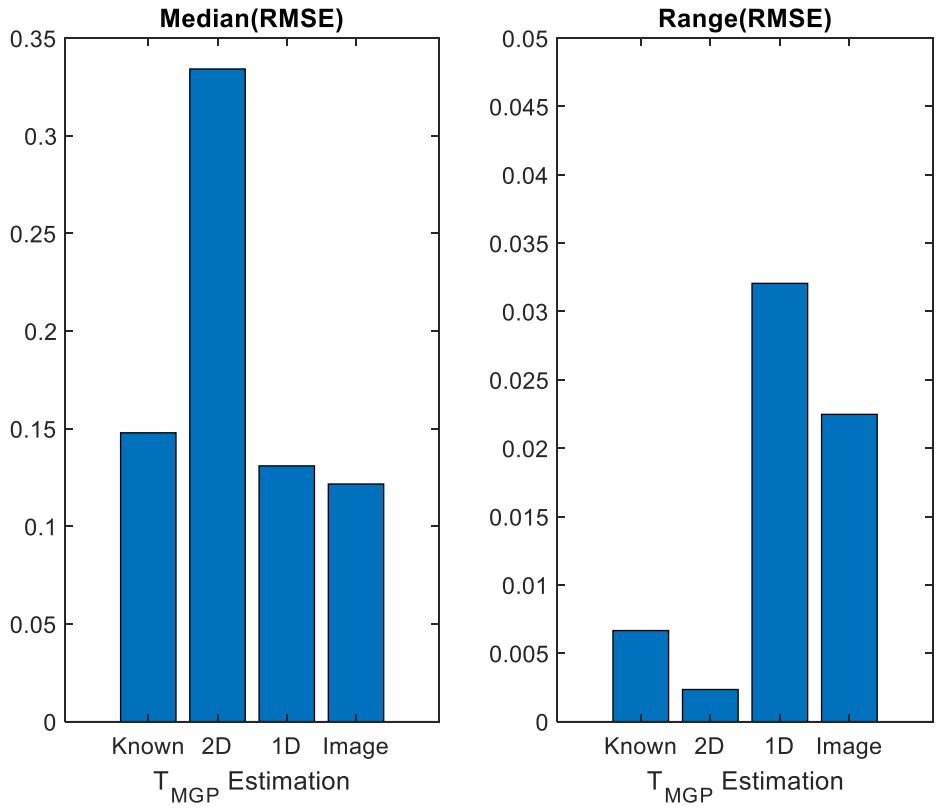


Figure 4.23 – Prediction of Coolant Flow Rate with Inferred T_{MGP}

The results in Figure 4.23 indicate that using an inferred value of T_{MGP} is also useful for predicting Coolant Flow Rate as the median RMSE values when inferring T_{MGP} using 1D

Avg or Image features are comparable to using the actual T_{MGP}. Figure 4.23 also indicates that using 2D Avg to infer T_{MGP} and then using this inferred value with the 2D Avg to predict Coolant Flow Rate results in poor accuracy. The reason is that the inferred T_{MGP} is a function of the 2D Avg and therefore, no new information is added to the model by its inclusion.

Up to this point, all analysis has been conducted on the ideal data set. It is important to determine how well these models perform on data with more variability in the ancillary factors. Therefore, application of these models to the original dataset is the topic of the next subsection.

4.2.5.2 Estimating $\dot{m}$ Using Original Dataset

In this subsection, the ability for the models fitted using the three feature extraction techniques to predict the Coolant Flow Rate when using the original dataset is analyzed. First, the performance of the models trained and tested on the ideal dataset are compared with those trained and tested on the original dataset. For all three feature extraction techniques, the accuracy of the model fitted with only the extracted features is compared with the accuracy of the model fitted with both the extracted features and T_{MGP}. When testing the models, the T_{MGP} and the ancillary sensor measurements ($T_{Cool}, \frac{dT_{ID}}{dt}, \frac{dT_{OD}}{dt}$) are assumed to be observable. Table 4.5 displays the predictors and response variables for each model. The tilde operators indicate that the model uses observations from the original dataset.

Table 4.5 – Comparison of Models between Ideal and Original Datasets

Model	Training Predictors	Response
2D Avg, Train Ideal, Test Ideal	x_{2D}	$\dot{m}_{Cool}$
2D Avg, Train Ideal, Test Ideal with T_{MGP}	$x_{2D}, T_{MGP}, T_{MGP}x_{2D}$	$\dot{m}_{Cool}$
2D Avg, Train Original, Test Original	$\tilde{x}_{2D}$	$\dot{m}_{Cool}$
2D Avg, Train Original, Test Original with T_{MGP}	$\tilde{x}_{2D}, \tilde{T}_{MGP}, \tilde{T}_{MGP}\tilde{x}_{2D}$	$\dot{m}_{Cool}$
1D Avg, Train Ideal, Test Ideal	$\boldsymbol{x}_{1D}^{T}$	$\dot{m}_{Cool}$
1D Avg, Train Ideal, Test Ideal with T_{MGP}	$\boldsymbol{x}_{1D}^{T}, T_{MGP}, T_{MGP}\boldsymbol{x}_{1D}^{T}$	$\dot{m}_{Cool}$
1D Avg, Train Original, Test Original	$\tilde{\boldsymbol{x}}_{1D}^{T}$	$\dot{m}_{Cool}$
1D Avg, Train Original, Test Original with T_{MGP}	$\tilde{\boldsymbol{x}}_{1D}^{T}, \tilde{T}_{MGP}, \tilde{T}_{MGP}\tilde{\boldsymbol{x}}_{1D}$	$\dot{m}_{Cool}$
Image, Train Ideal, Test Ideal	$\boldsymbol{x}_{Image}^{T}$	$\dot{m}_{Cool}$
Image, Train Ideal, Test Ideal with T_{MGP}	$\boldsymbol{x}_{Image}^{T}, T_{MGP}, T_{MGP}\boldsymbol{x}_{Image}^{T}$	$\dot{m}_{Cool}$
Image, Train Original, Test Original	$\tilde{\boldsymbol{x}}_{Image}^{T}$	$\dot{m}_{Cool}$
Image, Train Original, Test Original with T_{MGP}	$\tilde{\boldsymbol{x}}_{Image}^{T}, \tilde{T}_{MGP}, \tilde{T}_{MGP}\tilde{\boldsymbol{x}}_{Image}^{T}$	$\dot{m}_{Cool}$

The results are shown in Figure 4.24, where only the median RMSE is reported.

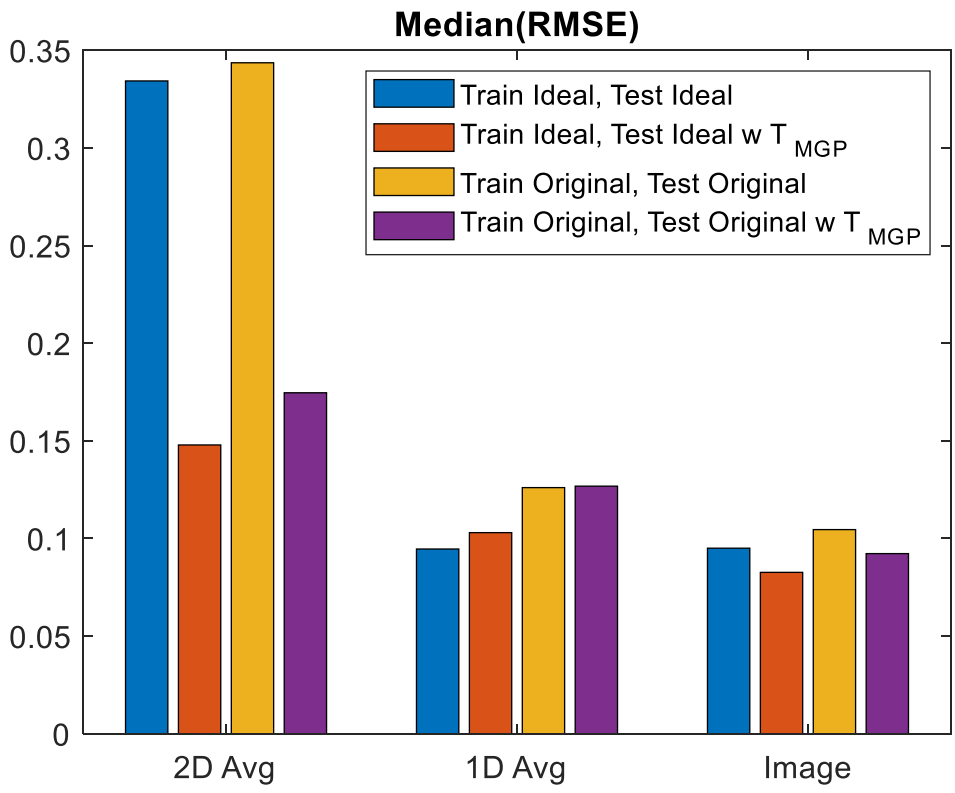


Figure 4.24 – Comparison of $\dot{m}$ Prediction Accuracy Between Ideal and Original Dataset

The bar chart in Figure 4.24 demonstrates that the prediction accuracy is lower when training and testing a model on the original dataset than when training and testing on the ideal dataset. This is expected due to heterogeneity in the ancillary factors. The improvement in the accuracy for the 2D Avg due to including T_{MGP} in the model is consistent for the ideal and the original dataset. However, this improvement is very minimal for the 1D Avg and Image features across both datasets. To account for the lower accuracy in predicting Coolant Flow Rate for the original dataset, the sensor measurements from the ancillary factors are incorporated into the models. To account for the complex relationship between the covariates and the temperature distribution of the blade, all possible interaction terms are included in the models and the Lasso penalty to determine

which interactions are significant. Let $x_2 = \tilde{T}_{MGP}, x_3 = T_{Cool}, x_4 = \frac{dT_{ID}}{dt}, x_5 = \frac{dT_{OD}}{dt}$. Table 4.6 and Table 4.7 show the predictors and the response variable for the additional models shown in Figure 4.25. Here, ancillary sensor measurements are called covariates (Cov).

Table 4.6 – Models with Ancillary Sensor Measurements

Model	Training Predictors	Response
2D Avg, Train Original, Test Original with Cov	$\tilde{x}_{2D}, x_3, x_4, x_5,$ $x_3\tilde{x}_{2D}, x_4\tilde{x}_{2D}, x_5\tilde{x}_{2D},$ $x_3x_4, x_3x_5, x_4x_5,$ $x_3x_4\tilde{x}_{2D}, x_3x_5\tilde{x}_{2D},$ $x_4x_5\tilde{x}_{2D}, x_3x_4x_5,$ $x_3x_4x_5\tilde{x}_{2D}$	$\dot{m}_{Cool}$
1D Avg, Train Original, Test Original with Cov	$\tilde{\boldsymbol{x}}_{1D}^T, x_3, x_4, x_5,$ $x_3\tilde{\boldsymbol{x}}_{1D}^T, x_4\tilde{\boldsymbol{x}}_{1D}^T, x_5\tilde{\boldsymbol{x}}_{1D}^T,$ $x_3x_4, x_3x_5, x_4x_5,$ $x_3x_4\tilde{\boldsymbol{x}}_{1D}^T, x_3x_5\tilde{\boldsymbol{x}}_{1D}^T,$ $x_4x_5\tilde{\boldsymbol{x}}_{1D}^T, x_3x_4x_5,$ $x_3x_4x_5\tilde{\boldsymbol{x}}_{1D}^T$	$\dot{m}_{Cool}$
Image, Train Original, Test Original with Cov	$\tilde{\boldsymbol{x}}_{Image}^T, x_3, x_4, x_5,$ $x_3\tilde{\boldsymbol{x}}_{Image}^T, x_4\tilde{\boldsymbol{x}}_{Image}^T, x_5\tilde{\boldsymbol{x}}_{Image}^T,$ $x_3x_4, x_3x_5, x_4x_5,$ $x_3x_4\tilde{\boldsymbol{x}}_{Image}^T, x_3x_5\tilde{\boldsymbol{x}}_{Image}^T,$ $x_4x_5\tilde{\boldsymbol{x}}_{Image}^T, x_3x_4x_5,$ $x_3x_4x_5\tilde{\boldsymbol{x}}_{Image}^T$	$\dot{m}_{Cool}$

Table 4.7 – Models with MGP Temperature and Ancillary Sensor Measurements

Model	Training Predictors	Response
2D Avg, Train Original, Test Original with T_{MGP} and Cov	$\tilde{x}_{2D}, x_2, x_3, x_4, x_5,$ $x_2\tilde{x}_{2D}, x_3\tilde{x}_{2D}, x_4\tilde{x}_{2D}, x_5\tilde{x}_{2D},$ $x_2x_3, x_2x_4, x_2x_5, x_3x_4,$ $x_3x_5, x_4x_5, x_2x_3\tilde{x}_{2D}, x_2x_4\tilde{x}_{2D}, x_2x_5\tilde{x}_{2D},$ $x_3x_4\tilde{x}_{2D}, x_3x_5\tilde{x}_{2D}, x_4x_5\tilde{x}_{2D}, x_2x_3x_4,$ $x_2x_3x_5, x_2x_4x_5, x_3x_4x_5, x_2x_3x_4\tilde{x}_{2D},$ $x_2x_3x_5\tilde{x}_{2D}, x_2x_4x_5\tilde{x}_{2D}, x_3x_4x_5\tilde{x}_{2D},$ $x_2x_3x_4x_5, x_2x_3x_4x_5\tilde{x}_{2D}$	$\dot{m}_{Cool}$
1D Avg, Train Original, Test Original with T_{MGP} and Cov	$\tilde{\boldsymbol{x}}_{1D}^{T}, x_2, x_3, x_4, x_5,$ $x_2\tilde{\boldsymbol{x}}_{1D}^{T}, x_3\tilde{\boldsymbol{x}}_{1D}^{T}, x_4\tilde{\boldsymbol{x}}_{1D}^{T}, x_5\tilde{\boldsymbol{x}}_{1D}^{T},$ $x_2x_3, x_2x_4, x_2x_5, x_3x_4,$ $x_3x_5, x_4x_5, x_2x_3\tilde{\boldsymbol{x}}_{1D}^{T}, x_2x_4\tilde{\boldsymbol{x}}_{1D}^{T}, x_2x_5\tilde{\boldsymbol{x}}_{1D}^{T},$ $x_3x_4\tilde{\boldsymbol{x}}_{1D}^{T}, x_3x_5\tilde{\boldsymbol{x}}_{1D}^{T}, x_4x_5\tilde{\boldsymbol{x}}_{1D}^{T}, x_2x_3x_4,$ $x_2x_3x_5, x_2x_4x_5, x_3x_4x_5, x_2x_3x_4\tilde{\boldsymbol{x}}_{1D}^{T},$ $x_2x_3x_5\tilde{\boldsymbol{x}}_{1D}^{T}, x_2x_4x_5\tilde{\boldsymbol{x}}_{1D}^{T}, x_3x_4x_5\tilde{\boldsymbol{x}}_{1D}^{T},$ $x_2x_3x_4x_5, x_2x_3x_4x_5\tilde{\boldsymbol{x}}_{1D}^{T}$	$\dot{m}_{Cool}$
Image, Train Original, Test Original with T_{MGP} and Cov	$\tilde{\boldsymbol{x}}_{Image}^{T}, x_2, x_3, x_4, x_5,$ $x_2\tilde{\boldsymbol{x}}_{Image}^{T}, x_3\tilde{\boldsymbol{x}}_{Image}^{T}, x_4\tilde{\boldsymbol{x}}_{Image}^{T}, x_5\tilde{\boldsymbol{x}}_{Image}^{T},$ $x_2x_3, x_2x_4, x_2x_5, x_3x_4,$ $x_3x_5, x_4x_5, x_2x_3\tilde{\boldsymbol{x}}_{Image}^{T}, x_2x_4\tilde{\boldsymbol{x}}_{Image}^{T}, x_2x_5\tilde{\boldsymbol{x}}_{Image}^{T},$ $x_3x_4\tilde{\boldsymbol{x}}_{Image}^{T}, x_3x_5\tilde{\boldsymbol{x}}_{Image}^{T}, x_4x_5\tilde{\boldsymbol{x}}_{Image}^{T}, x_2x_3x_4,$ $x_2x_3x_5, x_2x_4x_5, x_3x_4x_5, x_2x_3x_4\tilde{\boldsymbol{x}}_{Image}^{T},$ $x_2x_3x_5\tilde{\boldsymbol{x}}_{Image}^{T}, x_2x_4x_5\tilde{\boldsymbol{x}}_{Image}^{T}, x_3x_4x_5\tilde{\boldsymbol{x}}_{Image}^{T},$ $x_2x_3x_4x_5, x_2x_3x_4x_5\tilde{\boldsymbol{x}}_{Image}^{T}$	$\dot{m}_{Cool}$

The results for these models are displayed in Figure 4.25.

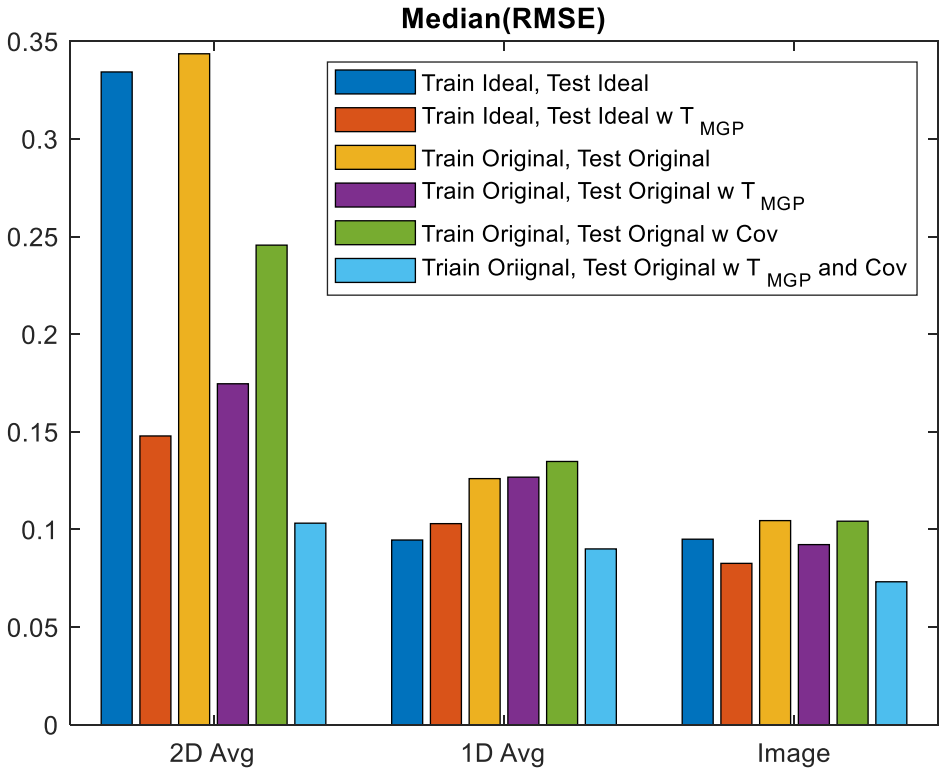


Figure 4.25 - Comparison of $\dot{m}$ Prediction Accuracy after Including Covariates

The results in Figure 4.25 indicate that including both T_{MGP} and the ancillary factors in the model provide more accurate predictions of the Coolant Flow Rate than when just using the ideal dataset. This is likely due to variability still present in the sensor values of the ancillary factors within the ideal case. Of course, the accuracy gained by incorporating these sensor measurements comes with the cost of increased sensing requirements. Furthermore, the optimal model requires knowledge of T_{MGP}. However, focusing on the models using all the image pixels as predictors, there is little difference between the model trained on the original dataset without T_{MGP} and covariates (yellow) and the model trained on the original dataset with covariates but without T_{MGP} (green). While the model trained on the original dataset with just the image features is not the most accurate overall, it is the

most accurate of the models trained on the original dataset that do not include T_{MGP} or ancillary factors. Therefore, this model is investigated further in the next subsection.

