

Development of a Data-driven Model for Marine Gas Turbine (MGT) Engine Health Monitoring

Daniel Maraini¹, Mark Simpson², Ronald Brown³, and Michael Poporad⁴

^{1,2} *McKean Defense Group, Philadelphia, PA 19112, USA*
dmaraini@mckean-defense.com
msimpson@mckean-defense.com

^{3,4} *Naval Surface Warfare Center, Philadelphia, PA, 19112, USA*
ronald.r.brown@navy.mil
michael.poporad@navy.mil

ABSTRACT

Throughout the United States Navy, marine gas turbines (MGT) are used for the production of propulsive and electric power aboard many surface ships. The operational availability of these ships during deployments is contingent on the health and reliability of the installed MGTs. Currently, to ensure the health and proper functionality of these turbines, a series of manual evaluations are employed with varying success (e.g., pre-deployment visual inspections, periodic inspections of predefined wear out modes, characteristic vibration surveys, Integrated Performance Analysis Reports). This paper examines historical records associated with the General Electric LM2500 MGT installed for propulsion aboard Guided Missile Destroyers (DDG) and Guided Missile Cruisers (CG) to develop a deployable model of a healthy engine for the automated, near-real time comparison of sensed data. In traditional gas-path analysis (GPA), parameters such as compressor discharge pressure (CDP), compressor inlet temperature (CIT), and exhaust gas temperature (EGT) are predicted as a function of engine speed using baseline engine data and correction factors. Implementation of GPA in the MGT environment is particularly challenging, as analysis will be limited to narrow operation bands and not account for influence factors such as engine load or the performance of critical subsystems on the engine. In this work, a multi-layer perceptron (MLP) regression model is developed in order to capture the nonlinear relationships between engine controller inputs, external loads, and ambient conditions to selected sensor outputs such as gas generator speed (N_{GG}), low pressure turbine speed (N_{PT}), and power turbine inlet pressure (P_{54}). Optimal inputs and outputs are chosen using both mutual in-

formation (MI) scores and input from subject matter experts. A healthy data set was created using data from 60 MGT's in service and time-synchronized failure records over a five year period from 2012 to 2017. An inter-engine model was trained from the healthy data set and used to generate model residuals, which show strong correlation with a variety of critical failure modes reported in maintenance history, thus enabling automatic fault detection and remote identification of asset health and reliability.

1. INTRODUCTION

The health and reliability of installed equipment aboard the United States Navy (USN) surface fleet is critical to ensuring operational availability and readiness. Chief among this ship-board equipment are the marine gas turbines (MGT) installed for the production of propulsive power aboard DDG-51 and CG-47 class ships, the General Electric (GE) LM2500. To achieve the necessary reliability for this vital piece of equipment, development and deployment of health monitoring solutions in order to automate shore-side analysis and identify abnormal behavior are essential.

1.1. Gas Turbine Health and Performance Monitoring

A comprehensive review of gas turbine performance-based health monitoring, diagnostics, prognostics, and condition-based maintenance (CBM) was presented in (Tahan, Tsoutsanis, et al., 2017). State-of-the-art techniques in fault detection and isolation (FDI), model-based methods, data-driven methods, and hybrid methods were all presented from different authors. The large majority of literature in this field is focused on aviation, although the techniques applied to aircraft engines, industrial gas turbines, and marine gas turbines share many similarities. Linear gas path analysis (GPA) is one of the earliest techniques developed for monitoring gas turbines

Daniel Maraini et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(Urban, 1969). The purpose of GPA is to estimate the state vector of the engine by mapping measured variables such as compressor discharge pressure (CDP) and spool speed (N_1) to state variables such as compressor efficiency (η_c) and flow capacity (FC_c) (Ganguli, 2012). Typically, influence coefficients that map engine measurements to states are determined from baseline data and least-squares regression. Linear GPA was used in both (Pinelli et al., 2012) and (Roumeliotis et al., 2017) demonstrating the effectiveness of the technique to identify performance shifts on actual gas turbines in service. A commercial application of linear GPA has been presented in Doel (1992), which provides an overview of General Electric's TEMPER software. Various techniques have been developed to overcome the limitations of linear GPA, which are particularly important due to the strong nonlinear relationships between measurements and state variables in gas turbines. An improved GPA technique using particle swarm optimization (PSO) was demonstrated on simulated MGT data in (Ying et al., 2015). A diagnostic method using GPA and fuzzy logic was developed and demonstrated on industrial two-shaft gas turbine simulation data in (Amare et al., 2017). The use of Kalman filter and neural network methodologies combined with GPA for identifying faults in simulated gas turbine data was presented in (Volponi et al., 2003). An adaptive gas path analysis tool was presented in (Li, 2010), which introduces a novel two-step process for estimating degraded engine performance.

Many advances in gas turbine health and performance monitoring have come from the application of various machine learning (ML) techniques, such as artificial neural networks (ANN) and support vector machines (SVM). There are three main uses for ML models in gas turbine monitoring: 1) a healthy engine model in which a regression model is used to map selected inputs to outputs (typically sensor measurements), 2) fault detection and diagnostics in which a model is used for classification by mapping sensor measurements or model residuals to fault classes, and 3) prognostics for estimating remaining useful life (RUL). Research in this area has received increased attention over the past 15 years, particularly in the application of ANN's. An ANN was used for system identification of a single-shaft gas turbine in (Asgari et al., 2013), based on simulated data from physics-based models. In (Barad et al., 2012), steady-state and transient ANN models were developed for the combined performance and mechanical health monitoring of a gas turbine. Healthy gas turbine engine models were trained from simulated gas turbine data in (Kumar et al., n.d.) and (Tahan, Muhammad, & Karim, 2017). Healthy ANN-based gas turbine models based on field data were developed in (Fast et al., 2009) and (Shirazi et al., 2016). Logistic regression and ANNs were used in (Allegorico & Mantini, 2014) in order to identify cold spots, hot spots, and anomalies in exhaust gas temperature (EGT) profiles. A deep learning (DL) methodology was presented in

(Yan & Yu, 2015), in which learned features extracted from EGT profiles were shown to out-perform hand-crafted features for combustor anomaly detection. In (Zhou et al., 2015), an SVM was trained with simulated data from a degraded engine model in order to diagnose faults. Other techniques have been implemented for fault detection, including the application of similarity-based models (Carricajo et al., 2013) and sparse Bayesian learning (Pu et al., 2013). Research in gas turbine prognostics has also received increased attention. A comparison of data-driven techniques for gas turbine prognostics was presented in (Baptista et al., n.d.), including support vector regression (SVR), ANNs, random forests, and k-nearest neighbors (kNN) regression. A Bayesian hierarchical model was implemented on fleet data in (Zaidan et al., 2015), and shown to accurately track degradation and RUL.

1.2. Marine Gas Turbine (MGT) Health and Performance Monitoring

Research in the area of Marine Gas Turbine (MGT) health and performance monitoring for surface combatants dates back over 40 years. Some of the earliest work was presented in (Kandl & Groghan, 1980), which includes the development and testing of a condition monitoring (CM) solution for LM2500 MGTs aboard DD-963 and FFG-7 class ships. An approach for trending LM2500 vibration data was presented in (Hartranft, 1995), demonstrating a technique to predict optimal maintenance inspection intervals. In (Thompson & Raczkowski, 1996), the development of a diagnostic tool for trouble-shooting LM2500 performance and controls problems was presented, based on the development of baseline performance models from test cell data. A prognostic tool for timing optimal crank-wash intervals was presented in (Kacprzynski et al., 2001) and (Scharschan & Caguiat, 2005), based on land-based fouling tests. Algorithms for identifying clogged fuel nozzles based on features extracted from gas path sensor measurements for Rolls Royce 501-K17 gas turbine generators (GTG) were presented in (Byington et al., 2003). An ANN-based algorithm for detecting LM2500 compressor stalls was presented in (Caguiat et al., 2006), based on data generated from land-based tests. In (Campora et al., 2017), a combined fault detection and diagnostics tool was demonstrated based on Mahalanobis distance (MD) and ANNs, using simulated data for the entire power plant of a gas turbine powered frigate.

