FAULT DETECTION AND PROGNOSIS OF AEROSPACE SYSTEMS USING LONG SHORT-TERM MEMORY BASED RECURRENT NEURAL NETWORKS

by

Vasanth Dhanagopal

Bachelor of Engineering, Anna University, Chennai,
India (2012)

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the program of

Aerospace Engineering

Toronto, Ontario, Canada, 2020

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

FAULT DETECTION AND PROGNOSIS OF AEROSPACE SYSTEMS USING LONG SHORT-TERM MEMORY BASED RECURRENT NEURAL NETWORKS

Vasanth Dhanagopal
Master of Applied Science, Aerospace Engineering, Ryerson University,
2020

Health monitoring and remaining useful life predictions for the aerospace systems is a challenging and complex task to accomplish. Internal or external complications in these aerospace systems (aircraft and satellites) may lead to extremely hazardous or catastrophic consequences to the entire mission involving human life and budget. Considering the severity and complexity of the problem, this thesis deals in developing a diagnosis and prognosis health management system (DPHM) for the attitude actuator control system that uses reaction wheels in pyramid configuration onboard Kepler spacecraft and for the fleet of air-breathing turbofan engines. The established model is comparatively effective and computationally light in managing the objective of fault detection and prognostics.

An advanced data-driven DPHM scheme with optimization techniques is developed and evaluated. Initially, a recurrent LSTM (Long Short-Term Memory) neural network model is established and assessed with the general dataset (Particulate Matter ($PM_{2.5}$)). Secondly, a statistical-based fault detection method with functional factors of Weibull and mathematical features of frictional parameters showed that reaction wheels 2 and 4 of Kepler spacecraft have an early sign (~2 months) of their respective failures. This statistical method is compared with the proposed LSTM model for validation. Thirdly, the prognostic approach for estimation of remaining useful life (RUL) of the C-MAPPS and PHM08 datasets is successfully achieved. Numerous preprocessing methods such as digital filters (Savitzky-Golay (S-G)), principal component analysis (PCA) are used for standardizing the data. Finally, the optimization tools such as genetic algorithm (GA) and particle swarm optimization (PSO) are merged with LSTM for fine-tuning the hyper-parameters. Overall, the optimized model performs with better accuracy and can be concluded as a promising algorithm for the health management of complex systems.

ACKNOWLEDGMENT

With boundless love and respect, I would like to express my sincere thanks and whole-hearted gratitude to my supervisor Dr. Krishna Dev Kumar. This work would not have been possible without his highly informative and valuable guidance throughout every phase of time. His continuous motivation and support extended until the completion of this report. I am so blessed for having such an enthusiastic and intellectual person as my supervisor. He always pushed me beyond my boundaries and is always successful in inspiring me to achieve the goal. I am so thankful for being able to work on different platforms which broadened my vision in solving real-time problems.

I would like to thank the Department of Aerospace Engineering, Ryerson University for selecting me as an International candidate for pursuing my career. The financial assistance from the school kept me focused and facilitated my stay without any concerns. The teaching assistant opportunities encouraged learning new concepts and helped in understanding the new perspectives of the educational system.

Over a period of two years, I was fortunate enough to interact with many knowledgeable researchers of Artificial Intelligence and Predictive Analytics for Aerospace Systems (AI-PAAS) laboratory and engineers throughout Ryerson university. I learned a lot from every one of them, they never failed to support me intellectually and morally in any difficult times. I would like to thank all my colleagues and friends for their kind help whenever the necessity arose.

My Parents Dhanagopal and Vijayalakshmi, are the ones who nurtured and bestowed upon me everything that I am today. I am truly blessed to have an astoundingly supportive family, I will always cherish the memories and be obligated to them until the last day of my life. This thesis is dedicated to my parents and my extended family, in the hope that it will make them proud of the person they raised.

I appreciate and thank all the friends that strengthened me in hard times and made my stay very pleasant in Canada whenever I felt homesick.

I am always thankful to almighty God and classic Tamil literature for keeping me unique, refreshed, energetic and creative all the time. I would also like to thank one and all who have directly or indirectly helped me in making this project a great success.

For my loving and caring

Dad and Mom

grandparents, uncles, and aunts

siblings, in-laws, cousins

nephews and nieces

Table of Contents

List of Tables

List of Figures

Abbreviations

AI	Artificial Intelligence
ADCS	Attitude Determination and Control System
ANN	Artificial Neural Network
AI-PAAS	Artificial Intelligence and Predictive Analytics for Aerospace Systems
AOCS	Attitude and Orbit Control System
C&DH	Command and Data Handling Subsystem
CMAPPS	Commercial Modular Aero-Propulsion System Simulation
CME	Coronal Mass Ejection
CMG	Control Moment Gyro
CONTOUR	COmet Nucleus TOUR
DPHM	Diagnosis and Prognosis Health Management
FOD	Foreign Object Damage
GA	Genetic Algorithm
GPS	Global Positioning System
IDC	International Data Corporation
IGV	Inlet Guide Vanes
IPACS	Integrated Power and Attitude Control System
LSTM	Long-Short Term Memory
MDP	Markov Decision Process
MSG	Meteosat Second Generation
MSAT	Mobile Satellite
NASA	National Aeronautics and Space Administration
NN	Neural Network
PCA	Principal Component Analysis
$PM_{2.5}$	Particulate Matter
PSO	Particle Swarm Optimization
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network

RUL	Remaining Useful Life
RW	Reaction Wheel
RWA	Reaction Wheel Assembly
SF	Score Function
SOHO	Solar & Heliospheric Observatory
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TDRS	Tracking and Data Relay Satellite
VGV	Variable Guide Vanes

Nomenclature

α	Pitch angle
ϕ	Roll angle
ϑ	Yaw angle
$\vec{R}$	Distance from the satellite to the center of Earth
θ	True anomaly, measured counterclockwise from the perigee
$x(t)$	Input signal in the spectrum analysis
Amp_o	Initial amplitude
Amp_{min}	Minimum amplitude
Amp_{max}	Maximum amplitude
ω	Angular frequency
c	ARMA model constant
σ_t	Random variable in ARMA model
ϑ_t	Noise factor in ARMA model
α_i , β_i	ARMA model factors
F(t)	Weibull distribution function
β, η	Shape factor and Scale factor in Weibull analysis
x_j	Input to the j^{th} neuron
y_i	Actual output of the NN model
dj	Desired output of the NN model
w_{ij}	Connection weight
I_t	Input vector of LSTM
O_t	Output vector of LSTM
W_f	Weight matrix of LSTM
b_f	Bias vector of LSTM
C_t, C_{t-1}	Cell state vector at current time and pervious step
f_t	Forget vector of LSTM
σ	Sigmoid activation function

$tanh$	Hyerpbolic tangent activation function
N_t	Cell state update vector of LSTM
τ	Dry Friction
z	Normal distribution of prediction interval
S	Sensor values of turbofan engines
S_{nor}	Normalized sensor values
cov	Covariance matrix
$\overrightarrow{x_i}$	Position vector of Euler integration
$\overrightarrow{v_i}$	Velocity vector of Euler integration
$\overrightarrow{p_g}$	Velocity of the best particle in Euler integration
ρ_1, ρ_2	Random constants for social and cognitive behavior of the particles

CHAPTER 1

1. Introduction

In the present world, data science has become the solitary approach for all the Engineering problems. Most of the systems are automated and health monitored at regular time intervals. Intelligent and autonomous structure has been a core attraction for the wide range of industries and the elite technological group. The data provides details of the mechanism, while the technology helps in diagnosing the interruptions. Data science is the unification of statistics and machine learning approaches for determining the features without the actual knowledge of the system [1].

Big data analysis is a tool to extract, analyze and evaluate the complex data sets with a data processing algorithm. Data is growing at a very faster rate since the increasing use of the internet of things such as mobiles, cameras, health fitness, and medical devices, surveillance and other wireless networks. On average, about 1.7 megabytes of data is generated every second of the day by every person online. The International Data Corporation (IDC) predicted that the global data volume will upsurge exponentially from 4 to 40 zettabytes between 2013 and 2020 [2]. This gives the overall forthcoming evolutions for the data based predictive analysis. Figure 1.1 shows the growth of data usage over the period.

There are various methods available to perform the diagnosis of the system, they are model-based, signal-based, knowledge-based and Hybrid methods (Figure 1.2) [3]. In a model-based approach, a small-scaled model is used to formulate the measured and estimated outcomes. The signal-based method works based on measurement signals from the system. Both these methods require prior knowledge of the system and have well-documented limitations which will be discussed in the course of the work. Whereas the data-driven technique only needs a large historical dataset. The advantage of this approach is that even a complex system without a plant model can be investigated [4].

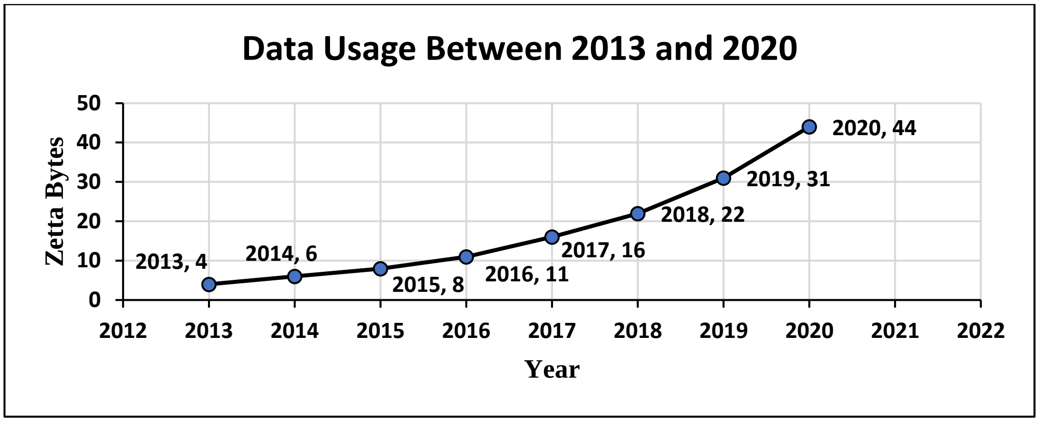

Figure 1.1 Growth of Data Usage [2]

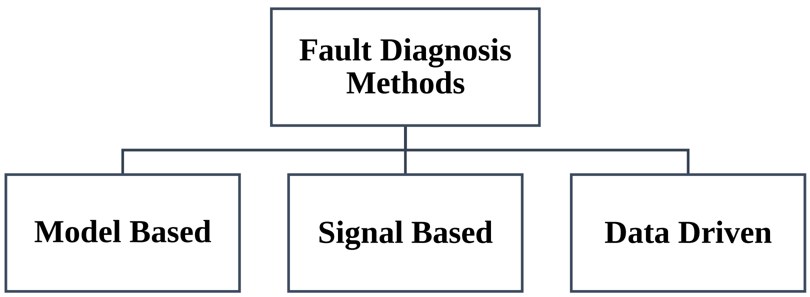

Figure 1.2 Approaches in Fault Detection [3]

In this work, health monitoring of the spacecraft, the remaining useful life of the Jet engines and future predictions of the environmental pollution factor are studied.

1.1 Motivation

The main goal of the thesis is to develop a data-driven model for the detection and isolation of fault, prediction of remaining useful life and health monitoring of the complex systems. The proposed algorithm with variations is applied on the Kepler spacecraft, C-MAPSS turbofan engine and for predicting particulate matter ($PM_{2.5}$) over three Canadian cities. Particulate matter is a fine particle pollutant (less than 2.5 microns) that affects the air quality.

1.1.1 Spacecraft Systems

Attitude dynamics of the orbiting body and its control has been a distinct interest among the elite community for a long time. A Satellite placed in its orbit is tagged to perform its specific mission such as forecasting, navigation, imaging and so on. However, the functionality of the satellite can be easily disturbed by internal and external sources. The internal factors are radiation thrust, mass expulsion, momentum shift, and the external elements includes environmental radiations, earth's gravity gradient, magnetic and aerodynamic torques [5].

To control the attitude of the vehicle; sensors, actuators and a specific algorithm based on the current and the desired attitude are required. Reaction wheels and control moment gyros are two of the commonly used attitude control systems. A resistant torque can be produced by accelerating or decelerating the rotation of the wheel since the initial angular velocity is zero for the reaction wheel, whereas the momentum wheels have initial angular velocity hence the magnitude of the resistant torque can be controlled by changing the angular velocity [6]. Figure 1.3 represents the motion of the orbiting spacecraft.

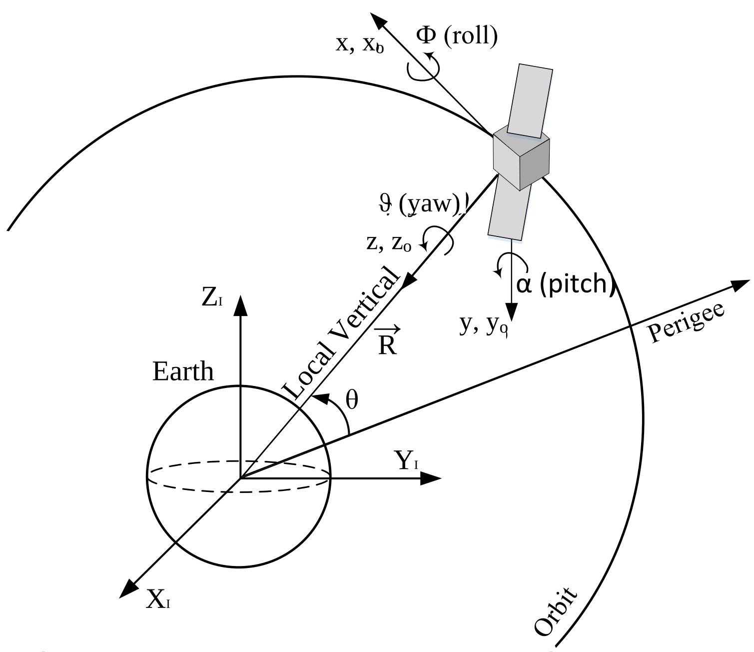

Figure 1.3 Attitude and Orbital Motion of the Rigid body Spacecraft [4]

Apart from the interruptions and external disturbances the air-borne vehicles and the spacecraft components are subjected to fatigue which results in wear and tear. Allied Market Research estimated that the usage of the small satellite will reach around $7 billion by the year 2020 [7], so it is about the huge budget spent on the launch. Besides everything, it is very difficult to perform maintenance in the outer space. Hence, the defects must be identified, isolated in a well advanced period and an appropriate control system must be calibrated.

1.1.1.1 Subsystems of the Spacecraft

A spacecraft comprises the following subsystems (Figure 1.4) [8] for its operational functionality [6]. Throughout the mission cycle, every subsystem of the spacecraft is prone to faults and failures. They are as explained,

a. Structure - An enclosure capable of withstanding stress and vibrations
b. Attitude and Orbit Determination and Control - Guidance and Navigation
c. Communication System - Exchanging signals
d. Command and Data Handling Subsystem (C&DH) -Telemetry system for downlink and uplink of data
e. Power System - Batteries, Fuel cells and solar panels
f. Thermal system - Protection from the hostile environment
g. Payload - Scientific equipment with a mission-specific functionality

1.1.1.2 Satellite Attitude Determination and Control System

Every man-made object in space such as spacecraft, space stations, and satellites are susceptible to many disturbances in space environment that create an undesirable translation and rotational motions of the spacecraft or satellites. Attitude Determination and Control System (ADCS) as shown in Figure 1.5 maintains the attitude of satellite against external disturbance torque such as solar radiation, aerodynamic drag and the impact of the earth's magnetic field. Thus, an attitude control torque device such as the momentum wheels and reaction wheels are mounted within vehicles to reduce attitude errors.

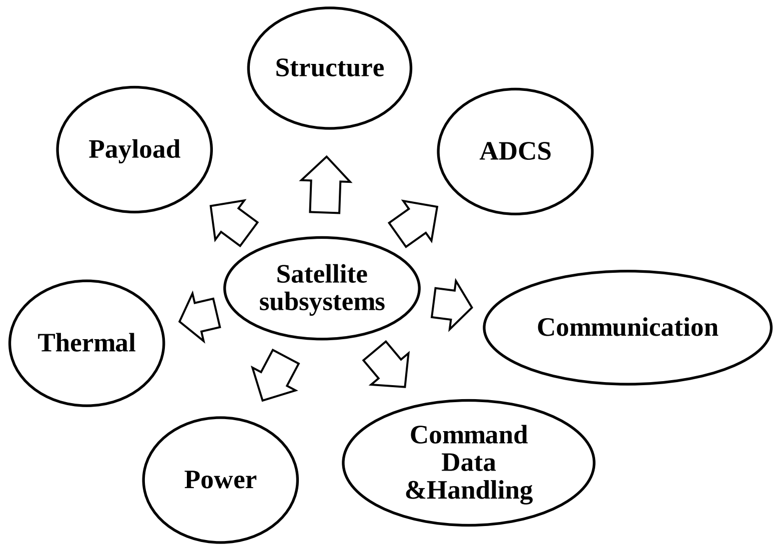

Figure 1.4 Subsystems of the Spacecraft [8]

The later development of a control torque generator that uses a momentum wheel principle is Control Moment Gyroscopes (CMGs). CMGs are commonly used to provide attitude control for a variety of vehicles, including spacecraft and satellites. CMGs generate attitude control torque in response to onboard or ground commands. Mostly the four double-gimbal control moment gyroscopes assembly is used to maintain the desired attitude. CMGs have many advantages; performance-wise, CMGs produce very minimal error as compared to other actuators that are up to 0.001°of pointing accuracy, but very expensive in terms of cost and mechanically complex. Thus, CMGs are normally preferred for high-cost missions, which require a high pointing accuracy [9].

Most spacecraft uses Reaction Wheel (RW) as the actuator, which is a three-axis attitude controller that operates at a constant rotation speed. They provide high pointing accuracy and are useful for rotating the spacecraft by a very small angle. The spacecraft is maneuvered when the torque is applied to one of its axes. Combinations of RWs integrated to the system can provide full axis attitude control and stability [10].

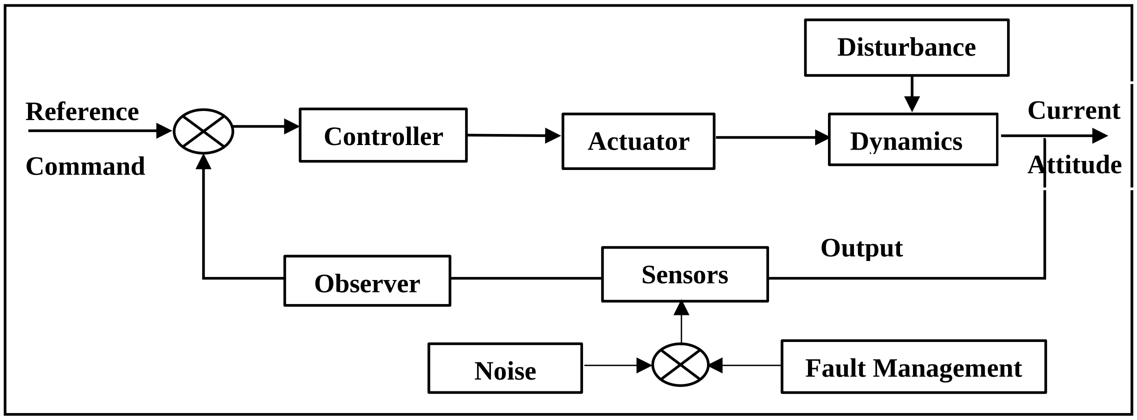

Figure 1.5 Schematic of Satellite Attitude Determination and Control System [11]

Use of chemical batteries to store excess energy generated by the solar panels during periods of exposure to the Sun. The primary problem with this approach is the cycle life of batteries and the additional power system mass required to control the charging and discharging cycles. An alternative to chemical batteries is the use of flywheels to store energy. This concept termed the Integrated Power and Attitude Control System (IPACS) has been studied since the 1960s, but it has been particularly popular since the 1980s. In fact, the use of flywheels instead of batteries to store energy on spacecraft was suggested as early as 1961 [12].

Also, these actuators such as reaction wheels, control moment gyros, and momentum wheels are electro-mechanical devices that are suspected to malfunction during the longer operating cycle.

1.1.1.3 Health Monitoring System

The main objective of the attitude controller is to generate a corrective torque based on the current and desired position. The reference position is attained from the inbuilt sensor measurements and the de-tumbling of the satellite to the desired position is performed by the actuators. In case of failure of any of those mechanical systems, the attitude will be lost, and trajectory will be out of control. Hence the health management system must be incorporated with the ACDS of the satellite as shown in Figure 1.6.

The performance of the system will be affected after the onset of the fault which will be detected by the integrated Diagnosis and Prognosis Health Management (DPHM) module. The algorithm must identify the type and locate the fault. Depending on the severity the ACDS system should be informed for further corrective actions, if the fault is inevitable then the algorithm should estimate the remaining useful life of the component [13].

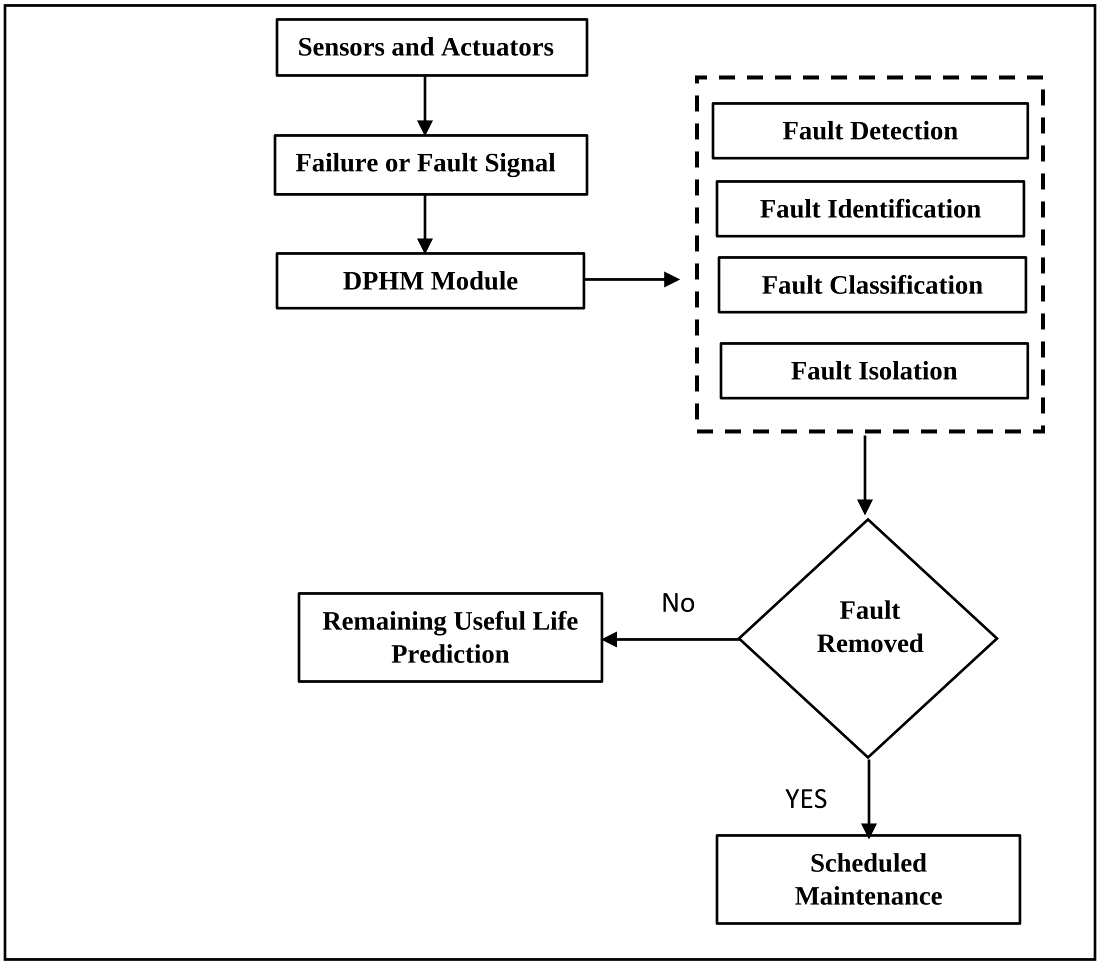

Figure 1.6 Diagnosis and Prognosis Health Management (DPHM) Module [11]

1.1.1.4 History of Failures of Spacecraft Systems

Malfunctioning and failures are unavoidable with any of the mechanical systems, critical failure may result in termination of the mission, one such is a failure in the attitude control actuators. The actuator wheels that are positioned outside the spacecraft body are subjected to various irregularities in the temperature, speed constraints and momentum shift during the maneuvers. Conventionally, faults and failures were dealt with redundant hardware, however, design and operational cost will be increased.