4.2.5.3 Investigation of $\dot{m}$ vs. Image Model

Let $\mathcal{X}_i, i = 1, \dots, N$ denote the ith sampled infrared image of the turbine blade, where

$$\mathcal{X}_i = \begin{bmatrix} x_{i,1} & \cdots & x_{i,694} \\ \vdots & \ddots & \vdots \\ x_{i,63} & \cdots & x_{i,756} \end{bmatrix} \tag{4.2.4}$$

since for this experiment, the images are 63×12. Let $\langle \mathcal{X}_i, \mathcal{B} \rangle = \sum_{j=1}^{756} x_j \beta_j$, then the Linear Regression with Lasso model can be reformulated as:

$$\min_{\beta_0, \mathcal{B}} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \langle \mathcal{X}_i, \mathcal{B} \rangle)^2 + \lambda \|\mathcal{B}\|_1 \tag{4.2.5}$$

In Equation 4.2.5, the regression coefficients are arranged into a 63×12 matrix whose elements correspond with the elements of the observed image $\mathcal{X}_i$. Thus, solving Equation 4.14 is akin to estimating this coefficient matrix. Since Lasso performs variable selection, this coefficient matrix is expected to be sparse. These sparse regions can indicate potential locations for placement of less expensive telemetry such as pyrometers. In Figure 4.26, the contour plot of the regression coefficient matrix is displayed for the $\dot{m}$ model trained on image features using the original dataset.

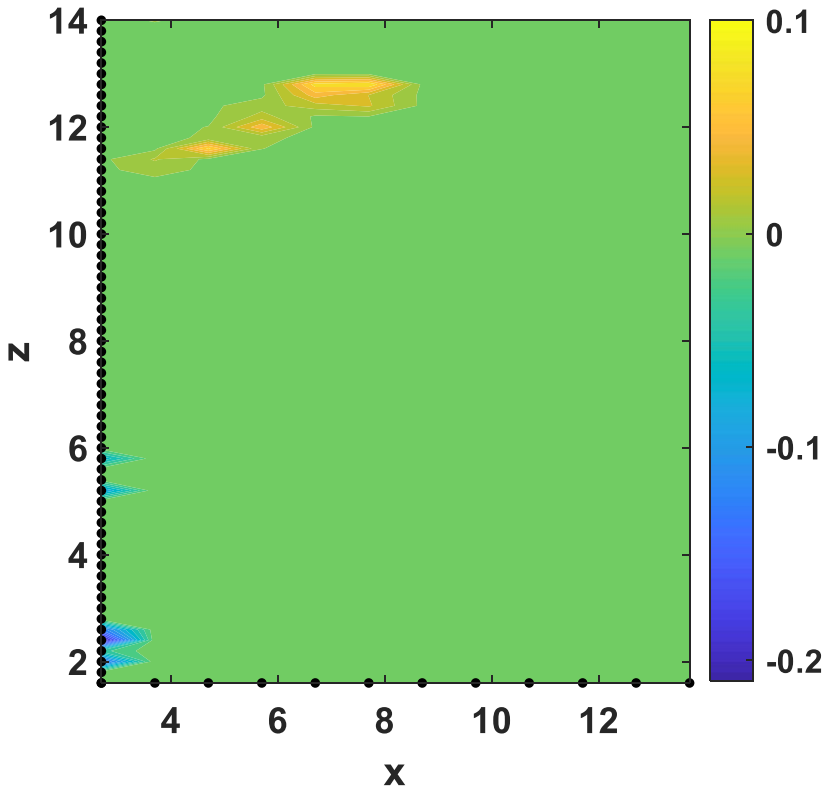


Figure 4.26 – Contour Plot of Regression Coefficient Matrix

In Figure 4.26, most of the coefficients are zero apart from a few sparse regions. This indicates that an entire infrared image is not necessary for estimating Coolant Flow Rate. Instead, focus only needs to be paid to areas of the blade corresponding to these sparse regions. Further investigation of Figure 4.26 shows that the locations selected refer to coldest (blue) and hottest locations (yellow) on the blade.

Although Linear Regression with Lasso revealed sparse regions for monitoring, the number of pixels being monitored is 33. To determine the tradeoff between the number of features retained in the model and the accuracy in estimating the Coolant Flow Rate, consider the following sequence of tuning parameters: $\lambda_{(1)}, \lambda_{(2)}, \ldots, \lambda_{(10)}, \ldots$, where $\lambda_{(p)}$ is

the largest value of the tuning parameter such that Linear Regression with Lasso selects p features for modeling. These values are obtained by performing 10-fold cross-validation and analyzing the sequence of tuning parameters for the number of features retained in the model. For each tuning parameter in the sequence, Linear Regression with Lasso is performed. This is repeated 250 times for various partitions of the original data into training and testing. The median RMSE values are reported up through $\lambda_{(10)}$ along with the median RMSE for the optimal model in Figure 4.27.

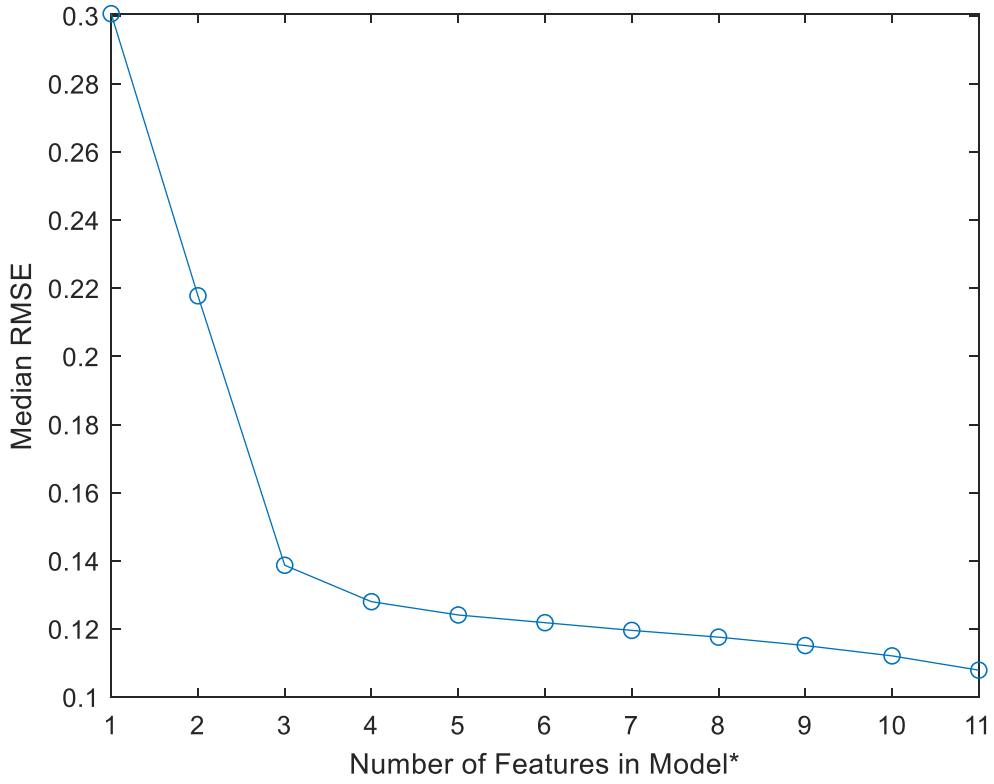


Figure 4.27 – Prediction Error of $\dot{m}$ as a Function of Number of Features in Model

The asterisk indicates that model 11 refers to the optimal model and not the model with 11 features. Figure 4.27 shows that once four features are used, the rate of improvement in the

prediction error diminishes. Therefore, an accurate estimate of the Coolant Flow Rate can be obtained using only four features. This offers opportunities for alternative, more affordable sensing technology that focuses on localized regions of the blade to diagnose Coolant Flow Rate (e.g. single-point pyrometers).

4.2.6 Conclusion

This section demonstrated several regression models for predicting Coolant Flow Rate using infrared imaging. This is important as no studies had been developed for monitoring the health of the coolant system. This methodology enables turbine operators to detect faults in the coolant even under various MGP temperatures. However, lurking variables had influence on the correlation between Coolant Flow Rate and the infrared images. Downsampling was employed to reduce these effects. Three techniques for extracting features from the images were evaluated and the 1D Avg and Image features were shown to perform better than the 2D Avg. The 2D Avg was shown to be a strong predictor of Coolant Flow Rate if the MGP Temperature was measurable. When including the observations from the original dataset, there was a reduction in accuracy from the ideal dataset. However, this reduction in accuracy was mitigated by including ancillary sensor measurements in the model. This demonstrates the importance of experimental design in removing the effect of lurking variables.

In the analysis of the regression coefficients for the model using the image as a predictor, the estimated coefficients were shown to be sparse. A study was conducted where the number of features in the model was progressively increased and the decrease in prediction error was observed. After four features, there was only marginal improvement

in the prediction error. Therefore, accurate prediction of the Coolant Flow Rate can be accomplished by sensing local regions of the gas turbine blade instead of the whole blade. This provides a cost effective solution to monitoring the health of the coolant system.

CHAPTER 5. PREDICTING REMAINING USEFUL LIFE IN COMPLEX SYSTEMS WITH MULTIPLE FAILURE MODES

5.1 Introduction and Literature Review

Modern industrial systems are comprised of multiple components with complex degradation processes. To address this complexity, modern systems are equipped with several sensors for condition monitoring, a process that utilizes sensor data for predicting the system's remaining useful life (RUL), where the RUL is the time remaining before system failure. This prognostics task is important because it drives maintenance decisions that aim to reduce downtime and prevent potentially catastrophic failures. However, prognostics is challenging for complex systems due to the large volume of data collected by these sensors and because complex systems are subject to multiple failure modes. High-dimensional (HD) data analytics holds potential for enabling the development of robust prognostics frameworks for complex systems due to its ability to process the large data structures generated by these sensors. Therefore, this chapter focuses on the development of an HD data analytics framework for predicting RUL in complex systems with multiple failure modes.

Traditionally, the literature on condition-based prognostics has revolved around modeling a single degradation signal. For example, Gebraeel et al. (2005) used an exponential model with random coefficients, Chen et al. (2015) modeled heterogeneous degradation signals using an inverse Gaussian process with a random drift parameter, and Shu et al. (2015) modeled cumulative degradation with random jumps using a Levy

process. These methodologies leverage their respective degradation models to develop closed-form RUL distributions. However, applying these methodologies to complex systems likely results in low prediction accuracy since a single sensor may not be sufficient for capturing the failure dynamics of a system. This was demonstrated by Chehade et al. (2018) when comparing their sensor fusion methodology to a methodology using only one sensor. Therefore, access to multiple sensors is valuable for effective prognostics.

According to Fang et al. (2017), multiple sensor applications tend to fuse all available sensor data using techniques such as neural networks and health index models. However, groups of sensors may be redundant with other sensors, while other sensors may have no correlation with the underlying degradation. Fusing redundant or noninformative sensors may lead to reduced RUL prediction accuracy. Therefore, a robust algorithm for selecting a subset of informative sensors for predicting RUL is necessary for improving prediction accuracy. Fang et al. (2017) combined location-scale regression models with group non-negative garrote penalty to select an optimal subset of informative sensors. They then fused this subset of sensors using a functional data analysis approach. The methodology by Fang et al. (2017) is designed for systems with one failure mode. Additional singular failure mode models can be found in comprehensive literature reviews by Jardine et al. (2006) and Sikorska et al. (2011). However, complex systems often suffer from multiple failure modes. Large RUL prediction errors are expected if a prognostics methodology fails to account for the possibility of alternative causes of failure. Therefore, this methodology explicitly considers multiple failure modes in its development.

Compared to the research on prognostic models for single failure modes, the research on multiple failure modes is relatively sparse. Prognostics approaches for multiple failure

modes can be characterized as model-based and data-driven. Model-based approaches attempt to develop mathematical models that characterize the underlying physics of the system. Examples of these include papers by Thumati et al. (2014) and Yu and Wang (2014). The advantages of these approaches are that their development does not depend on extensive historical data and that closed form analytic expressions are derived for modeling degradation and predicting RUL. However, their use is limited to the systems for which such expressions can be developed. A general prognostics approach for complex systems using graph theory and stochastic models was proposed by Blancke et al. (2018). In that paper, interdependencies between failure modes are captured by a directed acyclic graph developed using expert knowledge of the system dynamics. Though for general use, the accuracy of this model depends on the quality of available expert knowledge. Common to all model-based approaches is that their fidelity relies on an accurate representation of the physics of failure. However, as system complexity continues to grow, so too does the difficulty in mathematically characterizing the physics of failure processes of these systems.

Data-driven approaches are alternatives to model-based approaches that utilize combinations of observed times-to-failure, performance metrics, and sensor data to develop prognostics models. The advantage of data-driven approaches is that they are flexible to various types of systems since they do not require analytical expressions of the system dynamics. Instead, they rely on the statistical analysis of the data to predict RUL. Many of these approaches are focused on developing reliability models. The reliability literature consists of probabilistic models for systems comprised of components in series, parallel, or in various combined arrangements. Theoretical models for state-dependent

systems where the transition of one component from a healthy state to a faulted state increases the failure rate of the other component have also been developed (Ebling, 2010). These methods typically use time-to-failure (TTF) data to fit reliability distributions for system components individually and then use Markov analysis to derive system reliability. Another class of reliability models for systems with multiple failure modes are competing risk models. A competing risk is an alternative failure that occurs prior to the failure mode of interest. Therefore, the competing risk literature focuses on developing system reliability models for scenarios such as a single sudden failure superseding failure due to gradual system degradation (Bocchetti et al., 2009), multiple sudden failures competing both with each other and with the gradual system degradation (Couallier, 2008; Liu et al., 2013), and multiple degradation processes competing against each other while experiencing sudden shocks (Wang and Pham, 2012). However, these methodologies either validate their methodology using univariate sensor signals or with no sensor data.

Recently, HD data analytics have been incorporated into multiple failure mode prognostics models. Ragab et al. (2019) utilized time-series and wavelet features from a vibration signal to predict RUL in rotating machinery. In their methodology, the authors utilize Compensation Distance Evaluation Technique (CDET) to select informative features from the vibration signal. However, modern systems are equipped with multiple sensors for monitoring degradation. Compared to single sensor applications, the literature on multi-sensor applications is quite sparse. Hajiha et al. (2020) utilized multiple sensors for monitoring degradation and dynamic operating conditions in systems with competing degradation processes. Zhang et al. (2014) used multiple sensors to model the system failure density as a mixture of Weibull proportional hazard models. Chehade et al. (2018)

developed a sensor fusion methodology for predicting RUL in systems with multiple failure modes, whereas Kim and Liu (2020) developed a Bayesian deep learning framework that performs interval estimates of RUL of complex systems. These methodologies either utilize all available sensors or use visual inspection to remove uninformative sensors. However, information salient to degradation may only be embedded in a subset of the sensors, while the other sensors only contribute noise. Therefore, a robust sensor selection algorithm is developed. In this algorithm, an optimal subset of sensors for monitoring is selected for each failure mode. However, the subsets are not mutually exclusive, meaning that the same sensor can be used for monitoring various degradation processes.

Our prognostics methodology consists of two steps: sensor selection and RUL prediction. In the first step, a historical training set comprised of both complete multivariate sensor signals and observed failure times is utilized to develop a sensor selection algorithm that identifies optimal subsets of sensors for each failure mode. In this algorithm, it is assumed that each system fails independently and that its failure is dominated by one of K possible failure processes. Furthermore, it is assumed that the relationship between the TTF and the sensor signals is dependent on the failure mode. Since various failure modes each impact the shape of the sensor data, the distribution of the TTF given the observed signals is multimodal. This motivates the use of a mixture model for modeling the distribution of the TTF. To ensure that the TTF is positive, the functional relationship between the natural logarithm of the time-to-failure ($\ln TTF$) and the sensors is modeled as a mixture of Gaussian regressions (MGR). Since degradation signals are high-dimensional, features are extracted from the individual degradation signals and these features are used in the MGR

model. To perform sensor selection, the MGR model is combined with a regularization term. Städler et al. (2012) proposed combining MGR with adaptive lasso, which enables individual feature selection. However, their method does not account for the inherent group structure present due to multiple sensors. Therefore, the MGR model is combined with a regularization term that performs group variable selection: adaptive sparse group lasso (ASGL). This approach is labeled MGR-ASGL The ASGL penalty ensures that the regression coefficients for informative sensors have nonzero l2-norm, while the coefficients for uninformative sensors are zero. To fit this model, a novel Expectation-Maximization (EM) algorithm that first determines the responsibility of each failure mode on observed failures and then selects an optimal subset of sensors for each failure mode is proposed. This algorithm is unique in that it can perform diagnosis and sensor selection simultaneously. Since the EM algorithm searches for a local optimum, the solution to which the algorithm converges depends on the initial solution provided. To provide an initial clustering of the raw signals, Dynamic Time Warping (DTW) is combined with hierarchical clustering. DTW is a tool that measures distances between data structures of various lengths. Therefore, it is a valuable tool for clustering degradation signals that vary in length due to failure time variability.

In the second step, the RUL is predicted by first intelligently combining the information in the sensor signals selected via the sensor selection algorithm. This is achieved using Multivariate Functional Principal Component Analysis (MFPCA). MFPCA is a method for capturing dominant modes of variation for data structures where a single observation is a multivariate functional profile (Ramsay and Silverman, 2005). In MFPCA, multivariate profiles are represented as weighted combinations of multivariate

eigenfunctions. Through these weights, the MFPC scores, very high-dimensional data can be represented with a much smaller number of features. Fang et al. (2017) utilized MFPCA to fuse sensor signals selected via a penalized log-location scale regression model. This method differs in that MFPCA is first used to fuse the selected sensors across all failure modes to fit a supervised adaptive mixture of Gaussian classifier where the clustering from the MGR-ASGL model informs the labels for the classifier. The model is adaptive in that it is refitted every time new degradation is observed. For a degrading system, the classifier is used to map the MFPC scores extracted from the observed degradation data to one of the failure modes. Given this failure mode, MFPCA is used on the subset of sensors corresponding to the diagnosed fault. Using the MFPC scores, an adaptive penalized least squares model is fitted using training observations that failed due to the diagnosed fault. This model is used to map MFPC scores extracted from the observed degradation data to the $\ln TTF$. The RUL is then estimated by subtracting the current observed time from the predicted TTF.

The primary contribution of this paper is a generalizable prognostic methodology that automatically identifies a subset of sensors associated with each degradation mode, detects which failure mode is active, and predicts and updates the RUL in real time. The sensor selection algorithm revolves around modeling the $\ln TTF$ using a penalized mixture of regressions model named MGR-ASGL. To estimate the parameters of this model,]a novel EM algorithm is presented. This algorithm enables a simultaneous diagnosis of the failures in the training set and selection of optimal sensors for each failure mode. The failure mode labels and selected sensors inform an adaptive approach to predicting RUL that includes extracting informative features by fusing the selected sensors using MFPCA

and developing a classification and regression model using the MFPC scores. The remainder of the paper is outlined as follows. In Section 5.2, the sensor selection algorithm consisting of feature extraction and fitting of the MGR-ASGL model using the novel EM algorithm is presented. In Section 5.3, the fusion of informative sensors to diagnose faults and predict RUL is discussed. In Section 5.4, the methodology is demonstrated through a case study using the aircraft turbofan dataset created by NASA. The chapter is concluded in Section 5.5 with a summary of findings and discussion of some potential for future research.