Much of the MGT research has been focused on specific faults using synthetic or test cell data, as opposed to overall engine health monitoring developed from fleet-wide historic operational data. Currently, the USN collects data from sensors installed on the gas turbine, fuel system, control system, and lube oil system. These data are sent to a shore-side repository for remote analysis, which includes a combination of automated fault detection based on preset sensor thresholds and manual data analysis from equipment experts. In

this work, a healthy engine model is sought in order to automate shore-side analysis and identify abnormal behavior in the LM2500 gas turbine and sub-systems, with less user intervention. An inter-engine model approach is taken using a Multi-layer Perceptron (MLP) regression model trained on a subset of healthy engine data from 60 LM2500s in service. The model is used to predict healthy engine behavior (sensor measurements) for a five year period of operation, demonstrating the ability to identify critical faults, many of which remained undetected by present methods.

2. LM2500 MARINE GAS TURBINE (MGT)

The General Electric LM2500 Marine Gas Turbine (MGT) is a light-weight, high-powered main propulsion engine installed aboard United States Navy (USN) Guided Missile Cruisers (CG), Guided Missile Destroyers (DDG), and the Independence Class Littoral Combat Ships (LCS-2 Variant). Since its introduction into the surface fleet in 1975, this derivative of the CF6-6/TF-39 aircraft engine has become ubiquitous with over 350 engines currently installed and operational aboard Navy vessels. The LM2500 MGT consists of two major modules (Figure 1), the Gas Generator Assembly (GGA) and Power Turbine Assembly (PTA), and fuel and lubrication sub-systems. The GGA is composed of a variable geometry compressor, an annular combustor, and a High Pressure Turbine (HPT); the PTA houses the six-stage Low Pressure Turbine (LPT). Data used in the following analyses are limited to ships equipped with the Digital Fuel Controller (DFC), which is described in the following section.

2.1. Fuel and Speed Governing System

The LM2500 utilizes the DFC to control gas generator speed (N_{GG}) via regulation of fuel flow into the combustor through the Fuel Metering Valve/Actuator (FMV/A), which is adjusted based on the desired Throttle Input Command (TIC) to the engine. The position of the Variable Stator Vanes (VSV) within the variable geometry compressor are rotated and translated based on N_{GG} and GGA inlet temperature (T_2) to maintain satisfactory compressor performance. The LPT speed (N_{PT}) is not directly controlled, but is established by the gas stream energy. In addition to measurements related to the command and control (e.g. TIC, VSV position, FMV/A position, N_{GG} , N_{PT} , and LPT Torque (M_{load})), the LM2500 is equipped with built-in sensors monitoring the gas path, vibration levels, fuel oil system, and lubrication oil system to remotely determine equipment health during underway operation. Gas path sensors include T_2 , GGA inlet pressure (P_2), compressor discharge pressure (P_3), LPT inlet pressure (P_{54}), and LPT inlet temperature (T_{54}). Vibration signals are measured via two accelerometers, one mounted to both the GGA and PTA. Signal conditioners are used to output displacements at each location at the frequency of rotation for both the GG and LPT such that four vibration signals are

produced: GGA vibrations at N_{GG} (V_{11}), GGA vibrations at N_{PT} (V_{12}), PTA vibrations at N_{GG} (V_{21}), and PTA vibrations at N_{PT} (V_{22}). The lubrication oil system supplies the bearings (housed in A- through D-Sumps) and the accessory gearbox with adequate cool oil to prevent excessive friction and heat. Relevant supply and scavenged lubrication oil temperatures and pressures throughout the system are monitored via in situ sensors as diagrammed in Figure 2. Fuel oil system health is monitored via remote sensing of fuel manifold pressure (P_{F2}) and fuel filter differential pressure (P_{F1}).

3. MODEL DEVELOPMENT

3.1. LM2500 Engine Data Collection

For each installed engine, sensors are measured simultaneously and collected shipboard during underway operation via the Integrated Condition Assessment System (ICAS). Data are acquired once per hour during steady-state operation when the compressor speed (N_{GG}) is greater than 4000 RPM or once the equipment regains steady-state if N_{GG} changes by more than 500 RPM in the interim. Steady-state operations are characterized by nearly constant N_{GG} , TIC, and T_{54} over a five minute interval. Qualified sensor data in ICAS is replicated to the shore side data repository CMAS (Consolidated Machinery Assessment System) for storage and analysis. The data set used in these analyses, obtained via the CMAS data repository, includes historical operational data from 60 LM2500s installed aboard 15 ships recorded during a five year span between 2012 and 2017. It includes 921,652 rows of samples with each row containing readings from 28 unique sensors. Additionally, failure records, limited to "critical" events expected to significantly affect engine performance and covering the entire five year span of the data set, were supplied and time-synchronized with the CMAS sensor data. A comprehensive overview of LM2500 failures was provided in (Driscoll et al., 2011), covering a 35-year period of engine removals.

3.2. Healthy Engine Data Selection

The time-synchronized failure records were used to label and extract nominally "healthy engine data" for model training and validation. The subset of healthy engine data was selected by first removing samples from the raw operational data in the 90 day window leading up to reported failures. This step accounted for the removal of approximately 40 percent of the original data set. The failures, limited to "critical" events only, include major faults in the Fuel Metering Valve (FMV), Variable Stator Vanes (VSV), Gas Generator Assembly (GGA), Lube Oil (LO) system, Power Turbine Assembly (PTA), and instrumentation. Next, rows with missing or erroneous sensor readings were removed, as the modeling approach requires complete vectors at each time sample. Practical upper and lower sensor limits extracted from the technical

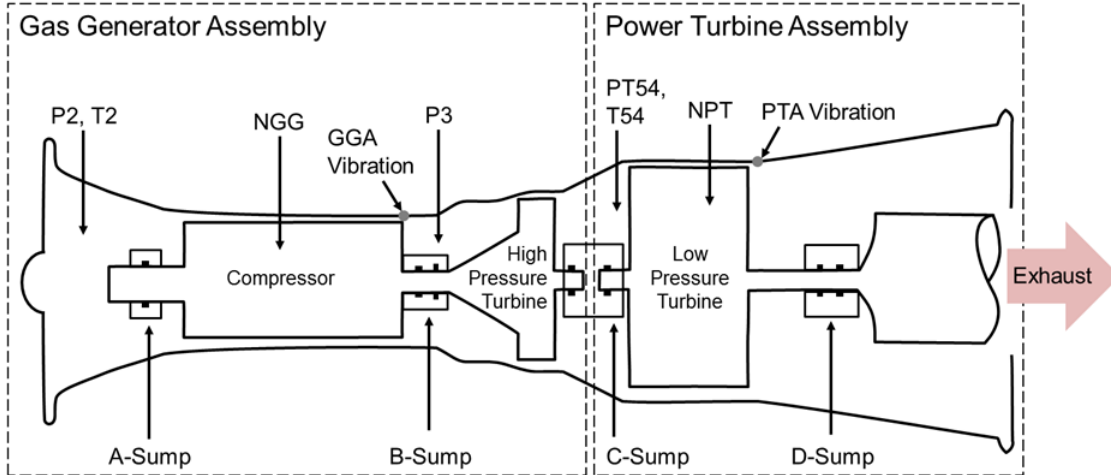


Figure 1. Schematic of LM2500 marine gas turbine.