The history of space mission failure emphasizes the importance of health monitoring for the control systems of the spacecraft. For the mission GPS BI-05, the reaction wheels 2 and 3 stopped completely with full motor voltage applied. In 2001, the backup momentum wheel of Radarsat 1 failed to lose the attitude. In 2002, there were anomalies in two of the thrusters of EchoStar VIII, the 2^{nd} generation Globalstar satellite experienced a reaction wheel failure, the satellite missions for MSG-1, TDRS I, Nozomi, CONTOUR, EchoStar V and VI were aborted in the same year. For the Cassini spacecraft in mid-2003 after 5 years of the launch, the operations team observed an anomalous drag torque in the bearings of the RWA-3 [14]. In the same year, 2003 various mechanical components malfunctioned for SOHO, MSAT1, Nimiq2, Thaicom3 missions which were terminated [6].

After two reaction wheel failures in 2004-05, the spacecraft HAYABUSA was switched to one wheel and thruster configuration. In October 2005, the satellite TOPEX could not perform attitude maneuvers due to the failure of the pitch reaction wheel. After the failure of two reaction wheels, the Far Ultraviolet Spectroscopic Explorer (FUSE) mission was altered using a hybrid controller, later in mid-2007 the last reaction wheel was also failed and the efforts to restart the spacecraft was unsuccessful. In 2007, NASA’s Thermosphere, Ionosphere, Mesosphere Energetics, and Dynamics (TIMED) satellite’s reaction wheel failed due to bearing problems [15].

In the recent past, various cases were observed, two reaction wheels of the Dawn spacecraft of NASA has failed in the year 2010 and 2012 due to excessive friction development. In July 2012 Kepler’s reaction wheel 2 failed unexpectedly and later in 2013 after ten months of the first failure the reaction wheel 4 was also futile, after which the primary mission was altered as K2 [16]. The failure analysis for this Kepler mission is studied in this research work.

The larger satellites can always be equipped with redundant wheels onboard in case of any failures or malfunctions of main Reaction wheel assemblies (RWA). Swift Gamma-Ray Burst Explorer (Swift) and the James Webb Space Telescope (JWST) employs six RWAs. But for the micro and nanosatellites, there is no room for accommodating additional hardware due to limited space and power budget. In such a case, the attitude control system and health monitoring systems should be effective. Any small anomaly in the behavior of the components should be immediately addressed. Figure 1.7 shows the details of subsystem failure and year wise ADCS failure rate from 1980-2005 [17].

Now the analytical procedure can very well be used avoiding the space and budget constraints to detect and isolate the failures. Because of the growing demand for reliability and safety in the control systems, there has been an increasing interest in fault diagnosis. Fault tolerance is an essential concern in the attitude control design.

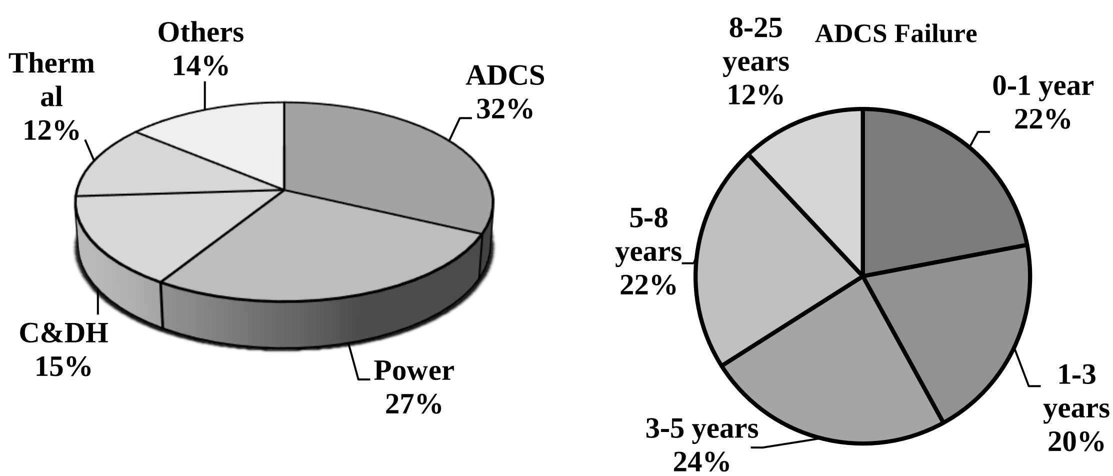

Figure 1.7 Failure Percentage of the Subsystems and ADCS Time to Failure from Launch (1980-2005) [17]

1.1.1.5 Faults in Attitude Determination Control Systems (ADCS)

From the above discussion of failure events and percentage of subsystem faults, it is clearly evident that components of the attitude control system are prone to malfunction with a higher rate and also more than 40 percent of them failed within 3 years of the launch. This study gives an insight into studying the components of ADCS.

The Attitude control system consists of sensors and actuators. The sensors are used to measure the attitude, which includes sun sensor, star sensor, horizon sensor, magnetometer, etc. The inputs from the available sensors are combined to a single solution and processed for determining the current attitude of the system. Sun sensors measure one or two angles between their mounting base and incident sunlight with a clear field of view. Star sensors are mainly trackers, they are used to scan multiple star fields for deriving the vehicle's attitude. Horizon sensors are infrared devices that are capable of detecting the contrast between cold space and the Earth's atmosphere. Magnetometers measure the magnitude and direction of Earth's magnetic field which then establishes the vehicle's relative attitude to the local magnetic field.

Actuators are selected based on the mission requirements. They produce torque required for orienting the spacecraft to the desired position. Some of the actuators used in the space vehicles are Reaction wheels, Momentum Wheels, Wheel drive electronics, control moment gyros, value drive electronics, magnetic torquer, gas jet propulsion system, etc. [18]. Failure of the control system can be from any of the above-mentioned components in ADCS. From Figure 1.8, it is evident that among all other components RWs, CMGs, and Momentum wheels constitute more ratio of failure percentage [17] and hence have to be monitored cautiously. In the present work, the most predominant section of the failure (reaction wheel fault) of the Kepler Mission in the ADCS is taken for consideration. Anomaly detection without apriori knowledge of the mission is carried out in a precise manner.

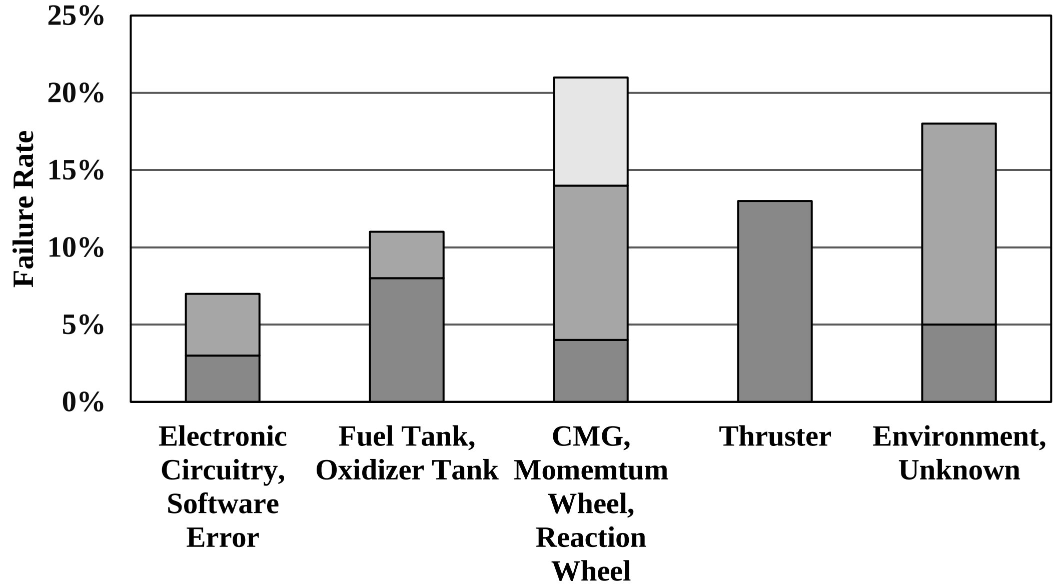

Figure 1.8 Component Wise Failure of ADCS (1998-2005) [17]

1.1.1.6 Fault Classification

All the mechanical components are prone to faults and failure, every system is set with periodic maintenance after its permissible cycle of operation. In the case of complex systems or space vehicles, the onboard maintenance is impossible or very difficult which will lead to failure of the entire mission. Failure/fault is caused by various physical phenomena such as fracture, creep, thermal shock, corrosion, wear/tear, and other external factors.

Faults can be abrupt, intermittent, transient and/or anticipated (Figure 1.9) that can degrade the actual performance of the system [11]. Abrupt faults in the system cannot be predicted by data-driven methods because it is sudden and there are no histories feed in training the model. Transient and intermittent faults are timely malfunctions that may or may not affect the performance of the system, therefore consideration of this type will result in the false signal. The prolonged incipient fault is the only condition that has the feature which can be mimicked while training the data-driven model.

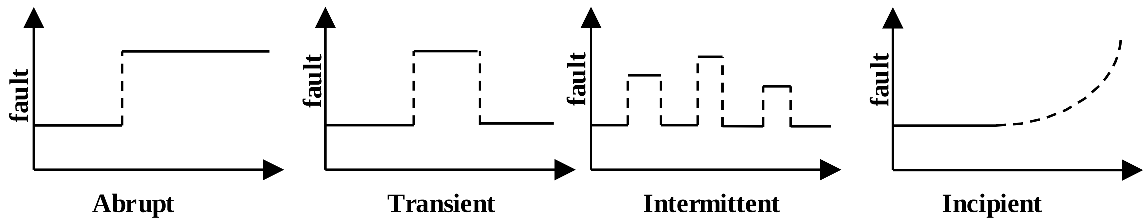

Figure 1.9 Types of Fault-Based on Time Characteristics [11]

Based on the operative behavior, the fault is either identified as additive or multiplicative [11]. Additive faults are observed in attitude control systems of the reaction wheels. The continuous malfunctioning leading to the accumulation of torques that changes the nominal action is categorized as a multiplicative fault (Figure 1.10).

Figure 1.10 Types of Fault-Based on Mathematical Operations [11]

The fault in the RWs are due to the failure of signal response, decreased reaction torque, increased bias torque. Analysis of systematic failure events and examining its cause and effects will reduce the risk of the severity. Fault diagnosis and prognosis methods for satellite systems continuously monitor the health and predict future events.

1.1.2 Jet Engines

The Jet Engines are air-breathing engines used for propelling the aircraft in the forward direction. The turbofan, turboprop, turboshaft engines are variants of turbojet engines for producing thrust through fans, propellers, shafts respectively. They have inlets, low pressure, and high compressors, combustion systems, turbines, and nozzles as the components. There are several performance parameters, and all are interrelated which in turn affects the overall functionality of the engine and the aircraft [19].

1.1.2.1 Performance Parameters

The operating conditions of the engines vary with different speeds (Mach number) and altitude. The thrust generated is the foremost parameter which is the balance of uninstalled force minus the drag. The second is the rate of fuel used per unit thrust produced and is known as specific fuel consumption. The other parameters are defined as efficiencies and are related to shaft power, kinetic energy, overall velocity, thermal energy, bypass ratio of the engine [19]. Every feature is dependent on local conditions, internal and external factors, a comprehensive study of cause and effects between each parameter is extensively compared with appropriate experimental results. Optimal shape, weight, material, speed, altitude was already prescribed by the technical group for efficient operation. The performance of the engine declines regardless of the health conditions of the individual components. Deterioration can occur due to change in the load or variation of the external ambient conditions.

Every fault mode has an effect on the performance, (i) Intake - Filter clogging will reduce the mass flow and power loss. (ii) Compressor and Turbine – fouling, corrosion, tip clearance, IGV/VGV malfunction, FOD will tend to increase the heat-rate, loss of power, loss of thrust and engine blowout. (iii) Combustion Chamber – liner cracking, clogging, fuel injectors malfunction may result in a lean mixture thereby reducing the thrust and power. (v) Nozzle – Surface Erosion increases the back pressure and thus affecting the overall efficiency of the engine [20].

1.1.2.2 Onboard Sensors

For measuring operating parameters, the modern aircraft engines are incorporated with numerous sensors. The output of the sensors is used to monitor the engine health conditions. The sensors are capable of operating at severe environmental conditions and supply signals to the cockpit indicators [21]. They are mounted at various locations of the engine to measure the critical variables such as total temperature, static and total pressure, fan and core speed, bypass ratio, pressure ratio, fuel flow, etc. Sensors can detect impending fault and alarm the system for corrective actions.

1.1.2.3 Health Monitoring Systems

Health monitoring is a continuous process of obtaining, measuring, filtering, analyzing and detecting the irregularity. This system uses the information from multiple sensors for the diagnosis of the engine. The data from the sensors are used both to monitor the performance in real-time and predict the remaining useful life of the engine and its components so that timely maintenance is scheduled to prevent the failure events. The failure can be sudden, random, uncertain; hence the diagnostic approach has to be a blend of expert systems and artificial intelligence. Observation of non-performance values like exhaust pollution, noise, vibration, debris, and clearances can provide additional attention to the safety and reliability of the system [22].

Model Based Approach – nonlinear and dynamic gas-path models should be considered with powerful computational platforms

Model-Free Methods – advanced AI techniques can be used for pattern recognition and fault isolation with automatic data collection

Hybrid Models – fusion of classical models will improve the prediction results

Modern methods should include the effects of the operating condition, such as power in the trend of degradation, in order to reduce uncertainty of simple model

Global Pattern Recognition from across the similar jet engines should be considered to include the preceding measurements and failure events

Figure 1.11 Modern Health Monitoring Framework [23]

A key objective is to find out the effects of the actual cause of the deterioration. The pattern of the faulty feature has to be identified and used for diagnostic purposes. Every diagnostic and prognostic technique has its own advantages and disadvantages thus a hybrid model will be an ideal framework for the hierarchical development of different techniques while alleviating their limitations [23]. Remaining useful life (RUL) is the central parameter in condition-based maintenance and prognostic health management. Figure 1.11 presents the recommendations for designing modern health monitoring and prognostic system.

1.1.3 Environmental Pollution

Several manmade sources like agricultural burning, residential wood burning, Industries, carbon vehicles, Airplanes and natural sources like dust storms, volcanic eruptions cause the pollutant contaminates to stay in the atmosphere. Long-term exposure leads to climatic changes and severe health issues to living beings [24]. The fine particles with 2.5µm or less were deadly, ranks the sixth leading risk factor globally and cause a 36% increase in lung cancer as it can penetrate deeper into the lungs.

In 2016, $PM_{2.5}$ exposure contributed to more than 4 million deaths worldwide which include heart disease, chronic lung disease and respiratory infections [25]. In this work, the trend of Particulate Matter $PM_{2.5}$ over the three Canadian cities is studied with data-driven techniques.

1.2 Data Science

Data science is a growing field that uses algorithms, frameworks, scientific approaches to extract meaningful features and distinct characteristics of the given data. The primary goal of the analysis is to get insight into both structured and unstructured forms of the given set [26]. The world is flooded with data and the growth level increases every day. Data analysis is the most powerful multi-disciplinary field that involves statistics, mathematics, computer science and deep learning for solving real-time problems in an efficient manner. The advancement in computer technology and the internet empowered the world with ease of accessibility and low-cost storage of data.

1.2.1 Features of Data

The current work is on data-driven modeling and hence the characteristics of the data have to be studied for utilizing them to the relevant application. Big data is a multi-variant collection of combinations of different sets from various sources. There are about 17 (V's) attributes identified so far to describe the internal feature of the data set. These extended findings will solve many problems and interrogations about Big Data Analytics.

I. Since the evolution of Big Data (1995), many attributes are added to evaluate the accountability of their usefulness. Among them, the investigation carried by Douglas Laney [27] recommended the 3 V's – Volume, Velocity, and Variety. This work is developed in 2004 with Gartner and is popularly known as Gartner's Interpretation.

1. Volume - Size of the data available for the analysis (Bytes)
2. Velocity - Speed of accessibility of the data
3. Variety - Type of the data: structured, unstructured, numeric, image, video, audio

II. In addition to the Douglas Laney's 3 Vs, data scientists of IBM introduced another attribute known as Veracity - the 4^{th} V.

4. Veracity - Indicates uncertainty of the captured data

III. For enhancing the data analysis business, Microsoft stretched the attributes to 6 Vs, they are variability and visibility.

5. Variability - Multi-dimensionality or complexity of the dataset
6. Visibility - Inclusiveness of the minute details of the data

IV. The sheer size along with the complexity of analysis, Oracle - in its study [28] included one more attribute named value.

7. Value - Represents the commercial importance derived from the given source

V. Further analysis of the big data challenges, invoked Kirk Borne [29] to define the feature in 10 V's.

8. Venue - Source and platform of the data
9. Vocabulary - Terminology used for describing the data models
10. Vagueness - Misconceptions of the existing data

VI. The world is operated with more sophisticated levels with the invent of more specifics of the datasets, in 2017 a formal study by a group of elites [30], discovered 7 more Vs to the existing list, making it 17.

11. Validity - Authenticity of the data
12. Volatility - Usefulness period of the stored data
13. Virality - Rate at which the data spreads among the users
14. Viscosity - Signifies the lag of the described event
15. Verbosity - Redundancy of the available information
16. Voluntariness - Relevant availability of the data according context
17. Versatility - Flexibility of the data to be used at various platforms with different functionalities.

For solving every distinct problem, the characteristics of the captured dataset have to be studied prudently with all the mentioned V's to predict better outcomes.

1.2.2 Data Mining

The term data-mining was coined in the 1990s previously used as 'data fishing', it is a process of obtaining useful features form the data to describe the unknown patterns, trends, and functionality of the system. This procedure is more prevalent after the advent of big data because the size of the data is larger, and information is more varied in its content. In the process of data mining, the data has to be sampled and transformed for the model to evaluate the outcome of the problem [1].

Deep learning is a subset of machine learning that works based on neural network algorithms. With the quick evolution of technology, machines are capable of replacing human intelligence with ease and accuracy. The multilayered network learns from the complex, unstructured, diverse data and solves for the reliable solution [31]. Deep learning is applied in a variety of businesses, banking, manufacturing, education, stock markets, retails, etc.

1.2.3 Machine Learning

Machine learning came to existence after the research study of Arthur Samuel (1959) and Tom M. Mitchell (1999). It is the field where the computer is trained to learn without a programming platform [32]. The machine learning methods are classified based on the type of learning; supervised (trained with inputs and desired outputs), unsupervised (without the output label) and nature of the output; binary classifiable outputs and regressive outputs.

The accuracy or correctness of a machine learning approach is validated based on the error of the untrained data. Based on the nature of the tasks, there are different types of machine learning algorithms used to accomplish the problem. Some commonly used methods are Linear/Logistic Regression, Least Squares Regression, Support Vector Machine (SVM), Decision Trees (supervised learning approach), Navie Bayes classification (probabilistic classifier based on Baye's theorem), K-means clustering, Nearest-neighbor mapping, Artificial Neural Networks, etc. [33].

A model based on any of the above algorithm is created to learn a specific feature form the data. Later a targeted variable is predicted based on the training given to the model. The present world uses machine learning methods to improve the decisions and productivity in the business and forecasting and detecting faults in the industries. Better tools and self-adaptive algorithms have to be framed with the exponential growth of technology to solve complex real-time problems. Machine learning is broadly applied in many fields such as fraud/anomaly detection, Image prediction, sentiment analysis (twitter opinion), face recognition, natural language processing, virus/spam detection, etc.

1.2.4 Artificial Intelligence

Artificial intelligence is the modern approach in science and technology and started gaining importance after the II world war, and the names were devised in 1956. For the AI algorithm to think humanly, the human brain has to be studied intensely. The insides of the actual working of the mind have to be introspected through psychological experiments.

Cognitive science is the interdisciplinary field that binds these human behavioral techniques together with the computer models.

AI is a thriving field that covers a variety of intellectual task-related fields such as face recognition, natural language processing, proving mathematical relations, sentence predictions, machine/robotic operations, diagnosing medical conditions, etc. [34]. This field has gained unbelievable growth in the past decades with a very high ability to handle and calculating complex problems. Artificial intelligence has to outdo the performance of humans to solve multiple complicated tasks. AI is the enclosure of deep learning and machine learning concepts together as shown in Figure 1.12.

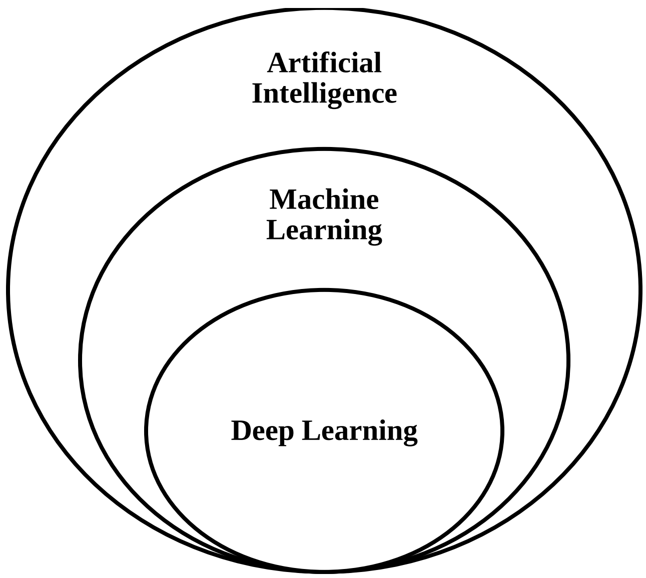

Figure 1.12 Pictorial Representation of AI Concept [34]

AI should act and think realistically and rationally, according to Turning Test [35], the computer should be capable of the following things,

I. Communicate successfully in natural language
II. To store the obtained data
III. Proper reasoning with new conclusions
IV. Should adapt, detect and extrapolate new trends
V. Needs a computer vision to perceive objects
VI. Manipulator to move the objects in the simulated direction

The different categories of AI algorithms include a. Reactive Machines (reactive but failed to use past data), b. limited Memory AI (uses static data for the process- self-driving), c. Theory of Mind (understanding the emotions of the human), d. Self-aware AI (supplement of theory of mind AI, configure about themselves), e. Artificial Narrow Intelligence (Weak AI: technology used in smart devices), f. Artificial General Intelligence (Strong AI: robotic technology), g. Artificial Superhuman Intelligence (Powerful AI: humanoid robot).

1.3 Limitations of the Approach

a. **Attributes of Data** - The measured data can be messy, inappropriate or unnecessary for the analysis. Attributes that are discussed in the previous section may not exist in most of the measurements and values thus making it more difficult for the application.
b. **External Factors -** The output of the model based on the training set containing history of events (faults and trends). Future predictions exclusively depend on the past source of data. Interaction of external factors, changes in the current situation, and unknown factors will affect the predictability of the network. The severity of these aspects will make the model to a complete unfeasible form.
c. **Tuning of Hyperparameters** - All the neural network and other mathematical models has its own specific characteristics. Selecting the effective combination of the various parametric functions manually by trial and error is a difficult and demanding task.
d. **Noisy Data** - Noise is a meaningless feature irrelevant in explaining the relationship between source and target. Noise in the tabular data are of three types; irregularities in the recorded item, unwanted features, records that won't follow the flow. This has to be removed from the data to improve accuracy. Filtering techniques, wrapper methods like recursive and backward elimination, embedded regularization methods are applied to remove these irrelevant functionalities from the dataset.
e. **Overfitting and Underfitting -** This is the most common issue in statistical fitting based problems. Overfitting occurs when a model learns unwanted features to an

extent that impacts the performance in a negative way. Underfitting is the case in which the model is incapable of capturing the underlying trends of the training data.

f. **Incomplete Data** - It is very difficult to obtain the complete set of data with all the dimensional features of the system. Data is mostly insufficient and less correspondent between the events and facts. Ethical issues like privacy, ownership, liabilities of the data are the major matter of concern.

1.4 Literature Review

Artificial Intelligence and Predictive Analytics for Aerospace Systems (AI-PAAS) laboratory and Space Systems Dynamics and Control Laboratory in the Department of Aerospace Engineering at Ryerson University directed by Dr. Krishna Dev Kumar carried out majority of the existing research in the field of fault detection, isolation and prognosis through model-based methods [4][11] and knowledge-based data-driven approaches [8]. An extensive literature review for model development, future forecasting, fault diagnosis, fault prognosis, and remaining useful life prediction along with optimization is detailly discussed in each of the below-mentioned sections.

1.4.1 Model Development

In the earlier days, the behavior of any system is studied by simulating its small-scaled model. The model-based approaches require the complete details about the plant and the outcomes of the system have to be mathematically formulated to estimate the actual output and to detect or isolate the existing faults and malfunctions. A survey on model-based methods in the complex plants using statistical testing, signature analysis is carried [36]. A technical process with conventional limit and trend checks is studied with various examples [37]. A review of process engineering using a basic quantitative model-based method is available in the paper [38]. A general framework on dynamic systems of model-based FDI techniques are structured with residual estimations [39]. To highlight the characteristics of aerospace systems an extensive health monitoring investigation is performed to detect sensor and actuator faults [40]. To avoid the potential hazards and detect the dangerous faults a thorough investigation with model and signal based methodologies [3].