5.2 Sensor Selection Algorithm

The salient information in high-dimensional (HD) data is often embedded in a few dimensions. For a multiple sensor setting, this means that a subset of the available sensors suffices for characterizing system degradation, while the other sensors only contribute noise. The question of which sensors are significant is made more challenging in a multiple failure mode setting since the behavior of sensor data can vary drastically in response to different active degradation processes. Therefore, this section focuses on the development of a sensor selection algorithm for systems whose failure is governed by multiple degradation processes. For this algorithm, it is assumed there exists a historical repository of multivariate sensor data used to monitor the health of a fleet of independent and identical systems from the moment an incipient fault occurs up to system failure. Furthermore, it is assumed that raw sensor signals can be synthesized into degradation signals. Finally, failures in the observed historical repository are assumed to be due to one of K failure modes, where K is known. This algorithm consists of three steps. First, the raw degradation signals are clustered using a combination of Dynamic Time Warping (DTW) and

hierarchical clustering. After clustering, the features are extracted from the degradation signals of each sensor by fitting their curves to an appropriate function. Since degradation signals can take on various forms, various degradation models in the literature are highlighted. Finally, the relationship between the failure times and the extracted failures is modeled using mixture of Gaussian regressions with adaptive sparse group lasso (MGR-ASGL). The model parameters are estimated using the novel Expectation-Maximization (EM) algorithm through which the failures in the dataset are clustered and an optimal subset of sensors is selected for each failure mode. This EM algorithm is initialized using the results of DTW and hierarchical clustering. The following subsection discusses the initial clustering.

5.2.1 Initial Clustering via DTW and Hierarchical Clustering

Due to unit-to-unit variability in degradation rates, the lengths of degradation signals are typically quite different, and their features are often misaligned. This creates a challenge when attempting to measure the distance between two degradation signals. DTW is a tool that measures the distance between two signals of varying lengths by first creating a lattice of pairwise distances. Then it finds the path through the lattice that minimizes the accumulated distance subject to constraints on both the direction the path can travel and on the starting and ending positions of the path. To formalize DTW, consider a training set consisting of N units monitored by P sensors. For unit i, $i = 1, \dots, N$, let $\boldsymbol{s}_i[t] = \left(s_{i,1}[t], s_{i,2}[t], \dots, s_{i,P}[t]\right)^T$ denote the vector of sensor observations for system i at time $t \in \{0,1, \dots, T_i - 1\}$ where T_i is the time-to-failure (TTF) of system i. To compute a

distance between the degradation signals of units i and j, $i, j \in \{1,2, \dots, N\}$, the following matrices are constructed:

$$\begin{matrix} \mathbb{S}_i = [\boldsymbol{s}_i[0] & \dots & \boldsymbol{s}_i[T_i - 1]] \\ \mathbb{S}_j = [\boldsymbol{s}_j[0] & \dots & \boldsymbol{s}_j[T_j - 1]] \end{matrix} \tag{5.2.1}$$

Then, the Euclidean distance between signals $\mathbb{S}_i$ and $\mathbb{S}_j$ at time t and t' respectively is denoted as:

$$\begin{aligned} d_{t,t'}(\mathbb{S}_i, \mathbb{S}_j) = \left\| \boldsymbol{s}_i[t] - \boldsymbol{s}_j[t'] \right\|_2^2, t \in [0,1, \dots, T_i - 1], t' \\ \in [0,1, \dots, T_j - 1] \end{aligned} \tag{5.2.2}$$

After computing distances for all pairwise points, the lattice matrix is constructed as follows:

$$\begin{bmatrix} d_{0,T_j-1}(\mathbb{S}_i, \mathbb{S}_j) & - & d_{T_i-1,T_j-1}(\mathbb{S}_i, \mathbb{S}_j) \\ | & \diagdown / & | \\ d_{t,t'}(\mathbb{S}_i, \mathbb{S}_j) & & \diagup \diagdown \\ | & \diagup & | \\ d_{0,0}(\mathbb{S}_i, \mathbb{S}_j) & - & d_{T_i-1,0}(\mathbb{S}_i, \mathbb{S}_j) \end{bmatrix} \tag{5.2.3}$$

DTW works by finding the minimum path from $d_{0,0}(\mathbb{S}_i, \mathbb{S}_j)$ to $d_{T_i-1,T_j-1}(\mathbb{S}_i, \mathbb{S}_j)$ where only right horizontal moves $(t, t') \rightarrow (t + 1, t')$, upward vertical moves $(t, t') \rightarrow (t, t' + 1)$ and upper-right diagonal moves $(t, t') \rightarrow (t + 1, t' + 1)$ are allowed. The warping distance is then determined by adding all distances along the optimal path $PathOpt$.

$$d_{i,j} = \sum_{t,t' \in PathOpt} d_{t,t'}(\mathbb{S}_i, \mathbb{S}_j) \tag{5.2.4}$$

Given these distances, the observations can be clustered using bottom-up hierarchical clustering. To start, all N observations are treated as separate clusters. The two observations with the smallest distance are clustered together and the overall number of clusters reduces to $N - 1$. This continues until the number of clusters is equal to K. The cluster assignments are to be used for obtaining an initial solution to the optimization problem used to perform sensor selection. However, informative features need to be extracted from the raw degradation signals.

5.2.2 Feature Extraction

Since degradation signals monitor the health of systems through their lifetime, the length of degradation signals can be quite large. Therefore, it is advantageous to extract salient features from these degradation signals. In Fang et al. (2017), Functional Principal Component Analysis was utilized for feature extraction. This method assumes that degradation signals are realizations of a stochastic process with a uniform mean function. However, different failures modes may elicit different trends in the sensor data. For example, one failure mode may cause the pressure to rise, while another failure mode may cause the pressure to drop. Therefore, the sensor data has different mean profiles for each failure mode and thus, the uniform mean assumption is not valid. Furthermore, since the failure modes are assumed to be unknown, it is challenging to estimate the multimodal mean function. Therefore, a proposition is to model the degradation signals individually by fitting them to an appropriate growth function. Several examples in the literature have been

proposed to model degradation data. These include linear models (Bae and Kvam, 2008), quadratic models (Chehade et al., 2018), exponential models (Gebraeel et al., 2005), and the Duane Growth Model (Ebling, 2010). In general, each degradation signal from each sensor is fitted individually using an appropriate model. The parameters of this model are then used as features for the MGR-ASGL model. As an example, suppose the degradation signals are modeled as quadratic in time.

$$s_{i,p}[t] = \mathbb{x}_{i,p,0} + \mathbb{x}_{i,p,1}t + \mathbb{x}_{i,p,2}t^2 + \epsilon_{i,p}[t], t \in \{0,1,\dots,T_i - 1\} \qquad (5.2.5)$$

where $\epsilon_{i,p}[t], t \in \{0,1,\dots,T_i - 1\}$ are independent Gaussian random variables with mean 0 and variance $\sigma_{i,p}^2$. For sensor p of unit i, $\mathbb{x}_{i,p} = \left(\mathbb{x}_{i,p,0}, \mathbb{x}_{i,p,1}, \mathbb{x}_{i,p,2}\right)^T$ is extracted and then standardized to obtain: $x_{i,p} = \left(x_{i,p,0}, x_{i,p,1}, x_{i,p,2}\right)^T$. Here, the standardized variables are used as features in the MGR-ASGL model. The standardization serves to enable interpretable regression coefficients in the MGR-ASGL model since the amplitude of the coefficients would represent the significance of the corresponding feature.

5.2.3 Sensor Selection Using MGR-ASGL

To perform sensor selection, the conditional distribution of the $\ln TTF$ given that a set of features from the degradation signals is observed is modeled. It is assumed that given the observed sensor features, the observed failure times are independent. Let $T_i = \exp(Y_i)$ denote the random variable corresponding to the TTF of system i. Then $Y_i = \ln(T_i)$. Furthermore, let $\boldsymbol{Z}_i$ denote a "one-hot" random vector where $\boldsymbol{Z}_i[k] = 1$ if system i failed due to failure mode $k, k = 1, \dots, K$ and equals zero otherwise. Also by assumption, given

that system i failed due to failure mode k, the conditional probability density function (PDF) of $Y_i|\boldsymbol{X}_{i,1},\dots,\boldsymbol{X}_{i,p}$ is the Gaussian PDF as follows:

$$f_{\vartheta_k}(y_i|\boldsymbol{x}_{i,1},\dots,\boldsymbol{x}_{i,P},\boldsymbol{z}_i[k]=1)=\frac{1}{\sqrt{2\pi}\sigma_k^2}\exp\left(-\frac{1}{2\sigma_k^2}\left(y_i-\beta_{0,k}-\sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\beta}_{p,k}\right)^2\right) \quad (5.2.6)$$

where $\vartheta_k=\{\beta_{0,k},\boldsymbol{\beta}_{1,k},\dots,\boldsymbol{\beta}_{P,k},\sigma_k^2\}$ are the model parameters consisting of the regression coefficients and variance respectively for component k. Using the Law of Total Probability, the marginal distribution of $Y_i|\boldsymbol{X}_{i,1},\dots,\boldsymbol{X}_{i,p}$ is obtained as follows:

$$f_{\vartheta}(y_i|\boldsymbol{x}_{i,1},\dots,\boldsymbol{x}_{i,P})=\sum_{k=1}^{K}\pi_k\left(\frac{1}{\sqrt{2\pi}\sigma_k^2}\exp\left(-\frac{1}{2\sigma_k^2}\left(y_i-\beta_{0,k}-\sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\beta}_{p,k}\right)^2\right)\right) \quad (5.2.7)$$

where

$\vartheta=\left\{\left\{\beta_{0,k},\boldsymbol{\beta}_{1,k},\dots,\boldsymbol{\beta}_{P,k}\right\}_{k=1}^{K},\sigma_1^2,\dots,\sigma_K^2,\pi_1,\dots,\pi_{K-1}\right\}\in\mathcal{R}^{K(\sum_{p=1}^{P}q_p+1)}\times\mathcal{R}_{>0}^{K}\times\Pi$ and $\Pi=\{\pi:\pi_k>0 \text{ for } k=1,\dots,K-1 \text{ and } \sum_{k=1}^{K-1}\pi_k<1\},\pi_K=1-\sum_{k=1}^{K-1}\pi_k$. In Equation 5.2.7, $\pi_1,\dots,\pi_K$ are the mixing coefficients and they represent the probability that $\boldsymbol{Z}_i[k]=1$ for $k=1,\dots,K$. Furthermore, q_p are the number of features extracted from sensor p. Therefore, $\mathcal{R}^{K(\sum_{p=1}^{P}q_p+1)}$ is the set of real numbers in dimension $K\left(\sum_{p=1}^{P}q_p+1\right)$. $\mathcal{R}_{>0}^{K}$ is the set of positive real numbers in dimension K. This is the PDF for the MGR model. From Städler et al. (2010), a scale-invariant form of the MGR model is achieved by introducing $\rho_k=1/\sigma_k$, $\psi_{0,k}=\beta_{0,k}/\sigma_k$ and $\boldsymbol{\psi}_{p,k}=\boldsymbol{\beta}_{P,k}/\sigma_k$ for $k=1,\dots,K$ and $p=1,\dots,P$. This alternative form enables the development of a convex optimization problem, which

simplifies parameter estimation. The reformulated mixture of Gaussian regressions is as follows:

$$Y_i|\boldsymbol{X}_{i,1},\dots,\boldsymbol{X}_{i,p} \text{ independent for } i=1,\dots,N$$

$$Y_i|\boldsymbol{X}_{i,1}=\boldsymbol{x}_{i,1},\dots,\boldsymbol{X}_{i,P}=\boldsymbol{x}_{i,p}\sim h_\theta\left(y_i|\boldsymbol{x}_{i,1},\dots,\boldsymbol{x}_{i,P}\right)dy_i \text{ for } i=1,\dots,N$$

$$h_\theta\left(y_i|\boldsymbol{x}_{i,1},\dots,\boldsymbol{x}_{i,P}\right)=\sum_{k=1}^{K}\pi_k\frac{\rho_k}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\rho_k y_i-\psi_{0,k}-\sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\psi}_{p,k}\right)^2\right) \tag{5.2.8}$$

$$\theta=\left\{\left\{\psi_{0,k},\boldsymbol{\psi}_{1,k},\dots\boldsymbol{\psi}_{P,k}\right\}_{k=1}^{K},\rho_1,\dots,\rho_K,\pi_1,\dots,\pi_{K-1}\right\}$$

$$\in\mathcal{R}^{K\left(\sum_{p=1}^{P}q_p+1\right)}\times\mathcal{R}_{>0}^{K}\times\Pi$$

$$\Pi=\left\{\pi:\pi_k>0 \text{ for } k=1,\dots,K-1 \text{ and } \sum_{k=1}^{K-1}\pi_k<1\right\},\pi_K=1-\sum_{k=1}^{K-1}\pi_k$$

Based on this formulation, the incomplete-data log-likelihood (IDLL) is constructed as follows:

$$\ell(\theta;\boldsymbol{Y})=\sum_{i=1}^{N}\ln\sum_{k=1}^{K}\pi_k\left(\frac{\rho_k}{\sqrt{2\pi}}\right)\exp\left(-\frac{1}{2}\left(\rho_k y_i-\psi_{0,k}-\sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\psi}_{p,k}\right)^2\right) \tag{5.2.9}$$

Equation 5.2.9 is incomplete because the cause of the observed failures is unknown. To perform variable selection, Städler et al. (2010) appended the IDLL with the adaptive lasso penalty. This penalty enables modeling the response variable with a sparse selection

of predictors. However, features are extracted from each sensor. The features extracted from a particular sensor represent a group since combined they form a low-dimensional representation of the signal generated from the sensor. To account for this group structure, the use adaptive sparse group lasso (ASGL) penalty is proposed. This penalty ensures that features from important sensors will have nonzero l2-norm whereas features from unimportant sensors will have zero l2-norm and thus, regression coefficients of zero. The adaptive part of the penalty accounts for both the number of features extracted for each sensor and the mixing coefficient for each failure mode. After including the penalty, the objective function to maximize is:

$$\mathcal{L}(\theta; \boldsymbol{Y}) - \lambda \sum_{k=1}^{K} \pi_k \sum_{p=1}^{P} \sqrt{q_p} \left[(1-\alpha) \left\| \boldsymbol{\psi}_{p,k} \right\|_2 + \alpha \left\| \boldsymbol{\psi}_{p,k} \right\|_1 \right] \tag{5.2.10}$$

Here $\lambda > 0$ and $\alpha \in [0,1]$ are tuning parameters. The ASGL penalty is a compromise between adaptive lasso ($\alpha = 1$) and adaptive group lasso ($\alpha = 0$). Intermediate values of α enable a sparse selection of sensors within each group, which is useful because its plausible that not all features within a group are correlated with the response.

Maximization of the objective function in Equation 2.10 is equivalent to minimizing the objective function with a sign change. Therefore, the optimization problem to solve to fit the MGR-ASGL model is:

$$\min_{\theta} -\sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k \left(\frac{\rho_k}{\sqrt{2\pi}}\right) \exp\left(-\frac{1}{2}\left(\rho_k y_i - \psi_{0,k} - \sum_{p=1}^{P} \boldsymbol{x}_{i,p}^T \boldsymbol{\psi}_{p,k}\right)^2\right) + \lambda \sum_{k=1}^{K} \pi_k \sum_{p=1}^{P} \sqrt{q_p} \left[(1-\alpha)\left\|\boldsymbol{\psi}_{p,k}\right\|_2 + \alpha\left\|\boldsymbol{\psi}_{p,k}\right\|_1\right] \quad (5.2.11)$$

Due to the summation within the logarithm in Equation 5.2.11, the objective function is nonconvex, thus the function in Equation 5.2.10 is nonconcave. Therefore, an Expectation-Maximization (EM) algorithm is developed to find a local maximum to Equation 5.2.10 by finding the local minimum to the problem in Equation 5.2.11. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$ denote the observed $\ln TTF$ and $\mathbb{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_N)$ denote the set of independent binary random variables representing the unknown failure mode for all observations. The first part of the EM algorithm is to construct the complete-data log-likelihood (CDLL). Compared to the IDLL, this function assumes knowledge of the underlying causes of failure. The CDLL is shown below:

$$\ell_C(\theta; \boldsymbol{Y}, \mathbb{Z}) = \sum_{i=1}^{N} \sum_{k=1}^{K} Z_i[k] \ln \pi_k \left(\frac{\rho_k}{\sqrt{2\pi}}\right) \exp\left(-\frac{1}{2}\left(\rho_k y_i - \psi_{0,k} - \sum_{p=1}^{P} \boldsymbol{x}_{i,p}^T \boldsymbol{\psi}_{p,k}\right)^2\right) \quad (5.2.12)$$

The benefit of the CDLL is that now, the logarithm is inside the summation and thus, Equation 5.2.12 is convex in the parameters θ. While the underlying cause of failure is unknown, the expectation of Equation 5.2.12 can be formulated and optimized instead. This is known as the E-step of the EM algorithm and it is described below.

5.2.3.1 E-step

In the E-step, the expectation of Equation 5.2.12 is computed with respect to the posterior distribution of $\mathbb{Z}$ given $\boldsymbol{Y}$ and θ_{old}, where θ_{old} are the parameter estimates from the previous iteration. Let $\mathcal{Q}(\theta|\theta_{old}) = -E_{\theta_{old}}[\ell_C(\theta; \boldsymbol{Y}, \mathbb{Z})|\boldsymbol{Y}]$ denote the negative expectation of the CDLL. The only term in the CDLL that depends on $\mathbb{Z}$ is $Z_i[k]$. Therefore, to compute $\mathcal{Q}(\theta|\theta_{old})$, $\gamma_{i,k} = E_{\theta_{old}}[Z_i[k]|\boldsymbol{Y}]$, which represents the responsibility of failure mode k to observation i, is computed using Bayes' formula.

$$\gamma_{i,k} = \frac{\pi_k^{old}\rho_k^{old}\exp\left(-\frac{1}{2}\left(\rho_k^{old}y_i - \psi_{0,k}^{old} - \sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\psi}_{p,k}^{old}\right)^2\right)}{\sum_{l=1}^{K}\pi_l^{old}\rho_l^{old}\exp\left(-\frac{1}{2}\left(\rho_l^{old}y_i - \psi_{0,l}^{old} - \sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\psi}_{p,l}^{old}\right)^2\right)} \tag{5.2.13}$$

A key outcome of the E-step is the clustering of observed failures into their respective failure modes. This results in class labels for the training set, which are leveraged when developing models for predicting RUL.