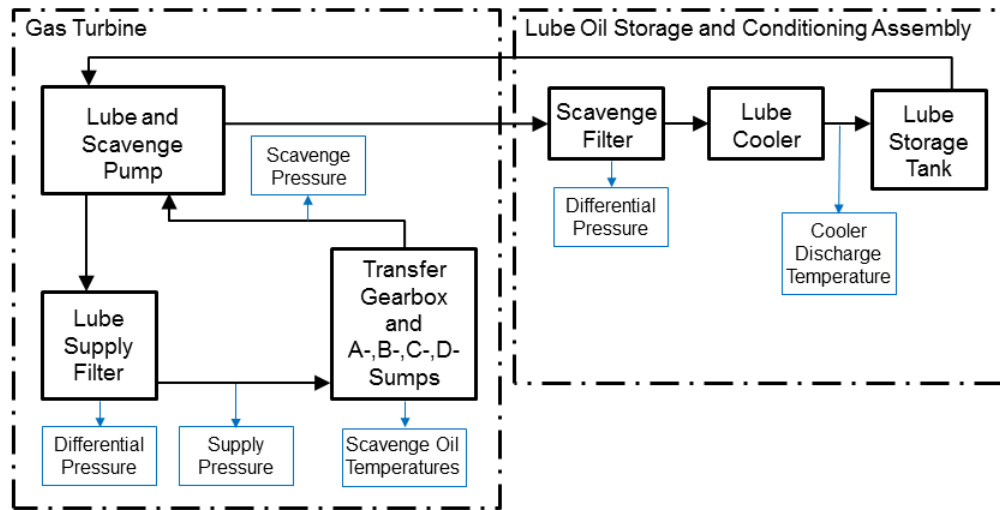


Figure 2. LM2500 lubrication system with corresponding sensors in blue.

manuals were used to filter out impossible sensor values a priori, such as vibration levels less than zero reported during operation. After filtering, 280,963 qualified samples were retained in the healthy data set, or approximately 30 percent of the original data set.

3.3. Model Input and Output Variable Selection

The input variables (see Table 1) of the model were chosen based on the operational description of the LM2500 with the aid of Mutual Information (MI) scores (Kappaganthu & Nataraj, 2011). The gas generator speed (N_{GG}) is directly controlled by the TIC setting, thereby making TIC an appropriate input to the model. Both T_2 and P_2 were also chosen as inputs, as they represent ambient conditions at the inlet of GGA. The power turbine load (M_{load}) was chosen as the final input and represents the torque load on the engine from driv-

Table 1. Model input variables.

Parameter	Symbol
Compressor inlet temperature	T_2
Compressor inlet pressure	P_2
Power Turbine Load	M_{load}
Throttle input command	TIC

ing the main reduction gear, which drives the ship's propeller shaft. The remaining sensors were considered as outputs of the model.

The dependencies amongst the various sensors were studied by computing the MI scores for all of the variables in the healthy data set. For two random variables, X and Y , the MI can be described as the reduction in uncertainty in X , given knowledge of Y . It is a measure of the dependency between the variables and equals zero if the two random variables are

independent. It can be calculated using the conditional and marginal entropy

$$I(X, Y) = H(X) - H(X|Y) \quad (1)$$

where the entropy is calculated from

$$H(X) = - \sum_x p(x) \log_2 p(x) \quad (2)$$

For discretely sampled data, estimates of $p(x)$ and $p(x|y)$ are required. Nearest neighbor density estimation was used in this work, which estimates the probability density by finding an appropriate volume containing a fixed number of points. Table 2 lists the selected model outputs and the three variables with the highest MI scores. For example, TIC has the highest MI score for the parameter N_{GG} , which is expected since N_{GG} is controlled directly by this variable.

3.4. Model Architecture

Figure 3 shows the model architecture for a single engine, which consists of a Multi-layer Perceptron (MLP) regression model that maps the inputs from Table 1 to the outputs listed in Table 2. The model is intended to be used in real-time, by continuously generating predictions for comparison with measurements as new data is recorded. This can be classified as an inter-engine model, as the combined historical operational data from 60 engines was used for training. The predicted output of the model is compared to the measured output and used to generate residuals, e , which are used to automatically detect deviations from "normal" behavior. In other words, this modeling approach aims to identify deviations away from typical fleet behavior.

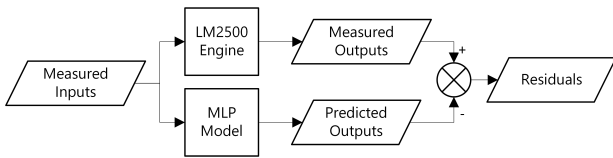


Figure 3. Model architecture.

3.4.1. Multi-layer Perceptron (MLP) Model

A basic multi-layer perceptron (MLP) model can be described as a nonlinear function that maps a set of input variables, (x_i) to a set of output variables (y_k) (Bishop, 2016). In this work, an MLP regression model was utilized with a single hidden layer and unit activation function in the output layer. The overall network function takes the form

$$y_k(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^M w_{kj}^{(2)} h \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right) \quad (3)$$

where D is the number of inputs, M is the number of hidden neurons, and $h(a)$ is the activation function. The rectified linear unit (RELU) was chosen as the activation function for the hidden layer and is given by

$$z_j = \max(0, a_j) \quad (4)$$

where

$$a_j = \sum_{i=0}^D w_{ji}^{(1)} x_i \quad (5)$$

The MLP model is trained by finding the weights, \mathbf{w} , that minimize the loss function

$$\Gamma = \frac{1}{2} \|\hat{y} - y\|_2^2 + \frac{\alpha}{2} \|\mathbf{w}\|_2^2 \quad (6)$$

where \hat{y} are the predicted values and α is the regularization hyper-parameter. In this work, optimization of Equation 6 was achieved using the Adam algorithm, an iterative algorithm that uses running averages of the gradients and second moment of the gradients. This algorithm was chosen for its computational efficiency and performance on large data sets, as demonstrated in (Kingma & Ba, 2014).

3.5. Model Training and Hyper-parameter Tuning

The optimal number of neurons (M) in the hidden layer of the MLP model was determined by examining the validation curves, which compares training and cross-validation error with respect to M . Before training, the healthy data set was standardized by removing the mean and dividing by the standard deviation for each variable. A digitized version of the throttle input command (TIC) was created using 25 linearly spaced bins to serve as the label for stratification, ensuring the folds preserve the percentage of samples in each bin. The healthy data set was then divided into training and validation data using a stratified cross-validation scheme with 10 folds. Training was repeated until each fold was used for validation, with the remaining 9 folds of data utilized for training. This was repeated for each value of M in order to compute the average cross-validation error, training error, and the standard deviation of both metrics. Figure 4 shows the training and cross-validation error plotted as a function of the number of neurons (M). The shaded bands indicate the standard deviation (confidence interval) computed across the ten folds for training and testing. The minimum cross-validation MSE was found to be 0.456 for $M = 20$, with a corresponding training MSE of 0.477. For $M > 20$, the training MSE continues to improve but the cross-validation error steadily degrades as the two curves diverge. The number of hidden neurons was therefore set to $M = 20$, as it achieves minimal error without

Table 2. Model output variables.