In the digital era, problems in the field of life sciences, finance, commerce, engineering, and technology are resolved efficiently by data-driven methods. The data mining techniques become popular since the 1960s and gained momentum after the innovations in computer systems after 1995. Numerous software tools were developed to incorporate the capability of the analysis to solve real-time problems [41]. A comprehensive review of data mining techniques from 2000-2017 is extensively studied [42][43].

Classical data-driven methods such as visualization, correlation, and discriminant analysis are initially applied to the medical field to extract information about the individual disease database [44]. To learn the user pattern of web page usage, an elaborate research effort using the WebSIFT system is trained and modeled [45]. Urban planning for different regions in Munich is proposed by extracting the entire road database [46]. The relationship between switch hitch and weather factor, temperature and electrical nets is exercised using statistical techniques. The crime scene investigation without the ability of actual samples is monitored and forensic samples have been recovered with best yielding results using the obtained data [47]. A vital study on voting patterns in the United States has been interpreted successfully using the online data based on the comments and opinion polls. Some mathematical techniques including association rule mining, t-weight calculations, and decision tree analysis are used for the outline study [48]. Similarly, accident investigation based on the different conditions is simulated with factor analysis and actions to avoid certain known situations are proposed [49]. A broad review of financial fraud detection [50] and cybersecurity intrusion detection [51] is executed. The Medical field is advanced with data mining techniques (Cross Industry Standard Process for data mining) to determine the factors related to chronic kidney disease [52] and cholera related mortality is reduced by studying combined satellite data with local environment and climate along with the field data [53].

The modern data-driven approach incorporates machine learning algorithms with neural network topology. From the public source data, every general analysis is shifted towards artificial neural network models. Electric load forecasting [54], rainfall estimation [55] and stock price predictions are successfully studied using artificial neural networks. Recurrent neural networks are powerful in solving time series problems [56] and certain long term dependency issues in the time-related data are reconsidered and a modified RNN known as LSTM [57] is currently applied with certain modifications.

1.4.2 Fault Diagnosis

Early fault detection and isolation [58] from the plant helps in avoiding anomalous event progression and improves reliability and safety [59]. The conventional methods use state parameters and characteristics data to process the detection [18]. The majority of the existing research in the field of fault diagnosis is performed through model-based methods and data-driven approaches.

The model-based fault detection methods [40] can be both stochastic and deterministic. Fault identification is achieved using filtering schemes in the stochastic approach [11] and deterministic methods use observers and parity relations [3]. A model-based diagnosis, prognosis, and health monitoring (DPHM) framework was developed and evaluated using a new fault detection algorithm, using Unscented Kalman filters (UKF) in conjunction with residual and innovation sequences [60]. Sliding mode observers estimate the state vector and a signal-residual is generated from the estimated output to detect the fault [61].

Classical data-driven methods for fault diagnosis especially for complex systems such as satellite component failure and aerospace-related matters are investigated extensively. Friction data methodology [62] was developed for monitoring potential failure of the reaction wheel of the Globalstar 2^{nd} generation satellite. A time algorithm with envelope learning and monitoring with limit sensing trends and adaptive limit-checking using regression tree learning [63] of the telemetry data from spacecraft is applied to detect the abnormal activities.

Modern data-driven methods applied to-date are very effective in detecting irregular behavior without knowing the apriori knowledge of the system. In this research work of time series examples, the main focus on diagnosis methods is using ANNs. FDI approaches for dynamic non-linear systems [64] [65] using AI techniques in comparison with model-based methods are surveyed. A dynamic neural network approach [66] trains the model torque command and voltage output to detect the RW failure. A similar work with input-output based ANN [67] for RADARSAT actuator fault is studied with sensor data.

Based on the literature review on model development, the latest and suitable fault diagnosis on time series systems will be LSTM methods. The recent works on LSTM include usage of stacked LSTM networks for fault detection on space systems, multi-sensor engine dataset and an exclusive Gated Neural Networks (GRU) for identifying the events reducing safety margins of flight operations. Invariant of the applications the latest common security hacking of the modern automobiles [68] and valuable resources of government websites are precisely detected using LSTM neural networks. The author has referred the latest spacecraft pilot anomalies detection algorithm (LSTM) with nonparametric dynamic thresholding with proper false mitigation strategies to build [69] the FDI model for this work.

1.4.3 Fault Prognosis

Fault prognosis is one of the top ten challenges in the aviation safety program. It is the process of assessing the forthcoming status of the system and estimating the remaining useful life of the degrading components. Similar to the diagnosis the prognosis methods are reviewed with model-based and data-driven approaches to determine the finest possible technique. Model-based requires mathematical equations explaining degradation pattern, a scientific approach on actuator dynamics of the aircraft [70] and a predictive approach for bearing analysis [71] is considered. The physics-based methods are always on top of the pyramid when compared to other methods due to its law of nature and are always precise, accurate, easy to validation and verification. After establishing the system model with all features, it can be applied for determining the remaining useful life; the gearbox prognostic module [70], residual-based failure in dynamic hydraulic systems and other works related to CMG system spin motor failure prognosis in the satellite, wiener degradation based life estimation algorithm [72] for small satellites were determined.

In spite of the advantages, the classical methods are time-consuming, expensive and intensive Hence conventional numerical methods are widely used due to its speedy implementation. Regression analysis, Bayesian approaches [73], support vector regression [74] are the most used algorithms for remaining useful life predictions. Some hybrid models [75] [76] combining the advantages of both the above-discussed techniques are applied for prognostic development.

Machine learning data-driven for prognostic problems become renowned after the 2008 PHM (PHM08) conference data challenge competition for complex engineering systems. These approaches are based on uncertainty and approximation; hence a robust algorithm has to be proposed for the unstructured, multivariant and noisy data. The generic methodology for structural health prognostics is introduced and parallelly multiple algorithms to achieve better performance are also added. Advanced RNNs with extended Kalman filter evolutionary algorithms, Hidden Markov Models and LSTMs, and deep convolution neural networks [77] are examined to estimate RUL. Recently, semi-supervised deep architecture for turbofan engine degradation [78] and battery cycle life capacity degradation are effectively implemented. In this work, the LSTM model with optimization techniques is proposed for RUL predictions.

1.4.4 Optimization Techniques

The suggested model in the present work for fault diagnosis and prognosis uses LSTM networks for analysis. There are several hyperparameters in the artificial neural network models which have to be fine-tuned befittingly to get convergence with the high level of accuracy. The authoritative articles discussing the optimization techniques are studied to decide the suitable scheme. Rahimi et al. [79] presented a methodology using adaptive filter (Particle Swarm Optimization) for detecting the fault in the reaction wheels onboard satellites with greater performance. Likewise, numerous studies with PSO over backpropagation for length of stay, non-linear channel equalization [80], in ANN and tuning of emotion recognition framework, stock forecast in LSTM resulted in superior predictions. Genetic Algorithm-optimized models [81] have also become popular for diagnostic and prognostic decision-making frameworks and hence both (PSO and GA) the optimization procedures are discoursed, outlined and compared at the end of this thesis.

1.5 Problem Statement

In the current research work, a superior data-driven model is developed, then the classical approach is refurbished with improvements and finally, optimization techniques are applied to the intended model. Sketching from the previous section on literature review, the following problems can be framed,

[PROBLEM 1] To the author's knowledge, the majority of the prevailing work on health management was model-based [4] which requires apriori knowledge of the system. Later, several modern data-driven methods are proposed [43] with neural network topology, but they are incompetent with overfitting and underfitting [31]. Also, those models require a complete historic dataset [55], for the seamless predictions.

[PROBLEM 2] From the published work, the classical data-driven methods are simple and widely used in varied fields from urban planning [46], medical [44], to crime investigations [47]. Apart from the measurement dataset, conventional methods such as the hidden Markov model, autoregressive integrated moving average model also requires extensive parameters of the system.

[PROBLEM 3] Based on the review, the available literature on the fault detection algorithms are limited by (1) use of simple dynamic systems without considering multi-level integrated units [67], (2) the detectivity of the model is limited by its pre-historic data and is ineffective with external feature interaction or when a different scenario occurs build [69].

[PROBLEM 4] To the best of the author's knowledge, there is no considerable literature available on model-based prognostic works with auto-tuning of the hyperparameters. Time series evolutionary algorithms consume time in choosing the appropriate values with the combinational sets for training and validation of the model.

1.6 Research Objectives

To resolve the identified problems, the subsequent objectives were established for the research work,

[OBJECTIVE 1] To address problem 1, a modern recurrent neural network model that is capable of remembering the long term dependencies has to be framed. The anticipated framework has to be (1) proficient with the available data (minimal information), (2) update the weights based on the significance of the input, (3) avoid the overfitting and underfitting curve fitting issues.

[OBJECTIVE 2] To tackle the problem 2, the statistical data-driven methods need to be refurbished with a new capability. Therefore, the featuring properties of the desired conventional methods have to be extracted and merged together into a better algorithm to outperform the other classical methods. The other most important characteristic is the ability to develop the model without expert knowledge about any system. The intended framework has to be easily accessible and executable to any user irrespective of the complexity of the field of application.

[OBJECTIVE 3] To encounter problem 3, an algorithm for handling all abnormal scenarios including transient, intermittent and incipient faults needs to be formulated. The new fault detection approach should (1) be intelligent in perceiving the behavioral pattern of dynamic nonlinear systems, (2) be able to predict the future trends well in advance without the complete history of data, (3) be adaptable to the changes incorporated in the system by any external means as new scenario or disturbances.

[OBJECTIVE 4] To overcome the problem 4, the neural network model has to be extended with the capability estimating the remaining useful life. Optimization techniques have to be integrated into the base model for improving the computational time and probability in selecting the finest parameters. Minimum two optimization tools need to be configured for the comparative study of their performance metrics to suit the assigned problem.

1.7 Main Contributions

The contributions for the formulated problems and the research objectives are explained as follows,

[CONTRIBUTION 1] A recurrent neural network (LSTM) model to predict future trends with minimal data requirement and adaptive adjustments of weights is developed. Multi-step predictions of standard pollution particulate matter ($PM_{2.5}$) for three Canadian cites ahead a month using a sliding window technique with supervised learning (different combinations of hyperparameters) is achieved. Dickey-Fuller test is applied to the data for removing the stationarity time differencing problems.

[CONTRIBUTION 2] An improvised statistical framework is formulated with the combinations of three significant methods. The predominant factors are extracted from the correlation functions, then the shape and scale factors are calculated with Weibull analysis, later the intercepts from the torque friction and rotor speed are utilized to detect the fault of the Kepler mission Reaction Wheel failures (RW2 and RW4).

[CONTRIBUTION 3] The LSTM model with elementary design requirements is modified to the Kepler dataset with the improvement of early detection capability. With the minimal data, the network is tuned and trained effectively to capture the outline of the telemetry data. The overfitting and underfitting issues are prevented with appropriate selection of regularizers and dropout values.

[CONTRIBUTION 4] An improvised prognostic model for estimating the remaining useful life of the aircraft turbofan engine is designed. The proposed recurrent neural network model is integrated with a genetic algorithm and particle swarm technique separately and compared with the best combinations based on the results extracted from merged parameters (PCA) and digital filters (Savitzky – Golay filter).

1.8 Thesis Outline

This thesis is organized as follows: in Chapter 2, terminologies and requirements of fault management system are discussed. Also, a comprehensive analysis of both conventional approaches and data-driven modern methods are explained. LSTM model is developed and validated for pollution dataset with single and multi-step sliding window prediction capability. In Chapter 3, the integrated and combined statistical techniques for fault detection of Kepler spacecraft reaction wheel scenarios are outlined. In the second part of Chapter 3, the enhanced statistical technique is validated using the LSTM model with simulations and case studies using coronal discharge (space weather data). In Chapter 4, the prognosis algorithm is derived and the remaining useful life of the same fleet of a turbofan engine is estimated for all the given operational settings. This prognostic model is enhanced with heuristic optimization tools such as genetic algorithm and particle swarm optimization to improve the results. In Chapter 5, the highlights, summary of contributions and results of the work are described, and the scope of future work is specified.

CHAPTER 2

2. Data-Driven Approach

The most important issues of the mechanical system are performance, reliability, and safety. In order to avoid minor issues or failures, these parameters need to be studied and monitored completely at every instant in time. Every approach has its own effectiveness and shortcomings in detecting and resolving the anomalies. Since the early days, the failure rate of the complex systems is declining steadily. The terrestrial systems are frequently tested for various modes of failure and are designed with a high level of precision. The faulty components can be serviced and will be recovered for its complete operation. On the other hand, airborne vehicles and space systems are hard to simulate and extremely vulnerable to unknown external disturbances. Based on the literature review all the fault detection methods are studied and explained. In this Chapter, a suitable approach with a high predictability factor is identified for the most complex systems.

Several fault prognosis and diagnosis methods available in the field successfully mitigates the issue and helps in monitoring the health of the system throughout its operating cycle. As a treated field from anomalies, it is important to describe the terminology of the factors [82] that are required for the health monitoring analysis,

Fault:

- It is defined as an unpermitted deviation of at least one characteristic property from the acceptable, standard condition.
- The smaller and hidden faults are difficult to detect.
- It can be related to manufacturing, design, assembly, maintenance or operators.
- It can be minor and may not affect the functionally of the system.
- It may cause the system to malfunction or collapse.

Malfunction:

- The intermittent irregularity in the fulfillment of the system's anticipated function is termed as malfunction.
- It is a temporary disruption of a regular operation.
- One or more faults result in a malfunction.
- It occurs after a prolonged stressing of the system.

Failure:

- Failure is defined as permanent obtrusion which breaks the ability to perform a desired process of operation.
- One or more faults result in failure.
- It also occurs after a prolonged stressing of the system.
- Failure can be systematic, random, deterministic or causal based on predictability.

2.1 Desired Requirements of a Fault Management System

An ideal fault diagnosis model as shown in Figure 2.1 should satisfy certain requirements in order to accomplish the desired task with high accuracy.

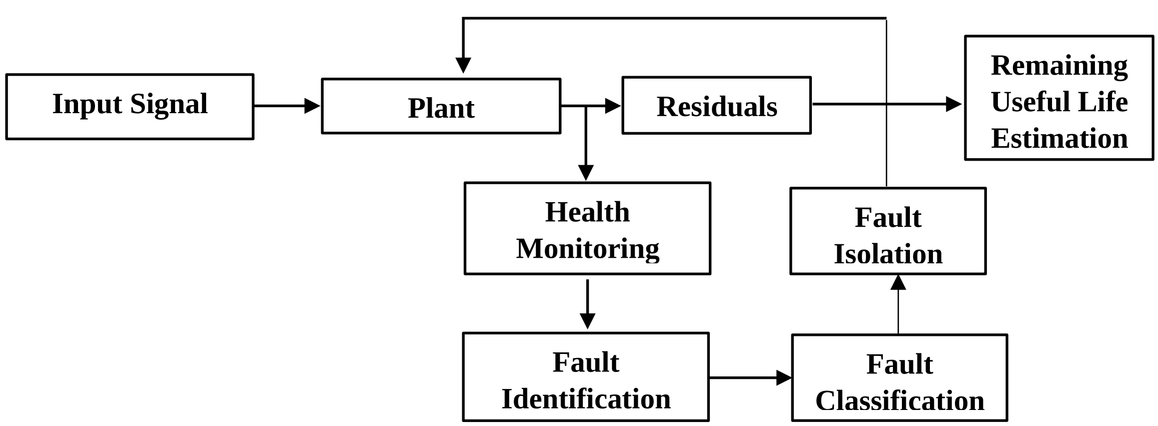

Figure 2.1 Schematic of Fault Diagnosis and Prognosis Model [8]

Some of the important general attributes of the DPHM module are listed below [83],

a) **Early Detection and Isolation:** For the system to function normally, it has to be uninterpreted and flawless. In the case of critical-systems, even a small initiation of the fault may propagate and breaks the mechanism. Hence the fault-tolerant module should be highly sensitive in identifying and eliminating the fault in the initial stage. The system should also be capable of avoiding false warnings.

b) **Fault Identifiability:** Recognizing the fault is most challenging to accomplish due to the presence of uncertainties in the model, external disturbances and turbulences, measurement noise and most importantly the coupled interactions between the potential fault sources. The type and severity of the fault has to be estimated.

c) **Unknown Fault Identifiability:** Detecting the unfamiliar fault is relatively easier than isolating and identifying them. Due to their irregularities and uncertainties, the new faults cannot be modeled. The most renowned failure analysis tools such as FMEA (failure mode and effects analysis) and FMECA (failure mode, effects, and criticality analysis) which are used in industry levels known for its accurate predictions, can even fail to identify these unacquainted malfunctions.

d) **Isolability:** The identified fault that occurred in the system has to be isolated from the cause, so distinguishing the roots of the fault from other potential sources is the next important capability of the tolerant-control system. Isolability is absolutely essential since the necessary counter-actions cannot be signaled without the source of the glitch. This attribute is also affected by the uncertainties discussed in the previous part.

e) **Observability:** This is the factor of how well the details of the fault can be inferred from the outputs and the known system parameters.

f) **Robustness:** It augments the reliability, consistency, and effectiveness of the health monitoring system. Uncertainties are unavoidable in the practical applications, thus robustness in every measurements and modeling is the utmost important attribute for the designed system.

g) **Multiple Fault Identification:** Even a simple system or a model can exhibit multiple faults, so the proposed diagnostic system should be capable of identifying and classifying the faults. Incorporating this attribute is problematic due to couplings and nonlinearities that exist between the source and the states.

h) **Fault Detailing:** The fault management system should be capable of identifying the exact location, reason, and explanation of how the fault affected the functionality of the operation.

i) **Adaptability:** Each system is exceptional and unique based on its application and works in a different operating condition. System and its components are always designed to function in its permissible range of vibrations, temperature fluctuations, stresses, and other external disturbances. Likewise, the fault-tolerant module has to adapt to the system's operating state in order to maintain competence.

j) **Computational Requirements:** For the real-time application, the fault detection module has to be computationally compatible, memory sufficient, low power consumable. It has to have an exceptional specific functionality for each model designed.

2.2 Fault Detection/Isolation Methods

The fault diagnosis algorithms are modeled with the above-mentioned characteristics for effectively finding the magnitude and locating the faults at the component level. Fault can be from the sensors, actuators or even the processes. The proposed algorithm should be capable of notifying every single detail of the irregularity such as proportions, location, severity, type and time of detection.

There exist numerous fault detection methods starting from the invent of mechanical systems. A comprehensive analysis of both conventional approaches and data-driven modern methods is explained with its advantages and limitations. For aircraft and spacecraft, the severity and consequences are very much higher, due to their direct connection with human safety, constraints of time and financial budgets.

2.2.1 Conventional Methods

The conventional methods require apriori knowledge about the system that is dealt with. Large scale complex systems such as space stations, satellites, launchers, aircraft require the highest safety. Therefore, these systems are cautiously designed with an uppermost factor of safety with a fault diagnosis module as a part of the unit. In this section, the different types of fault-diagnosis methods used from earlier days to the present world are discussed. They are broadly classified as model-based, signal-based and knowledge-based methods.

2.2.1.1 Model-Based Methods

Model-based methods use the mathematical dynamic equations and generate a computational model of the system with their actual features, then the output is simulated and compared with the target system to detect irregularities and reason of cause about the anomalies. The complete parameters and structure have to be precisely known about the system. In the Mid ’90s, Groundbreaking work in the model-based approach was developed for NASA’s DeepSpaceOne (DS-1) mission. Most of the earlier methods were based on reasoning and quantitative models.

I. Detection by Parity Relations (Deterministic Approach)

The key idea of this method is to generate and evaluate the residuals. These residuals are compared between the real and processed model to describe the behavior. If the system is functioning normally, then it is said to be non-faulty. The consistency [84] of the mathematical equations by using the actual values is checked for any signal differences as shown in Figure 2.2.

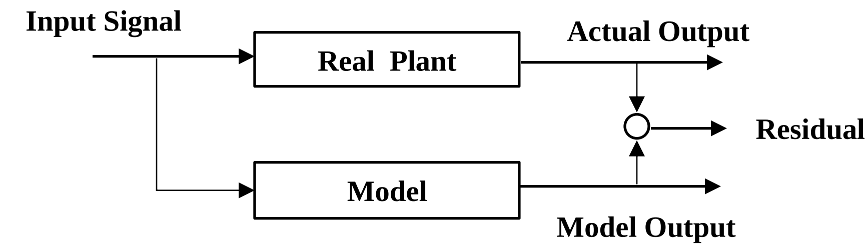

Figure 2.2 Parity Relations Approach [84]

II. Detection by State observers/ Estimators (Deterministic Approach)

If the system is not completely controllable or directly observable, then the state feedback cannot be obtained. In such cases, a new subsystem known as an observer/estimator is designed to duplicate the state vectors. An observer computes the states of the dynamic system based on the model outputs. Changes in the input or the output lead to the variations in the state variables. The function of the observer is to correct the output using the estimation error as a residual for the detection of the fault. The general schematic for the feedback control system with an observer is as shown in Figure 2.3.

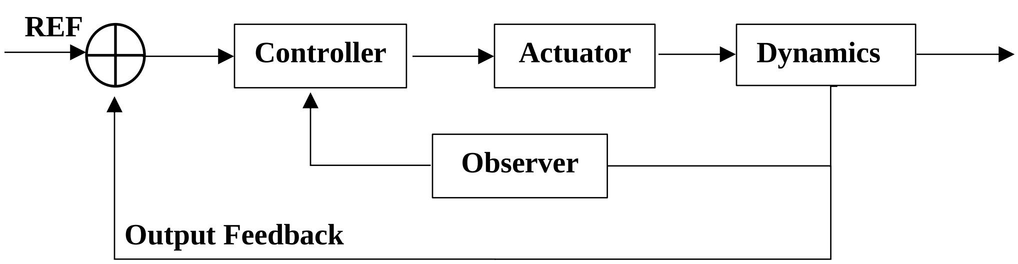

Figure 2.3 Schematic of the Output Feedback Controller with an Observer [11]

2.2.1.2 Signal Based Methods

Signal-based methods work on the principle of processing the measurement signals, it does not explicitly require the details about the system. The output signal is analyzed and changes in them are related to the fault in the process. Either time-domain or frequency-domain is used for processing the amplitudes, phases, and spectrum of the signals. Vibrational signal measurements use the acoustic-based frequency domain for the diagnosis and time-domain methods are mostly used in electric motor applications [82].

I. Spectrum Analysis

The faulty sceneries within a certain bandwidth of the signal in the spectrum are restricted to the amplitude densities. The frequency of the signal (Eqn 2.1) is generated using the Fast Fourier Transform (FFT) algorithm. The amplitude (Amp) signal will have pre-defined limit values for a normal operational condition. When the amplitude crosses the restricted range (Eqn 2.2), the proposed model will sense the fault and initiate the corrective actions.

$$x(t) = Amp_o + \sum_{i=1}^{N} Amp_i \sin \omega t \tag{2.1}$$

$$Amp_{min} \leq |Amp_i| \leq Amp_{max} \tag{2.2}$$

where x(t) is the input signal, Amp_o is the initial amplitude and ω is the frequency.

II. Parametric Model (ARMA)

The autoregressive-moving-average model is a type of parametric signal approach used mainly for the time-series datasets. The model describes the process in terms of two forms, one for regression and the other for the error term. These models are sensitive to very small changes in frequency. ARMA(p,q) model is defined in Eqn (2.3) with p as the order of AR term and q as the order of MA term.

$$x(t) = c + \sigma_t + \sum_{i=1}^{p} \beta_i X_{t-1} + \sum_{i=1}^{q} \alpha_i \sigma_{t-1} + \vartheta_t \tag{2.3}$$

with parameters β_i and α_i, c is the model constant, σ_t is the random variable and ϑ_t is noise.

2.2.1.3 Knowledge-Based Methods

The trend of fault diagnosis methods moved towards knowledge-based algorithms (also known as rule-based) to solve complex systems. Qualitative methods work on the basis of a certain set of rules developed for the system using historic data and prior understanding. While the quantitative knowledge methods are based on both statistical and non-statistical analysis approaches [85].

I. Expert Systems

In the earlier days, this method is considered as the most significant accomplishment of artificial intelligence. Most of the research work in space systems developed in the early 90s is exclusively based on the invent of this scheme. With a high level of experience and expertise about the system, some common set of rules about the cause and effects of the anomalies and their relationship between the warnings and failures is vigilantly established. The advantages of this scheme are that the rules can be easily amended, changed or removed with suitable explanations, it outperforms the model-based methods in terms of power and capability. The most significant issues of this technique are that the physical properties are either not fully available or costlier to acquire, inability to detect unknown phenomenon, inefficient inconsistency, incapability to learn from their errors.