5.2.3.2 <u>M-step</u>

For the M-step, the ASGL penalty is appended to $\mathcal{Q}(\theta|\theta_{old})$ resulting in the following optimization problem:

$$\begin{aligned}\min_{\theta} -\sum_{i=1}^{N}\sum_{k=1}^{K}\gamma_{i,k}\ln\pi_k\left(\frac{\rho_k}{\sqrt{2\pi}}\right)\exp\left(-\frac{1}{2}\left(\rho_k y_i - \psi_{0,k} - \sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\psi}_{p,k}\right)^2\right) \\ + \lambda\sum_{k=1}^{K}\pi_k\sum_{p=1}^{P}\sqrt{q_p}\left[(1-\alpha)\left\|\boldsymbol{\psi}_{p,k}\right\|_2 + \alpha\left\|\boldsymbol{\psi}_{p,k}\right\|_1\right]\end{aligned} \tag{5.2.14}$$

Solving this optimization problem requires minimizing the objective function with respect to each of the parameters.

a) Improvement with respect to $\boldsymbol{\pi}$

To update the mixing coefficients, the following problem is solved for $k = 1, \dots, K$:

$$\min_{\pi_k} -\sum_{i=1}^{N} \gamma_{i,k} \ln \pi_k + \lambda \pi_k \sum_{p=1}^{P} \sqrt{q_p} \left[(1-\alpha) \left\| \boldsymbol{\psi}_{p,k} \right\|_2 + \alpha \left\| \boldsymbol{\psi}_{p,k} \right\|_1 \right] \quad (5.2.15)$$

$$s.t. \sum_{l=1}^{K} \pi_l = 1$$

Using a Lagrange multiplier η, the problem to solve becomes:

$$\min_{\pi_k} -\sum_{i=1}^{N} \gamma_{i,k} \ln \pi_k + \lambda' \pi_k - \eta \left(\sum_{l=1}^{K} \pi_l - 1 \right) \quad (5.2.16)$$

Differentiating the objective function with respect to π_k and setting it equal to zero yields:

$$\pi_k (\eta - \lambda') = -\sum_{i=1}^{N} \gamma_{i,k} \quad (5.2.17)$$

Summing over k on both sides of the equation yields $= \lambda' - N$. Then, substituting η_k in Equation 5.2.17 with $\lambda' - N$ yields:

$$\pi_k^{new} = \frac{1}{N}\sum_{i=1}^{N}\gamma_{i,k} \tag{5.2.18}$$

Therefore, the mixing coefficients are updated as the average of the N responsibilities.

b) Improvement with respect to $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)$ and $\Psi = \left\{\psi_{0,k}, \left\{\boldsymbol{\psi}_{p,k}\right\}_{p=1}^{P}\right\}_{k=1}^{K}$

Equation 5.2.14 decouples into K distinct optimization problems of the form:

$$\begin{aligned}\min_{\rho_k, \psi_{0,k}, \left\{\boldsymbol{\psi}_{p,k}\right\}_{p=1}^{P}} & -\left(\sum_{i=1}^{N}\gamma_{i,k}\right)\ln\rho_k + \frac{1}{2}\sum_{i=1}^{N}\gamma_{i,k}\left(\rho_k y_i - \psi_{0,k} - \sum_{p=1}^{P}\boldsymbol{x}_{i,p}^T\boldsymbol{\psi}_{p,k}\right)^2 \\ & + \lambda\pi_k^{new}\sum_{p=1}^{P}\sqrt{q_p}\left[(1-\alpha)\left\|\boldsymbol{\psi}_{p,k}\right\|_2 + \alpha\left\|\boldsymbol{\psi}_{p,k}\right\|_1\right], k \\ & = 1, \dots, K\end{aligned} \tag{5.2.19}$$

For each k, this is solved using the convex optimizer in MATLAB.

To summarize, an EM algorithm is developed for finding a local maximum to Equation 5.2.10. Through this algorithm, failure diagnosis and sensor selection for each failure mode can be achieved simultaneously. Diagnosis is performed in the E-step where Equation 5.2.13 is used to compute the expectation of the CDLL. Then in the M-step, sensor selection is performed by updating the parameter estimates through minimization of Equation 5.2.14. First, $\boldsymbol{\pi}$ is updated using the average of the responsibilities obtained in the E-step. Then for $k = 1, \dots, K$, the convex optimizer is used to estimate the model parameters: $\left\{\rho_k, \psi_{0,k}, \left\{\boldsymbol{\psi}_{p,k}\right\}_{p=1}^{P}\right\}$. Successive iterations of the EM algorithm are performed until the model converges. For $k = 1, \dots, K$, the sensor groups whose coefficients have nonzero Euclidean norm are said to be significant for failure mode k.

5.3 Signal Data Fusion, Diagnosis, and RUL Prediction

In this section, the cluster labels and sensor selection from the previous step is leveraged to develop both an adaptive classifier and adaptive regression model that are used together to 1) detect the failure mode most likely contributing to the degradation and 2) predict the system RUL. The degradation of a fielded system is monitored using the sensors selected by the MGR-ASGL model. After observing degradation up to some time t^*, two data fusion techniques are implemented, both incorporating Multivariate Functional Principal Component Analysis (MFPCA), a functional data analysis technique that captures dominant modes of variation in multivariate signal data. Using the training set, an MFPCA model is fitted to all the multivariate degradation signals that have survived up to time t^*. The MFPC scores are used to fit a supervised mixture of Gaussians model, which serves as the classifier. Still using the training set, a functional regression model for each failure mode is then developed, where the signals from the sensors selected for that failure mode are used as a multivariate functional predictor. Using MFPCA, the functional regression model is reduced to a traditional multiple linear regression model where the predictors are the MFPC scores. The regression model is appended with the lasso penalty to select the most informative MFPC scores for predicting RUL. For a fielded system, the MFPCA model fitted with all the sensors is used to extract MFPC scores. Then, classification is performed by selecting the Gaussian component most likely to have generated the MFPC scores. Following classification, the MFPC scores are extracted using the MFPCA model for the diagnosed failure mode. These scores are used as inputs into the corresponding regression model to predict the RUL. Prior to fitting these models, data quality issues that could compromise the effectiveness of this methodology are addressed.

5.3.1 Data Curation

Sensor data is often corrupted by noise and outliers caused by intermittent sensor faults. If not addressed, these data quality issues can compromise the diagnostic and prognostic accuracy of the methodology. Therefore, a data curation strategy that addresses data quality issues is important. For this methodology, the problem of noise and intermittent outlier is addressed by smoothing the sensor data using Robust Locally Weighted Regression (Cleveland, 1979). In this methodology, points that largely deviate from the smoothed signal are assigned a weight of zero and the smoothing is repeated with updated weights. Successive iterations of this procedure are performed until there are no detected outliers. The details of Robust Locally Weighted Regression are as follows: let $\{s_{i,p}[t]\}_{t=0}^{T_i-1}$ denote the noisy degradation signal recorded by sensor $p, p = 1, \dots, P$ from unit $, i = 1, \dots, N$. Robust Locally Weighted Regression uses kernel functions as weights to perform smoothing. Define $u_{i,p}^{v} \coloneqq \frac{v-t}{\omega_{i,p}}$ as the input into a kernel function, where $\omega_{i,p}$ is the bandwidth parameter for the kernel function, $v \in \{0,1, \dots, T_i - 1\}$ is a time epoch at which the signal is sampled, and t is the time epoch at which the kernel function is evaluated. A kernel function $\mathcal{K}(u_{i,p}^{v})$ is defined by the following properties:

1. $\int_{-\infty}^{\infty} \mathcal{K}(u_{i,p}^{v}) du_{i,p}^{v} = 1$
2. $\mathcal{K}(-u_{i,p}^{v}) = \mathcal{K}(u_{i,p}^{v})$
3. $\int_{-\infty}^{\infty} (u_{i,p}^{v})^2 \mathcal{K}(u_{i,p}^{v}) du_{i,p}^{v} < \infty$

Local kernel quadratic smoothing is performed at time t by solving the following optimization problem:

$$\left\{\hat{b}_{i,p}^{a}(t)\right\}_{a=0}^{2} = \arg\min_{\left\{\hat{b}_{i,p}^{a}(t)\right\}_{a=1}^{2}} \sum_{v=0}^{T_i-1} \mathcal{K}\left(u_{i,p}^{v}\right)\left\{s_{i,p}[v] - \sum_{a=0}^{2} b_{i,p}^{a}(t)v^{a}\right\}^{2} \quad (5.3.1)$$

Then for $t \in \{0,1,\dots,T_i-1\}$, $s_{i,p}[t]$ is estimated as:

$$\hat{s}_{i,p}[t] = \sum_{a=0}^{2} \hat{b}_{i,p}^{a}(t)t^{a} \quad (5.3.2)$$

Given this formulation, Robust Locally Weighted Regression is implemented as shown in Algorithm 1.

Algorithm 1: Robust Locally Weighted Regression (Cleveland, 1979)

Input: $s_{i,p}[t], t = 0,1,\dots,T_i-1, max\_iter$

Output: $\hat{s}_{i,p}[t], t = 0,1,\dots,T_i-1$

$iter = 1$

Choose a bandwidth parameter and a kernel function and perform local kernel polynomial smoothing for $t = 0,1,\dots,T_i-1$

While $iter \leq max\_iter$ **do**

1. Calculate the residuals of the smoother $r_{i,p}[t] = s_{i,p}[t] - \hat{s}_{i,p}[t]$ for $t = 0,1,\dots,T_i-1$
2. Compute the robust weights for each data point within the bandwidth

$$\delta(t) = \begin{cases} \left(1 - \left(\frac{r_{i,p}[t]}{6MAD}\right)^{2}\right)^{2} & \left|r_{i,p}[t]\right| < 6MAD \\ 0 & \left|r_{i,p}[t]\right| \geq 6MAD \end{cases}$$

   $MAD = median\left(\left|r_{i,p}[0]\right|,\dots,\left|r_{i,p}[T_i-1]\right|\right)$
3. Smooth the data again using Equations 3.1 and 3.2 but replacing $\mathcal{K}\left(u_{i,p}^{v}\right)$ with $\delta\left(t_{i,j}\right)\mathcal{K}\left(u_{i,p}^{v}\right)$
4. $iter = iter + 1$

End

This algorithm is applied to all selected sensor signals for all observations in the training set. It is also applied to signals from fielded systems by replacing $T_i - 1$ with $t^* - 1$. By

removing the noise and intermittent outliers from the signals, highly informative features can be extracted from the sensor data. This is accomplished using MFPCA as discussed in the following subsection.

5.3.2 Sensor Fusion via Multivariate Functional Principal Component Analysis

MFPCA is a functional data analysis tool for capturing dominant modes of variation in multi-dimensional functional data. It works by projecting multi-dimensional signals onto an orthonormal basis comprised of multi-dimensional eigenfunctions of the signals' covariance function. For this application, MFPCA serves two primary purposes. First, it provides a means for fusing important degradation signals into a low-dimensional, but highly informative set of MFPC scores that can be used to predict RUL (Fang et al, 2017). Second, it enables the reduction of functional regression to multiple linear regression. The details of this are described in Section 5.3.4. Without loss of generality, redefine the sensor index as $\{1,2,\dots,\mathcal{P}\}$, where $\mathcal{P}$ is the number of sensors selected across all failure modes using the MGR-ASGL model. Furthermore, the index i is replaced with (i) to denote the training signals being sorted in descending order of their TTFs. In MFPCA, the multi-dimensional signal model is as follows:

$$\hat{\boldsymbol{s}}_{(i)}[t] = \boldsymbol{\mu}[t] + \sum_{\hbar=1}^{\infty} \zeta_{(i),\hbar}\boldsymbol{\varpi}_{\hbar}[t] + \boldsymbol{\epsilon}_{(i)}[t], t = 0,1,\dots,t^* - 1 \qquad (5.3.3)$$

Here, $\boldsymbol{\mu}[t] = (\mu_1[t],\dots,\mu_{\mathcal{P}}[t])^T \in \mathcal{R}^{\mathcal{P}}$ is the mean vector at time t, $\boldsymbol{\epsilon}_{(i)}[t] \sim N(\mathbf{0}, \varrho^2\mathbb{I}_{\mathcal{P}\times\mathcal{P}})$ are independent Gaussian errors with variance ϱ^2, and $\boldsymbol{\varpi}_{\hbar}[t]$ is a $\mathcal{P}$-dimensional function evaluated at time t. $\boldsymbol{\varpi}_{\hbar}$ is the $\hbar$th eigenfunction of the covariance function

$$\mathbb{C}[t,t'] = E[(\hat{\boldsymbol{s}}[t] - \boldsymbol{\mu}[t])(\hat{\boldsymbol{s}}[t'] - \boldsymbol{\mu}[t'])^T], t, t' = 0,1,\dots,t^* - 1 \qquad (5.3.4)$$

corresponding to the $\hbar$th ordered eigenvalue $\chi_\hbar$, where $\chi_1 > \cdots > \chi_\hbar > \cdots$. Here, the covariance function is a $\mathcal{P} \times \mathcal{P}$ block matrix whose elements are the cross-covariance functions.

$$\mathbb{C}[t,t'] = \begin{bmatrix} \mathbb{C}_{1,1}[t,t'] & \cdots & \mathbb{C}_{1,\mathcal{P}}[t,t'] \\ \vdots & \ddots & \vdots \\ \mathbb{C}_{\mathcal{P},1}[t,t'] & \cdots & \mathbb{C}_{\mathcal{P},\mathcal{P}}[t,t'] \end{bmatrix} \qquad (5.3.5)$$

where $\mathbb{C}_{p,p'}[t,t'] = E\left[\left(\hat{s}_p[t] - \mu_p[t]\right)\left(\hat{s}_{p'}[t'] - \mu_{p'}[t']\right)^T\right]$. The MFPC scores, $\zeta_{(i),\hbar}, \hbar = 1,2,\dots,\mathcal{K}$ can be found by projecting the mean subtracted signals onto the first $\mathcal{K}$ eigenfunctions where $\mathcal{K}$, the number of MFPC scores can be chosen using fraction of variance explained criteria. The equation for computing the MFPC scores is:

$$\zeta_{(i),\hbar} = \sum_{t=0}^{t^*-1} \left(\hat{\boldsymbol{s}}_{(i)}[t] - \boldsymbol{\mu}[t]\right)^T \boldsymbol{\varpi}_\hbar[t], \hbar = 1,2,\dots,\mathcal{K} \qquad (5.3.6)$$

In practice, MFPCA is performed by concatenating the smoothed degradation signals from all selected sensors into a single vector and performing Principal Component Analysis (PCA) on these observations. In this application, Singular Value Decomposition is used to obtain the eigenfunctions. Suppose the number of surviving units at time t^* is N_{t^*}. Let

$$\boldsymbol{S} = \begin{bmatrix} \left(\hat{\boldsymbol{s}}_{(1)} - \boldsymbol{\mu}\right)^T \\ \vdots \\ \left(\hat{\boldsymbol{s}}_{(N_{t^*})} - \boldsymbol{\mu}\right)^T \end{bmatrix}$$

denote the matrix whose row vectors are the mean subtracted signals with $\hat{\boldsymbol{s}}_{(i)} = \left(\hat{s}_{(i),1}[0], \dots, \hat{s}_{(i),1}[t^* - 1], \dots, \hat{s}_{(i),\mathcal{P}}[0], \dots, \hat{s}_{(i),\mathcal{P}}[t^* - 1]\right)^T$ and $\boldsymbol{\mu} = (\mu_1[0], \dots, \mu_1[t^* - 1], \dots, \mu_{\mathcal{P}}[t^* - 1], \dots, \mu_{\mathcal{P}}[t^* - 1])^T$. The Singular Value Decomposition of $\boldsymbol{S}$ is shown below:

$$\boldsymbol{S} = \mathbb{U}\Sigma\mathbb{W}^T \tag{5.3.7}$$

Here, $\mathbb{U}$ is an $N_{t^*} \times N_{t^*}$ matrix whose columns are the orthogonal left singular vectors of $\boldsymbol{S}$, Σ is a $N_{t^*} \times t^*\mathcal{P}$ rectangular diagonal matrix of the singular values of $\boldsymbol{S}$, and $\mathbb{W}$ is a $t^*\mathcal{P} \times t^*\mathcal{P}$ matrix whose columns are the concatenated discretized eigenfunctions. Using only the first $\mathcal{K}$ eigenfunctions, The MFPC scores can be computed as:

$$\mathrm{Z} = (\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{\mathcal{K}}) = \boldsymbol{S}\mathbb{W}^{\mathcal{K}} \tag{5.3.8}$$

where $\boldsymbol{\zeta}_{\mathcal{k}} = \left(\zeta_{1,\mathcal{k}}, \dots, \zeta_{N_{t^*},\mathcal{k}}\right)^T$ and $\mathbb{W}^{\mathcal{K}}$ is comprised of the first $\mathcal{K}$ columns of $\mathbb{W}$. The mean signal $\boldsymbol{\mu}$ is estimated by computing the sample average of the N_{t^*} degradation signals.

5.3.3 Classification Using Supervised Mixture of Gaussians

For a system monitored up to time t^*, MFPCA is first performed on all sensor signals selected by the MGR-ASGL model. Then, the failure mode most likely contributing to the observed degradation is diagnosed. To perform diagnosis, the MFPC scores are used to fit

a mixture of Gaussians model. Since the features are updated after every time epoch, the model is fitted in an adaptive fashion. For the mixture of Gaussians model, the marginal distribution of $\boldsymbol{\zeta}_{(i)}$ is modeled as follows:

$$f\left(\boldsymbol{\zeta}_{(i)}\middle|\{\pi_k', \boldsymbol{\mu}_k', \Sigma_k'\}_{k=1}^{K}\right) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{\zeta}_{(i)}|\mu_k', \Sigma_k'\right) \tag{5.3.9}$$

where $\mathcal{N}\left(\boldsymbol{\zeta}_{(i)}|\boldsymbol{\mu}_k', \Sigma_k'\right)$ is the PDF of the multivariate Gaussian distribution with mean μ_k' and covariance Σ_k'. Let $\theta' = \{\pi_k', \boldsymbol{\mu}_k', \Sigma_k'\}_{k=1}^{K}$. Since the cluster assignments are known due to the sensor selection algorithm in Section 5.2, the complete-data log-likelihood for the mixture of Gaussians can be formulated as follows:

$$\mathcal{L}\left(\theta'; \boldsymbol{\zeta}_{(1)}, \dots, \boldsymbol{\zeta}_{(N_{t^*})}\right) = \sum_{(i)=1}^{N_{t^*}} \sum_{k=1}^{K} \boldsymbol{Z}_{(i)}[k] \ln\left(\pi_k \mathcal{N}\left(\boldsymbol{\zeta}_{(i)}|\boldsymbol{\mu}_k', \Sigma_k'\right)\right) \tag{5.3.10}$$

Let N_k denote the number of training observations that fail due to failure mode $k, k = 1, \dots, K$. The model parameters are estimated using the Maximum Likelihood estimates as follows:

$$\begin{aligned} \hat{\pi}_k' = \frac{1}{N_k} \sum_{(i)=1}^{N_{t^*}} \boldsymbol{Z}_{(i)}[k]\,, \hat{\boldsymbol{\mu}}_k' = \frac{1}{N_k} \sum_{(i)=1}^{N_{t^*}} \boldsymbol{Z}_{(i)}[k]\boldsymbol{\zeta}_{(i)}\,, \hat{\Sigma}_k' \\ = \frac{1}{N_k} \sum_{(i)=1}^{N_{t^*}} \boldsymbol{Z}_{(i)}[k]\left(\boldsymbol{\zeta}_{(i)} - \hat{\boldsymbol{\mu}}_k'\right)\left(\boldsymbol{\zeta}_{(i)} - \hat{\boldsymbol{\mu}}_k'\right)^T \end{aligned} \tag{5.3.11}$$

To perform diagnosis, first extract MFPC scores for the fielded system using Equation 5.3.6. Let $\boldsymbol{\zeta}_*$ denote the MFPC scores for the fielded system. Using these scores, the failure mode k with the maximum likelihood given the observed scores is determined using the following equation:

$$k^* = \arg\max_{k=1,\dots,K} \pi_k \mathcal{N}(\boldsymbol{\zeta}_* | \boldsymbol{\mu}_k', \Sigma_k') \tag{5.3.12}$$

Given the diagnosis, a model for predicting the system RUL is developed.