Variable	Symbol	Top MI Ranked Predictors
Cooling air temperature	T_{clg}	T_d, T_b, T_{gb}
Fuel metering valve position	FMV	P_3, VSV, P_{54}
Fuel filter differential pressure	P_{F1}	P_3, P_{F2}, VSV
Fuel manifold pressure	P_{F2}	P_3, VSV, FMV
Gas generator vibration at N_{GG}	V_{11}	V_{21}, V_{22}, P_{L2}
Gas generator vibration at N_{PT}	V_{12}	V_{22}, N_{PT}, P_{L1}
Lube oil cooler outlet temperature	T_{loc}	T_a, T_{gb}, T_d
Lube oil scavenge filter differential pressure	P_{L4}	P_{L3}, P_{L2}, V_{12}
Lube oil scavenge gearbox temperature	T_{gb}	T_a, T_c, T_d
Lube oil scavenge pump pressure	P_{L3}	N_{GG}, P_3, P_{F2}
Lube oil sump A temperature	T_a	T_c, T_b, T_{gb}
Lube oil sump B temperature	T_b	T_a, T_c, T_{gb}
Lube oil sump C temperature	T_c	T_a, T_{gb}, T_c
Lube oil sump D temperature	T_d	T_c, T_{gb}, T_a
Lube oil supply filter differential pressure	P_{L2}	V_{12}, V_{22}, P_{L3}
Lube oil supply pressure	P_{L1}	P_{F2}, VSV, FMV
Gas generator speed	N_{GG}	TIC, VSV, FMV
Low pressure turbine speed	N_{PT}	P_3, VSV, FMV
Compressor discharge pressure	P_3	VSV, FMV, P_{54}
Low pressure turbine inlet pressure	P_{54}	P_3, FMV, M_{load}
Power turbine vibration at N_{GG}	V_{21}	V_{11}, V_{22}, P_2
Power turbine vibration at N_{PT}	V_{22}	V_{12}, P_{L2}, N_{PT}
Low pressure turbine inlet temperature	T_{54}	FMV, TIC, N_{GG}
Variable stator vane position	VSV	P_3, N_{GG}, FMV

over-fitting and also achieves the lowest standard deviation of the cross-validation error.

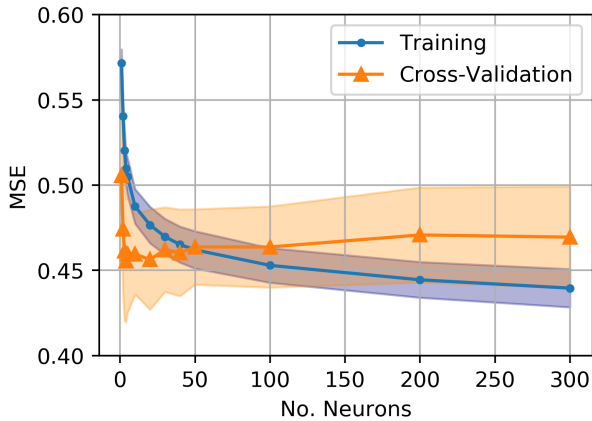


Figure 4. Training and cross-validation mean-squared error (MSE) vs. number of hidden neurons for MLP model.

3.5.1. Performance of MLP with Multiple Hidden Layers

An MLP with a single hidden layer was chosen for its simplicity, low computational expense, and performance. However, the performance of an MLP model with additional hidden layers was also studied. Since the hyper-parameter space for a multi-layer network can be large, a genetic algorithm (GA) was used to find the optimal number of hidden layers and nodes in each layer for comparison with the single-layer MLP. The parameter space was limited to a maximum of five

hidden layers and 50 hidden neurons per layer. Optimization was carried out as follows:

Genetic Algorithm:

1. Create initial population containing random hidden layer configurations within search subspace. Each member of the population is described by a five gene sequence with each gene corresponding to the number of hidden neurons.
 - (a) For each member of the population, compute the fitness score using the cross-validation MSE from 10-fold cross-validation
 - (b) Select the "parents" based on the fitness scores, the cross-validation MSE
 - (c) Select the "elite children" with the highest fitness scores
 - (d) Create "crossover children" by randomly combining genes from parents
 - (e) Create "mutation children" by randomly selecting/mutating genes from parents. Mutation was achieved by adding random integers between -5 and 5 to the genes
 - (f) Replace current population with new generation of children
2. Steps (a) to (f) are repeated until the difference in the best scores from successive populations is less than the tolerance

The GA was run with the parameters listed in Table 3. The algorithm was run repeatedly, using the best "children" from previous runs as part of the initial population for successive

runs. Table 4 lists the best scores from optimization as well as the results from optimization of the MLP with a single hidden layer. The additional hidden layers do not seem to offer any performance advantage, yielding comparable cross-validation MSE scores to the single-layer MLP. It was also determined that the three or four layer networks consistently scored better than two or five layer networks. Architectures with more than four layers were usually discarded after the first iteration. Each run converged in less than 50 iterations, often arriving within 10% of the final MSE after the first iteration.

Table 3. Genetic algorithm parameters.

Parameter	Value
Population	30
Number of Parents	4
Number of Elite Children	4
Crossover Fraction	0.8
Tolerance	5e-4

Table 4. Genetic algorithm results.

Model architecture	Best MSE score
4 - 44 - 46 - 43 - 43 - 24	0.458
4 - 28 - 26 - 26 - 24	0.463
4 - 20 - 24	0.457

3.5.2. Final MLP Model Training

The MLP model was trained with the model parameters listed in Table 5 using the healthy data set and a 70/30 training/testing split, yielding 196,674 training samples and 84,289 testing samples. The model reached convergence after 108 epochs with a final training MSE of 0.469 and testing MSE of 0.457.

Table 5. MLP Model Parameters.

Parameter	Value
Network architecture	4 - 20 - 24
Activation unit	RELU
Output activation unit	Linear
First moment decay rate (β_1)	0.9
Second moment decay rate (β_2)	0.99
Learning rate (α)	0.001
Momentum	0.9
Epsilon (ϵ)	1e-8
Solver	Adam

The performance of the model was examined by comparing the predicted outputs with the measured outputs for the healthy data set. Table 6 lists the model uncertainty for each of the predicted outputs expressed in physical units and as a percentage of the full scale range (% FS). The model is capable of predicting gas path parameters such as N_{GG} , VSV, P_3 , FMV, T_{54} , and P_{54} with the lowest prediction error and tends to perform worst on lube oil and vibration variables. This is

to be expected, as both lube oil and vibration sensors show a much larger band of process noise and variability when compared to the gas path parameters.

Figure 5 shows six performance curves typically used to assess gas turbine performance, including N_{GG} vs TIC, VSV vs N_{GG} , P_3 vs N_{GG} , FMV vs N_{GG} , T_{54} vs N_{PT} , and P_{54} vs N_{PT} . Each plot shows the measurements and model predictions for the entire healthy data set with 60 engines. The measured and predicted output show strong agreement in each of these plots, with only a very small subset of points falling outside of the prediction envelope.

Figure 6 shows the predicted and measured lube oil sump temperatures for sumps A, B, and C. For these variables, there are significantly more points that fall outside of the prediction envelope, particularly in the data for sump A and sump C. The model prediction curve for each sump tends to fall within the middle of measured range, thereby capturing the general behavior of the fleet of engines without over-fitting to the points in regions of lower density. The spread in data around this prediction envelope can be accounted for by the general variation amongst the engines, process noise, and outliers or erroneous readings that were not effectively filtered during healthy data selection. This spread is not of particular concern, as the general trend representative of the fleet of engines was sought, such that the model residuals become indicative of deviations away from typical fleet behavior, as opposed to specific engine behavior.

Figure 7 shows three plots that are typically used in the vibration assessment of the engine. The first plot, V_{22} vs. N_{PT} , shows the presence of high speed coupling shaft (HSCS) imbalance through elevated vibration levels in between 3000 and 3300 RPM. The second plot, V_{12} vs. N_{PT} , shows the presence of LPT imbalance through the peak located at 1575 RPM. The third plot, V_{21} vs. N_{GG} , shows the presence of GG rotor imbalance through the peak located at 7500 RPM. Similar to the case with the lube oil temperature sensors, the model fails to predict behavior in regions of lower density, including the peaks that result from imbalance of various engine components. These values were not filtered from the healthy data set as they still fall within the alarm limits of the engine. The large size of the training data as well as the model order selected prevent the model from over-fitting to these points. The peaks will therefore create high residual values indicative of anomalous behavior, which is desired.