II. Fuzzy Logic

The Fuzzy logic theory had been studied since the 1920s and was later termed by Lotfi Zadeh in 1965 while proposing the fuzzy set theory. Based on the available information and existing data the fuzzy logic controller is designed to detect the anomalies. Boolean logic only has two variables either true (1) / fault or false (0) / no-fault, in contrast, the fuzzy logic has many variables between 0 and 1 representing the range of intensity of the faults [86].

Fuzzification, Rule execution, Defuzzification are the major steps involved in the Fuzzy logic controller. Fuzzification is the process of tagging the input variables with random values between 0 and 1 with membership functions. The variables with 0s and near 0s are not a part of the fuzzy set, on the other hand, the variables with 1s and near 1s are considered to be in the fuzzy set. Each value between the interval [0,1] has a significant degree of certainty. Secondly, based on the rules and reasonings the process is evaluated and executed for computing the output. Finally, defuzzification - converting the fuzzy values to its original form is carried out to decide the state of the system based on the prearranged conditions [87].

2.2.2 Data-Driven Methods

All the conventional fault detection methods discussed earlier requires the detailed historic knowledge about the model. In other words, they are based on empirical reasoning processes for future predictions. On the other hand, fault diagnosis methods based on data-driven analysis only requires past telemetry data with all possible combinations of operational conditions. They are popularly known as Machine learning or Data Mining approaches and has recently gained popularity and drawn attention in almost every field. With the advancements in internet and technology, acquiring, handling or storing data has become much easier and simpler in the present world. Many advanced algorithms have been developed with a large volume of available data and high computational power. Data is considered to be a critical asset to the business world, the growth in each sector has made a leap forward with the highest advancement in retails and location-based services.

New challenging and attractive data-driven solutions are introduced every day and are changing the insights of the markets. The important challenge in big data analytics is extracting meaningful features from both physical and digital environments. Successful utilization of data requires five elements as shown in Figure 2.4.

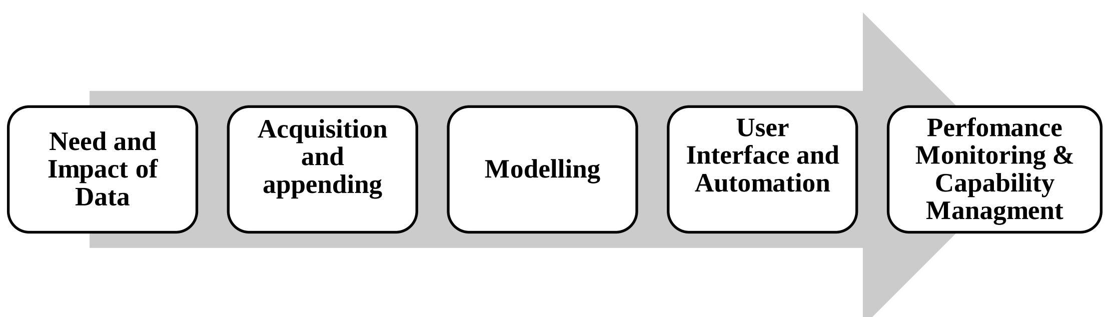

Figure 2.4 Transformation Process of Data and Analytics [34]

To demonstrate the detailing about the process or model, an appropriate statistical measure has to be framed using the generated data. The competence of these data-driven methods depends on the objectives and independent parameters required to describe the system. The characteristics of the data such as mean, variance, standard deviation remains the same unless a fault occurs in the model [88].

The measured data has to be pretreated to remove unwanted features, signal noise, and discontinuities. Further, it has to be scaled and reshaped for input to the model to perform regressive training. Lastly, the trained model is tested for fault, classification or future prediction applications for the real world.

2.2.2.1 Classical Data-Driven Approaches

In this section, the conventional methods formulated in the early days of data analysis are discussed in detail. These standard procedures are integrated with modern algorithms for improving redundancy and reliability. Numerous classical fault detection approaches were applied across different anomaly types, application domains, and various data types. Some of them are as discussed below,

I. Limit Checking

Limit checking is the most fundamental and widely used fault detecting method. The nominal range for each sensor value is determined by engineers and designers. The telemetry data received from the system is monitored continuously for its nominal values with a specified upper and lower limit. When the signal deviates from the pre-determined threshold value then the component is considered to be defective or malfunctioning as represented in Figure 2.5 and its application is discussed in Chapter 3.

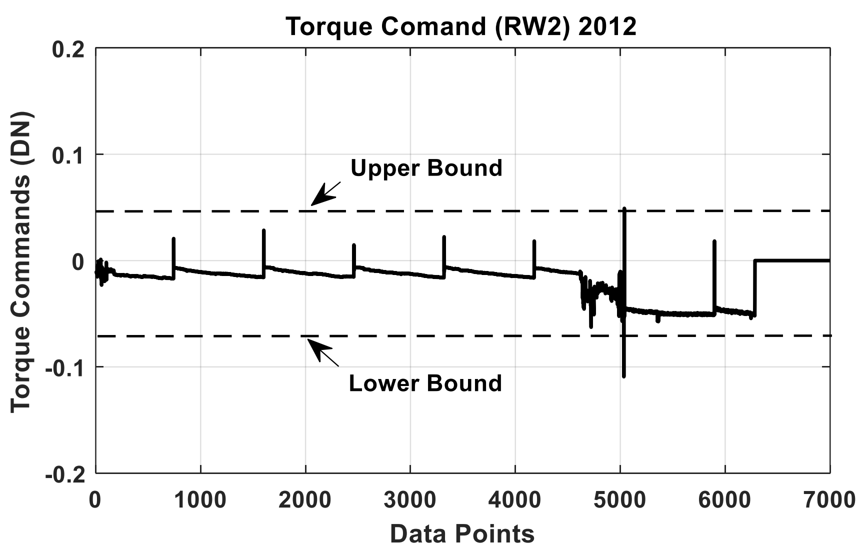

Figure 2.5 Threshold Limits for Spacecraft Telemetry Data

This method is still used in many complex systems due to its significant features. One is simplicity - the ease of implementing; second is the success rate of detecting the fault and the third factor is that the limits values can be repeatedly adjusted by the technicians based on the local condition. However, this approach fails in detecting the new faults and the types of failure that are associated with various factors other than limit volitions.

II. Principal Component Analysis (PCA)

PCA is mainly applied in the multivariant data field to transform the vector space into a subspace maintaining the maximum variance of the actual space with dimensionality reduction. It is the linear transformation technique to convert the correlated data into an uncorrected form that explains the trend of the process as shown in Figure 2.6 [89].

The nonlinear space systems and aircraft will have a multitude of sensor values, this method helps in mapping the telemetry values of each time segment into a high dimensional feature space by the kernel (polynomial) and detects the fault as the deviation of the principal direction of the data. This method is successfully applied in merging the sensor values of the turbojet engine (Chapter 4) for obtaining the principal direction.

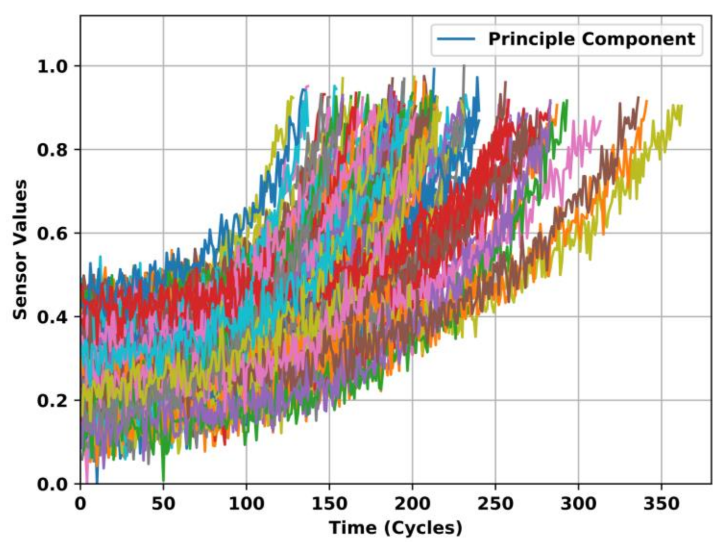

Figure 2.6 PCA Components of the Multiple-Sensor Values of the Turbojet Engine

III. Correlation Analysis

Correlation analysis is a simple formulation to find the strength of the association between the design variables. The correlation coefficient (ranges between -1 and +1) quantifies the direction and strength of the connection. This concept is applied primarily in the Kepler dataset (Chapter 3) for understanding the defect causing factor and to reduce the dimensions. Figure 2.7 shows the positive correlation between the two telemetry values (Rotor Speed and Torque Command). The figure of merit that is mostly used to evaluate the health condition of the wheel is torque command and speed.

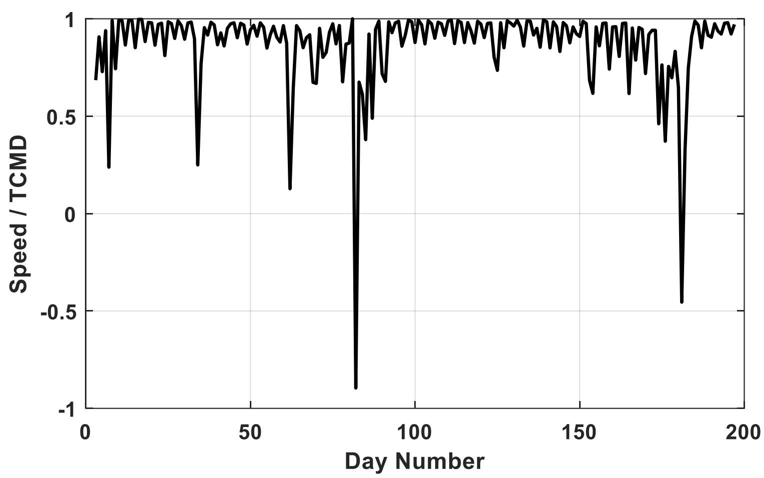

Figure 2.7 Correlation Between Speed and Torque Command – Kepler Mission

IV. Weibull Analysis

The Weibull distribution can be used to predict failure even for small sample sizes. In most cases the two-parameter model is preferable and a three-parameter model is used to describe failure behavior when there is a time period where no failure can occur. Weibull distribution with the two-parameter model is mathematically expressed as,

$$F(t) = 1 - e^{-(t/\eta)\beta} \tag{2.4}$$

The distribution has two factors, scale factor (η) and shape factor (β):

For the shape factor,

If β is less than 1, the sample fails earlier than expected.

If β is equal to 1 shows random failures

If β is greater than 1, specifies the sign of wear out failures

The second parameter η also is known as the scale parameter represents the time at which the specimen will fail, the higher the scale parameter the better the durability. The lifespan of any product can be determined by fitting the statistical distribution to its sampled data unit. The probability density function (pdf) describes the mathematical representation and the model estimates three parameters [90].

After obtaining the parameters, factors like mean life, failure rate, reliable life with confidence bounds are calculated. The Weibull distribution is applied for the same Kepler dataset (Chapter 3) and Figure 2.8 is a plot that represents unreliability versus time.

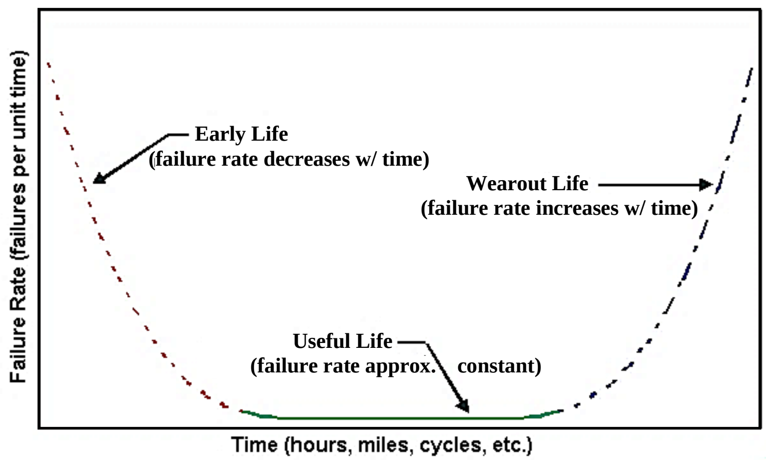

Figure 2.8 Plot Representing Early Life, Useful Life, and Wear out Life [90]

V. Friction Model

The review of failure analysis reveals that bearing-related anomalies are the major cause of actuator failure in the spacecraft. The Frictional analysis of the components based on the bearing friction data will be a major breakthrough for the fault diagnosis zone. This method has the most powerful impact on detecting irregularities [62]. Dry friction, calculated from the total friction-induced torque and rotor speed (both available from the telemetry data) is plotted (Figure 2.9). The intercept and slope values of the plot are analyzed for discrepancies with the nominal friction values (Chapter 3).

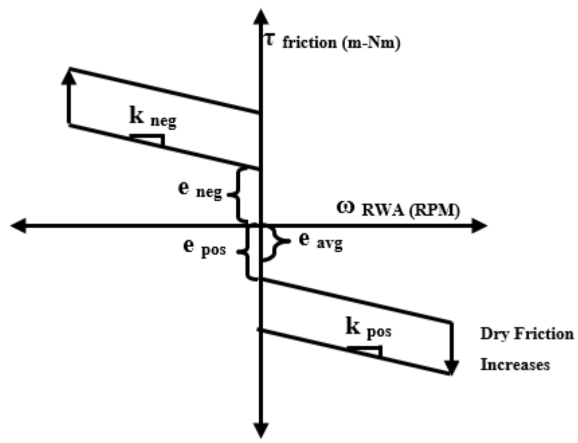

Figure 2.9 Classical Frictional Model [62]

2.2.2.2 Statistical Based Data-Driven Methods

In statistical-based methods the given data set is scrutinized with statistical features such as mean, median, bias, variance, standard deviation, etc., to solve the problem. Statistical learning will identify the risk factors, classifies the variables that cause discrepancies, emphasizes the uncertainty and interpretability of the model. Linear regression is an elementary statistical technique used for predicting the target variable by fitting the linear relationship. Some other famous tools are classification analysis, resampling methods, subset selection predictors, shrinkage analysis, tree-based methods, etc., Two of the wide-spread models are discussed below in detail.

I. Hidden Markov Model (HMM)

The Hierarchical Hidden Markov Model inspects both the observable events and hidden units of the error causing function. The model is embedded with Markov chains that are structured with transition probability state variables [91]. Initially, the probability of the occurrence with all the possible combinations is calculated and the Viterbi decoding algorithm is applied to find the hidden states. Deciding the number of independent features is the initial step for the HMM, then the normal distribution values for each segment is computed. The boundaries and neighboring values are altered based on the z-score. Upon iteration, the performance parameters are estimated.

II. Autoregressive Integrated Moving Average Model (ARIMA)

This model is a popular statistical approach for time series analysis which captures a standard temporal structure in the data. In the algorithm, the dependent association with current and past observation is inspected. Then, the data is converted to a stationary form using differencing techniques to eliminate the trend and seasonality that affects the regression model. Finally, the dependency factor is established from a residual error [92]. The parameters such as the number of lag observations (p), degree of differencing (d) and order of moving average are defined, and the model is trained to forecast the time associated applications. The model has the number of variations VARIMA (Multiple Variables), SARIMA (Seasonality effects), FARIMA (Long-range dependences).

2.2.2.3 Advanced AI Methods

Artificial intelligence methods help the computer to mimic the human intelligence for solving any problem types such as classification, image or speech recognition, prediction or new data generation. It is a well-established cutting edge technology for the scientific investigation. Artificial learning involves a number of techniques or procedures that recognize patterns and relations in huge amounts of complex data. Figure 2.10 characterizes the features of the advanced AI algorithms.

Techniques such as support vector machine, Markov decision models, k-mean clustering are in use for decades. Falling into the latter class are artificial neural networks - a model inspired by the connectivity of neurons in the human brain.

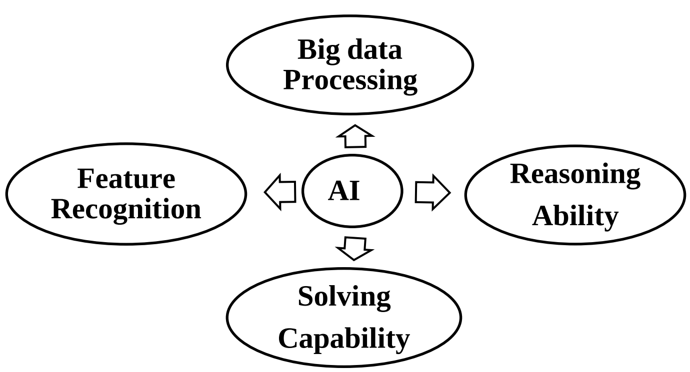

Figure 2.10 Process Involved in Artifical Intelligent Methods [34]

I. Support Vector Machine (SVM)

Support Vector Machine got its name from the data points that are directly on either of the supporting lines. Support Vector Machine is a simple machine learning algorithm that is used mostly in classification based problems. The objective of the algorithm is to locate the hyperplane in N-dimensional space (N distinct features) to classify the data points (Figure 2.11) [93]. For nonlinear multidimensional problems, the kernel-based mathematical tool is used for classifying the data points.

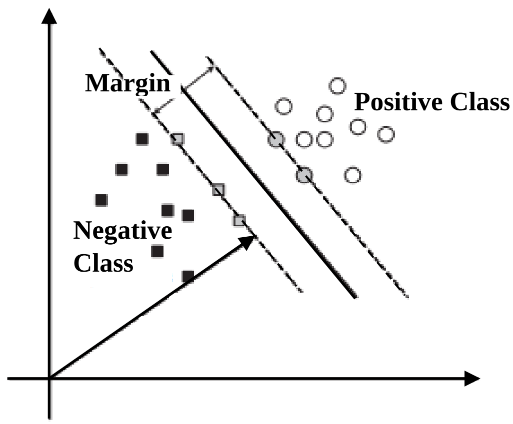

Figure 2.11 Classification Using SVM [93]

The dimension of the classified vector used for training the SVM does not have an influence on the performance and hence it is capable of handling large features and has better generalization properties than the conventional classifiers [93]. SVM will recognize the pattern and classifies the fault of the nonlinear system using kernel Hilbert spaces. SVM algorithm is fully automatic, robust and outperforms with substantial improvements over the other methods. Recently developed techniques such as coordinate descent and sub-gradient in SVM classifiers are more efficient with large and sparse datasets.

II. Markov Decision Process

A Markov Decision Process (MDP) is a modeling technique with the outcome that is partly random and partly based on the input of the decision-maker. The main goal of MDP is to find a policy for a problem and to indicate further required action to normalize the situation. The model contains the following parts,

a. A set of possible world states

b. A set of possible actions

c. Transition probabilities - effects of each action

d. Direction towards planning - reward function

Markov decision process (Figure 2.12) is defined by the five-tuple (x, y, y(.), p, r) where x is the state space, y denotes the action space, y(x) is a set of admissible actions in the state x, p(x,y)(z) is the probability of transitioning from state x to z and r denotes the reward when y action is taken in the state x [94]. In the MDP process, deterministic and stochastic actions have to be represented with a solution and a policy (Л) for mapping sets to actions. Subsequently, the assigned policy has to be followed up for the new state and evaluated for an optimal value. Later, an objective function maps infinite sequences to a single real number representing utility.

Figure 2.12 Markov Decision Process [94]

III. Artificial Neural Networks

Intelligence in the machine can be created by repeated learning and adaptivity. This concept is inspired by the human brain which has a highly interconnected network of neurons for communication. ANNs (Figure 2.13) are modeled by simulating the network of model neurons with weights and activation functions. The output of the feed-forward network is based on the assigned weights and is evaluated by mean squared error metrics. The same steps are repeated for all training data and until a pre-defined optimized value is obtained. Neural networks can be trained to store, process and retrieve information based on past observations.

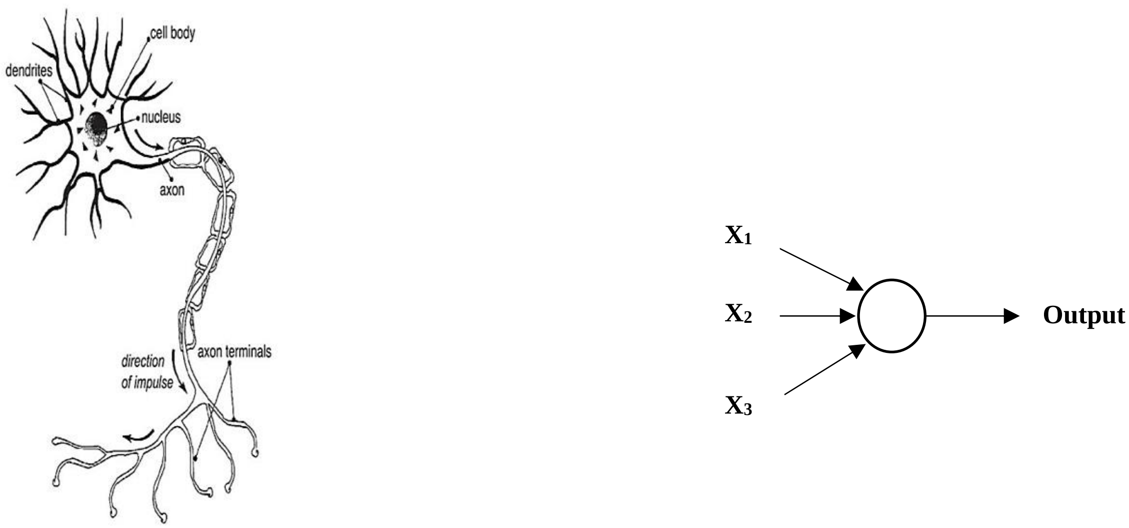

Figure 2.13 Graphical Representation of Biological Neuron (left) and an Artificial Neuron (right) [95]

In the biological brain, learning consists of strengthening or weakening the bonds between different neurons to remember or forget things accordingly. In the case of ANN, learning is just updating the weights at every time step during the training process.

The structure and features of the neural networks can be organized in several ways based on its application. Components of an artificial neural network are,

a. Input layers
b. Neurons
c. Activation Functions and Learning Rate
d. Hidden Layers
e. Output Layers

Different types of neural networks with its variations and wide range of applications are discussed in detail in the upcoming section 2.3.

2.2.2 Hybrid Methods

Both classical and data-based methods are discussed in detail in the previous sections, a major problem in the conventional methods is that the system model has to be accurate. To manage this uncertainty, a hybrid approach with a probabilistic inference and an inductive estimation process based on telemetry data is created (Figure 2.14).

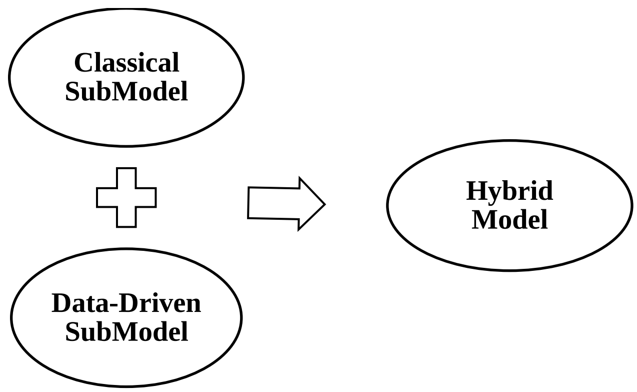

Figure 2.14 Schematic of Hybrid Model

Dynamic Bayesian Network (DBN) is a modified version of Bayesian Network (BN) proposed to model numerous dynamic systems. DBN is a state space approach using the Kalman filter and Hidden Markov model with the uncertain parameters learned from the past data. This method performs fault analysis by sequentially estimating the unobservable variables from the observable sets [63]. In the future, hybrid methods with the combinations of best features from different models will play a prominent role in the fault detection and diagnosis field.

2.3 List of Neural Network Models

There are different types of neural network models based on the problem type and the output, some of the major categories for binary classification, time series prediction, image processing, etc., are discussed below,

I. Feed-Forward Neural Network

Artificial Neural networks are a notion in the field of artificial intelligence. This concept is inspired by the sensory processing model of the human brain. The ANN is created by simulating a network of model neurons with input, output and hidden layers as shown in Figure 2.15.