5.3.4 Remaining Useful Life Prediction Using Functional Regression

A functional regression model that maps multivariate sensor signals to the $\ln TTF$ is used to predict the RUL. While the sensor signals selected across all failure modes were used for classification, only use the sensors selected for the diagnosed failure mode are used for prognostics. Without loss of generality, let $\{(1)^{k^*}, (2)^{k^*}, \dots, (N_{t^*})^{k^*}\}$ denote the indices of training set of units that both survived up to time t^* and failed due to failure mode k ranked in descending order of their TTFs.. Let $\hat{\boldsymbol{s}}_i^{k^*}[t] \in \mathcal{P}_{k^*}$ denote the vector of smooth degradation signals selected for estimated failure mode k^* and sampled at time $t \in \{0,1, \dots, t^* - 1\}$. Here, $\mathcal{P}_{k^*}$ is the number of sensors selected for estimated failure mode k^*. To fit a functional regression model, the following least squares problem is solved.

$$\min_{\phi_0, \{\boldsymbol{\phi}[t]\}_{t=0}^{t^*-1}} \sum_{i=(1)}^{(N_{t^*})} \left(y_i - \phi_0 - \sum_{t=0}^{t^*-1} \boldsymbol{\phi}[t]^T \hat{\boldsymbol{s}}_i^{k^*}[t] \right)^2 \tag{5.3.13}$$

where $y_i = \ln(T_i)$, ϕ_0 is the bias term, and $\boldsymbol{\phi}[t] \in \mathcal{R}^{\mathcal{P}_k}$ is the multivariate coefficient function of dimension $\mathcal{P}_{k^*}$ sampled at time t. In MFPCA, the discretized multivariate eigenfunctions of the signals' covariance function form an orthonormal basis in $\mathcal{R}^{t^* \times \mathcal{P}_{k^*}}$. Let $\boldsymbol{\varpi}_{\hbar}^{k^*}[t] \in \mathcal{R}^{\mathcal{P}_{k^*}}$ denote the $\hbar$th eigenfunction sampled at time t for the estimated failure mode k^*. An orthonormal expansion is used to represent the regression coefficient function in Equation 5.3.13 as a weighted sum of the $\mathcal{K}_{k^*}$ eigenfunctions, where $\mathcal{K}_{k^*}$ are the number of MFPC scores selected for the estimated failure mode k^*.

$$\boldsymbol{\phi}[t] = \sum_{\hbar=1}^{\mathcal{K}_{k^*}} c_{\hbar}^{k^*} \boldsymbol{\varpi}_{\hbar}^{k^*}[t] \tag{5.3.14}$$

Likewise, the multivariate sensor signal is expanded using the eigenfunctions.

$$\hat{\boldsymbol{s}}_i^{k^*}[t] = \boldsymbol{\mu}^{k^*}[t] + \sum_{\hbar=1}^{\mathcal{K}_{k^*}} \zeta_{i,\hbar}^{k^*} \boldsymbol{\varpi}_{\hbar}^{k^*}[t] \tag{5.3.15}$$

Since the eigenfunctions are orthonormal, substituting Equations 3.14 and 3.15 into 3.13 yields:

$$\min_{c_0^{k^*}, c_1^{k^*}, \ldots, c_{\mathcal{K}_{k^*}}^{k^*}} \sum_{i=(1)}^{(N_{t^*})} \left(y_i - c_0^{k^*} - \sum_{\hbar=1}^{\mathcal{K}_{k^*}} \zeta_{i,\hbar}^{k^*} c_{\hbar}^{k^*} \right)^2 \tag{5.3.16}$$

where $c_{i,0}^{k^*} = \phi_0 + \sum_{t=0}^{t^*-1} \sum_{\hbar=1}^{\mathcal{K}_{k^*}} c_{\hbar}^{k^*} \boldsymbol{\varpi}_{\hbar}^{k^*}[t]^T \boldsymbol{\mu}^{k^*}[t]$. Since not all MFPC scores may be informative, the objective function in Equation 5.3.16 is appended with the lasso penalty to select a subset of MFPC scores for prediction. Thus, the following problem is solved:

$$\min_{c_0^{k^*},c_1^{k^*},\dots,c_{\mathcal{K}_{k^*}}^{k^*}} \sum_{i=1}^{N_{t^*}} \left(y_i - c_0^{k^*} - \sum_{\hbar=1}^{\mathcal{K}_{k^*}} \zeta_{i,\hbar}^{k^*} c_{\hbar}^{k^*} \right)^2 + \lambda \sum_{\hbar=1}^{\mathcal{K}_{k^*}} \left| c_{\hbar}^{k^*} \right| \tag{5.3.17}$$

MFPCA enables the reduction of the functional regression problem to a multiple linear regression problem. Therefore, instead of estimating $t^*\mathcal{P}_{k^*} + 1$ parameters, only $\mathcal{K}_{k^*} + 1$ parameters need to be estimated, which is much less computationally expensive. Lasso regression is a well-studied problem (Tibshirani, 1996) that can be solved using Least Angle Regression (LARS) algorithm (Efron et al., 2004). After fitting the regression coefficients, the MFPC scores for the test signal $\zeta_{*,1}^{k^*}, \dots, \zeta_{*,\mathcal{K}_{k^*}}^{k^*}$ are extracted. These scores are then used to predict the RUL using the formula:

$$RUL_* = \exp\left(\hat{c}_0^{k^*} + \sum_{\hbar=1}^{\mathcal{K}_{k^*}} \zeta_{*,\hbar}^{k^*} \hat{c}_{\hbar}^{k^*} \right) - t^* \tag{5.3.18}$$

To summarize, the MGR-ASGL model was used to cluster the failures in the training set and to select optimal subsets of sensors for each failure mode. To predict RUL at time t^*, MFPCA is first performed using the smooth training signals from sensors selected across all failure modes. The MFPC scores along with the cluster assignments are used to train a mixture of Gaussians model in a supervised fashion. Then, the eigenfunction basis from MPCA is used to transform the sensor signals from the system currently degrading to a small number of MFPC scores. Using the mixture of Gaussians model, the cause of degradation in the system is diagnosed. Following this, MFPCA is performed again, but only using the signals from sensors selected for the diagnosed failure mode. Furthermore, only units that failed due to the diagnosed failure mode are included in the analysis. To

predict the TTF, a functional regression model is fitted. It was demonstrated that through MFPCA, the functional regression model can be transformed to a multiple linear regression model. Using this model, the MFPC scores extracted from the currently monitored system are mapped to the RUL via Equation 5.3.18. In the next section, the proposed methodology is demonstrated using a dataset from an aircraft engine failure simulation.

5.4 Case Study – Turbofan Engine

In this paper, a methodology for predicting RUL in systems with multiple failure modes was developed. To demonstrate the proposed methodology, a simulated aircraft turbofan engine dataset generated by the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) at NASA is utilized. This is a tool designed to realistically simulate the dynamics of a commercial turbofan engine of 90,000 lb. thrust class (Saxena et al., 2008). A diagram of the engine components is shown in Figure 5.1.

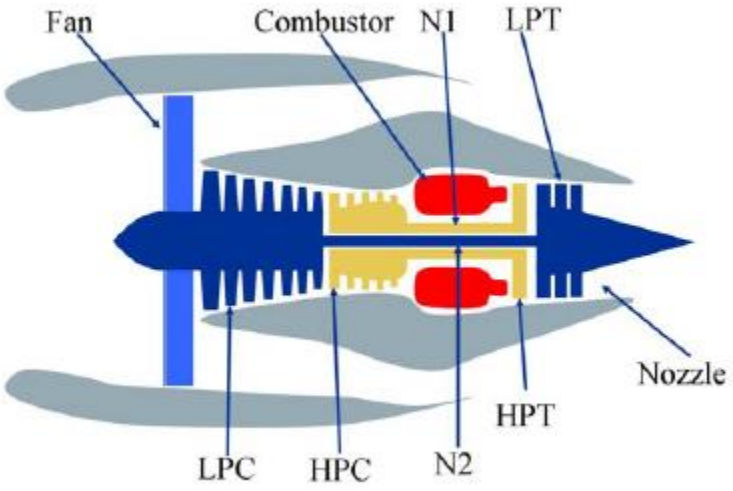


Figure 5.1 – Simplified Diagram of Engine Simulated in C_MAPSS (Saxena et al., 2008)

Four publicly available datasets have been generated using this simulation. For this work, C-MAPSS dataset 3 is utilized. In this dataset, the engine has two failure modes: the high-pressure compressor (HPC) and the fan. The role of the compressor is to compress air into a high-pressure stream that mixes with fuel and combusts providing a very hot fuel/air gaseous mixture that turns the turbine blades, powering the engine. Most of this power is used to turn the fan, which acts a propeller for aircraft. A smaller proportion of power is used to power the various compressor stages. A damaged compressor may result in a temporary loss of power or complete loss of compression. Meanwhile, a fan can experience cracks and chips which act as shrapnel that propagate into the engine and damage key turbine components. Failure of an engine aircraft incurs major economic and personal loss. Therefore, there is high value in the ability to predict the RUL of these engines.

In the simulation, the engines are monitored by 21 sensors. The names of these sensors along with their functionality is displayed in Table 5.1.

Table 5.1 – C-MAPSS Sensors

Symbol	Description	Units
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	Psia
P15	Total pressure in bypass duct	Psia
P30	Total pressure at HPC outlet	Psia
Nf	Physical fan speed	Rpm
Nc	Physical core speed	Rpm
epr	Engine pressure ratio (P50/P2)	
Ps30	Static pressure at HPC outlet	Psia
Phi	Ratio of fuel flow to Ps30	Pps/psi
NRf	Corrected fan speed	Rpm
NRc	Corrected core speed	Rpm
BPR	Bypass ratio	
farB	Burner fuel air ratio	
htBleed	Bleed enthalpy	
Nf_dmd	Demanded fan speed	Rpm
PCNfR_dmd	Demanded corrected fan speed	Rpm
W31	HPT coolant bleed	Lbm/s
W32	LPT coolant bleed	Lbm/s

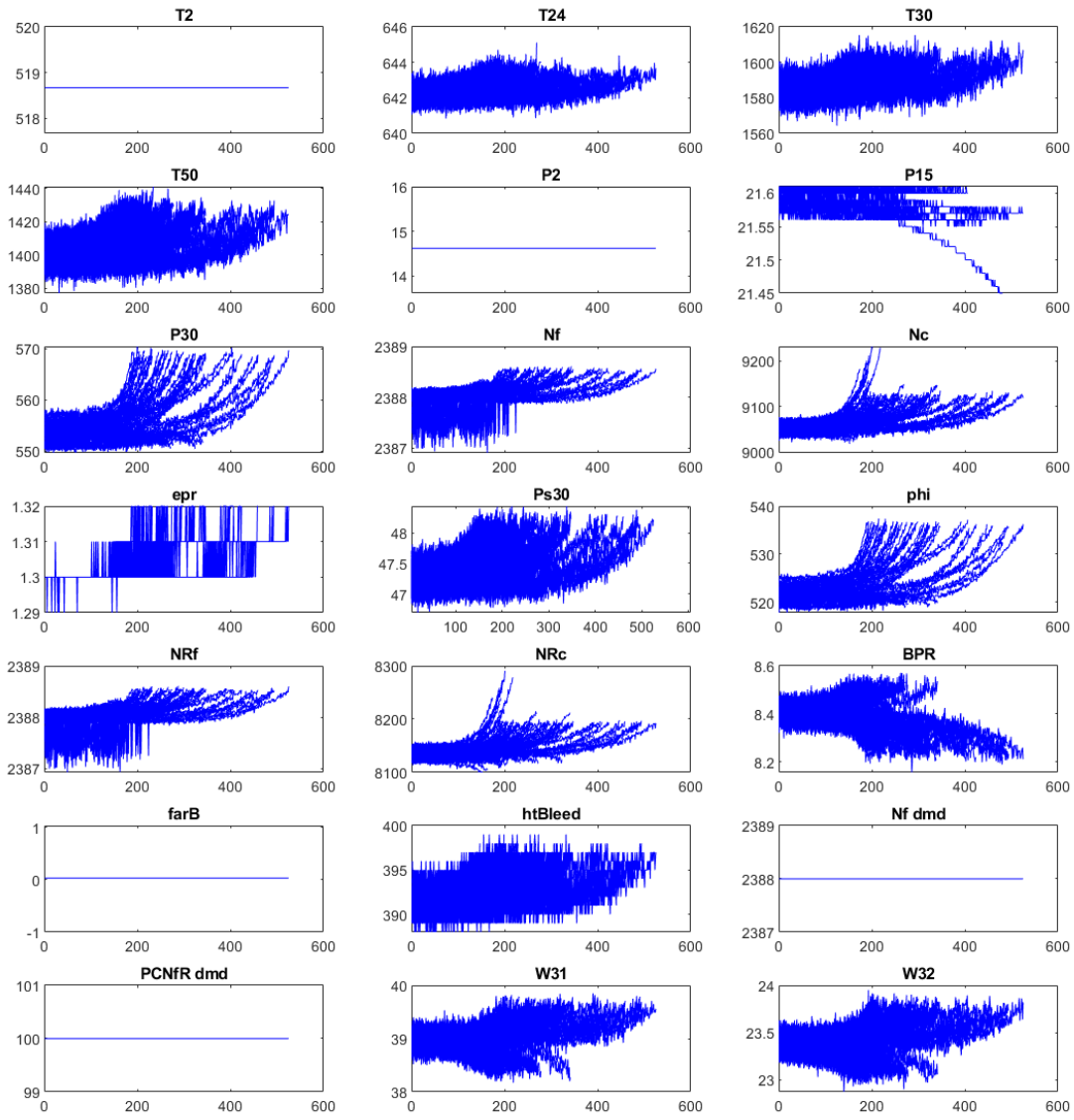


Figure 5.2 – C-MAPSS Sensor Signals

Figure 5.2 depicts the complimentary signals of the sensors present in the system. Note that T2, P2, P15, farB, NF_dmd, and PCNfR_dmd show virtually no trend or variation. To test the robustness of our sensor selection methodology to the presence of noninformative sensors, independently and identically distributed white noise with unit variance is added to these sensors.

This dataset consists of a training set and a test set. The training set includes 100 samples of degradation signals and the TTF for these observations is inferred from the length of the signal. However, the cause of failure is unknown. The test set also includes 100 samples. However, the test set only includes partially degraded systems and the RUL. Like with the training set, the failure mode is also unknown. For each test engine, the goal is to predict the RUL given the degradation history of the system.

The first step of the proposed methodology is to utilize the sensor selection algorithm to cluster the data and select an optimal subset of sensors for each failure mode. Fitting this model requires initial cluster assignments and features on which to regress the $\ln TTF$. Note that clustering and feature extraction can be done in parallel. To obtain initial cluster assignments, DTW, as described in Section 5.2.1, is used to compute a distance between all pairs of multidimensional signals. These distances are used by the hierarchical clustering algorithm. Since there are two failure modes in the dataset, the number of failure modes is $K = 2$. The result of initial clustering is shown in Figure 5.3.

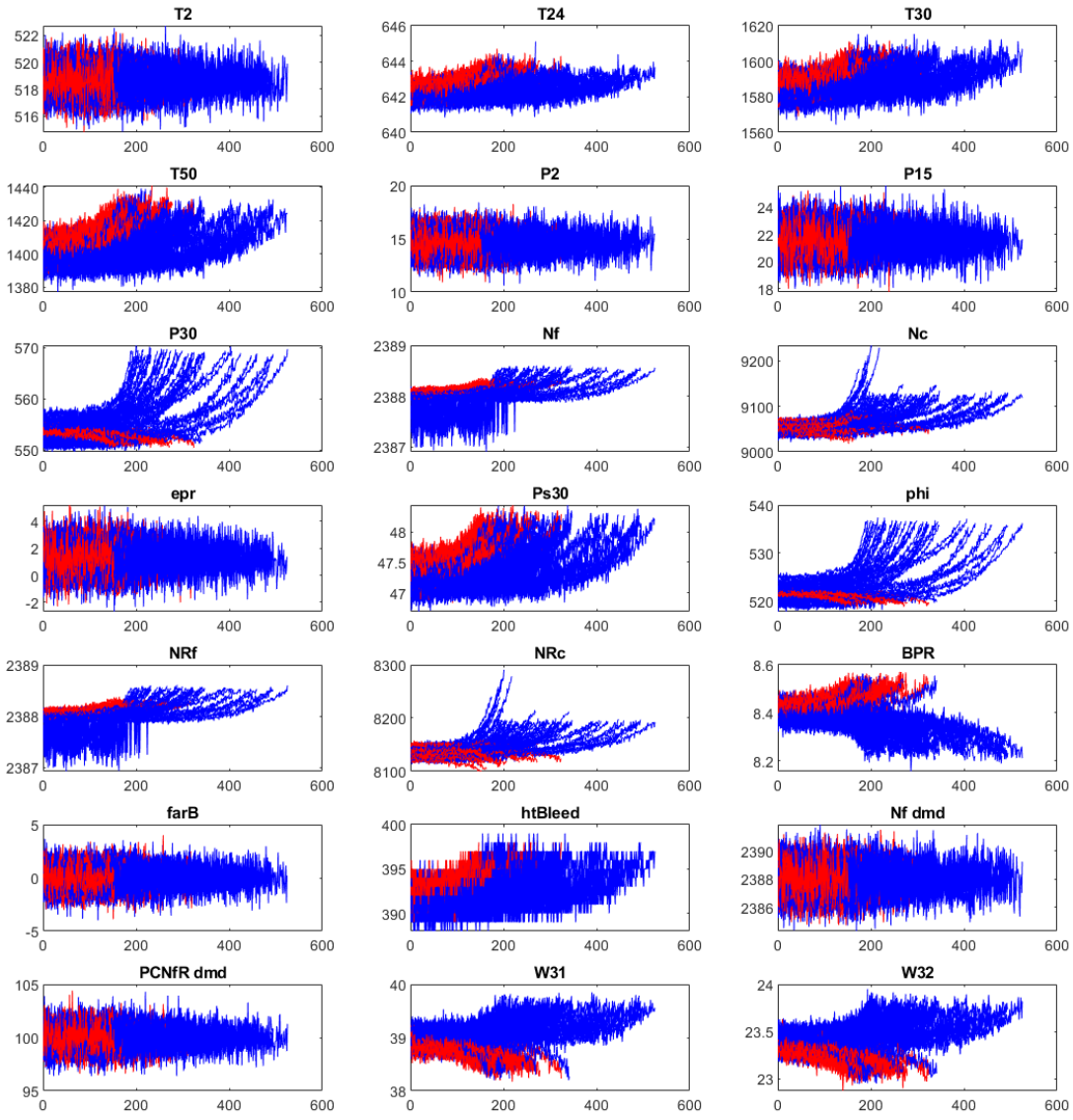


Figure 5.3 – Hierarchical Clustering of Sensor Signals

To extract features from these signals, a quadratic regression model is fitted to each individual signal as described in Section 5.2.2. The relationship between the $\ln TTF$ and the quadratic regression coefficients is shown in Figure 5.4.