4. RESULTS AND ANALYSIS

The original raw data set containing 921,652 samples was filtered by removing all rows containing missing data or values deemed erroneous a priori. The remaining data set of 624,276 samples contained unlabeled healthy and non-healthy operational data. Using the trained model, predictions were generated after the standardization (scaling) of the inputs and these

Table 6. Model Uncertainty for Predicted Variables.

Variable	Symbol	Uncertainty (% FS)	Units
Cooling air temperature	T_{clg}	12.8 (5.4%)	°F
Fuel metering valve position	FMV	1.46 (1.4%)	%
Fuel filter differential pressure	P_{F1}	3.45 (4.6%)	PSID
Fuel manifold pressure	P_{F2}	19.4 (2.5%)	PSIG
Gas generator vibration at N_{GG}	V_{11}	0.52 (6.7%)	mils
Gas generator vibration at N_{PT}	V_{12}	0.70 (4.1%)	mils
Lube oil cooler outlet temperature	T_{loc}	16.8 (1.9%)	°F
Lube oil scavenge filter differential pressure	P_{L4}	1.28 (6.8%)	PSID
Lube oil scavenge gearbox temperature	T_{gb}	15.2 (1.8%)	°F
Lube oil scavenge pump pressure	P_{L3}	2.93 (3.6%)	PSIG
Lube oil sump A temperature	T_a	16.0 (1.9%)	°F
Lube oil sump B temperature	T_b	13.9 (4.6%)	°F
Lube oil sump C temperature	T_c	15.7 (1.8%)	°F
Lube oil sump D temperature	T_d	23.9 (2.8%)	°F
Lube oil supply filter differential pressure	P_{L2}	0.56 (0.9%)	PSID
Lube oil supply pressure	P_{L1}	3.31 (4.2%)	PSIG
Gas generator speed	N_{GG}	22.8 (0.3%)	RPM
Low pressure turbine speed	N_{PT}	176 (4.8%)	RPM
Compressor discharge pressure	P_3	2.88 (1.1%)	PSIA
Low pressure turbine inlet pressure	P_{54}	2.06 (3.4%)	PSIA
Power turbine vibration at N_{GG}	V_{21}	0.37 (10%)	mils
Power turbine vibration at N_{PT}	V_{22}	0.33 (1.9%)	mils
Low pressure turbine inlet temperature	T_{54}	38.7 (2.5%)	°F
Variable stator vane position	VSV	0.95 (0.7%)	%

results were used to test model ability for fault and anomaly detection. Figures 8 and 9 compare measured and predicted time series for an engine operating normally, illustrating the strong agreement between the model and measurements.

To quantify sensor deviation from model predictions, residuals were generated using the following

$$e_k = y_k - \hat{y}_k \quad (7)$$

where y_k is the measured output and \hat{y}_k is the predicted output. The magnitudes and trending of the resulting residuals contain more diagnostics information than the raw signals alone. Current remote monitoring and detection techniques for installed LM2500s utilize only raw signals with universal alarm thresholds resulting in a significant number of undiagnosed faults and failures. A dashboard was developed for visualization of the residual signals, time series model predictions, and performance curves. Several examples of anomalous behavior detected through residual signal analysis are included in the following sections, illustrating the capabilities of the model for automating analyses. Smoothed residual signals are generated via a moving average with window size $n = 20$.

4.1. LPT Pressure (P_{54}) Degradation

Figure 10 shows the P_{54} raw and smoothed residual signals for an engine during several years of operation. A step change in the residual signal is observed around sample 2400 followed by a return to normal operation. Shortly thereafter, a second step change was observed at sample 2900, with the

residual signal remaining at -10 PSIA since. Reduction in P_{54} with respect to the model prediction can signal a loss in the engines ability to make power and hinder the platforms operational readiness. As the LPT speed (N_{PT}) is not directly controlled, but is established by the gas stream energy quantified by P_{54} entering the PTA, accurate modeling and monitoring of this sensor is essential. During this same operational period, further interrogation of the remaining gas path variables and their respective residuals found them to be within range with respect to the model, suggesting an instrumentation issue which went unnoticed for several months. Review of maintenance histories showed no known issues reported during this time period.

4.2. Lube Oil System Fault Detection

Figure 11 displays the raw and smoothed residuals of the scavenged lube oil temperature from sump A, T_A , over a period of five years for a single engine. Residuals as high as 600 °F were observed in the first several hundred samples, followed by a small interval of normal behavior before once again degrading. Sump A, located at the front of the GGA (Figure 1), contains the roller bearing that supports the GGA shaft ahead of the compressor. Lube oil to each of the four sumps (A through D) are supplied by a single storage tank and supply pump (Figure 2) suggesting a nominally constant lube oil temperature supplied to each bearing. Therefore, an elevated measurement of one lube oil scavenge temperature, while the remaining sump temperatures are found to be within range, can be indicative of impending bearing failure. Review of maintenance records during the operational

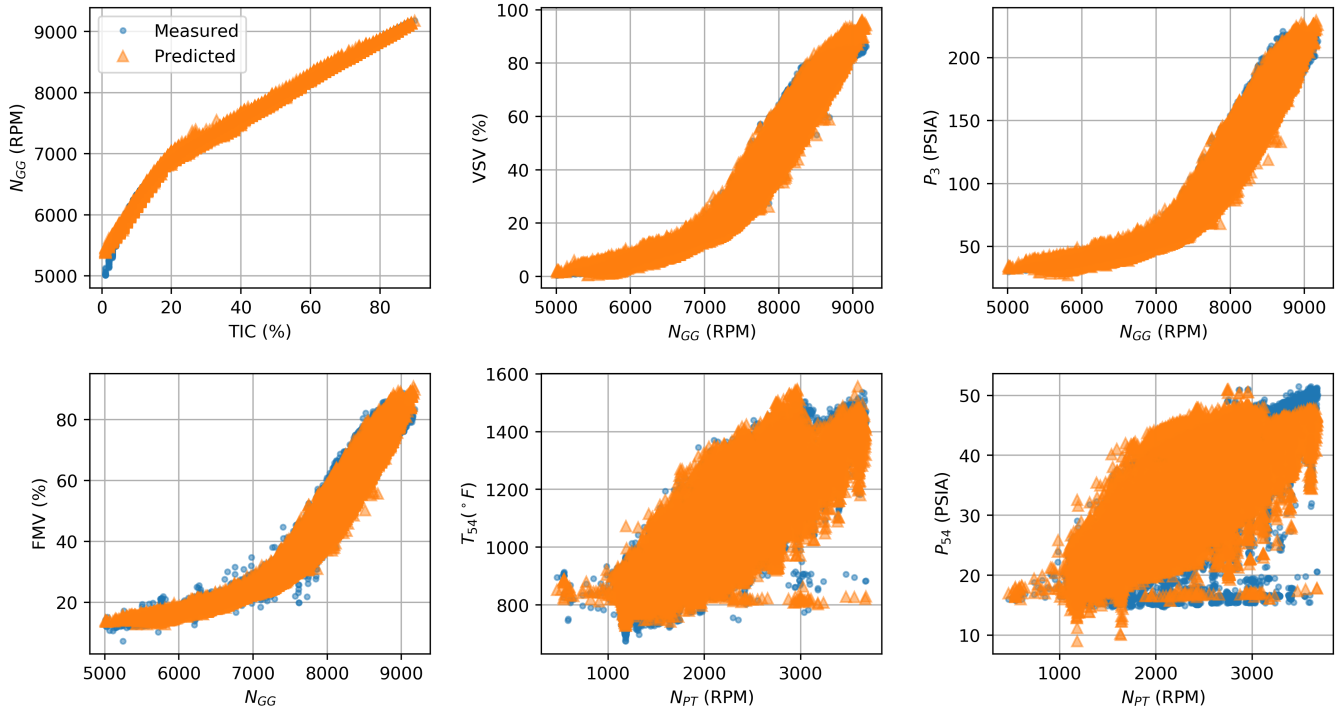


Figure 5. Comparison of measured and predicted performance curves generated from healthy data set.

time period when the deviation demonstrated in Figure 11 occurred revealed that the Sump A resistance temperature detector (RTD) was faulty. The faulty RTD sensor had gone undetected by current monitoring practices for three months. Maintenance and timely detection of faults associated with these RTDs are essential, as they are the primary indicator of bearing problems but tend to exhibit poor reliability.