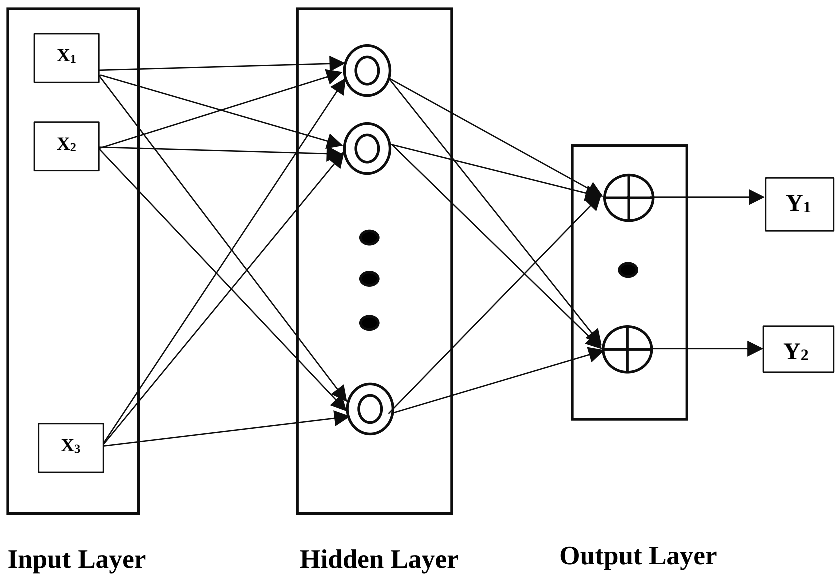

Figure 2.15 Multilayer Feedforward Neural Network

The system is trained repeatedly to imitate the real-time process termed as learning. Like the human brain, for the neural network knowledge is acquired through the learning process and connection strengths named as weights are used to store the information. Furthermore, the process of training ANN has many types, they are Perceptron, Backpropagation, Self-Organizing Map and Delta [96].

The neural network is developed from a popular machine learning approach named perceptron. In this method, the signal is allowed to travel only in one direction from input to output without a feedback loop. Feed-forward neural networks are the quintessential model that are similar to linear models used to solve the simple and most important commercial applications. The main goal is to approximate the function that maps the inputs with the classifier category. They represent many different functions together and hence are often called networks. The model is associated with the behavioral characteristics of the function, let the functions $f(1)$, $f(2)$, $f(3)$ are connected in a form $f(x) = f(3)(f(2)(f(1)(x)))$ where 1,2,3 represents the layer numbers. The overall length of the chain represents the depth of the model, thus it is named as "deep learning". The measured data with noise and imperfections are preprocessed to evaluate the output function at every iteration. Each hidden layer in the model is vector-valued with appropriate activation functions embedded in each of its neurons [31].

Neuroscience is the inspiration for choosing the number of neurons, hidden layers, learning rate, and other functions to compute the output. This is the elementary model developed in the 1950s and has many limitations such as limited capabilities (linear functions), the volume of past data, no feedback loop, difficult to compute the gradients of complicated functions. The back-propagation algorithm and its modern generalizations are effectively used to overcome the shortcomings.

Backpropagation is a supervised learning method introduced by Paul Werbos in 1974, then made popular by Rumelhart and McClelland in 1986. The value from the output layer is fed back to the input layer with changing the weights as shown in Figure 2.16. It is an efficient method of finding the change in each segment weight in a multi-layer network to minimize the error [97].

Let i be the input neuron and j the output neuron, w_{ij} denoted the connection weight, x_j represents the input to the j^{th} neuron, and y_j signifies the corresponding output and dj is the desired output,

Total Input $$x_j = \sum y_i w_{ij} \tag{2.5}$$

Output (Sigmoid Function) $$y_j = \frac{1}{1+e^{-x_j}} \tag{2.6}$$

This algorithm reduces the global error for the given set of weights, the error derivatives for all weights will be computed by feeding backward from the output units. The process of presenting epochs of training samples to the network continues until the average error reached a defined error goal.

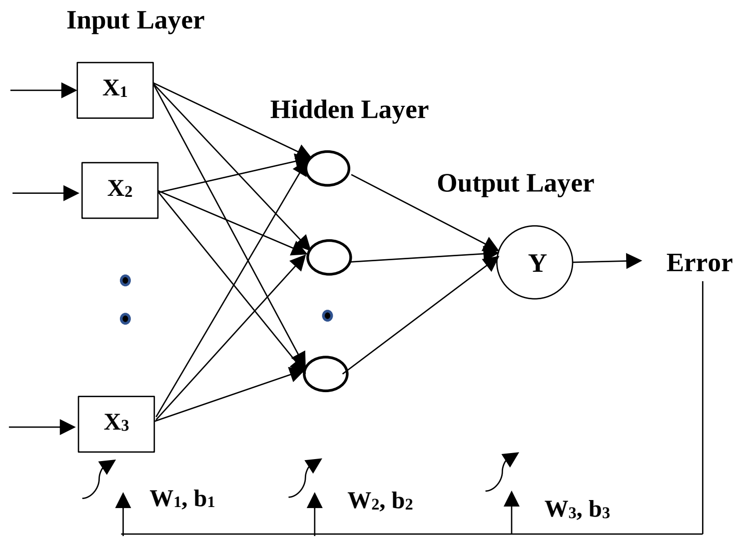

Figure 2.16 Neural Network with Backpropagation

II. Recurrent Neural Network (RNN)

Feedforward neural networks with a feedback connection is known as recurrent neural networks. They are capable of handling variable-length sequence input by having recurrent hidden states with dependent activations. RNNs are two-way communication networks with one to many or many to many configurations. This network can address issues related to storing and retrieving information at any instant in time [31].

In this chain-like structure, the contents of the output vector are influenced by both the present input and the history of past inputs. The adjustment of magnitude and direction of weights is estimated by running the back-propagation algorithm with the differential operations. The model parameters are updated along with the gradient direction to increase the target score. This process is repeated over many times until the solution converges with a consistent prediction.

Recurrent neural networks are specialized for processing sequential values mainly time series. This model produces an output at each time step and have connections between the hidden units. A generalized back-propagation algorithm is used to compute the gradient for RNN. The RNN model is unrolled and application to it is known as back-propagation through time (BPTT). An RNN is a multiple layer network with levels structurally being copies of each other and each layer passes information to the next layer as shown in Figure 2.17. RNN has been successfully applied to time series sequences, speech recognition, language modeling, translation, and image captioning. This feedback loop that records information at every time step has to be stored, but the RNN is incapable of doing them. The basic problem is that gradients propagated over many stages tend to either vanish or explode.

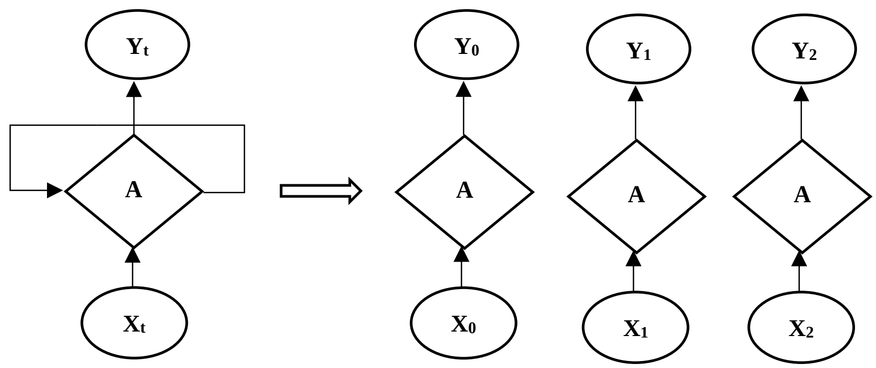

Figure 2.17 Structure of RNN with Unrolled View [31]

III. Long Short Term Memory (LSTM)

Recurrent neural networks are incapable of storing long term dependencies due to insufficient decaying error backflow. An efficient approach introduced by Hochreiter in 1991 addressed the gradient-based method termed as Long Short-term Memory. LSTM networks can learn past information and solve vanishing gradient problem with excess of 1000 time steps by imposing constant error flow through carrousels within the special units [57]. The architecture allows constant error to flow through the special and self-connected units without any limitations of the previous methods.

LSTM networks are four times complex compared to the simple RNN and effective in recognizing patterns outlines across time. A multiplicative input and output gates are introduced to protect the stored memory contents and avoid perturbation by irrelevant values respectively. It is also embodied with a multifaceted memory cell around a central linear unit with a fixed self-connection. Due to the model’s speed, proficiency in solving complex problems and time lag tasks, LSTM is considered to be the most relevant method for this work and thus is applied throughout all the case-studies with suitable justifications.

IV. Convolutional Neural Network (CNN)

CNN is a class of feed-forward neural network which is mainly used for image processing, object and video recognition, natural language processing, etc., The structure of ConvNet is inspired by the organization of visual cortex in the human brain for image recognition. A ConvNet successfully captures the spatial and temporal dependencies in an image through the various layers and filters in the architecture [98].

A typical model consists of an input, output and multiple hidden layers, in CNN the hidden layers are made of series convolutional layers with pooling and fully connected units. The CNN consists of a set of layers (Figure 2.18) with one or more planes to recognize the complex multidimensional image,

Convolution Layer

a. Based on the given image (m x m), a number of filters of size (n x n) are applied with a particular stride size (s).
b. The filter is mapped to the input image and mathematical operations are computed and the obtained value is stored in the center of the cell.
c. The applied filter is moved with the pixel and is repeated for the entire portion of the image.
d. A feature map for the given filter is attained with the size (m-n+s) x (m-n+s) and the same procedure is applied with padding to every filter for obtaining the final stacked convoluted map.
e. The first ConvLayer captures the low-level features such as edges, colors, orientations, etc. Further, the high-level features are acquired with the added layers.

Pooling Layer

a. The purpose of the pooling layer to further reduce the spatial size of the convoluted feature which decreases the computational time and power.

b. There are two types of pooling, max-pooling (returns maximum value) and average-pooling to extract the dominant features which are rotational and positional invariant.

Fully Connected Layer

a. This Layer is used for learning the non-linear combinations of high-level features of the output in kernel space.
b. Activation functions such as ReLu or SoftMax is applied to classify the dominating and unwanted features from the image.

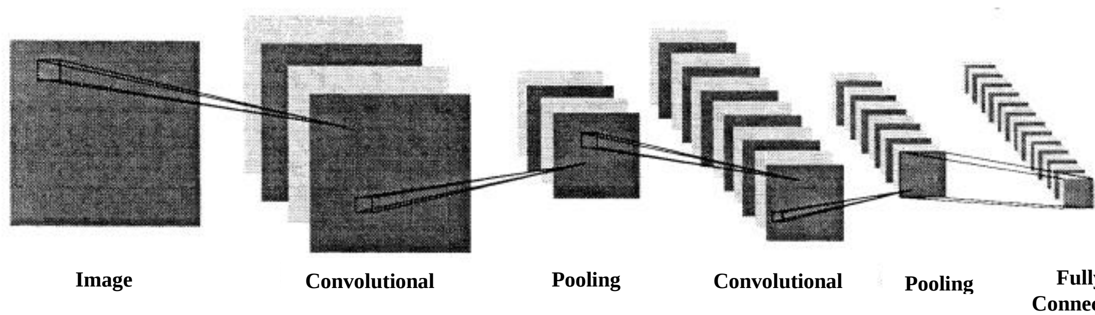

Figure 2.18 A Typical Convolutional Neural Network [98]

2.4 Steps Involved in Data Modeling

The steps involved in analyzing the problem using data analysis is discussed in detail, executing these segments will improve and simplify the investigating procedure in terms of efficiency and computational time.

a. Data Acquisition

The primary step is to examine the obtained data for its attributes as mentioned in section 1.2.1, the required feature for the related problem has to be taken into consideration. The collected data has to be organized will all the necessary details with proper units and source notes. Time frame, criticality factor, linear dependencies, correlative behavior has to be measured initially for consistency and applicability.

b. Data Preprocessing

Real-world datasets are mostly unstructured and unformatted that should be cleaned for imperfections like out-of-range values, NANs (Not A Number) and the missing values have to be filled with an appropriate substitute. The range of input dimensions may vary from problem to problem and it is laborious to change the setup every time, therefore these values have to be normalized based on the statistical features to generalize the situation. Techniques like Principal Component Analysis (PCA), Singular Value Decomposition (SVD) are proposed for removing insignificant items to reduce the dimensionalities.

c. Filtering Techniques

Apart from the visual features, there are some unwanted inbuilt characteristics like noise and impurities (outliners) within the data which has to filter to extract the actual response. Most complicated filtering is for time-series data which is analogous to signal processing. The high-frequency fluctuations are removed using the filters to improve the factor of predictability [99]. Wiener, Kalam, and Savitzky-Golay filters are the most commonly used signal processing techniques used in data mining applications.

d. Model Generation and Analysis

After manipulating the data in all possible ways, the input is reshaped to the model readable format with sample size, dimensions, timesteps, and distinct features. Based on the nature of the problem, one of the above discussed data-driven neural network models is designed. The input and output given to the model may be of supervised learning or unsupervised learning format. This proposed network is compiled with a portion of the training dataset and validated with the test-set using appropriate optimizers and loss functions to generate accurate future predictions of the given type of problem.

e. Interpretation of Results

Finally, from the iterative runs of the model, the convergence error is studied between train and test results for improving the efficiency. After determining the proficiency of the model, the future values are predicted with some confidence interval. In spite of regressive processing and analysis, the data-driven models may not be accurate due to proper model selection, overfitting and underfitting issues, improper tuning of hyperparameters.

2.5 Model Selection and Hyperparameters

Artificial neural networks have the greatest ability to learn and predict complex and nonlinear relationships. Different models discussed in the previous section 2.3 are applied accordingly to real-time problems such as binary classification, pattern recognition, future prediction, anomaly detection, image processing, and others.

The Feed-forward networks are employed in solving classification problems like true/false predictions, consumer activities in marketing, weather forecasting, speech, vision and handwriting recognition.

Time-series problems are mostly solved by recurrent neural networks due to its recursive properties. Any time-based data such as remaining useful life prediction of a system, telemetry data from the satellite or word predictions in a sentence, natural language processing are effectually solved by RNNs. The limitations in solving the above-listed issues are efficiently resolved by the special case RNN known as LSTM. The case studies in the current work fall in this category and thus an improvised LSTM model is solidly considered and successfully applied.

CNN's are used to identify the conceptual structure of an image by breaking down into overlapping tiles. This network gives better accuracy and boosts the performance due to its local connectivity with shared weights and is mainly applied in face recognition, scene labeling, image classification, action recognition, human pose estimation, document analysis, natural language processing, and others.

Each of the networks has many different parameters and each has a significant effect on the convergence of the solution. Some of the common parameters in the ANN models are,

a. Number of Neurons
b. Number of Hidden Layers
c. Input and Output Size
d. Learning Type
e. Activation Functions
f. Batch Size

g. Number of Iterations (Epoch)
h. Optimizer
i. Learning Rate
j. Decay Parameter
k. Dropout Value
l. L1/L2 Regularization

Tuning all the parameters manually results in a total of more than 6 00 000 combinations, which is timing consuming and laborious. To improve the accuracy, predicting factor, efficiency, power, and computational time some of the optimization techniques along with filters are applied.

2.6 Data-Driven Modelling: LSTM

A base LSTM model with all the essential parameters is created and validated with Particulate Matter ($PM_{2.5}$).

2.6.1 LSTM Structure

The structure of LSTM is similar to that of RNN with a difference in repeating modules. Instead of a single layer, there are layers interacting in a very special way as shown in Figure 2.19. It is composed of cell states, gates (to regulate the flow of information) and memory cells [100].

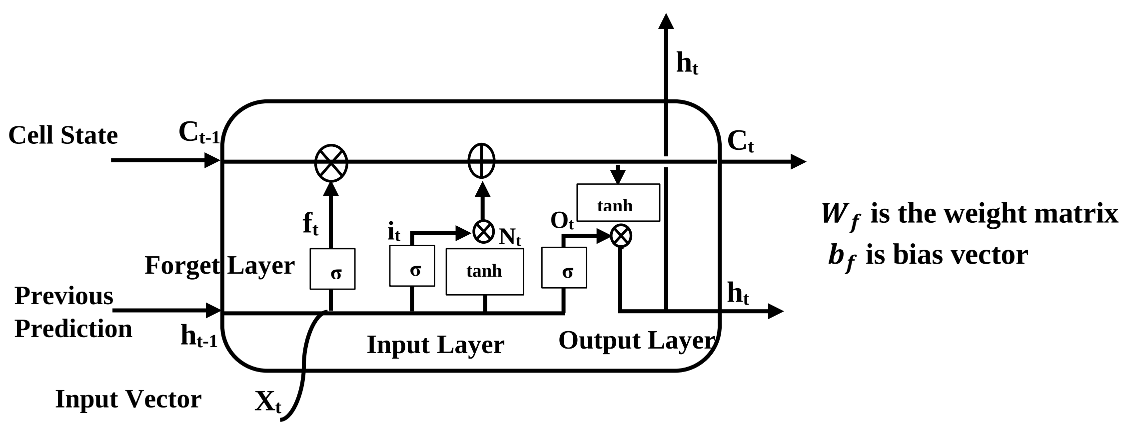

Figure 2.19 LSTM with Interacting Layers [57]

The cell state runs through the entire chain with linear interactions, forgetting old memory and adding new memory. Gates let/prevent information through it with am activation layer and an elementwise operation. An LSTM has three gates; input, output and forget gate.

The cell state vector at the current time is C_t and the previous time step is C_{t-1}.
Forget gate layer: Looks at h_{t-1} and x_t and outputs a number 0 and 1(sigmoid layer) for each number in the cell state C_{t-1}.

$$\text{Forget vector } f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2.7}$$

Input gate layer: Decides the update on the new information with 2 parts, a sigmoid layer (what to update) and a tanh layer (what new value to add).

$$\text{Input Vector } i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2.8}$$

$$\text{Cell state update vector } N_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{2.9}$$

Then all the old cell state C_{t-1} has to be updated with the new cell state C_t.
Output gate layer: Decides what is going out based on the filtered version of the cell state. Sigmoid layer runs in which part goes out and puts the cell state through the tanh layer.

$$Output\ Vector\ O_t = \sigma(W_o.[h_{t-1}, x_t] + b_o),\ h_t = O_t * \tanh(C_t) \tag{2.10}$$

Other popular variants of LSTM are Peephole connections, coupled forget and input gates and Gated Recurrent Unit (GRU). Among them, the GRU is simpler and growing increasingly popular in the current days.

2.6.2 Preprocessing

In the current case, the multistep pollutant prediction is studied. Accordingly, the input sample is thoroughly inspected and stacked in such a way the file only has the hourly interval of appropriate data with the absolute numbers reformatted to values between 0 and 1 with the inbuilt scaler function and the missed values are filled with their mean. The given data ($PM_{2.5}$) is converted to the supervised format by shifting the data with a given time step into inputs and outputs with the required interval.

Input: $PM_{2.5}$ (t-n) , Output: $PM_{2.5}$ (t+n)

where t is time step and n is number of future intervals

The training dataset (50%) is used to train and fit the model. In this zone, the model will capture, store the feature of the sample and learns weights, biases. The second set (25%)is used to fine-tune the hyperparameters and validate the proposed model. Finally, testing (25%) is the new set of data used for validating the model with the performance metrics.

2.6.3 Hyperparameters of the Model

This method minimizes the error and delivers the predicted value based on the fine-tuning of the hyperparameters of the model as explained,

a. Batch size, Neurons, Hidden layers:

Batch size is defined as the number samples that will get updated with the weights before the next iteration, it has to chosen to avoid convergence and time. For better memory, the number is always chosen in the powers of 2.

Batch Size = near the power of 2 [of (Total training set/1000)]]

For selecting number of neurons there is no direct thumb rule, to start with the upper bound has to be always within the range of N_H

$$N_H = \frac{No.\ of\ samples\ in\ the\ training\ data}{[scaling\ factor(2-10)\ *\ (No.\ of\ IP\ +\ No.\ of\ OP)]} \tag{2.11}$$

The number of hidden layers depends on the size of training data, an optimum number will avoid overfitting and bias problems. Two are preferable for this kind of problem and can represent an arbitrary boundary with smooth mapping.

b. Activation Function:

The nonlinear activation function includes logistic sigmoid, tan-hyperbolic, rectified linear (ReLU) and SoftMax/Softplus. The only requirement to choose the function is that it should be continuously differentiable. For the logistic sigmoid the output range is (0,1) it prevents jumps in the output values, the range for tanh is (-1,1), it is zero centered making it easier to model the negative values. ReLu (0, max) is computationally efficient function. SoftMax/Softplus is mostly used in the case of multiple-layered categories.

$$Sigmoid\ f(x) = \frac{1}{1+e^{-x}}\ ; \quad Tanh\ f(x)\frac{1-e^{-2x}}{1+e^{-2x}}; \tag{2.12}$$

$$ReLU\ f(x) = \begin{cases} 0, x < 0 \\ x, x \geq 0 \end{cases}; \quad SoftMax\ f(x) = \log_e(1+e^{x}) \tag{2.13}$$

c. Optimizers and Regularizers:

Optimizers are the inbuilt algorithm that is used to update the weights after every iteration. It helps to optimize the objective function of the given set. ADAM (Adaptive Moment Estimation) is a second-order optimizer that has a faster convergence rate, the capability of tuning the learning rate and keeps an exponentially decaying average of its past gradients and is used accordingly. Furthermore, underfitting and overfitting issues are avoided by adding dropout layers, it acts as a regularizer and removes the unwanted connections. L1 (Lasso Regression) and L2 (Ridge Regression) are regularizers wherein the high values of weights are curbed and is used when the model overfits ever after using the dropout value. Finally, the total number of iterations is regulated with the early stopping criteria. Figure 2.20 represents the LSTM architecture including all the above-mentioned parameters that are used in the modeling.

```
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 10, 50)            10400
_________________________________________________________________
dropout_1 (Dropout)          (None, 10, 50)            0
_________________________________________________________________
lstm_2 (LSTM)                (None, 50)                20200
_________________________________________________________________
dropout_2 (Dropout)          (None, 50)                0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 51
=================================================================
Total params: 30,651
Trainable params: 30,651
Non-trainable params: 0
```

Figure 2.20 Descriptive LSTM Model Plot

2.6.4 Stationarity of the Data

The model that exhibits correlation and tends to output (t+1) time index value close to time (t) is known as the persistence model. In order to avoid such a downside, the need to be checked for its trend over time and is changed to stationary standards with statistical properties such as mean, variance, standard deviation using the Dickey-Fuller test [101] and the obtained results are shown in Table 2.1.

Inference of Dickey-Fuller test :

i. P-value <= 0.05 the data does not have unit root and is stationary.

ii. The more the test statistic value is negative the data is stationary.

Table 2.1 Results of Dickey-Fuller Test

Number of Observations Used: 34990	
Test Statistic Value	-19.911217
P-Value	0.0000000
Critical Value (1%)	-3.4305370
Critical Value (5%)	-2.8616230
Critical Value (10%)	-2.5668140

2.6.5 Results and Discussion

The accuracy and efficiency of the LSTM model is validated for real-world data. Both one-step prediction (supervised learning) and multi-step predictions using the sliding window technique ahead a week and a month respectively are achieved. The model is evaluated with Root Mean Square Error (RMSE) and Mean Percentage Error.

2.6.5.1 LSTM - One Step Prediction

The LSTM model developed was used to predict the $PM_{2.5}$ values for the next time step. The parameters of the model are as given in Table 2.2.