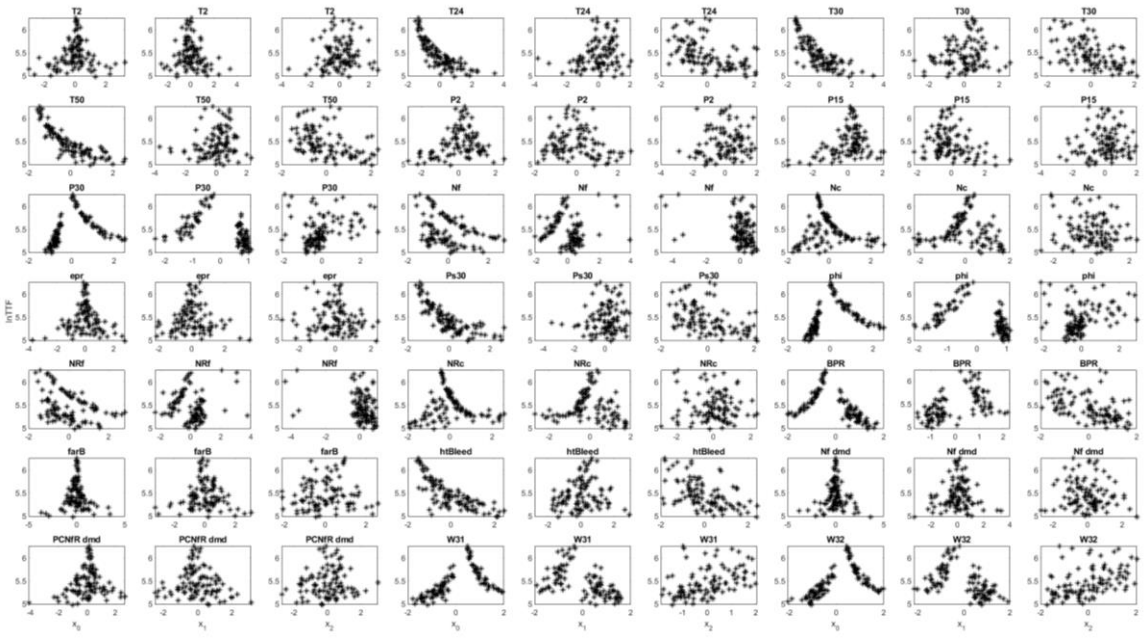


Figure 5.4 – ln TTF vs. Quadratic Regression Coefficients

After obtaining initial cluster assignments and extracting features, the ln TTF is regressed on the quadratic regression coefficients using the MGR-ASGL model described in Section 5.2.3. To fit the model, the E-step is initialized using the cluster assignments from hierarchical clustering. Based on these assignments, the model parameters are estimated using the M-step equations. The algorithm requires an initial estimate of the regression coefficients which can be a matrix of random numbers whose number of rows is equal to the number of predictors in the model plus one and whose number of columns is equal to the number of failure modes. For this case study, the matrix is 64×2 since three features are extracted per sensor and there are 21 sensors. There is also the bias term for each cluster. Successive iterations of the EM algorithm are performed until the decrease in the expected complete data log-likelihood is less than 1. To select the tuning parameters α and λ, 3-fold cross-validation is employed, where the combination of tuning parameters that minimizes the average cross-validated mean squared error between the actual failure

times and the predicted failure times obtained using the MGR-ASGL model is used. The $\ln TTF$ is predicted using the expectation of the marginal distribution of Y:

$$y^* = \sum_{k=1}^{K} \pi_k^{new} \left(\frac{\left(\psi_{0,k}^{new} + \sum_{p=1}^{P} \boldsymbol{x}_{i,p}^T \boldsymbol{\psi}_{p,k}^{new} \right)}{\rho_k^{new}} \right) \tag{5.4.1}$$

Then, the TTF is computed by evaluating exp (y^*).

The tuning parameters tested were $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ and $\lambda \in \{1, 4.22, 7.44, 10.67, 13.89, 17.11, 20.33, 23.56, 26.78, 30\}$. The results of cross-validation are shown in Figure 5.5.

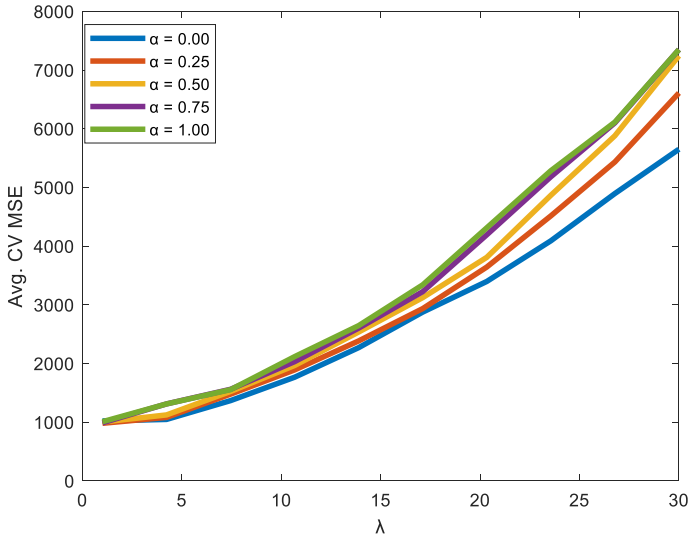


Figure 5.5 – Cross-validation Error

The minimum average MSE occurs when $\lambda = 1$ and $\alpha = 0.25$. However, the small penalty term results in a dense model. Therefore, the One Standard Error Rule, which selects the most parsimonious model with an average MSE at most one standard error greater than the minimum average MSE, is utilized. For $\alpha = 0.25$, this is achieved at $\lambda = 7.44$. To compare

the sensor selection using both the minimum average MSE and One Standard Error Rule, the l2-norms of the coefficients for each of the 21 groups are displayed in Table 5.2. The pink highlights indicate sensors with an l2-norm greater than 0.001. All other sensors are noninformative.

Table 5.2 – l2-norm of Sensor Group Coefficients

Sensor, p	FM 1		FM 2	
	min AvgMSE	1SE	min AvgMSE	1SE
T2	0.347	0.000	0.087	0.000
T24	0.411	0.000	0.339	0.482
T30	1.061	0.720	0.866	0.182
T50	1.166	0.248	10.743	1.522
P2	0.810	0.000	0.576	0.000
P15	1.060	0.000	0.250	0.000
P30	0.493	0.000	1.239	0.000
Nf	0.000	0.213	2.400	0.000
Nc	1.135	0.068	2.058	0.000
Epr	0.463	0.000	0.345	0.000
Ps30	1.668	1.219	1.926	0.723
Phi	0.000	0.000	0.000	0.000
NRf	5.646	0.381	4.449	0.000
NRc	0.000	0.000	0.000	0.000
BPR	3.227	0.497	1.422	0.023
farB	0.014	0.000	0.285	0.056
htBleed	0.979	0.246	1.929	0.440
Nf_dmd	0.526	0.000	1.169	0.000
PCNfR_dmd	0.457	0.000	0.494	0.000
W31	0.350	0.000	1.068	0.000
W32	0.046	0.000	0.383	0.000

From Table 5.2, it can be seen that only phi (ratio of fuel flow to Ps30) and NRc (corrected fan speed) are completely removed when using the minimum average MSE model. However, using the One Standard Error Rule results in a much more parsimonious model with sensors T30, T50, Nf, Nc, Ps30, NRf, BPR and htBleed being selected for cluster 1 and sensors T24, T30, T50, Ps30, BPR, farB, and htBleed being selected for cluster 2. Note

that there is strong overlap in the sensor selection indicating that some sensors are capable of monitoring multiple degradation processes. Furthermore, only one of the noninformative sensors (htBleed: bleed enthalpy) was selected as significant using the One Standard Error Rule. The clustering of the relationship between the $\ln TTF$ and the sensor features is shown in Figure 5.6.

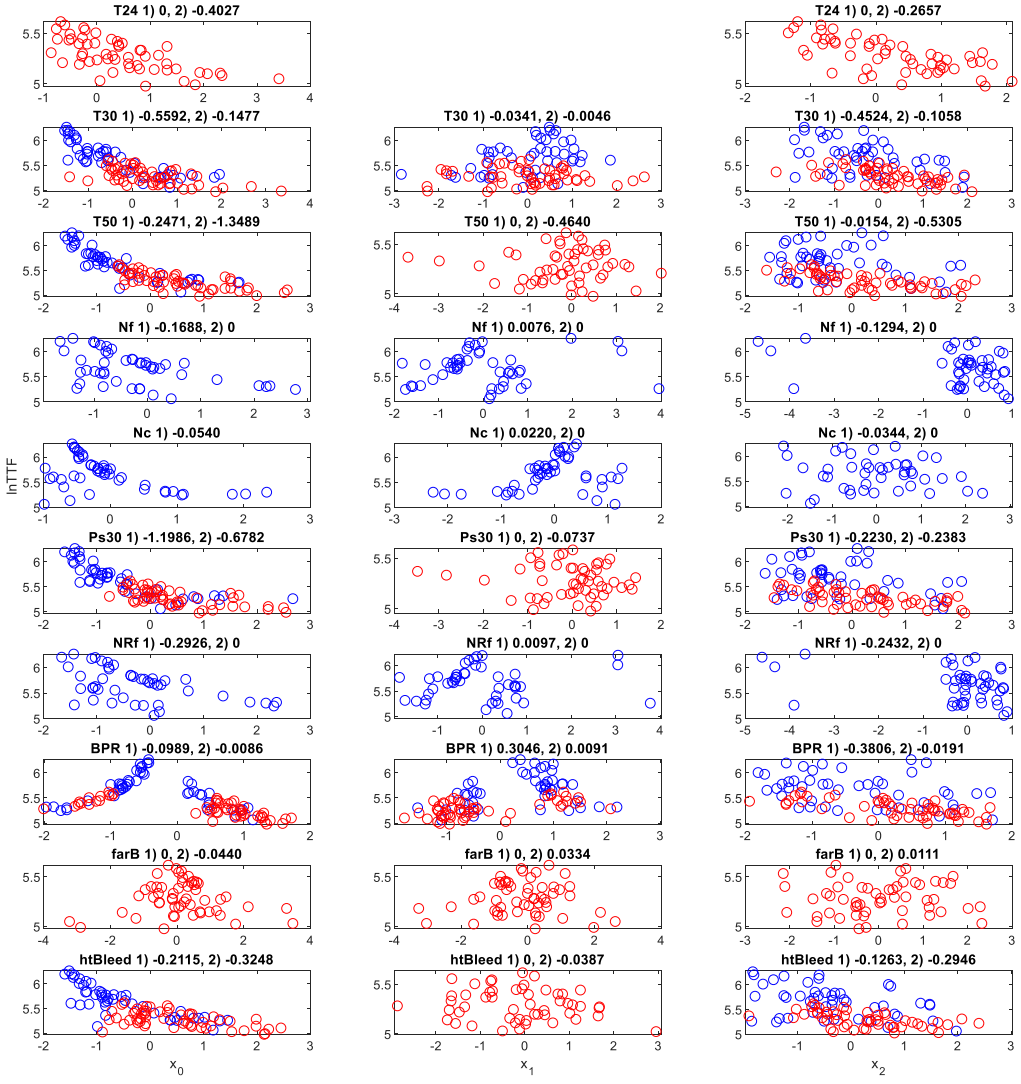


Figure 5.6 – Clustering via MGR-ASGL

In Figure 5.6, blue points are those assigned to cluster 1, while the red points are those assigned to cluster 2. Each graph displays the relationship between the $\ln TTF$ and the quadratic regression coefficients from Section 5.2.2 (labeled for the htBleed row). There are large regression coefficients in amplitude for Ps30, the static pressure at the HPC outlet. Furthermore, a high degree of curvature in the overall trend between the $\ln TTF$ and the quadratic regression coefficient x_0 is noticed. Based on the clustering, the algorithm converged to a solution that captured this curvature as opposed to a solution that separated the data by failure mode. However, clustering and sensor selection can also be performed as a two-step approach. It is clear by visual inspection that the failure modes can be identified by bisecting the $\ln TTF$ vs. x_0 relationship at $x_0 = 0$ for the BPR (Blade Pressure Ratio). These approaches are compared when predicting RUL using the methodology outlined in Section 5.3 to alternative approaches that do not address the attributes of complex systems.

The proposed methodology performs clustering and sensor selection using the MGR-ASGL model. Following the clustering, MFPCA is performed to fuse the selected sensors into informative predictors. However, the data is first smoothed using Robust Locally Weighted Regression as described in Algorithm 1. The choice of kernel utilized is the Bi-square kernel defined as follows:

$$\mathcal{K}\left(u_{i,p}^{v}\right) = \begin{cases} \left(1 - {u_{i,p}^{v}}^{2}\right)^{2}, \text{for } \left|u_{i,p}^{v}\right| < 1 \\ 0, \text{for } \left|u_{i,p}^{v}\right| \geq 1 \end{cases} \tag{5.4.2}$$

Consider a window with a width where $u_{i,p}^{v} \coloneqq \frac{v-t}{\omega_{i,p}}$, with $\omega_{i,p}$ equaling the absolute distance between t and its hth nearest neighbor. For this application, the number of points in the window is equal to 25% of the total length of the signal. If the signal is 100 time epochs long, then $\omega_{i,p}$ is the distance between t and the 25^{th} closest time epoch. After smoothing, the classification and regression models are implemented as described in Section 5.3.

To compare the results of the proposed methodology, first consider the effect of sensor selection and incorporating clustering. Two alternative models are proposed. The first does not implement the MGR-ASGL model. Instead, the initial clustering is used to label the training set. All 21 sensors are used in the methodology described in Section 5.3 to predict RUL. The second model treats all observations like they are in the same cluster. Multiple linear regression with ASGL penalty is utilized to perform sensor selection and the methodology described in Section 5.3 is used for predicting RUL. The average relative errors for these two models along with the proposed methodology are shown in Figure 5.7. The relative error is calculated as follows:

$$E_r = \frac{|RUL_{Actual} - RUL_{Predict}|}{TTF} \tag{5.4.3}$$

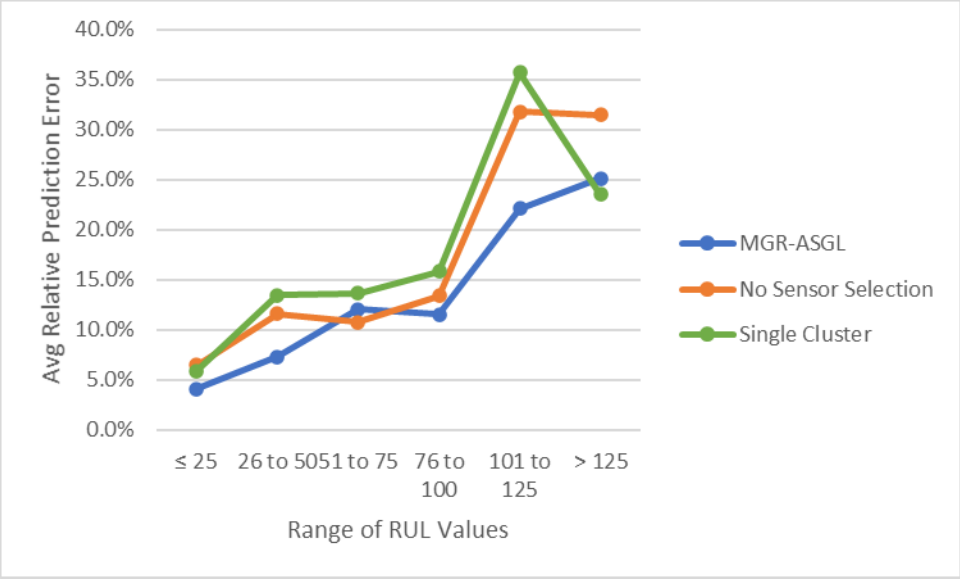


Figure 5.7 – Avg. Relative Error for Prognostics Models

In Figure 5.7, the abscissa displays 5 ranges of RUL. For the category $RUL \leq 25$, the average relative prediction error is computed for all testing units with RUL less than 25. The plot shows that outside of the units farthest from failure, ignoring multiple failure modes results in larger errors than if multiple failure modes are accounted for. Furthermore, the proposed sensor selection method results in lower average errors than not using sensor selection for all but the $50 < RUL \leq 75$ category. This indicates that improved prediction error can be obtained through a simultaneous clustering and sensor selection methodology.

The proposed methodology is also compared to the case where the cause of failures in the training set are known a priori, or they can be trivially diagnosed. For this, consider two alternative models. The first is a supervised approach of the MGR-ASGL model. In this case, we the CDLL can be formulated and optimized with respect to the model parameters without performing an E-step. Diagnostics and prognostics are performed using the methodology outlined in Section 5.3. The other alternative is a signal fusion

methodology proposed by Chehade et al. (2018). In their methodology, the authors used trends from HPC failures in the C-MAPSS 1 dataset to distinguish between HPC and Fan failures. To select sensors, they remove any sensor from the C-MAPSS 3 dataset that does not have a discernable trend. This leaves 14 sensors, whose signals they use to fuse into a health index that can be used to diagnose failure modes and predict RUL. The average relative errors for these methodologies are shown in Figure 5.8 below.

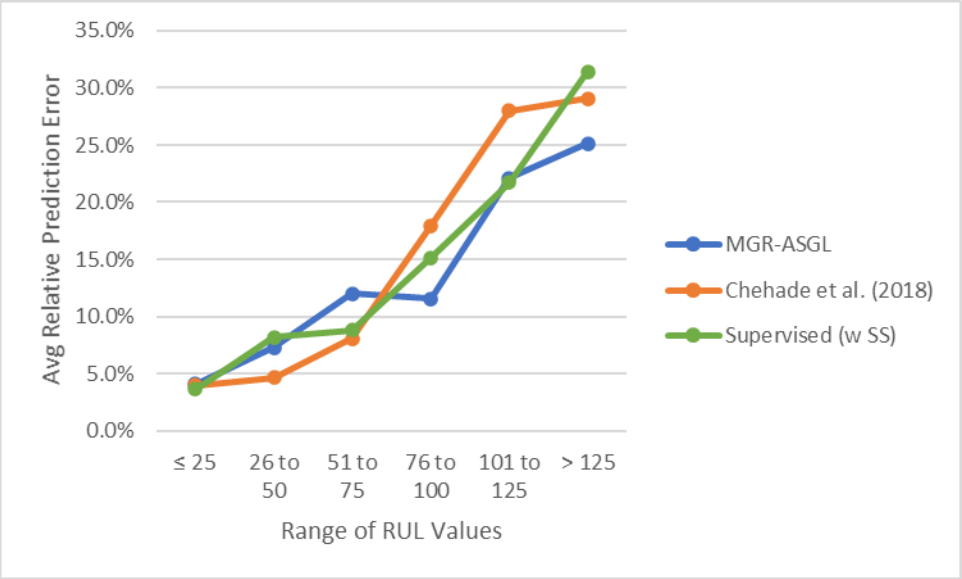


Figure 5.8 – Comparison of Avg. Relative Error with Failure Modes Known

The MGR-ASGL (blue) and supervised (green) methodologies both perform comparatively to each other. The root mean squared error of the test set for the MGR-ASGL is 34.49 whereas it is 33.60 for the supervised model, a 2.65% difference. However, this difference is primarily due to a large outlier in the MGR-ASGL model for the region where RUL is between 51-75. If the outlier is removed, the MGR-ASGL model has a root mean squared error 0.43% less than that of the supervised approach (MGR-ASGL: 33.45, Supervised: 34.60). The advantage of the MGR-ASGL is that it achieves this improvement

without a priori knowledge of the historical failure modes. However, classifying observations in the test set consists of assigning degradation signals to a cluster that may not map one-to-one with the actual cause of failure. Thus, diagnostic interpretation may be limited. The supervised approach alleviates this concern by defining the clusters based on the actual failure modes. Of course, this requires knowledge of the failure modes, which may not be trivially inferred in general. When comparing with Chehade et al. (2018), first note that both the MGR-ASGL and supervised approaches have lower root mean squared error for the test set (MGR-ASGL: 34.49, Supervised: 33.60, Chehade et al. (2018): 35). In comparing the mean relative errors for each range of RUL, it can be seen that the proposed methodologies perform better in the early stages of the system's life, but worse as failure becomes more imminent. However, upon looking at the median values in Figure 5.9, a different trend is noticed.