4.3. Vibration Analysis

Figure 12 shows the gas generator vibration (V_{12}), measured at the rear of the GGA (Figure 1), tracked at low pressure turbine speed (N_{PT}), with the data filtered to include only N_{PT} speeds in between 1550 and 1650 RPM. This speed range corresponds to a structural resonance in the engine that tends to amplify the synchronous vibrations associated with low pressure turbine imbalance. This engine operated with peak-to-peak residual vibration levels of 2 mils for an extended period, including several instances of peak-to-peak residual vibration levels of 4 mils. The LM2500 operational alarm level for the raw vibration signals tracked to the power turbine rotor speed is 7.0 mils with a shutoff level of 10.0 mils. These levels are significantly larger than any measured value within the current data set and, therefore, no alarm was ever triggered. Upon review of maintenance records, recognition of a potential PT imbalance and subsequent trim balancing occurred following a periodic boroscope inspection of the engine. The results of the trim balancing are clearly visible in

the residual signal, indicated by the step change around the 1500th sample to a nominal residual value of zero. Continued operation of an unbalanced LPT in the 1600 RPM speed range has been shown to result in premature degradation of the engine and accessory mount links in addition to damaging the bleed manifold (Thompson & Grobler, 2012). It is therefore imperative to track and monitor vibration levels above normal in order to prevent prolonged operation with rotor imbalance.

4.4. Fuel Oil System Fault Detection

Figure 13 shows the fuel filter differential pressure (P_{F1}) raw and smoothed residual signals for a single engine operating over a five year span. The fuel filter is a high pressure filter mounted on the fuel pump to prevent large contaminants from entering the fuel controller. Above 27 PSID, a high differential pressure alarm is triggered indicating the fuel filter is blocked and above 35 PSID a relief bypass valve is opened until P_{F1} returns to less than 27 PSID. The large step change in the residual signal observed in Figure 13 at the 1200th sample is indicative of a clogged fuel filter. Following the brief increase in P_{F1} , a step change return to nominally zero residual is suggestive of corrective maintenance occurring to rectify issues with either the fuel filter or fuel filter differential pressure transducer.

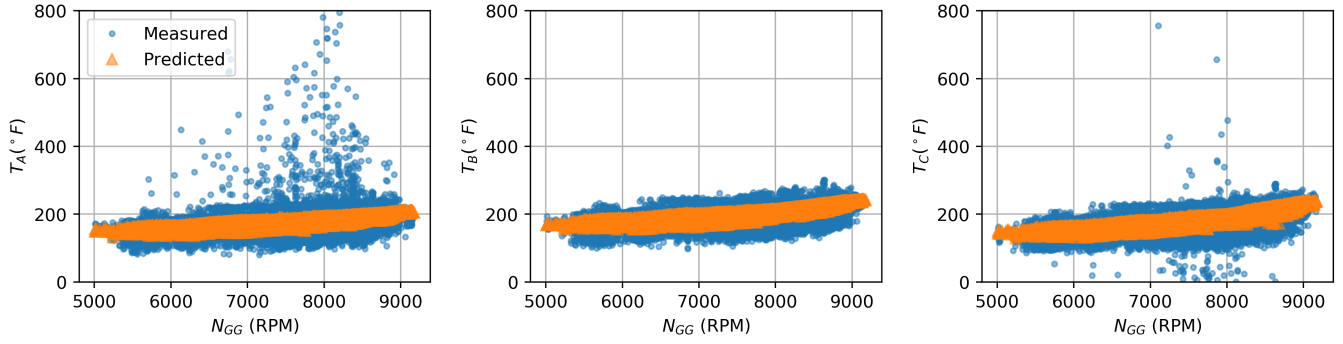


Figure 6. Comparison of measured and predicted lube oil sump temperatures generated from healthy data set.

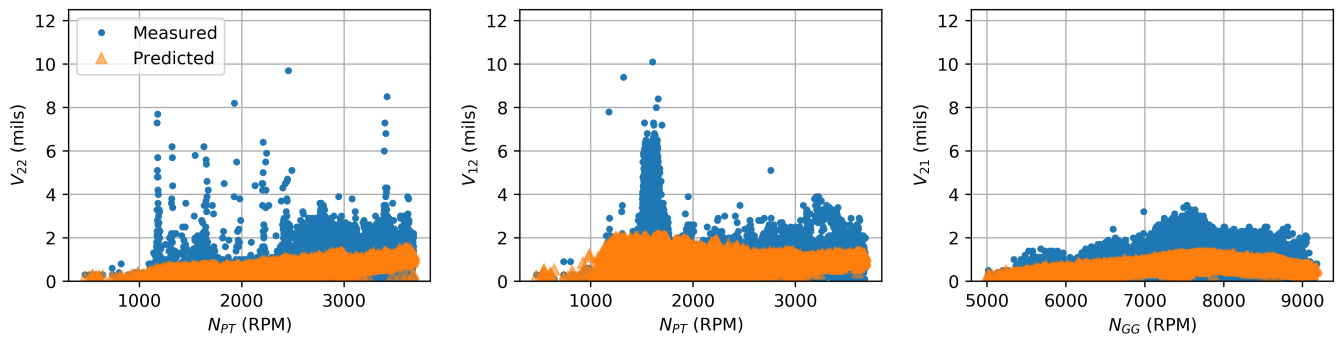


Figure 7. Comparison of measured and predicted vibration parameters generated from healthy data set.

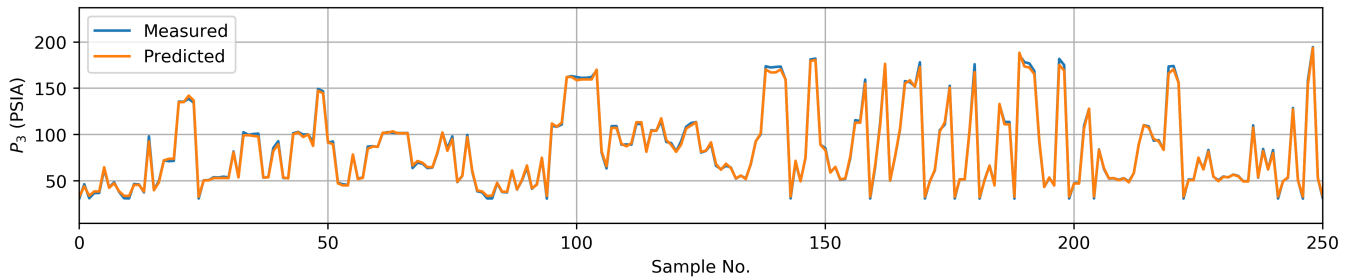


Figure 8. Comparison of measured and predicted compressor discharge pressure (P_3) for an engine operating normally.