Table 2.2 LSTM Model Parameters for One-Time Step

Parameters	Values
Number of Neurons	50, Dropout Value of 0.2
Number of Hidden Layers	1
Batch Size	64
Input shape	1,1
Output shape	1
Optimizer	ADAM
Loss Function	Mean Squared Error, Mean Absolute Error

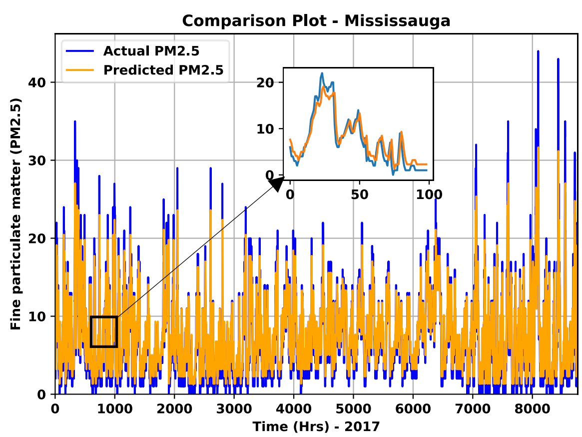

Figure 2.21 Actual and Predicted Values of Fine Matter Particulate ($PM_{2.5}$) [Mississauga]

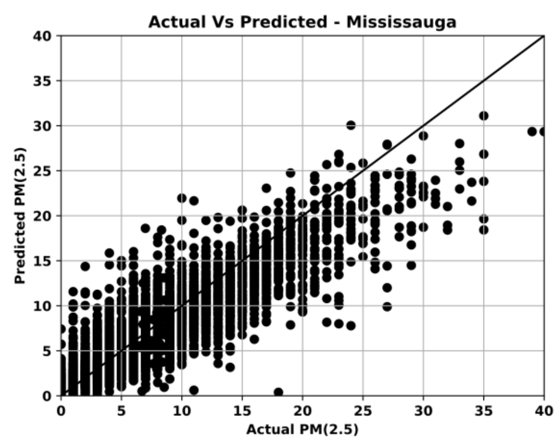

Figure 2.22 Variations of Actual and Predicted Values [Mississagua]

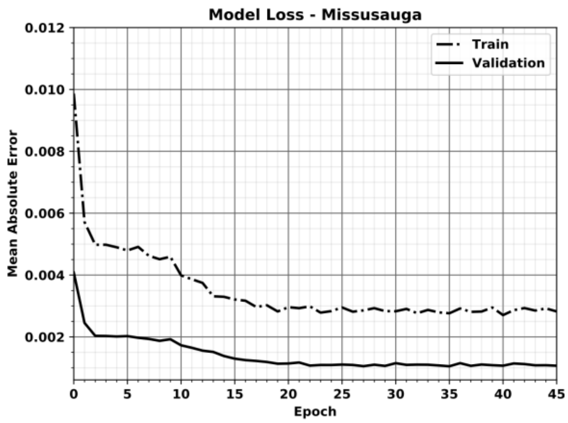

Figure 2.23 MAE Plot for Train and Validation Dataset [Mississagua]

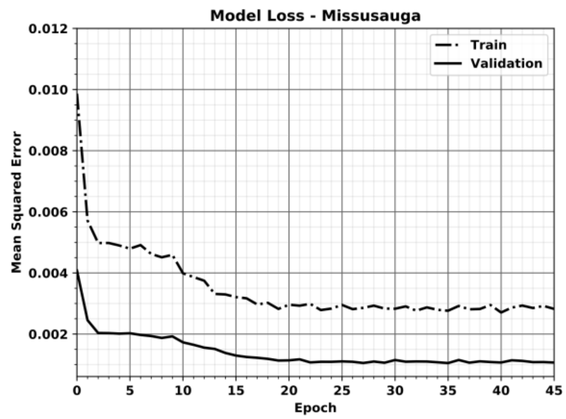

Figure 2.24 MSE Plot for Train and Validation Dataset [Mississagua]

Figure 2.21 shows the real and predicted values of Mississauga over the year 2017 - test set (never used in training or validation) as it almost overlaps with each other. Being the model learned all parameters, it is used to predict the test sequences one at a time. This is persistence prediction and is implemented by using the last value of the training data and history accumulated by walk-forward validation for predicting the next step value. Figure 2.22 shows the comparison of actual and predicted values, Figures 2.23 and 2.24 are convergence of error plots and is seen that the rate of the validation error is slightly above the training error which is considered to be a good fit. Table 2.3 represents the results obtained for all the cities.

Table 2.3 Model Metrics for the Results of One-Step Prediction

Performance Scores	Mississauga	Toronto Downtown	Ottawa
Root Mean Square Error (RMSE)	2.030	2.351	1.541
Mean Percentage Error	21.47	20.82	17.814
Total Computational Time (s)	101.43	73.39	86.16

2.6.5.2 LSTM - Multi-Step Prediction

This model is created with the aim to predict future events based on the fixed lag in time to predict the window of the same size. Since the model always inputs the sequences of known inputs it eliminates the built-up error. The algorithm was developed to one-week values to predict the next week's values. The prediction is iterated constantly every time to forecast the next week's corresponding value using the previous week's data. All the hourly data is resampled to a daily time interval before the analysis. The parameters of the model for both 1 week and 15 days predictions are as given in Table 2.4.

Table 2.4 LSTM Model Parameters for Multi Time Step Prediction

Parameters	Values
Number of Neurons	200, Dropout Value of 0.2
Number of Hidden Layers	1
Batch Size	16
Input Shape	1,1
Output Shape	1
Optimizer	ADAM
Loss Function	Mean Squared Error

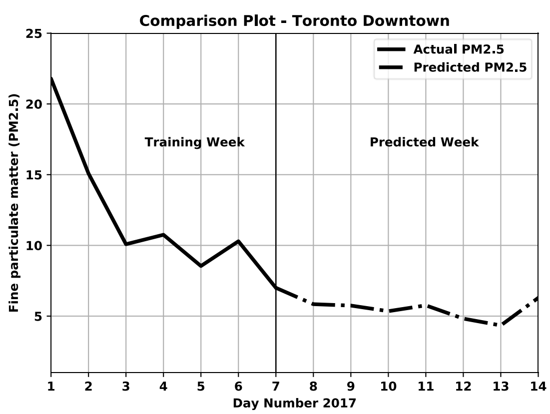

Figure 2.25 Actual and Predicted Values of Fine Particulate Matter ($PM_{2.5}$) 1-Week Ahead [Toronto]

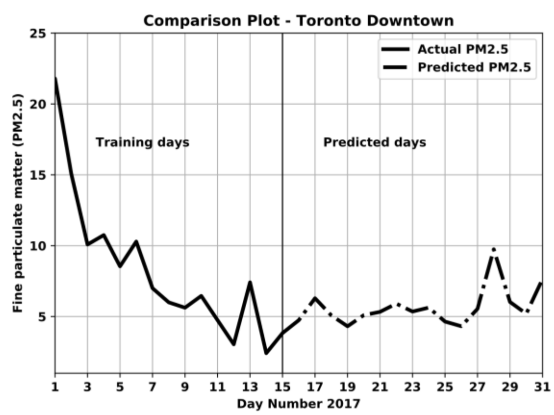

Figure 2.26 Actual and Predicted Values of Fine Particulate Matter ($PM_{2.5}$) 15 Day Ahead [Toronto]

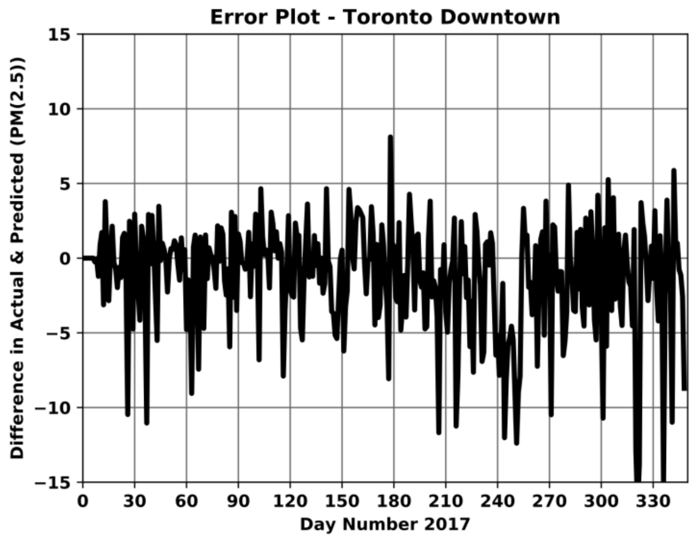

Figure 2.27 Error Plot of Actual and Predicted Values($PM_{2.5}$) [Toronto]

Figure 2.27 error plot of the actual and predicted values of pollutant particles. Table 2.5 represents the performance metric scores and computational time of the model taken by all the three different data sets for both 1 week and 15 days prediction.

Table 2.5 Score Metrics of Multi-Step Prediction

Performance Scores	**Mississauga**		**Toronto Downtown**		**Ottawa**	
	1 week	15 days	1 week	15 days	1 week	15 days
Root Mean Square Error (RMSE)	3.922	7.823	4.023	4.916	4.152	4.658
Mean Percentage Error	23.57	44.18	28.64	48.51	23.65	46.25
Total Computational Time (s)	77.458	105.98	110.14	137.66	156.15	145.67

2.6.5.3 Sliding Window - Multistep Prediction

The sliding window or moving window forecasting method is used to predict the future in multiple time steps without using the source data. The model will learn the input in a particular sequence and predicts the output, later this predicted output is used to forecast the future values in multiple time steps ahead. After training and validating the model same as explained in the previous sections a series of 360 hours (1 week) of data were used as the input t(0), t(1), t(2)….t(359)] to predict the 361st value. Every instant in time the first data will be removed and the predicted data will be added with the fixed window size, by this, the model builds on predicted values.

For achieving the accurate predictions, the hyper-parameters of the model are varied with many combinations and the results for the dataset are presented below with their results. The performance of the sliding window model is tested in the first month of January (2017) – from 760th hour (31st day) to the next 30 days (1400th hour) and the best-scored plot is as shown in Figure 2.28.

The common parameters of the model are as given in Table 2.6.

Table 2.6 Model Metrics for the Results of Multi-Step Prediction

Parameters	Values
Number of Neurons	50
Number of Hidden Layers	2
Regularizer	L1 = 0.01, L2 = 0.01
Batch Size	16
Input Shape	1,1
Output Shape	1
Optimizer	ADAM
Loss Function	Mean Squared Error

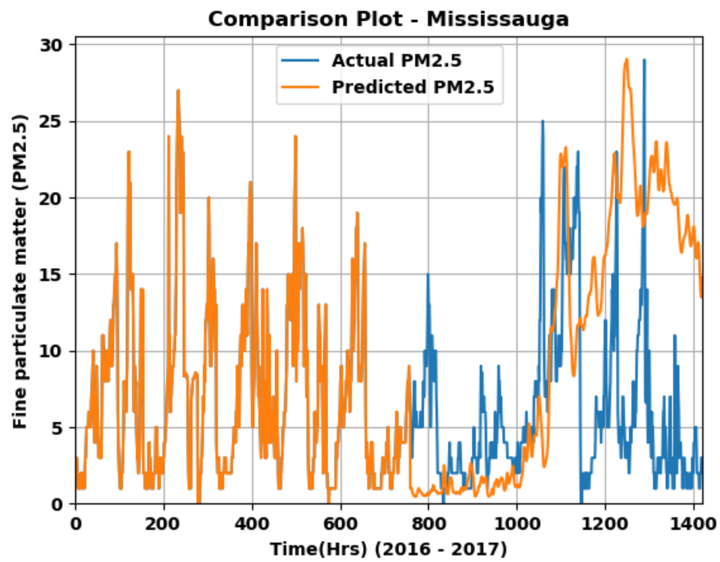

Figure 2.28 Fine Particulate Matter ($PM_{2.5}$) [Predictions 760th hr to 1400th hr]

Finally, Table 2.7 represents the summary of results for different variations in the hyperparameter choices.

Table 2.7 The Fixed Parameters of the LSTM Model

	Variations in the Parameters	RMSE	% Error	Computational Time (s)	Predictions Ahead
Case 1	Activation Function: Tanh, ReLU, Sigmoid. Dropout value : 0.5,0.5	13.64	44.16	481.56	~ 48 hrs (2 days)
Case 2	Activation Function: Tanh, ReLU, Tanh. Dropout value : 0.3,0.5	14.56	43.59	490.56	~ 340 hrs (14 days)
Case 3	Activation Function: Tanh, Tanh, Tanh. Dropout value : 0.5,0.5	17.56	45.78	416.18	100 hrs (4 days)
Case 4	Activation Function: Sigmoid, Tanh, Tanh. Dropout value : 0.3,0.3	12.56	33.16	381.15	~ 440 hrs (18 days)
Case 5	Activation Function: Tanh, Sigmoid, Tanh. Dropout value : 0.3,0.5	13.87	35.99	394.58	~ 440 hrs (18 days)
Case 6	Activation Function Tanh, Sigmoid, Tanh. Dropout value : 0.3,0.5 RMSprop - Optimizer	15.78	46.78	498.12	~ 340 hrs (14 days)

2.6.5.4 Conclusions

Data-driven modeling is a promising route for prognostics and diagnostics of any system. The viability of the LSTM models is demonstrated for forecasting the Fine particulate matter ($PM_{2.5}$). The first model developed is a persistence model and provided accurate predictions over time and is capable only with t+1 step predictions. The Second model is capable of predicting values ahead of 1 week and 15 days respectively. Finally, the sliding window algorithm is modeled with the notion of predicting the values ahead of 1 month with finite sequences of inputs. Key-areas for improvements will be on fine-tuning the hyper-parameters for future predictions of larger intervals.

There are no reference systems to compare and validate the efficiency of the designed model and choosing the best combination of the parameters by trial and error is time-consuming and laborious.

CHAPTER 3

3. Fault Detection of Reaction Wheels Onboard Kepler Spacecraft

3.1 Introduction - Kepler Mission

In this Chapter faults of the reaction wheels onboard Kepler Spacecraft are analyzed using both the telemetry data received from the Kepler mission and the correlated weather data. The yearly sets that include the critical parameters of the mission namely speed, torque commands, rotor temperature, torque friction, and attitude error are used for analysis. Initially, the classical approach such as correlation method (feature extraction), Weibull analysis (appropriate with normal distribution) and classical friction theory model with linear function is applied and later reinforced with data-driven methods to compare the results.

The Kepler spacecraft was launched on March 7, 2009, into an Earth trailing orbit with the intention of discovering Earth-sized planets in the habitable zone around sun-like stars. The mission goal of detecting planets is accomplished using the transit method. When a planet passes the line of sight of the telescope, the reduction of light from the planet is measured by the spacecraft's telescope. To perform this the onboard instrument should be capable of detecting changes in brightness on the order of 30 ppm. The duration and size of the dip in light can tell the orbital period and size of the planet. To date over 3,000 planet candidates are acknowledged by the Kepler program [102].

The attitude control of the Kepler spacecraft was made possible using four reaction wheels along with the set of fine guidance control sensors along with charge-coupled devices (Figure 3.1). All the four reaction wheels are Goodrich TW-16B200 and they are arranged in a pyramid configuration shown in Figure 3.1 with details presented in Table 3.1. Each wheel has a determined momentum capacity of 16.6 Nms, with a reaction torque capability of ±0.2 Nms which corresponds to a speed of 5100 RPM. Each of them has an upper limit (above 7000 RPM) on its speed and is controlled by the Overspeed electronic circuit. The life expectancy of the reaction wheels is approximately 10 years, and at least three should be operational for the system to remain stable [103].

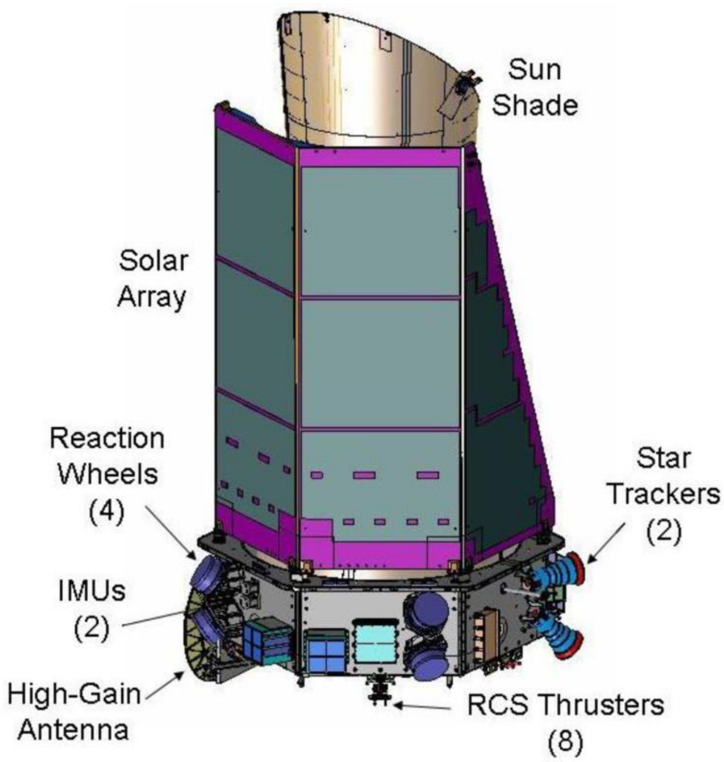

Figure 3.1 Kepler Spacecraft with Attitude Control Subsystem Components [103]

The reaction wheel is one of the satellite attitude control actuators with a flywheel and an electric motor. These wheels spun at variable speeds to turn the vehicle to the desired direction. For a spacecraft to be controlled about all three directions there must be one at each axis along with a redundant wheel to back them up. For the satellite to remain stable at least three wheels must be operational. Reaction wheel 2 of the Kepler spacecraft failed unexpectedly in July 2012, followed by the failure of reaction wheel 4 in May 2013. The yearly sets of data which include the critical parameters of the mission namely speed, torque commands, rotor temperature, torque friction, and attitude error are used for analysis.

Table 3.1 Reaction Wheel Spin Axis Unit Vectors [103]

	RW1	RW2	RW3	RW4
X	0.573526	-0.573526	0.573526	-0.573526
Y	0.484684	0.484684	0.484684	0.484684
Z	0.660328	0.660328	-0.660328	-0.660328

3.2 Data Processing Methodology

The telemetry data for the Kepler Spacecraft is provided for all the four reaction wheels and is a multivariate time series information with a five minute time interval. The data is available for the first 100 days after the launch in 2009, for the 275 days of 2012 in the year which reaction wheel 2 failed and 100 days in 2013 till the failure of reaction wheel 4. Each file includes TCMD (Torque Command in Nm), Speed (RPM), RW motor temperature (deg), Torque friction (Nm) and attitude errors (radians). All the critical data for the analysis is given and it is known that the most common cause of the fault is the frictional behavior of the rotating components. Altogether we have one input (TCMD) and 4 output data.

3.2.1 Kepler Mission Dataset

To begin with the analysis, all the parametric data are plotted with time for the given intervals. The first set of plots are for the early days of the year 2012 (Figures 3.2(a) to 3.2(e)) representing a healthy condition. It is to be noted that the following set of figures describes the telemetry data from 09th April 2012 to 10th May 2012, apart from the minor correction made on April 30th (around data points 6100 to 6300) all the parameters remain normal.

ACKNOWLEDGMENTS: *I Would like to thank Dr. Marcie Smith, Mission Director, and Dr. Sobeck Charles, Project Manager of The Kepler Mission, Nasa's Ames Research Center, California, USA for providing us the Kepler data for this study.*

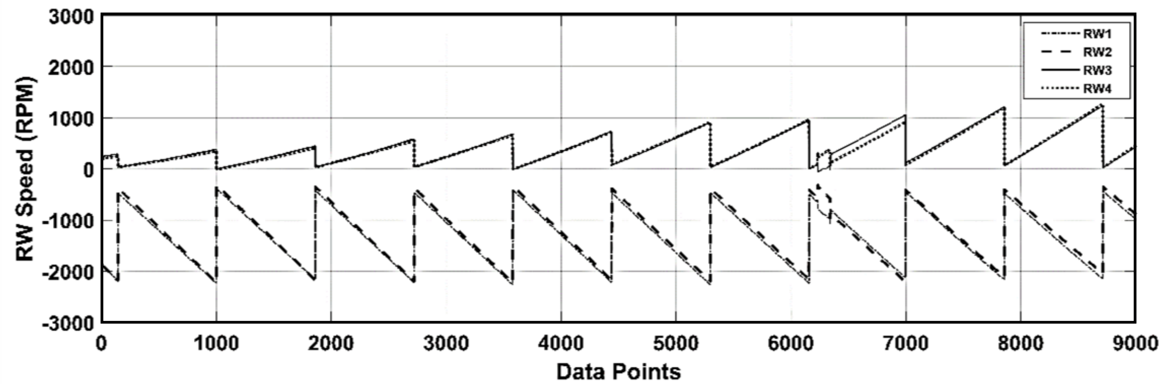

Figure 3.2 (a) RW Speed for 30 Days Sample Set, 2012 (Healthy Condition)

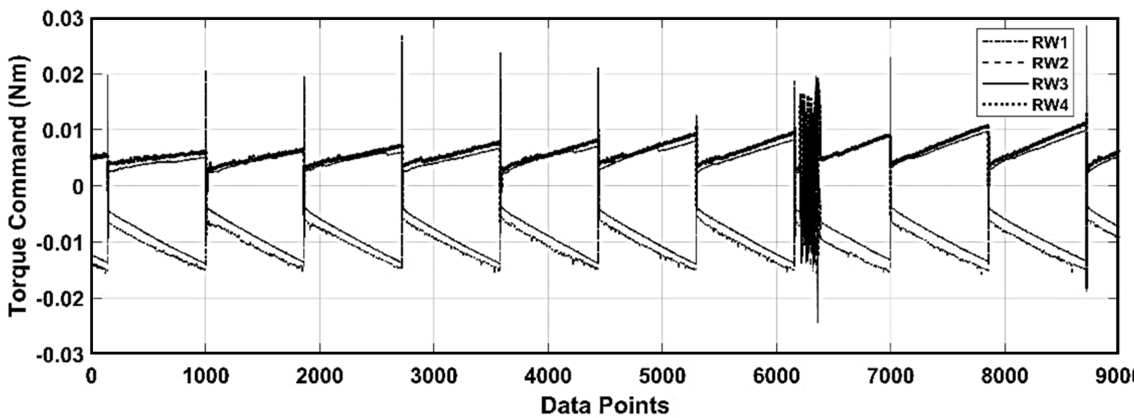

Figure 3.2 (b) Torque Command for 30 Days Sample Set, 2012 (Healthy Condition)

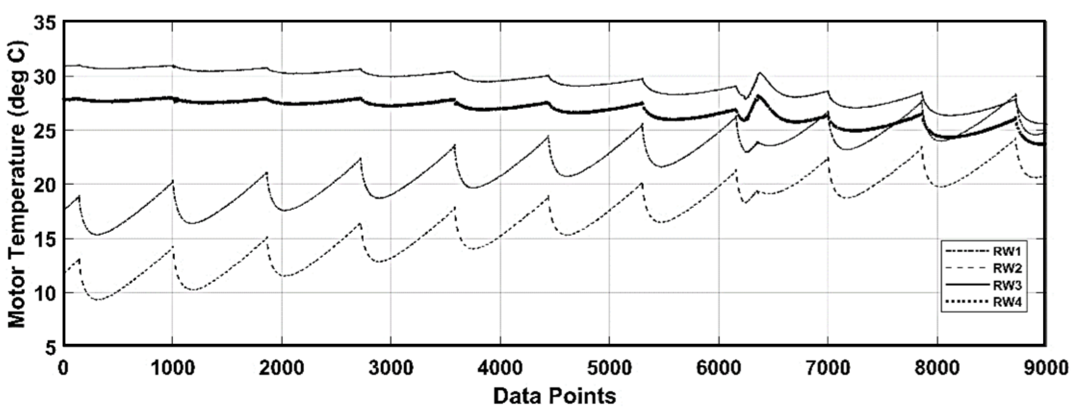

Figure 3.2 (c) Temperature for 30 Days Sample Dataset, 2012 (Healthy Condition)

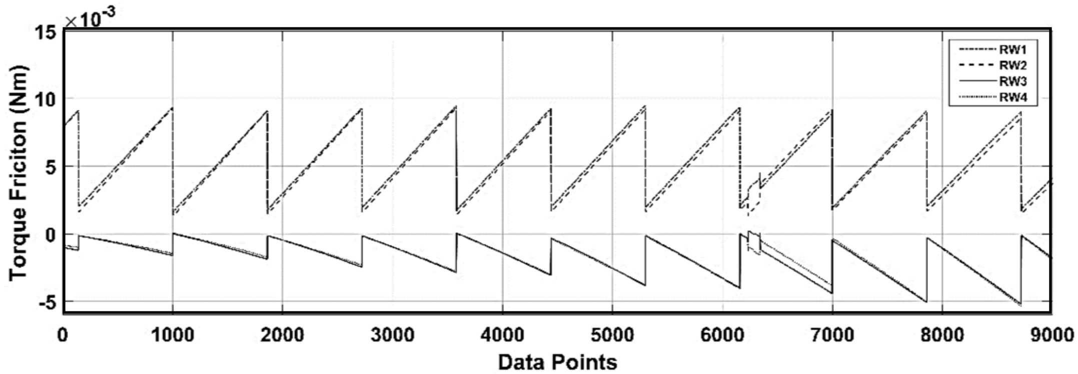

Figure 3.2 (d) Torque Friction for 30 Days Sample Dataset, 2012 (Healthy Condition)

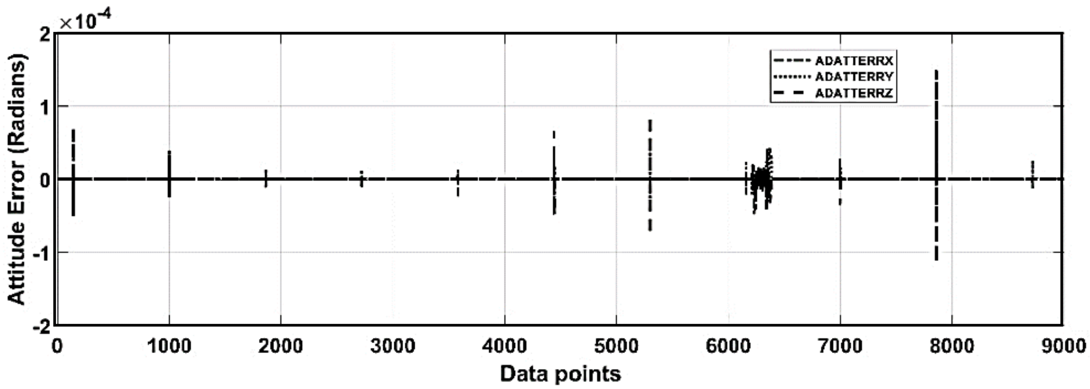

Figure 3.2 (e) Attitude Error for 30 Days Sample Dataset, 2012 (Healthy Condition)

Next, the second set of plots is considered (Figures 3.2(f) to 3.2(j)); this includes data from 28th June 2012 to 29th July 2012. Furthermore, this set has 15 days of data, ahead and after the failure event (14, July 2012) of the reaction wheel 2 (RW 2). A small anomaly is observed in the rotor speed of all the wheels, before the day of failure of reaction wheel 2 as seen in Figure 3.2(f).