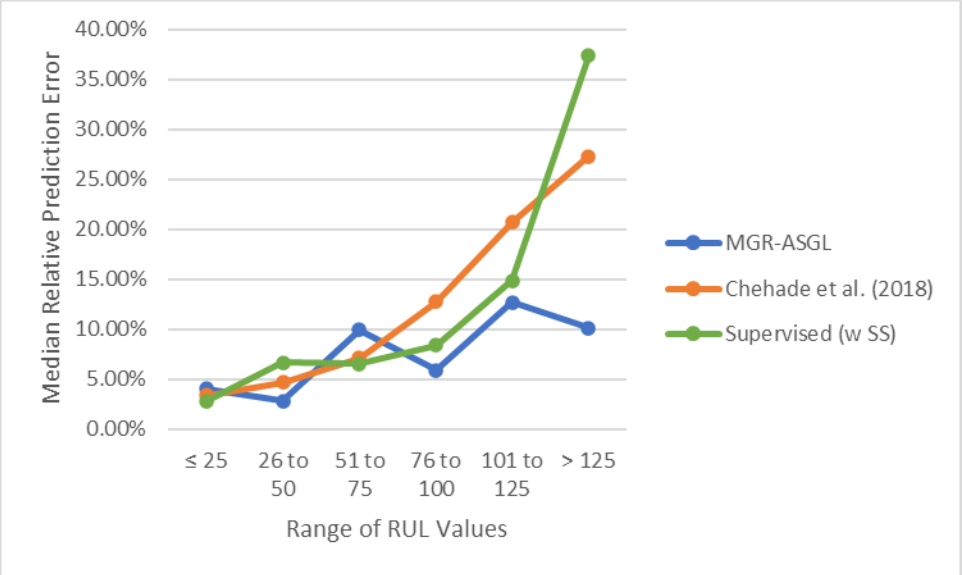


Figure 5.9 – Comparison of Median Relative Error with Failure Modes Known

This difference is due to the presence of outliers in the MGR-ASGL model that inflate the average. Suppose the smallest and largest errors for each RUL range of all three methodologies are removed. The effect on the average relative errors is shown in Figure 5.10.

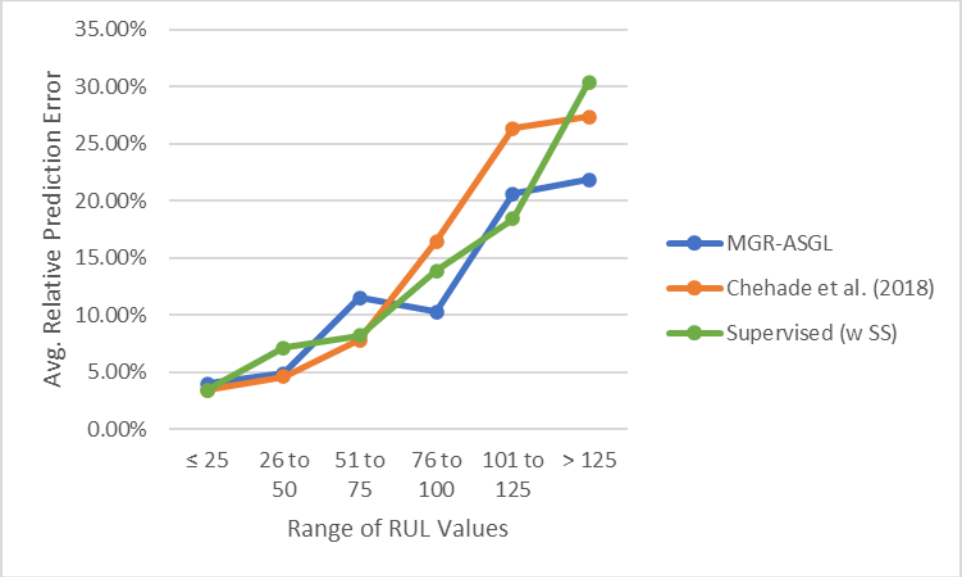


Figure 5.10 – Comparison of Avg. Relative Error with Extreme Values Removed

Figure 5.10 demonstrates that the average relative error for the MGR-ASGL methodology is on par with the methodology by Chehade et al. (2018) when the true RUL is less than or equal to 50. When the true RUL is between 51 and 75, the MGR-ASGL is prone to larger errors than the other two models. However, for ranges where the true RUL is greater than 76, the MGR-ASGL methodology has an average relative error no less than 5.49 percentage points lower than Chehade et al. (2018). For the supervised version, the errors when the true RUL is between 51 and 75 are 3.31 percentage points less than the MGR-ASGL methodology and only 0.40 percentage points more than the methodology by Chehade et al. (2018). However, it is the least accurate model when the true RUL is

between 26 and 50. Once the true RUL is less than 26, the average relative error for the supervised model is only 0.01 percentage points less than the methodology by Chehade et al. (2018). Like with the MGR-ASGL, the supervised approach has higher accuracy for true RUL greater than 75. However, it is less accurate than the other two methodologies when the true RUL is greater than 125. While both the MGR-ASGL and the supervised approach can provide improvements in accuracy in some true RUL intervals, there are other intervals where the performance is not as strong. However, the advantage of these methodologies is that they reduce the sensor load while still providing comparable accuracy within true RUL intervals and better accuracy overall.

5.5 Conclusion

The problem of predicting RUL for systems with multiple failure modes is important, but also challenging. Instrumenting a system with several sensors provides the benefit of highly informative data. However, processing this data is also challenging. In this paper, a prognostics methodology that utilizes signals from multiple sensors to predict the RUL of systems with multiple failure modes is proposed. The methodology consisted of modeling the $\ln TTF$ using a mixture of Gaussian regressions model with adaptive sparse group lasso penalty. To fit the MGR-ASGL model, a novel EM algorithm that simultaneously clusters the sensor data and selects optimal subsets of sensors for each cluster is utilized. However, the clustering algorithm did not separate the dataset into distinct failure modes. Instead, the algorithm seemed to converge to a model that captures nonlinear relationships between the $\ln TTF$ and the signal features. This could be due to the presence of homogenous, but nonlinear trends being present in some sensor features.

To predict RUL, the smoothed signals from the selected sensors are fused into informative predictors using MFPCA. The proposed modeling approach is more accurate than approaches that either do not consider multiple failure modes or do not perform sensor selection. The proposed methodology is also compared to the case where the failure modes in the training set are known a priori. For this, the proposed methodology is implemented in a supervised fashion where cluster assignments are determined by visual inspection of one of the Blade Pressure Ratio signal features. Both these methodologies had comparable accuracy with a signal fusion methodology published in the literature. The advantage of the proposed methodology is that similar accuracy can be achieved with a smaller number of sensors which could reduce the cost of collecting condition monitoring data. For future work, the approach to performing clustering and sensor selection simultaneously can be addressed. In the proposed approach, an optimization problem that seeks to minimize the error between the mixture of regressions model and the observed data is developed. This optimization problem may need to be amended to account for the goal of simultaneous clustering and sensor selection. However, the proposed methodology still performed comparably to the supervised approach in terms of prediction accuracy.

CHAPTER 6. CONCLUSION

This dissertation presents a framework for monitoring the SoH of complex industrial systems. The fundamental idea behind this framework is that for high-dimensional data, information salient to condition monitoring is often embedded in a low-dimensional space. Therefore, the framework combined statistical transforms with regularization techniques to isolate a sparse selection of features to perform condition monitoring tasks such as fault detection, diagnosis, and prognostics.

A key motivation for the topics discussed in this dissertation is to unite technical communities, such as gas turbine experts, that have a firm understanding of the physical dynamics of industrial systems with statisticians that develop generalized theoretical models for these systems. A benefit of this union is that the efficacy of statistics models for condition monitoring of real-world systems can be assessed. Furthermore, the modern challenges associated with these systems can motivate the development of new condition monitoring methodologies. For technical communities, the statistical models developed can provide physically interpretable models that can enable learning how degradation affects system dynamics, determine optimal sensor placement, and perform diagnostics and prognostics accurately.

For the remainder of this chapter, the contributions related to each topic are briefly summarized and the potential for future work is discussed.

6.1 Severity-based Diagnosis for Vehicular Electric Systems with Multiple, Interacting Fault Modes

The primary contribution of this chapter is a methodology that diagnoses both the presence and the severity of multiple, interacting fault modes in applications involving high-dimensional data. A vehicle-engine start test-rig is leveraged to collect time-resolved current and voltage profiles during the engine-cranking process. To process these functional data profiles, a data-analytics framework that consists of aligning the features of the profiles through curve registration, performing statistical transforms on the data for feature extraction, and then developing classifiers that utilize regularization to select a subset of these features for diagnosis was developed. Furthermore, this chapter presents a methodology for mapping the inferred fault states of the individual component severity levels to an overall system degradation level. Potential for future work would involve including more than two fault modes.

6.2 Condition Monitoring of Combustor

The transition from nonpremixed combustion to lean premixed combustion has rendered the gas turbine combustor susceptible to operational faults such as lean blowout and combustor instability that results in degradation to combustor components. This chapter presented a data curation algorithm that transforms a nonstationary signal to a white noise process. The characteristics of this white noise process are monitored using an EWRMS control chart. The alarms emitted from this control chart are imbued with a measure of risk of a blowout event. This measure of risk can be used by engine operators to inform decisions regarding how lean to operate. Future work is focused on the

development of a controller that automatically adjusts the fuel-to-air ratio based on this risk measure.

This chapter also presented a hierarchical feature selection algorithm that can be used to select a subset of optimal sensors for monitoring degradation of the combustor centerbody. The results indicated that this methodology had higher classification accuracy than a baseline methodology that utilized no variable selection. The hierarchical variable selection approach was also found to be more robust to reduction in the sample size of the training set. Furthermore, the sensors mounted on the premixer tubes were found to be more effective at classifying the degradation of the centerbody than the transverse sensors mounted on the tubes attached to the lab wall. Future work would include attempting to diagnose more granular levels of degradation than those used in the study. Furthermore, an alternative to hierarchical variable selection is to combine Adaptive Sparse Group Lasso (used in Chapter 5) and Multi-class Logistic Regression to select optimal sensors and optimal features for the selected sensors for classification.

6.3 Condition Monitoring of Turbine

This chapter presents data analytics methodologies for addressing challenges related to cooling mechanisms of the gas turbine. These mechanisms include the combination of rim seal geometry and cooling flows to create a seal to protect the wheel space from ingestion from the main gas path and coolant holes manufactured into the turbine blade. To predict sealing effectiveness, two methodologies using time-resolved pressure signals recorded either near the rim seal or on the outer casing were proposed. These methodologies included a data-driven approach and a two-step approach that leveraged

expert knowledge. The latter methodology was shown to be more accurate at predicting sealing effectiveness. Future work includes migrating this modelling effort from the laboratory to a real-world gas turbine.

Coolant holes are manufactured into the blade to prevent the blade temperatures from exceeding material melting points. This chapter proposed the novel idea of monitoring the coolant system through infrared imaging of the blade. The efficacy of this monitoring strategy was demonstrated for various feature extraction techniques. A key finding is that sparse regions corresponding to cold and hot regions in the turbine blade were found to be highly correlated with Coolant Flow Rate. This indicates there is potential for installing a small number of more cost-effective thermocouples at these locations for monitoring. A future step is to test this hypothesis by comparing the accuracy of sensing in sparse locations with the accuracy of using the infrared imaging.

6.4 Predicting Remaining Useful Life in Complex Systems with Multiple Failure Modes

This chapter presents a data analytics methodology for predicting remaining useful life in complex systems with multiple failure modes. The primary contribution of this methodology is that it identifies a subset of sensors associated with each failure mode, detects which failure mode is active, and predicts and updates the remaining useful life in real time. This is achieved by first modelling the conditional distribution of the natural logarithm of the time-to-failure given the sensor features as a mixture of Gaussian regression. Then, the log-likelihood is formulated and appended with the adaptive sparse

group lasso penalty to perform sensor selection. Finally, an EM algorithm is proposed to estimate the parameters of the model.

When applying the methodology to a simulated dataset, the clustering did not occur in an intuitive way. One hypothesis is that the presence of sensors whose relationship between the time-to-failure and sensor features appear to be homogenous may be interfering with the clustering. Therefore, future work is to address this issue. One potential solution is to partition the sensors into those that are useful for clustering and those that are not. Then, a joint likelihood between the time-to-failure and the set of useful for clustering sensors can be fitted. While the proposed methodology did not properly cluster the data, it did result in higher prediction accuracy than when treating the data as homogeneous. Furthermore, the proposed methodology had comparable prediction accuracy as a two-step approach that first clusters the data and then performs sensor selection.

REFERENCES

Admin. (2011). *Siemens H-class Gas Turbine Exceeds Targeted Efficiency: Power Industry News*. Power Industry News | Power industry News, Articles, Stories, Case Studies, Information about Power Industry. https://powerservices.lakho.com/2011/06/02/siemens-h-class-gas-turbine-exceeds-targeted-efficiency/.

Angello, L., Lieuwen, T., Noble, D., and Poole, B. (2017). System and method for anomaly detection. US Patent No. 9,752,960 B2. Washington, DC: U.S. Patent and Trademark Office.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Blancke, O., Tahan, A., Komljenovic, D., Amyot, N., Lévesque, M., and Hudon, C. (2018). A Holistic Multi-failure Mode Prognosis Approach for Complex Equipment. *Reliability Engineering & System Safety*, *180*, 136–151. https://doi.org/10.1016/j.ress.2018.07.006.

Bocchetti, D., Giorgio, M., Guida, M., and Pulcini, G. (2009). A Competing Risk Model for the Reliability of Cylinder Liners in Marine Diesel Engines. *Reliability Engineering & System Safety*, *94*(8), 1299–1307. https://doi.org/10.1016/j.ress.2009.01.010.

Bogard, D. G., and Thole, K. A. (2006). Gas Turbine Film Cooling. *Journal of Propulsion and Power*, *22*(2), 249–270. https://doi.org/10.2514/1.18034.

Bohn, D., Johann, E., and Krüger, U. (1995). Experimental and Numerical Investigations of Aerodynamic Aspects of Hot Gas Ingestion in Rotor-stator Systems with Superimposed Cooling Mass Flow. *Volume 1: Turbomachinery*. https://doi.org/10.1115/95-gt-143.

Bunker, R. S. (2017). Evolution of Turbine Cooling. *Volume 1: Aircraft Engine; Fans and Blowers; Marine; Honors and Awards*. https://doi.org/10.1115/gt2017-63205.

Chehade, A., Song, C., Liu, K., Saxena, A., and Zhang, X. (2018). A Data-level Fusion Approach for Degradation Modeling and Prognostic Analysis Under Multiple Failure Modes. *Journal of Quality Technology*, *50*(2), 150–165. https://doi.org/10.1080/00224065.2018.1436829.

Chen, L., Zhuang, Y., Zhang, J., and Wang, J. (2017). An End-to-end Approach For Bearing Fault Diagnosis Based On A Deep Convolution Neural Network. *Neural Information Processing Lecture Notes in Computer Science*, 101-109. doi:10.1007/978-3-319-70096-0_11.

Chen, N., Ye, Z.-S., Xiang, Y., and Zhang, L. (2015). Condition-based Maintenance Using the Inverse Gaussian Degradation Model. *European Journal of Operational Research*, *243*(1), 190–199. https://doi.org/10.1016/j.ejor.2014.11.029.

Clark, K., Barringer, M., Johnson, D., Thole, K., Grover, E., and Robak, C. (2018). Effects of Purge Flow Configuration on Sealing Effectiveness in a Rotor–stator Cavity. *Journal of Engineering for Gas Turbines and Power*, *140*(11). https://doi.org/10.1115/1.4040308.

Clark, K., Barringer, M., Thole, K., Clum, C., Hiester, P., Memory, C., and Robak, C. (2016). Using a Tracer Gas to Quantify Sealing Effectiveness for Engine Realistic Rim Seals. *Volume 5A: Heat Transfer*. https://doi.org/10.1115/gt2016-58095.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038.

Couallier, V. (2008). A Competing Risks Model for Degradation and Traumatic Failure Times. *Statistical Models and Methods for Biomedical and Technical Systems*, 83–93. https://doi.org/10.1007/978-0-8176-4619-6_7.

Cox, J. R., Anusonti-Inthra, P., and Arnold, S. (2015). Using Gas Turbine Engine Casing Accelerometer Measurements for Rotor Blade Health Monitoring. *51st AIAA/SAE/ASEE Joint Propulsion Conference*. https://doi.org/10.2514/6.2015-4194.

Daubechies, I. (1992). Ten Lectures on Wavelets. https://doi.org/10.1137/1.9781611970104.

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. *Multiple Classifier Systems Lecture Notes in Computer Science,* 1-15. doi:10.1007/3-540-45014-9_1.

Ebeling, C. E. (2010). *An Introduction to Reliability and Maintainability Engineering* (2nd ed.). Waveland Press.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, *32*(2). https://doi.org/10.1214/009053604000000067.

Emerson, B., O'Connor, J., Juniper, M. and Lieuwen, T. (2012). Density Ratio Effects on Reacting Bluff-body Flow Field Characteristics. *Journal of Fluid Mechanics*, 706, pp.219–250.

Emerson, B., Perullo, C., Lieuwen, T., Sheppard, S., Kee, J., Noble, D., and Angello, L. (2018). Combustion Dynamics Monitoring Considerations for Systems with Autotuning. *Volume 4B: Combustion, Fuels, and Emissions*, 1–9. https://doi.org/10.1115/gt2018-77072.

Erickson, R. R. and Soteriou, M.C. (2011). The Influence of Reactant Temperature on the Dynamics of Bluff Body Stabilized Premixed Flames. *Combustion and Flame*, 158(12), pp.2441–2457.

Fang, X., Paynabar, K., and Gebraeel, N. (2017). Multistream Sensor Fusion-based Prognostics Model for Systems with Single Failure Modes. *Reliability Engineering and System Safety, 159*, 322-331. doi:10.1016/j.ress.2016.11.008.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1). https://doi.org/10.18637/jss.v033.i01.

Gao, R. X., and Yan, R. (2011). *Wavelets: Theory and Applications for Manufacturing*. New York: Springer. doi:10.1007/978-1-4419-1545-0.

García Márquez, F. P., Tobias, A. M., Pinar Pérez, J. M., and Papaelias, M. (2012). Condition Monitoring of Wind Turbines: Techniques and Methods. *Renewable Energy*, *46*, 169–178. https://doi.org/10.1016/j.renene.2012.03.003.

Gebraeel, N. Z., Lawley, M. A., Li, R., and Ryan, J. K. (2005). Residual-life Distributions from Component Degradation Signals: A Bayesian Approach. *IIE Transactions*, *37*(6), 543–557. https://doi.org/10.1080/07408170590929018.

Ghimire, R., Sankavaram, C., Ghahari, A., Pattipati, K., Ghoneim, Y., Howell, M., and Salman, M. (2011). Integrated Model-based and Data-driven Fault Detection and Diagnosis Approach for an Automotive Electric Power Steering System. *2011 IEEE Autotestcon*. doi:10.1109/autest.2011.6058760.

Goel, N., Kumar, A., Narasimhan, V., Nayak, A., and Srivastava, A. (2008). Health Risk Assessment and Prognosis of Gas Turbine Blades by Simulation and Statistical Methods. *2008 Canadian Conference on Electrical and Computer Engineering*. https://doi.org/10.1109/ccece.2008.4564705.

Goy, C. J., James, S. R., Rea, S., Lieuwen, T., and Yang, V. (2005). Monitoring Combustion Instabilities: E.ON UK's Experience. In *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms and Modeling* (pp. 163–175). Book chapter, American Institute of Aeronautics and Astronautics.

Hajiha, M., Liu, X., and Hong, Y. (2020). Degradation under Dynamic Operating Conditions: Modeling, Competing Processes and Applications. *Journal of Quality Technology*, 1–22. https://doi.org/10.1080/00224065.2020.1757390.

Halderman, J. (1988). *Automotive Electrical and Electronic Systems*. Englewood Cliffs, N.J.: Prentice-Hall.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC.

Hoang, D., and Kang, H. (2017). Convolutional Neural Network Based Bearing Fault Diagnosis. *Intelligent Computing Theories and Application Lecture Notes in Computer Science*, 105-111. doi:10.1007/978-3-319-63312-1_9.

Hyndman, R.J. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 27(3).

Islam, M. M., and Kim, J. (2019). Reliable Multiple Combined Fault Diagnosis of Bearings using Heterogeneous Feature Models and Multiclass Support Vector Machines. *Reliability Engineering and System Safety,184*, 55-66. doi:10.1016/j.ress.2018.02.012.