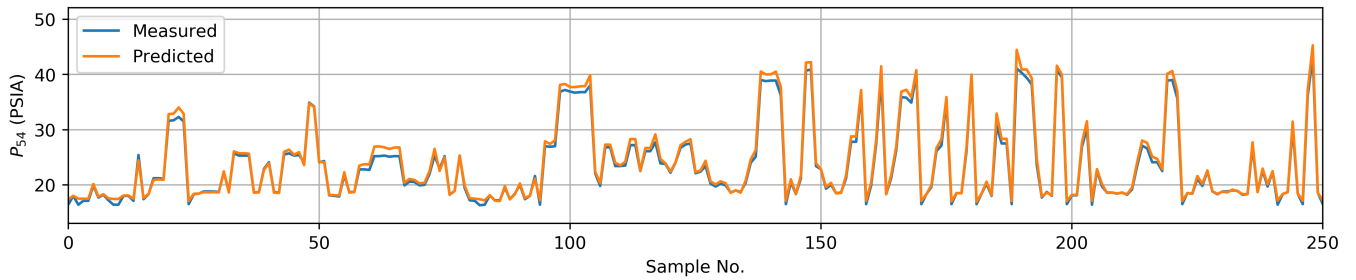


Figure 9. Comparison of measured and predicted low pressure turbine inlet pressure (P_{54}) for an engine operating normally.

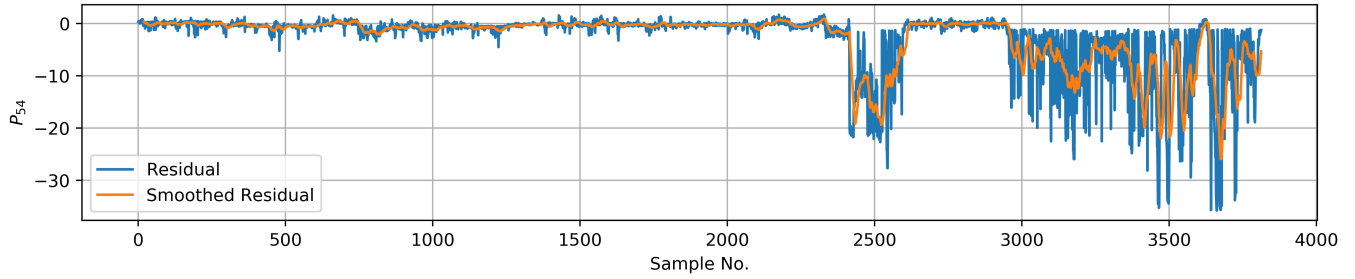


Figure 10. P_{54} raw and smoothed residual signals for single engine vs. sample number.

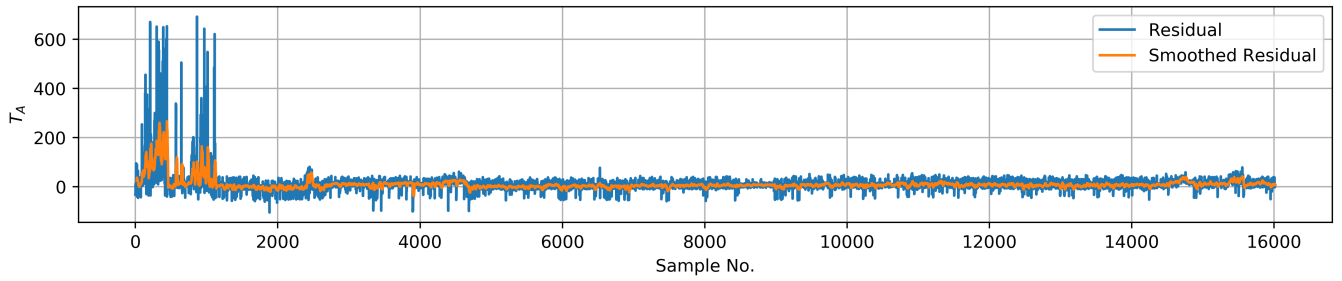


Figure 11. T_A raw and smoothed residual signals for a single engine vs. sample number.

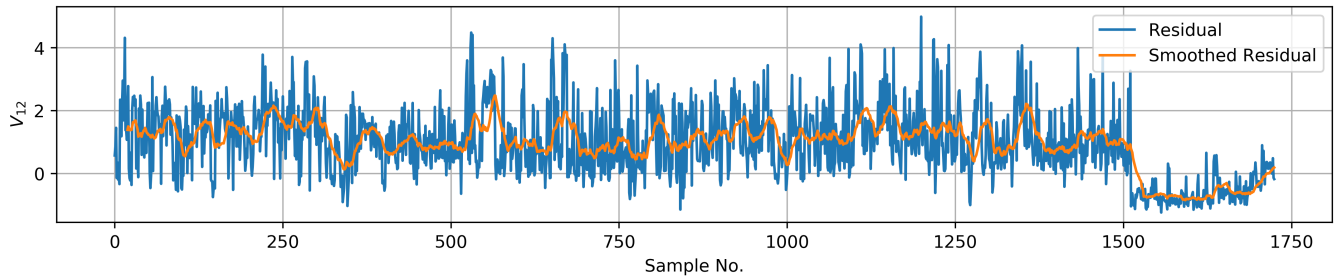


Figure 12. V_{12} raw and smoothed residual signals with N_{PT} speed range of 1550 to 1650 RPM for a single engine vs. sample number.

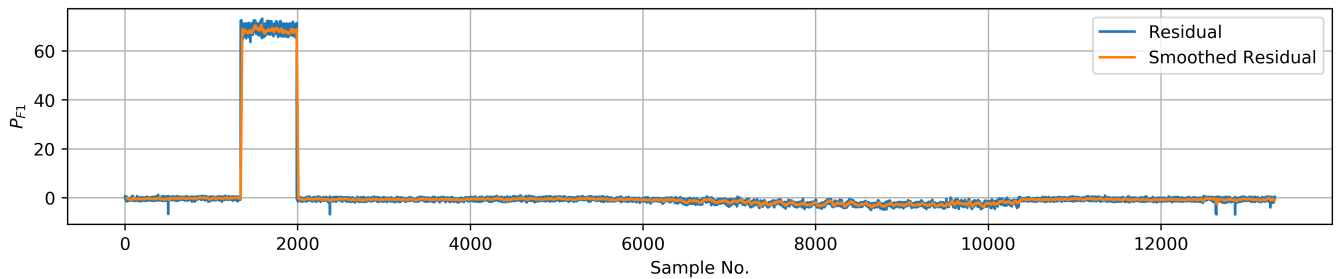


Figure 13. Fuel filter differential pressure (P_{F1}) raw and smoothed residual signals for a single engine vs. sample number.

5. CONCLUSION

A technique for monitoring the performance and health of LM2500 marine gas turbine (MGT) engines aboard CG-47

and DDG-51 class ships has been presented, using a Multi-layer Perceptron (MLP) regression model trained from a subset of healthy engine data from 60 MGTs in service. The

trained inter-engine model was used to generate predicted outputs for a period of five years from 2012 to 2017, on 15 ships. Residual signal analysis was performed in order to demonstrate the effectiveness of the model in identifying anomalies in the fuel system and gas path, vibration analysis, and instrumentation, corroborated by the failure history from the engines. A practical and scalable tool was developed for automated analyses of sensor data, reducing the hours required analyzing raw sensor data by equipment experts. Future work will focus on diagnostic and prognostic techniques applied to the residual signals, using failure mode history as the data class labels, with the aim of automating fault isolation, diagnostics, and prognostics for MGTs.