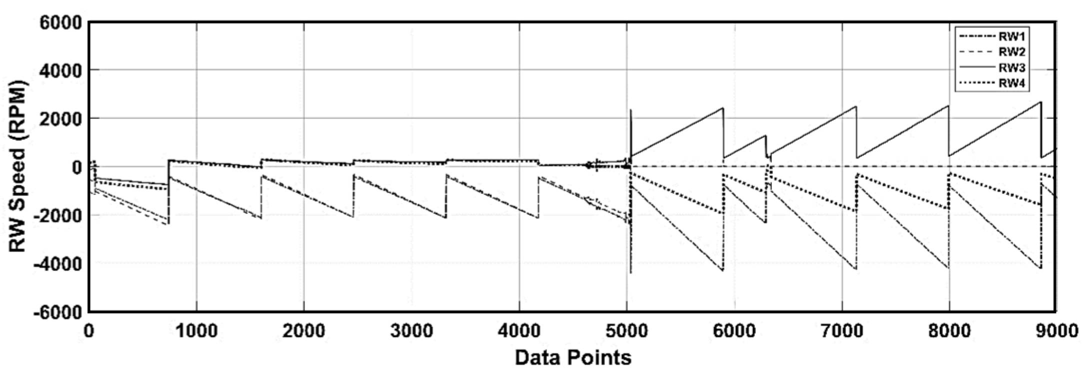

Figure 3.2 (f) RW Speed for 30 days Sample Dataset, 2012 (Faulty Condition)

From Figure 3.2(g), it is noted that before the day of the failure event, the torque command variations for reaction wheel 2 is higher than the other wheels.

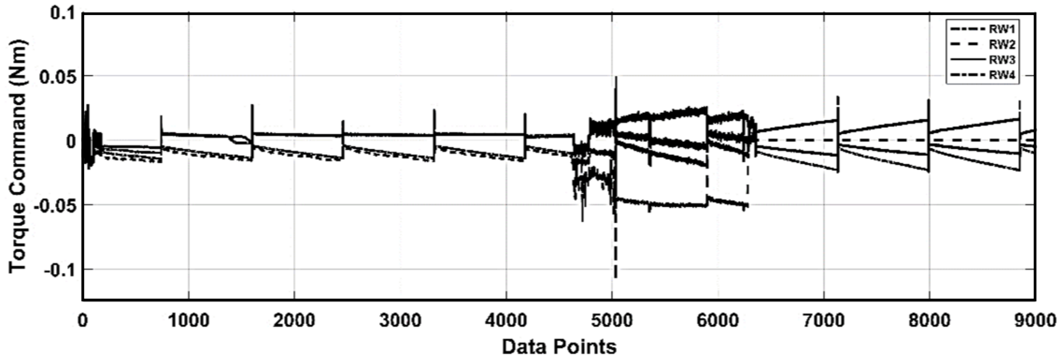

Figure 3.2 (g) Torque Command for 30 Days Sample Dataset, 2012 (Faulty Condition)

As such, there is no evident signature in the temperature profile of the wheels (Figure 3.2(h)) prior to the event of failure. A similar variation to rotor speed is exhibited in the torque friction profile (Figure 3.2(i)).

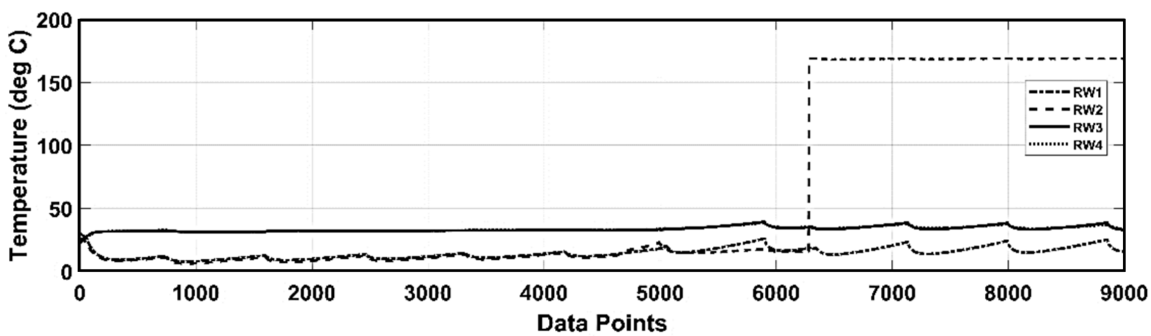

Figure 3.3 (h) Temperature for 30 Days Sample Dataset, 2012 (Faulty Condition)

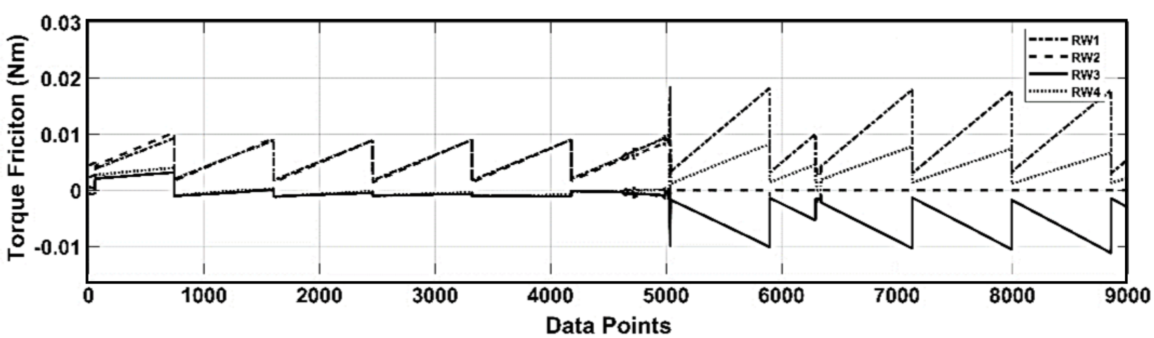

Figure 3.2 (i) Torque Friction for 30 Days Sample Dataset, 2012 (Faulty Condition)

The attitude errors of the spacecraft about all three axes are significantly high (see Figure 3.2 (j)) during the period of 10 days starting from 4700 data point (corresponds to 10th of June 2012) to 6500 data point (corresponds to 14th of June 2012). The attitude errors increase before the failure event (4000 data point (corresponds to June 2012) and reach high at the time of failure and then decrease and reach high again after 5 days.

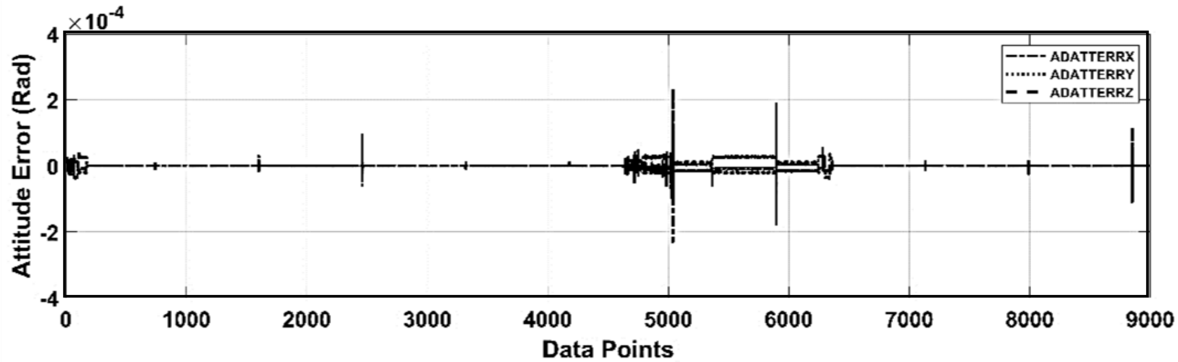

Figure 3.2 (j) Attitude Error for 30 Days Sample Dataset, 2012 (Faulty Condition)

3.3 Classical Data-Driven Approaches - Results

The comprehensive analysis is detailed in the conference paper$ and patent application* submitted by the supervisor and the author. The methodologies presented in the previous section are applied to analyze the Kepler spacecraft data, starting from the simpler data plot to linear intercept friction analysis; there are many inferences as explained below:

3.3.1 Raw Data

The rotor speed exhibited some small changes 2 days before the complete failure of the RW2 in 2012. A similar trend is observed for the RW4 in the year 2013, 8 days before the failure as shown in Figures 3.3 and 3.4.

$ Dhanagopal, V., Kumar, K. D. (2019). Fault Detection and Remaining Useful Life Prediction of Reaction Wheels Onboard Kepler Spacecraft. *15th International Space Conference of Pacific-basin Societies (ISCOPS) (AAS 18-713)*, (10-13, July 2018), Montréal, Québec, Canada.

* Dhanagopal, V., Kumar, K. D. "Fault Detection and Remaining Useful Life Prediction of Reaction Wheels Onboard Spacecraft," Ryerson University Invention Disclosure, AI-PASS Document No.: 2019-10-04, Oct. 04, 2019; to be submitted to Canadian Intellectual Property Office (CIPO), Date to be filed: October 2019.

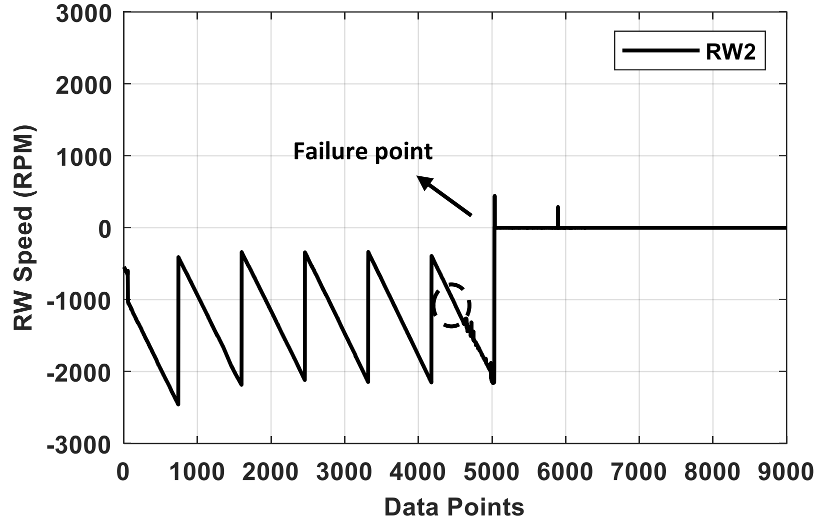

Figure 3.3 Speed Response of RW2 with a Prior Signature

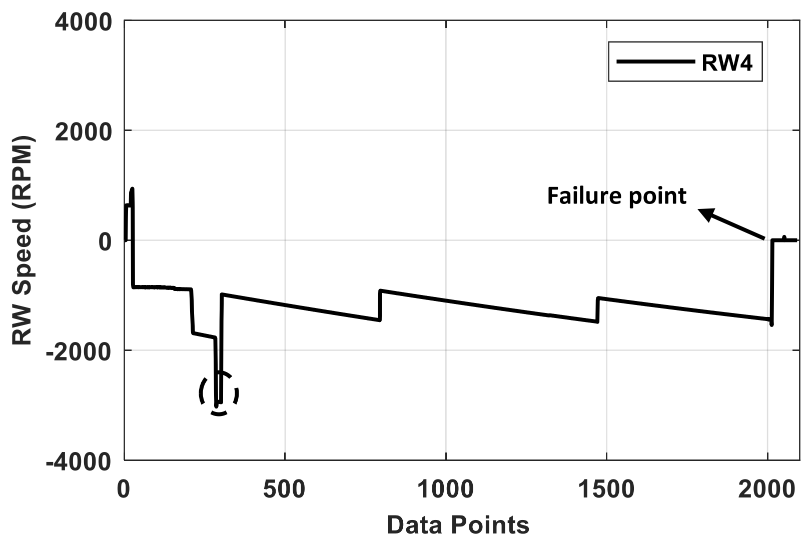

Figure 3.4 Speed Response of RW4

3.3.2 Correlation

As expected the plot for 2009 was normal. In the failure year 2012, the correlation plot exhibited a change in trend for rotor temperature with torque friction prior to 90 days, as shown in Figure 3.5. For the next event in 2013 for RW4, the trend is different with a significant shift prior to a month of the failure as shown in Figure 3.6.

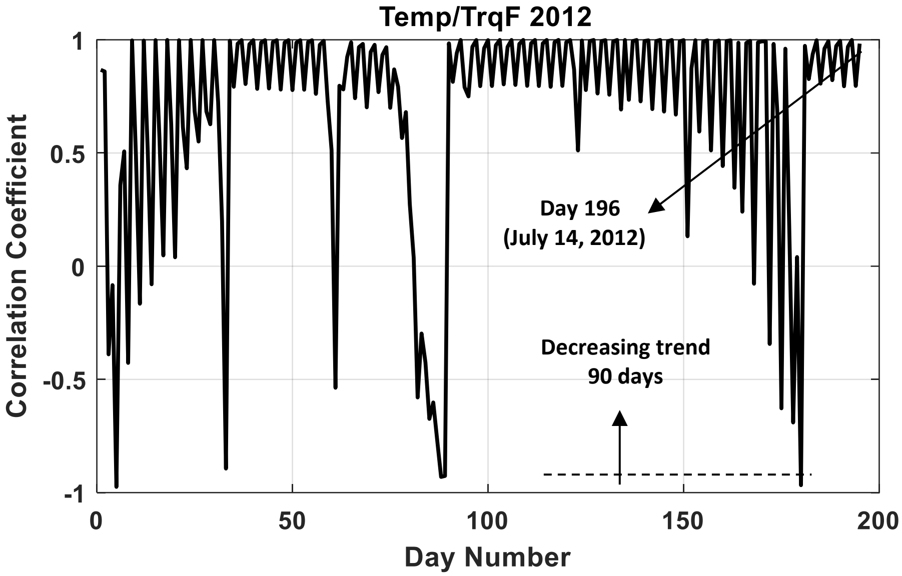

Figure 3.5 Correlation Between Rotor Temperature and Torque friction, RW2

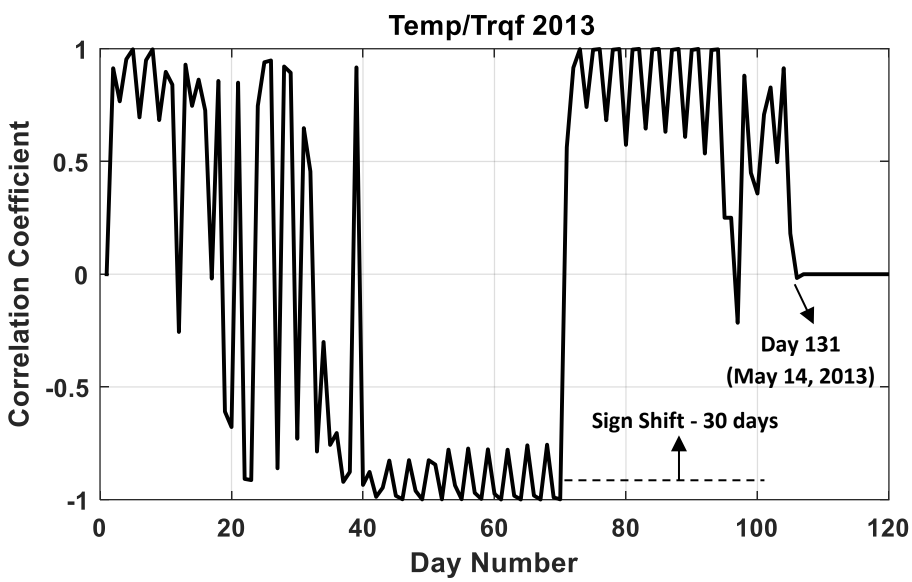

Figure 3.6 Correlation Between Rotor Temperature and Torque friction, RW4

Similar plots for rotor speed with torque friction in the year 2012 and 2013 for RW2 and RW4 respectively, exhibited the change in correlation coefficient before 3 months of the event as revealed in Figure 3.7 (a) and Figure 3.7 (b).

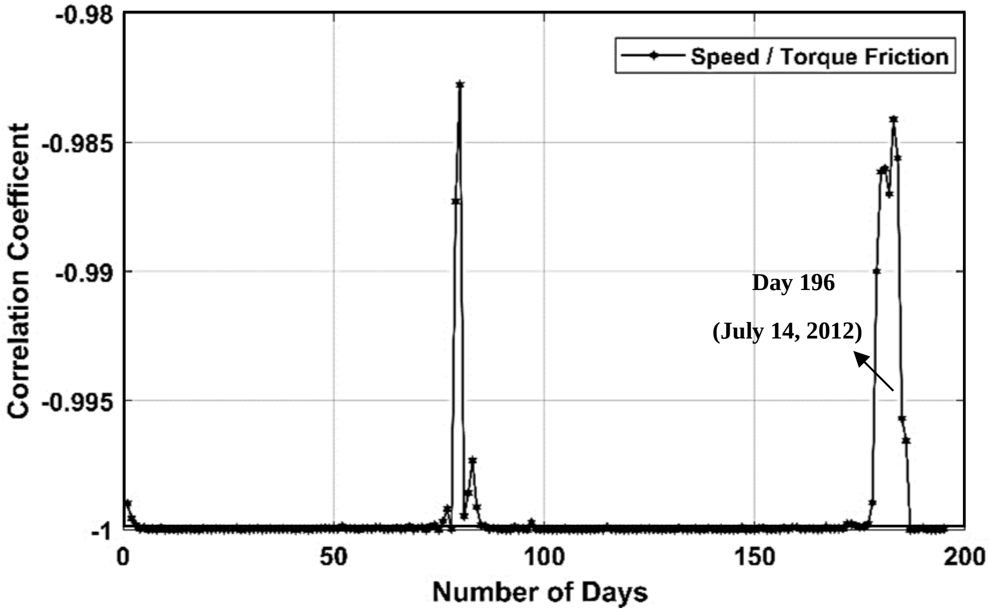

Figure 3.7 (a) Correlation Between Speed and Torque Friction, RW2

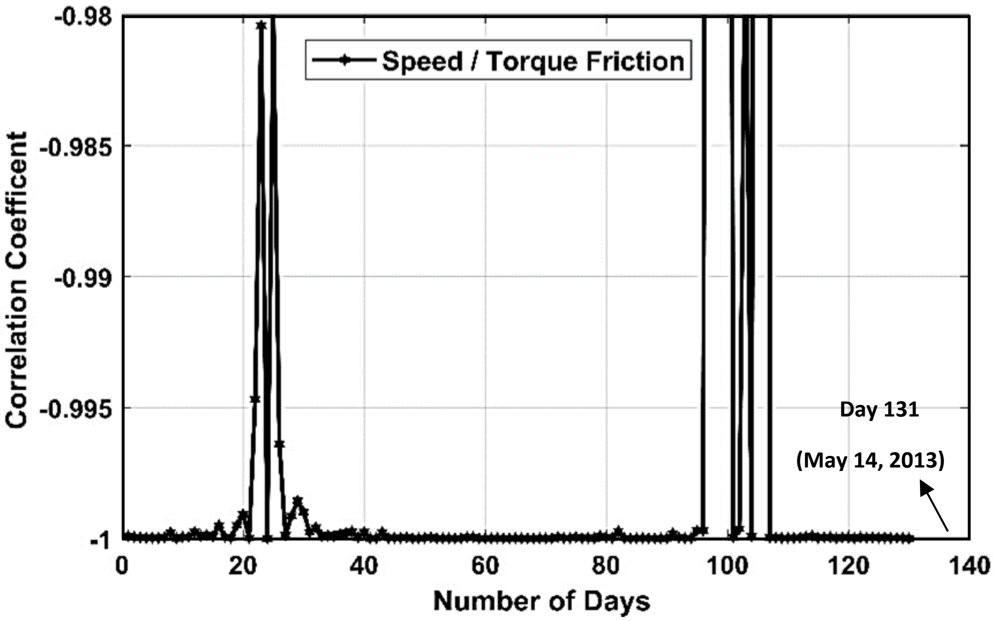

Figure 3.7 (b) Correlation Between Speed and Torque Friction, RW4

3.3.3 Weibull

The probability plots with mean rank, the median rank, and symmetric cumulative distribution functions are used. Lilliefors test and least R-squared test based on Kolmogorov-Smirnov is performed to check the normality of the data. Two parameter variant Weibull function is used in commercial software [104] to extract the shape and scale parameters of the given data. Since the plot is suitable even for smaller samples, the size of the sample is reduced from 30 to 16 days (case 1 - 2012) and 5 days (case 2 - 2013). The shape factors for torque command (input), torque friction (output), wheel speed (input) and rotor temperature (output) for all the reaction wheel assemblies are extracted from the sample sets.

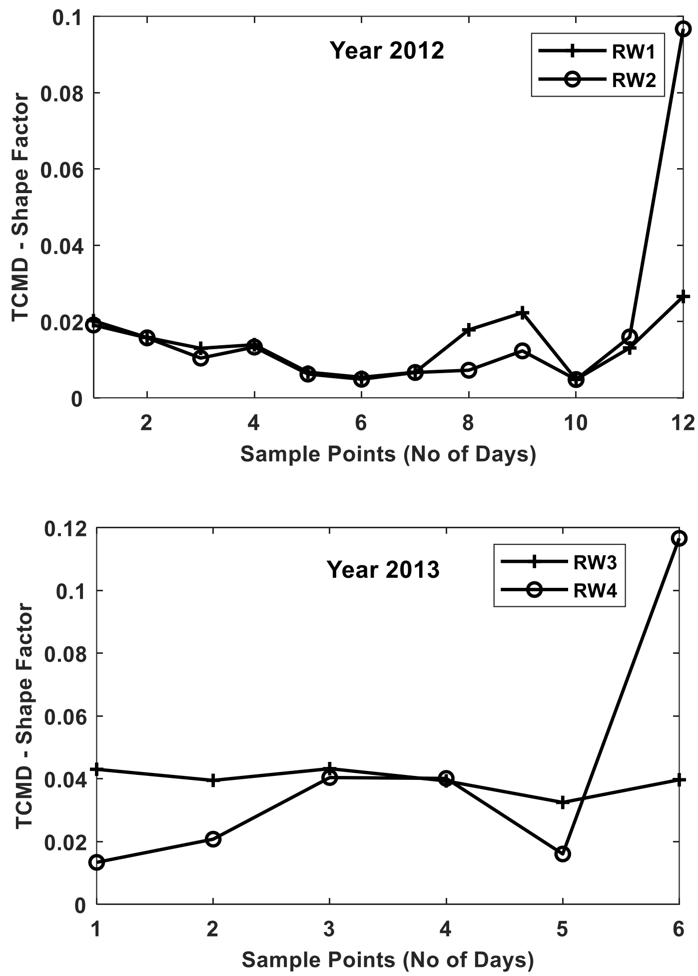

Figure 3.8 (a) Time Series Plot of Torque Command - Shape Factor for RW 1,2 and RW 3,4

From the observations the torque command was increased after the 11^{th} sample set which is 16 days before the failure of the RW2, similarly, for the second case in 2013, it was increased before 5 days of the RW4 failure as shown in Figure 3.8 (a) respectively.

All the 4 reaction wheels are mounted on the baseplate (exterior) of the vehicle. Throughout the Kepler quarter, they are transitioned from sunlit to dark area and vice versa. By this orientation, the wheels are kept warmer, but there were large variations of temperature depending on the time of the quarter [102]. The rotor temperature profile for the RW2 started to increase from approx. 2 months before the anomaly (Figure 3.8 (b)). After this failure event, each reaction wheel pair was coupled with a heater and controlled by flight software, thus there were no variations in the rotor temperature from RW4 prior to the failure (Figure 3.8(b)).