Jaramillo, V. H., Ottewill, J. R., Dudek, R., Lepiarczyk, D., and Pawlik, P. (2017). Condition Monitoring of Distributed Systems Using Two-stage Bayesian Inference Data Fusion. *Mechanical Systems and Signal Processing, 87*, 91-110. doi:10.1016/j.ymssp.2016.10.004.

Jardine, A. K., Lin, D., and Banjevic, D. (2006). A Review on Machinery Diagnostics and Prognostics Implementing Condition-based Maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483-1510. doi:10.1016/j.ymssp.2005.09.012.

Johnson, B., Mack, G., Paolillo, R., and Daniels, W. (1994). Turbine Rim Seal Gas Path Flow Ingestion Mechanisms. *30th Joint Propulsion Conference and Exhibit*. https://doi.org/10.2514/6.1994-2703

Kazemi, Z., Safavi, A., Pouresmaeeli, S., and Naseri, F. (2019). A Practical Framework for Implementing Multivariate Monitoring Techniques into Distributed Control System. *Control Engineering Practice, 82*, 118-129. doi:10.1016/j.conengprac.2018.10.003.

Kim, M., and Liu, K. (2020). A Bayesian Deep Learning Framework for Interval Estimation of Remaining Useful Life in Complex Systems by Incorporating General Degradation Characteristics. *IISE Transactions*, *53*(3), 326–340. https://doi.org/10.1080/24725854.2020.1766729.

Kodali, A., Zhang, Y., Sankavaram, C., Pattipati, K., and Salman, M. (2013). Fault Diagnosis in the Automotive Electric Power Generation and Storage System (EPGS). *IEEE/ASME Transactions on Mechatronics, 18*(6), 1809-1818. doi:10.1109/tmech.2012.2214397.

Koff, B. L. (2004). Gas Turbine Technology Evolution: A Designer's Perspective. *Journal of Propulsion and Power*, *20*(4), 577–595. https://doi.org/10.2514/1.4361.

Kumar, R. M., Peters, B., Emerson, B., Paynabar, K., Gebraeel, N., and Lieuwen, T. (2020). Data Driven Fault Detection of Premixer Centerbody Degradation in a Swirl Combustor. *Volume 4A: Combustion, Fuels, and Emissions*. https://doi.org/10.1115/gt2020-14222.

LeMieux, D. H., and Jonnalagadda, V. (2006). On-line Thermal Barrier Coating (TBC) Monitor for Real-time Failure Protection and Life Maximization. *Proceedings of the 2006 International Conference on Quantitative InfraRed Thermography*. https://doi.org/10.21611/qirt.2006.043.

Li, W., Suozzo, C., Onori, S., Rizzoni, G., Salman, M. A., and Zhang, F. (2008). Experimental Calibration and Validation of Fault Diagnosis and Prognosis Algorithms for Automotive Electric Power Generation and Storage System. *ASME 2008 Dynamic Systems and Control Conference, Parts A and B,* 1317-1324. doi:10.1115/dscc2008-2238.

Li, X., Dong, S., and Yuan, Z. (1999). Discrete Wavelet Transform for Tool Breakage Monitoring. *International Journal of Machine Tools and Manufacture*, *39*(12), 1935–1944. https://doi.org/10.1016/s0890-6955(99)00021-8.

Li, X., Tso, S. K., and Wang, J. (2000). Real-time Tool Condition Monitoring Using Wavelet Transforms and Fuzzy Techniques. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, *30*(3), 352–357. https://doi.org/10.1109/5326.885116.

Lim, M. H., and Leong, M. S. (2010). Improved Blade Fault Diagnosis Using Discrete Blade Passing Energy Packet and Rotor Dynamics Wavelet Analysis. *Volume 3: Controls, Diagnostics and Instrumentation; Cycle Innovations; Marine*. https://doi.org/10.1115/gt2010-22218.

Liu, J., Hu, Y., Wang, Y., Wu, B., Fan, J., and Hu, Z. (2018). An Integrated Multi-sensor Fusion-based Deep Feature Learning Approach for Rotating Machinery Diagnosis. *Measurement Science and Technology*,*29*(5), 055103. doi:10.1088/1361-6501/aaaca6.

Liu, X., Li, J., Al-Khalifa, K. N., Hamouda, A. S., Coit, D. W., and Elsayed, E. A. (2013). Condition-based Maintenance for Continuously Monitored Degrading Systems with Multiple Failure Modes. *IIE Transactions*, *45*(4), 422–435. https://doi.org/10.1080/0740817x.2012.690930.

Macgregor, J.F. and Harris, T.J. (1993). The Exponentially Weighted Moving Variance. *Journal of Quality Technology*, 25(2), pp.106–118.

Madhavan, S., Jain, R., Sujatha, C., and Sekhar, A. S. (2014). Vibration Based Damage Detection of Rotor Blades in a Gas Turbine Engine. *Engineering Failure Analysis*, *46*, 26–39. https://doi.org/10.1016/j.engfailanal.2014.07.021.

Markham, J., Cosgrove, J., Scire, J., Haldeman, C., and Agoos, I. (2014). Aircraft Engine-mounted Camera System for Long Wavelength Infrared Imaging of In-service Thermal Barrier Coated Turbine Blades. *Review of Scientific Instruments*, *85*(12), 124902. https://doi.org/10.1063/1.4903266.

Meher-Homji, C. B., and Gabriles, G. (1998). Gas Turbine Blade Failures - Causes, Avoidance, and Troubleshooting. *Proceedings of the 27th Turbomachinery Symposium*, 129–180. https://doi.org/https://doi.org/10.21423/R1RD4R.

Mondal, S., De, S., Mukhopadhyay, A., Sen, S., and Ray, A. (2020). Early Prediction of Lean Blowout from Chemiluminescence Time Series Data. *Combustion Science and Technology*.

Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*. Hoboken Wiley Copyright © By John Wiley and Sons, Inc.

Nair, S. and Lieuwen, T. (2005). Acoustic Detection of Blowout in Premixed Flames. *Journal of Propulsion and Power*, 21(1), pp.32–39.

Nair, S. and Lieuwen, T. (2007). Near-blowoff Dynamics of a Bluff-body Stabilized Flame. *Journal of Propulsion and Power*, 23(2), pp.421–427.

Noble, D., Angello, L., Sheppard, S., Kee, J., Emerson, B., and Lieuwen, T. (2019). Investigation into Advanced Combustion System Health Monitoring. *Proceedings of the ASME Turbo Expo 2019: Turbomachinery Technical Conference and Exposition. Volume 3: Coal, Biomass, Hydrogen, and Alternative Fuels; Cycle Innovations; Electric Power; Industrial and Cogeneration; Organic Rankine Cycle Power Systems.*, 1–10. https://doi.org/10.1115/gt2019-91973.

Owen, J. M. (2010a). Prediction of Ingestion Through Turbine Rim Seals—Part I: rotationally-induced ingress. *Volume 3: Heat Transfer, Parts A and B*. https://doi.org/10.1115/gt2009-59121.

Owen, J. M. (2010b). Prediction of Ingestion Through Turbine Rim Seals—Part II: externally induced and combined ingress. *Journal of Turbomachinery*, *133*(3). https://doi.org/10.1115/1.4001178.

Owen, J. M. (2012). Theoretical Modelling of Hot Gas Ingestion Through Turbine Rim Seals. *Propulsion and Power Research*, *1*(1), 1–11. https://doi.org/10.1016/j.jppr.2012.10.002.

Owen, J. M., Wu, K., Scobie, J. A., Sangan, C. M., Cho, G. H., and Lock, G. D. (2014). Use of Pressure Measurements to Determine Effectiveness of Turbine Rim Seals. *Journal of Engineering for Gas Turbines and Power*, *137*(3). https://doi.org/10.1115/1.4028395.

Park, J., Ha, J. M., Oh, H., Youn, B. D., Choi, J., and Kim, N. H. (2016). Model-based Fault Diagnosis of a Planetary Gear: A Novel Approach Using Transmission Error. *IEEE Transactions on Reliability, 65*(4), 1830-1841. doi:10.1109/tr.2016.2590997.

Pencer, M. (2016). *Feature Selection and Post-selection Statistical Inference in Multinomial Models* (dissertation).

Peters, B., Yildirim, M., Gebraeel, N., and Paynabar, K. (2020). Severity-based Diagnosis for Vehicular Electric Systems with Multiple, Interacting Fault Modes. *Reliability Engineering & System Safety, 195*, 106605. https://doi.org/10.1016/j.ress.2019.106605.

Pham, H. T. and Yang, B. S. (2010). Estimation and Forecasting of Machine Health Condition using ARMA/GARCH Model. *Mechanical Systems and Signal Processing*, 24(2), pp.546–558.

Poushter, J. (2015). Car, Bike or Motorcycle? Depends on Where You Live. Retrieved from https://www.pewresearch.org/fact-tank/2015/04/16/car-bike-or-motorcycle-depends-on-where-you-live/.

Prakash, S., Neumeier, Y. and Zinn, B. (2006). Investigation of Mode Shift Dynamics of Lean, Premixed Flames. *44th AIAA Aerospace Sciences Meeting and Exhibit*.

Ragab, A., Yacout, S., Ouali, M.-S., and Osman, H. (2016). Prognostics of Multiple Failure Modes in Rotating Machinery Using a Pattern-based Classifier and Cumulative Incidence Functions. *Journal of Intelligent Manufacturing*, *30*(1), 255–274. https://doi.org/10.1007/s10845-016-1244-8.

Ramsay, J. O., and Silverman, B. W. (2005), Functional Data Analysis, 2nd ed., Springer, New York, Ch. 7.

Ratcliff, P., Garbett, P., and Fischer, W. (2007). The New Siemens Gas Turbine SGT5-8000H for more customer benefit. 87. 128-132.

Rock, N., Chterev, I., Emerson, B., Won, S. H., Seitzman, J., and Lieuwen, T. (2019). Liquid Fuel Property Effects on Lean Blowout in an Aircraft Relevant Combustor. *Journal of Engineering for Gas Turbines and Power*, *141*(7). https://doi.org/10.1115/1.4042010.

Rock, N., Emerson, B., Seitzman, J. and Lieuwen, T. (2020). Near-lean Blowoff Dynamics in a Liquid Fueled Combustor. *Combustion and Flame*, 212, pp.53–66.

Sangan, C. M., Pountney, O. J., Zhou, K., Owen, J. M., Wilson, M., and Lock, G. D. (2012a). Experimental Measurements of Ingestion Through Turbine Rim Seals—Part I: Externally Induced Ingress. *Journal of Turbomachinery*, *135*(2). https://doi.org/10.1115/1.4006609.

Sangan, C. M., Pountney, O. J., Zhou, K., Owen, J. M., Wilson, M., and Lock, G. D. (2012b). Experimental Measurements of Ingestion Through Turbine Rim Seals—Part II: Rotationally Induced Ingress. *Journal of Turbomachinery*, *135*(2). https://doi.org/10.1115/1.4006586.

Sankavaram, C., Pattipati, B., Pattipati, K. R., Zhang, Y., and Howell, M. (2014). Fault Diagnosis in Hybrid Electric Vehicle Regenerative Braking System. *IEEE Access, 2*, 1225-1239. doi:10.1109/access.2014.2362756.

Saxena, A., Celaya, J. R., Roychoudhury, I., Saha, S., Saha, B., and Goebel, K. (2012). Designing Data-driven Battery Prognostic Approaches for Variable Loading Profiles: Some Lessons Learned. *European Conference of Prognostics and Health Management Society*.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage Propagation Modeling for Aircraft Engine Run-to-failure Simulation. *2008 International Conference on Prognostics and Health Management*. https://doi.org/10.1109/phm.2008.4711414.

Scacchioli, A., Rizzoni, G., Salman, M. A., Li, W., Onori, S., and Zhang, X. (2014). Model-based Diagnosis of an Automotive Electric Power Generation and Storage System. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(1), 72-85. doi:10.1109/tsmcc.2012.2235951.

Scobie, J. A., Sangan, C. M., Michael Owen, J., and Lock, G. D. (2016). Review of Ingress in Gas Turbines. *Journal of Engineering for Gas Turbines and Power*, *138*(12). https://doi.org/10.1115/1.4033938.

Sewell, J. B., Sobieski, P. A., Lieuwen, T., and Yang, V. (2005). Monitoring of Combustion Instabilities: Calpine's Experience. In *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms and Modeling* (pp. 147–162). Book chapter, American Institute of Aeronautics and Astronautics.

Shanbhogue, S.J., Husain, S. and Lieuwen, T. (2009). Lean Blowoff of Bluff Body Stabilized Flames: Scaling and Dynamics. *Progress in Energy and Combustion Science*, 35(1), pp.98–120.

Shu, Y., Feng, Q., and Coit, D. W. (2015). Life Distribution Analysis Based on Lévy Subordinators for Degradation with Random Jumps. *Naval Research Logistics (NRL)*, *62*(6), 483–492. https://doi.org/10.1002/nav.21642.

Sikorska, J., Hodkiewicz, M., and Ma, L. (2011). Prognostic Modelling Options for Remaining Useful Life Estimation by Industry. *Mechanical Systems and Signal Processing*, *25*(5), 1803–1836. https://doi.org/10.1016/j.ymssp.2010.11.018.

Simon N, Friedman J, and Hastie T. (2012) A Blockwise Descent Algorithm for Group-penalized Multiresponse and Multinomial Regression. arXiv:1311. 6529.

Städler, N., Bühlmann, P., and van de Geer, S. (2010). ℓ1-Penalization for Mixture Regression Models. *TEST*, *19*(2), 209–256. https://doi.org/10.1007/s11749-010-0197-z.

Stringer, D. B., Sheth, P. N., and Allaire, P. E. (2012). Physics-based Modeling Strategies for Diagnostic and Prognostic Application in Aerospace Systems. *Journal of Intelligent Manufacturing, 23*(2), 155-162. doi:10.1007/s10845-009-0340-4.

Tahan, M., Tsoutsanis, E., Muhammad, M., and Abdul Karim, Z. A. (2017). Performance-based Health Monitoring, Diagnostics and Prognostics for Condition-based Maintenance of Gas Turbines: a Review. *Applied Energy*, *198*, 122–144. https://doi.org/10.1016/j.apenergy.2017.04.048.

Tamilselvan, P., and Wang, P. (2013). Failure Diagnosis Using Deep Belief Learning Based Health State Classification. *Reliability Engineering & System Safety, 115*, 124-135. doi:10.1016/j.ress.2013.02.022.

Thumati, B. T., Feinstein, M. A., and Jagannathan, S. (2014). A Model-based Fault Detection and Prognostics Scheme for Takagi–Sugeno Fuzzy Systems. *IEEE Transactions on Fuzzy Systems*, *22*(4), 736–748. https://doi.org/10.1109/tfuzz.2013.2272584.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., and Tripot, G. (2012). A Data-driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models. *IEEE Transactions on Reliability*, *61*(2), 491–503. https://doi.org/10.1109/tr.2012.2194177.

Unni, V. R., and Sujith, R. I. (2016). Precursors to Blowout in a Turbulent Combustor Based on Recurrence Quantification. In Proceedings of 52nd AIAA/SAE/ASEE Joint Propulsion Conference. https://arc.aiaa.org/doi/abs/10.2514/6.2016-4649.

Valeo United Kingdom and Republic of Ireland. (2014). Valeo presents micro-hybrid Stop Start i-StARS reversible alternator and ReStart reinforced starter [Video file]. Retrieved from https://www.youtube.com/watch?v=VJSEQaxAh4c.

Velayudham, A., Krishnamurthy, R., and Soundarapandian, T. (2005). Acoustic Emission Based Drill Condition Monitoring During Drilling of Glass/Phenolic Polymeric Composite using Wavelet Packet Transform. *Materials Science and Engineering: A, 412*(1-2), 141–145. https://doi.org/10.1016/j.msea.2005.08.036.

Wang, Y., and Pham, H. (2012). Modeling the Dependent Competing Risks with Multiple Degradation Processes and Random Shock Using Time-varying Copulas. *IEEE Transactions on Reliability, 61*(1), 13–22. https://doi.org/10.1109/tr.2011.2170253.

Watson, S. J., Xiang, B. J., Yang, W., Tavner, P. J., and Crabtree, C. J. (2010). Condition Monitoring of the Power Output of Wind Turbine Generators Using Wavelets. *IEEE Transactions on Energy Conversion*, *25*(3), 715–721. https://doi.org/10.1109/tec.2010.2040083.

Wu, C., Jiang, P., Ding, C., Feng, F., and Chen, T. (2019). Intelligent Fault Diagnosis of Rotating Machinery Based on One-dimensional Convolutional Neural Network. *Computers in Industry, 108*, 53-61. doi:10.1016/j.compind.2018.12.001.

Yan, Y., Luh, P. B., and Pattipati, K. R. (2014). Fault Diagnosis Framework for Air Handling Units Based on the Integration of Dependency Matrices and PCA. *2014 IEEE International Conference on Automation Science and Engineering (CASE)*. doi:10.1109/coase.2014.6899463.

Yang, W., Tavner, P. J., Crabtree, C. J., and Wilkinson, M. (2010). Cost-effective Condition Monitoring for Wind Turbines. *IEEE Transactions on Industrial Electronics*, *57*(1), 263–271. https://doi.org/10.1109/tie.2009.2032202.

Yi, T., and Gutmark, E. J. (2007). Real-time Prediction of Incipient Lean Blowout in Gas Turbine Combustors. Aiaa J. 45 (7):1734–39. doi:10.2514/1.25847.

Yu, M., and Wang, D. (2014). Model-based Health Monitoring for a Vehicle Steering System with Multiple Faults of Unknown Types. *IEEE Transactions on Industrial Electronics*, *61*(7), 3574–3586. https://doi.org/10.1109/tie.2013.2281159.

Zarei, J., and Poshtan, J. (2007). Bearing Fault Detection Using Wavelet Packet Transform of Induction Motor Stator Current. *Tribology International*, *40*(5), 763–769. https://doi.org/10.1016/j.triboint.2006.07.002.

Zhang, F., Liu, Y., Chen, C., Li, Y., and Huang, H. (2014). Fault Diagnosis of Rotating Machinery Based on Kernel Density Estimation and Kullback-Leibler Divergence. *Journal of Mechanical Science and Technology, 28*(11), 4441-4454. doi:10.1007/s12206-014-1012-7.

Zhang, Q., Hua, C., and Xu, G. (2014). A Mixture Weibull Proportional Hazard Model for Mechanical System Failure Prediction Utilizing Lifetime and Monitoring Data. *Mechanical Systems and Signal Processing*, *43*(1-2), 103–112. https://doi.org/10.1016/j.ymssp.2013.10.013.

Zhang, Q., Shanbhogue, S.J. and Lieuwen, T. (2010). Dynamics of Premixed H2/CH4 Flames Under Near Blowoff Conditions. *Journal of Engineering for Gas Turbines and Power*, 132(11).

Zhang, X., Uliyar, H., Farfan-Ramos, L., Zhang, Y., and Salman, M. (2010). Fault Diagnosis of Automotive Electric Power Generation and Storage Systems. *2010 IEEE International Conference on Control Applications*. doi:10.1109/cca.2010.5611179.

Zhang, Y., Rajagopalan, S., and Salman, M. (2010). A Practical Approach for Belt Slip Detection in Automotive Electric Power Generation and Storage System. *2010 IEEE Aerospace Conference*, 1-7. doi:10.1109/aero.2010.5446832.

Zhu, K., Wong, Y. S., and Hong, G. S. (2009). Wavelet Analysis of Sensor Signals for Tool Condition Monitoring: a Review and Some New Results. *International Journal of Machine Tools and Manufacture*, *49*(7-8), 537–553. https://doi.org/10.1016/j.ijmachtools.2009.02.003.

Zou, H. (2006). The Adaptive Lasso and its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735.