REFERENCES

- Allegorico, C., & Mantini, V. (2014). A data-driven approach for on-line gas turbine combustion monitoring using classification models. In *Second european conference of the prognostics and health management society* (pp. 92–100).
- Amare, F., Gilani, S., Aklilu, B., & Mojahid, A. (2017). Two-shaft stationary gas turbine engine gas path diagnostics using fuzzy logic. *Journal of Mechanical Science and Technology*, 31(11), 5593–5602.
- Asgari, H., Chen, X., Menhaj, M. B., & Sainudiin, R. (2013). Artificial neural network-based system identification for a single-shaft gas turbine. *Journal of Engineering for Gas Turbines and Power*, 135(9), 092601.
- Baptista, M., Nascimento Jr, C. L., Prendinger, H., & Henriques, E. (n.d.). A case for the use of data-driven methods in gas turbine prognostics.
- Barad, S. G., Ramaiah, P., Giridhar, R., & Krishnaiah, G. (2012). Neural network approach for a combined performance and mechanical health monitoring of a gas turbine engine. *Mechanical Systems and Signal Processing*, 27, 729–742.
- Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer-Verlag New York.
- Byington, C. S., Watson, M., Roemer, M. J., Galie, T. R., McGroarty, J. J., & Savage, C. (2003). *Prognostic enhancements to gas turbine diagnostic systems* (Tech. Rep.). IMPACT TECHNOLOGIES LLC STATE COLLEGE PA.
- Caguiat, D., Scharschan, J., Zipkin, D., & Nicolo, J. (2006). Applied neural network for navy marine gas turbine stall algorithm development. In *Aerospace conference, 2006 IEEE* (pp. 15–pp).
- Campora, U., Cravero, C., & Zaccone, R. (2017). Marine gas turbine monitoring and diagnostics by simulation and pattern recognition. *International Journal of Naval Architecture and Ocean Engineering*.
- Carricajo, T., Kripper, F., Orchard, M. E., Yacher, L., & Paredes, R. (2013). Anomaly detection in gas turbine compressor of a power generation plant using similarity-based modeling and multivariate analysis. In *Annual conference of the prognostics and health management society*.
- Doel, D. L. (1992). Temper: a gas-path analysis tool for commercial jet engines. In *Asme 1992 international gas turbine and aeroengine congress and exposition* (pp. V005T15A013–V005T15A013).
- Driscoll, M., McFetridge, E., Patterson, J., & See, C. (2011). Removals for cause: A 35 year assessment of Im2500 engine. *ASME Paper No. GT2011-45031*.
- Fast, M., Assadi, M., & De, S. (2009). Development and multi-utility of an ann model for an industrial gas turbine. *Applied Energy*, 86(1), 9–17.
- Ganguli, R. (2012). *Gas turbine diagnostics: Signal processing and fault isolation*. CRC press.
- Hartranft, J. J. (1995). Description of a vibration diagnostic trending approach for a condition monitoring system for the.
- Kacprzyński, G. J., Gumina, M., Roemer, M. J., Caguiat, D. E., Galie, T. R., & McGroarty, J. J. (2001). A prognostic modeling approach for predicting recurring maintenance for shipboard propulsion systems. In *Asme turbo expo 2001: Power for land, sea, and air* (pp. V001T02A003–V001T02A003).
- Kandl, M., & Groghan, D. (1980). Us navy, Im2500 gas turbine condition monitoring development experience. In *Asme 1980 international gas turbine conference and products show* (pp. V01BT02A064–V01BT02A064).
- Kappaganthu, K., & Nataraj, C. (2011). Feature selection for fault detection in rolling element bearings using mutual information. *Journal of vibration and acoustics*, 133(6), 061001.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kumar, A., Srivastava, A., Banerjee, A., & Goel, A. (n.d.). Performance based anomaly detection analysis of a gas turbine engine by artificial neural network approach.
- Li, Y.-G. (2010). Gas turbine performance and health status estimation using adaptive gas path analysis. *Journal of engineering for gas turbines and power*, 132(4), 041701.
- Pinelli, M., Spina, P. R., & Venturini, M. (2012). Gas turbine health state determination: methodology approach and field application. *International Journal of Rotating Machinery, 2012*.
- Pu, X., Liu, S., Jiang, H., & Yu, D. (2013). Sparse bayesian learning for gas path diagnostics. *Journal of Engineering for Gas Turbines and Power*, 135(7), 071601.
- Roumeliotis, I., Aretakis, N., & Alexiou, A. (2017). Industrial gas turbine health and performance assessment with field data. *Journal of Engineering for Gas Turbines and Power*, 139(5), 051202.

- Scharschan, J., & Caguiat, D. (2005). Development and improvement of compressor performance prognostics for us navy gas turbine engines. In *Aerospace conference, 2005 IEEE* (pp. 3472–3478).
- Shirazi, A. Z., Hatami, M., Yaghoobi, M., & Chabok, S. J. S. M. (2016). An intelligent approach to predict vibration rate in a real gas turbine. *Intelligent Industrial Systems*, 2(3), 253–267.
- Tahan, M., Muhammad, M., & Karim, Z. A. (2017). A multi-nets ann model for real-time performance-based automatic fault diagnosis of industrial gas turbine engines. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 39(7), 2865–2876.
- Tahan, M., Tsoutsanis, E., Muhammad, M., & 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.
- Thompson, B. D., & Grobler, J. (2012). Development us navy gas turbine vibration analysis expertise: Lm2500 vibration and trim balance. In *Asme turbo expo 2012: Turbine technical conference and exposition* (pp. 243–253).
- Thompson, B. D., & Raczkowski, R. (1996). Development of a diagnostic tool to trouble shoot lm2500 performance and controls problems. In *Asme 1996 international gas turbine and aeroengine congress and exhibition* (pp. V002T03A002–V002T03A002).
- Urban, L. A. (1969). *Gas turbine engine parameter interrelationships*. Hamilton Standard Division of United Aircraft Corporation.
- Volponi, A. J., DePold, H., Ganguli, R., & Daguang, C. (2003). The use of kalman filter and neural network methodologies in gas turbine performance diagnostics: a comparative study. *Journal of engineering for gas turbines and power*, 125(4), 917–924.
- Yan, W., & Yu, L. (2015). On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach. In *Proceedings of the annual conference of the prognostics and health management society*.
- Ying, Y., Cao, Y., Li, S., & Li, J. (2015). Nonlinear steady-state model based gas turbine health status estimation approach with improved particle swarm optimization algorithm. *Mathematical Problems in Engineering*, 2015.
- Zaidan, M. A., Harrison, R. F., Mills, A. R., & Fleming, P. J. (2015). Bayesian hierarchical models for aerospace gas turbine engine prognostics. *Expert Systems with Applications*, 42(1), 539–553.
- Zhou, D., Zhang, H., & Weng, S. (2015). A new gas path fault diagnostic method of gas turbine based on support vector machine. *Journal of Engineering for Gas Turbines and Power*, 137(10), 102605.

BIOGRAPHIES

Daniel Maraini is a Senior Data Scientist at McKean Defense Group, focused on the development of prognostics and health management (PHM) solutions for United States Navy ship systems. Before joining McKean, Daniel served as the Director of Condition Monitoring Engineering at Amsted Rail Company, Inc., where he led a team focused on the development of algorithms for remote monitoring of freight rail vehicle component health. He holds an MS degree in Mechanical Engineering from Villanova University and a BS degree in Physics from West Chester University. He is also a part-time PhD student in engineering at Villanova University, where his research is focused on model-based diagnostics.

Mark Simpson is a Data Scientist at McKean Defense Group, focused on the development of prognostics and health management (PHM) solutions for United States Navy ship systems. He holds a Ph.D. and MS degree in Mechanical Engineering from the Georgia Institute of Technology and a BS degree in Aerospace engineering from Penn State University.

Ronald Brown is the Engineering Manager for Data Utilization for Remote Monitoring and Condition Based Maintenance Systems at the Naval Surface Warfare Center - Philadelphia Division.

Michael Poporad is a Mechanical Engineer for Surface Combatant/Gas Turbine Engineering at the Naval Surface Warfare Center - Philadelphia Division.