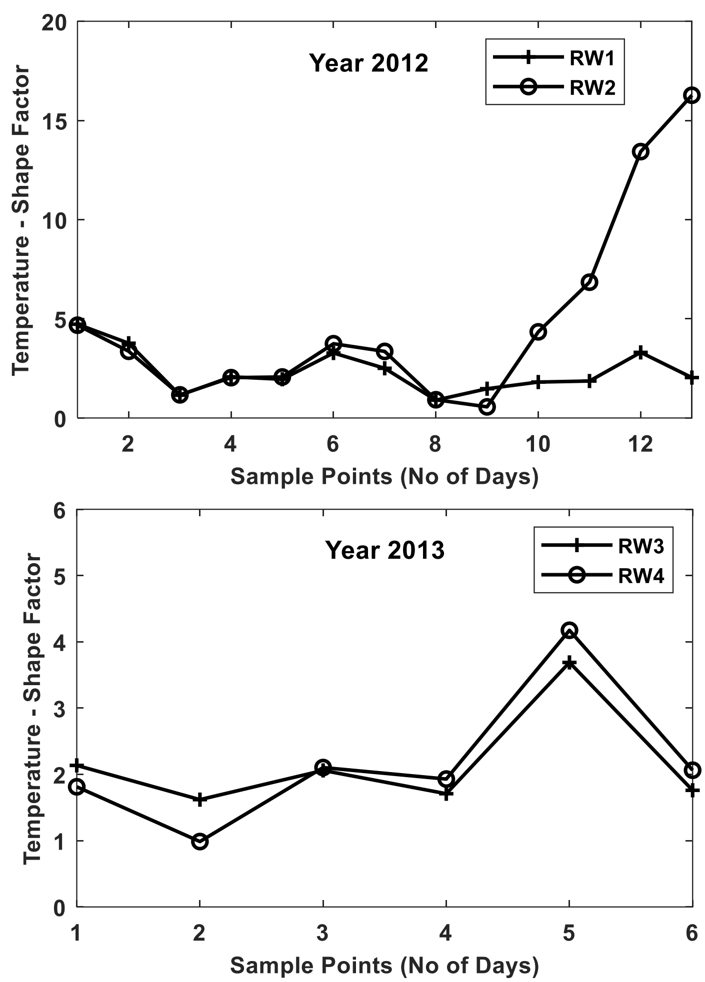

Figure 3.8 (b) Time Series Plot of Temperature - Shape Factor for RW 1,2 and RW 3,4

The first failure leads the team to examine the other three wheels with and an intense effort. In April of 2013, before 10 days of the failure warning signs of torque friction started to peak up in RW4 as shown in Figure 3.8 (c).

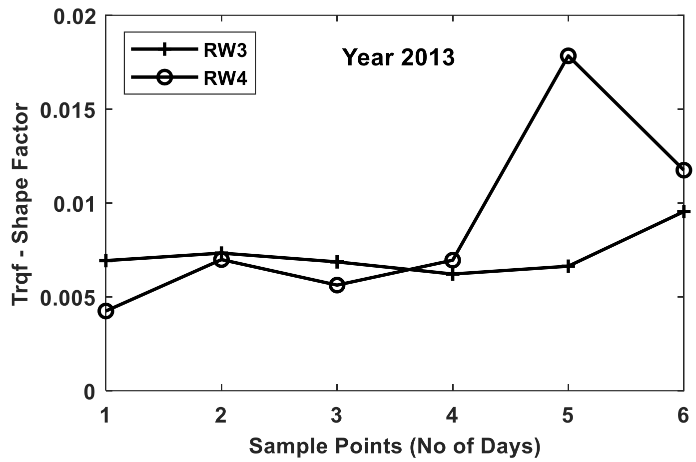

Figure 3.8 (c) Time Series Plot of Torque Friction - Shape Factor for RW 3,4

3.3.4 Friction Model

The dry friction is not available from the telemetry data, instead, it has to be calculated based on the total frictional torque (TRQF) with the speed of the wheel (SPD).

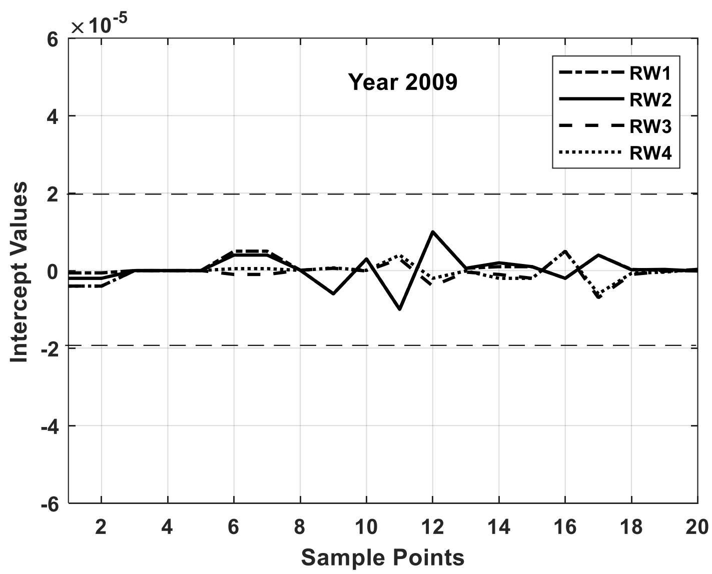

Figure 3.9 Intercept Plots for all the RW Assemblies in 2009

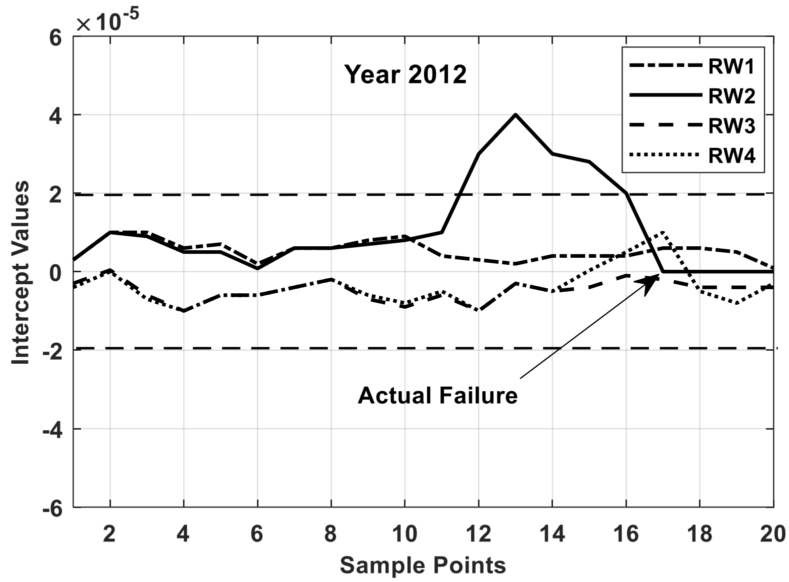

Figure 3.10 Intercept Plots for all the RW Assemblies in 2012

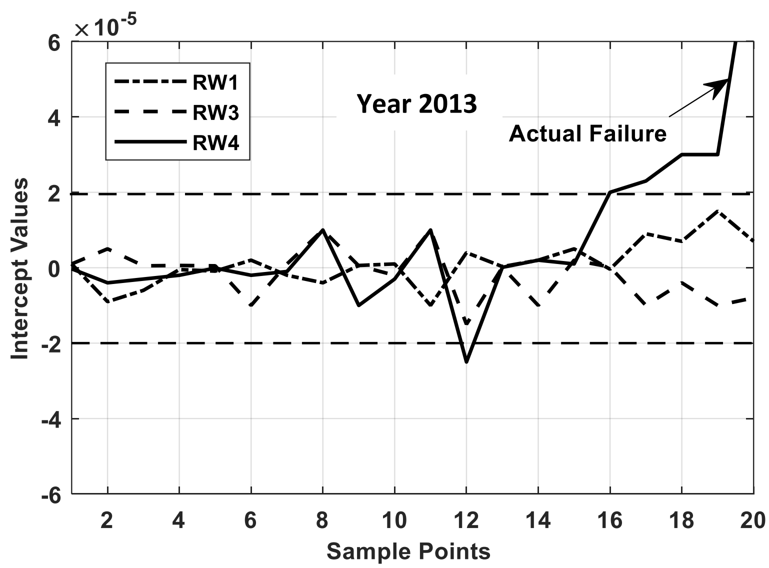

Figure 3.11 Intercept Plots for all the RW Assemblies in 2013

Speed vs T-friction plot is a line on either side of zero speed depending on the configuration. Intercept from Torque friction and speed is obtained for the same sample used for Weibull analysis. From Figures 3.9, 3.10, 3.11, it is clearly noticeable that, if a maximum and minimum trend line is fixed at 2E-5 and -2E-5. For the year 2009 everything was under control, but for 2012 (sample size - 16 days) the intercept of friction for the reaction wheel 2 crossed the limit approximately 2 months earlier of the failure and similar trend was noticed for the year 2013 (sample size – 5 days) approximately 1.5 months ahead of the reaction wheel 4 failure.

Dry friction (τ)= | Torque Friction | - |Slope $_{avg}$ * Wheel Speed| - Intercept $_{avg}$

Table 3.2 Dry Friction Values for the Given Telemetry Set

		RW1	RW2	RW3	RW4
2009	Min τ	0.44	0.41	0.47	0.46
	Max τ	11.39	8.03	8.67	10.33
2012	Min τ	0.61	0.46	0.67	0.61
	Max τ	11.38	18.13	10.41	12.13
2013	Min τ	0.68	N/A	0.57	0.77
	Max τ	11.69	N/A	12.13	19.11

The average slope and intercept values are calculated from the friction vs speed graphs and the baseline dry friction for a healthy Reaction wheel from Table 3.2 are 0.41 mNm and 12.13 mNm.

3.4 Remaining Useful Life – Classical Method

A system is said to be failed if it completely stops functioning its task. The satellites are more vulnerable to interruptions and external disturbances, also it is very difficult to perform maintenance in the outer space. Prediction of remaining useful life enables the team members to reconfigure the functionality of the system promptly and eliminates the total loss of mission.

In this case, the initial set of data (20 days) after 100 days of the launch (2009) is taken for consideration, the apriori knowledge is extracted and is extrapolated to the next 100 days. It is clear from Figure 3.12 that the value for RW2 approaches the threshold (2E-5) relatively faster than the other wheels, which can be taken as an initial warning sign for the failure.

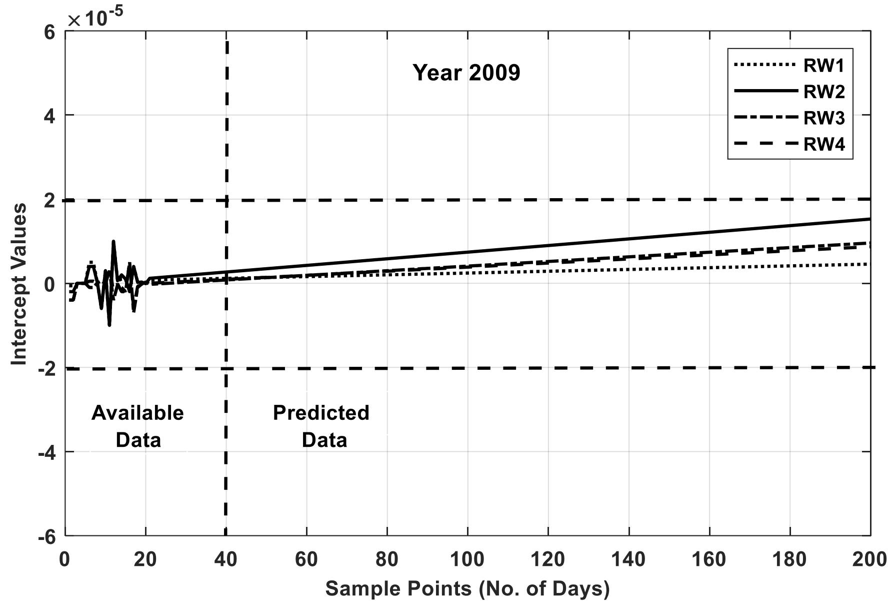

Figure 3.12 Degradation of the Intercept Values for all the RW Assemblies

From the friction model, the day at which intercept value crossed the threshold limit is established, from that point in time the data is extracted for RUL prediction. Primarily, the data is normalized with the difference in the mean value, such that every sample set from the onset of failure will depict the same degradation.

The cycle of rising and dip is every 20 days in the data for reaction wheel 2, thus 20 days sample set is used and is extrapolated for the next 80 days using the curve fit details.

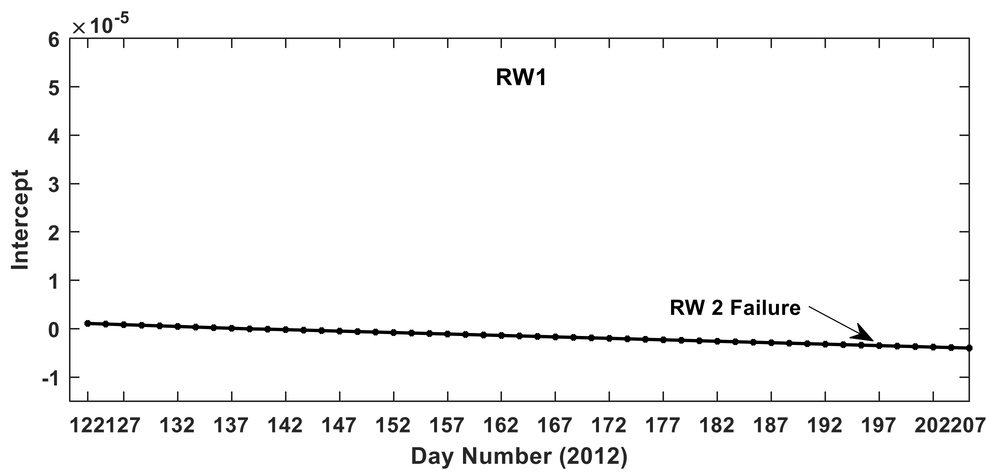

Figure 3.13 Prediction of Healthy Reaction Wheel - RW1

For RW 1 and RW2, the sample is taken from 120th day till 140th day, the intercept values remain almost the same for RW1 as shown in Figure 3.13, for RW2 it crossed the limit on 170th day as shown in Figure 3.14.

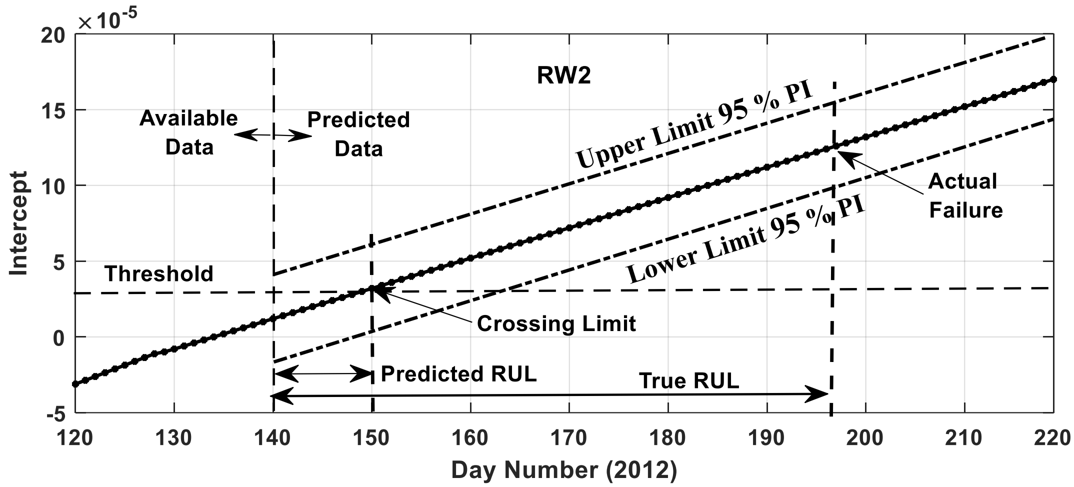

Figure 3.14 Remaining Useful Life for RW2 - 2012

Similarly, there is a cycle change every 10 days in the data for reaction wheel 4, thus 10 days sample set is used and extrapolated for the next 40 days using the curve fit details. For RW 3 and RW4, the sample is taken from 100th day till 109th day, the intercept values remain almost the same for RW3 as shown in Figure 3.15, for RW4 the limit is crossed on the 119th day as shown in Figure 3.16.

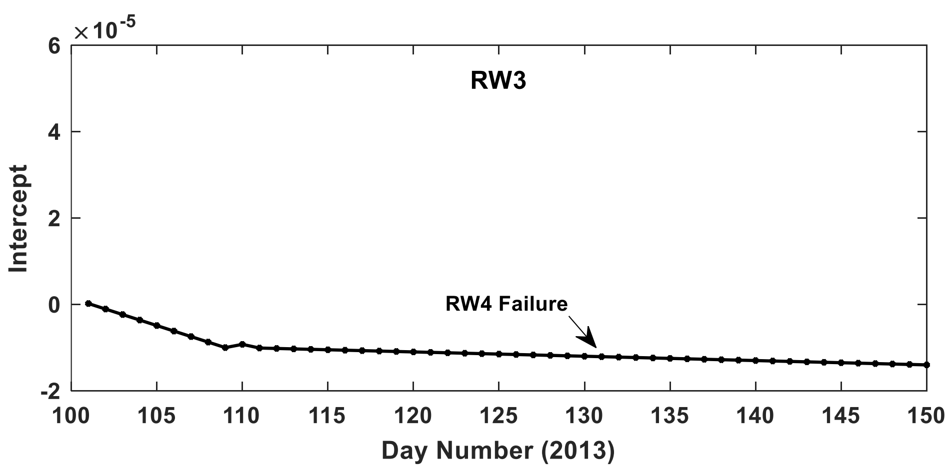

Figure 3.15 Prediction of Healthy Reaction Wheel - RW3

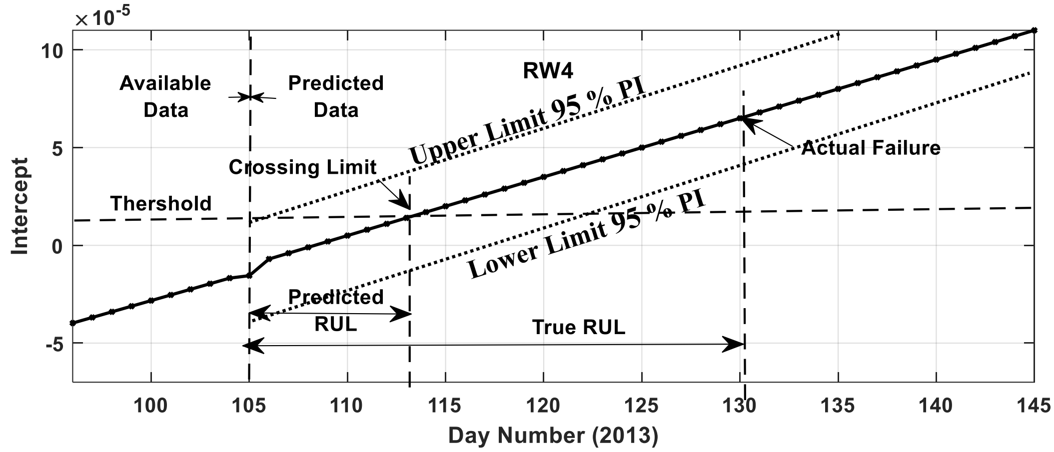

Figure 3.16 Remaining Useful Life for RW4 - 2013

A polynomial function was fit into the data and the remaining useful life was predicted and compared with the true RUL. In the statistical perspective, prediction Interval (PI) is used to measure the upper and lower bounds of confidence intervals for future observations.

$$PI = Mean \pm z * standard\ deviation$$

where z is the normal distribution curve value (for this case 95% confidence interval is used and thus the value of z is 1.96).

The actual and estimated life is represented with the percentage error in Table 3.3 as shown.

Table 3.3 Remaining Useful Life After Fault Detection: (Classical Method)

Reaction Wheel Assembly	True RUL (days)	Estimated RUL (days)	Percentage Error (%)
RW2	56	31	44.64
RW4	22	10	54.54

The most common problem for data analysis is the overfitting problem. The model has to be created with an accurate degree of the polynomial. For the appropriate selection, the root mean square (RMS) error for the functions with degree 1 to 8 is calculated. From the test performed, the error is comparatively lesser for the 1st degree as shown in Figure 3.17, for all the years. And thus 1st-degree polynomial is used for prediction analysis.

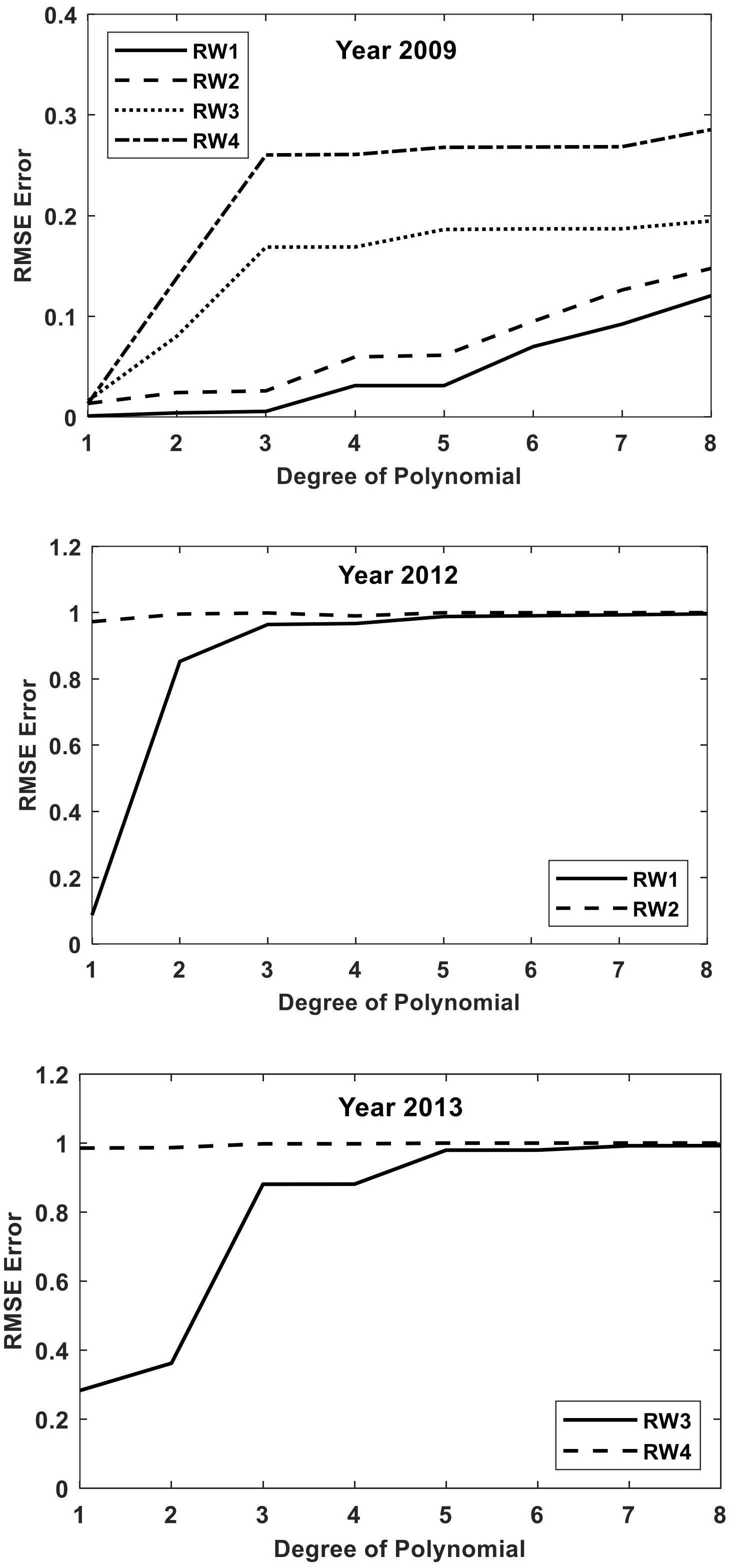

Figure 3.17 Justification for 1st Degree of Polynomial Selection

The actual failure date from the telemetry data set and the predicted failure date from the analyses are formulated and represented in Table 3.4 as shown below.

Table 3.4 Remaining Useful Life Prediction Results for RWAs

Reaction Wheel Assembly	Actual Date of Failure	Predicted Date of Failure
RW1	Operational	Beyond 5 years
RW2	July 14, 2012 (196th day)	May 11, 2012 (132nd day)
RW3	Operational	Beyond 5 years
RW4	May 11, 2013 (131st day)	March 22, 2013 (81st day)

3.5 Long-Short Term Memory (LSTM) - Results

A Long-Short Term Memory (LSTM) variant of Recurrent Neural Network (RNN) was used to implement supervised time-series forecasting using the sliding window technique to predict future values and faulty conditions.

3.5.1 Model Validation - Healthy Data

The model is trained and validated with the healthy dataset (year-2009) by comparing the train and test dataset for all the given sensor values such as rotor speed, temperature, torque command, torque friction. Being the model learned all parameters, it is used to predict the test sequences one at a time. Figures 3.18 to 3.21 represent the error plots and are seen that the rate of the testing error is lesser than the training error which is considered to be a good fit and is later used to predict the fault that occurred in Kepler's mission.

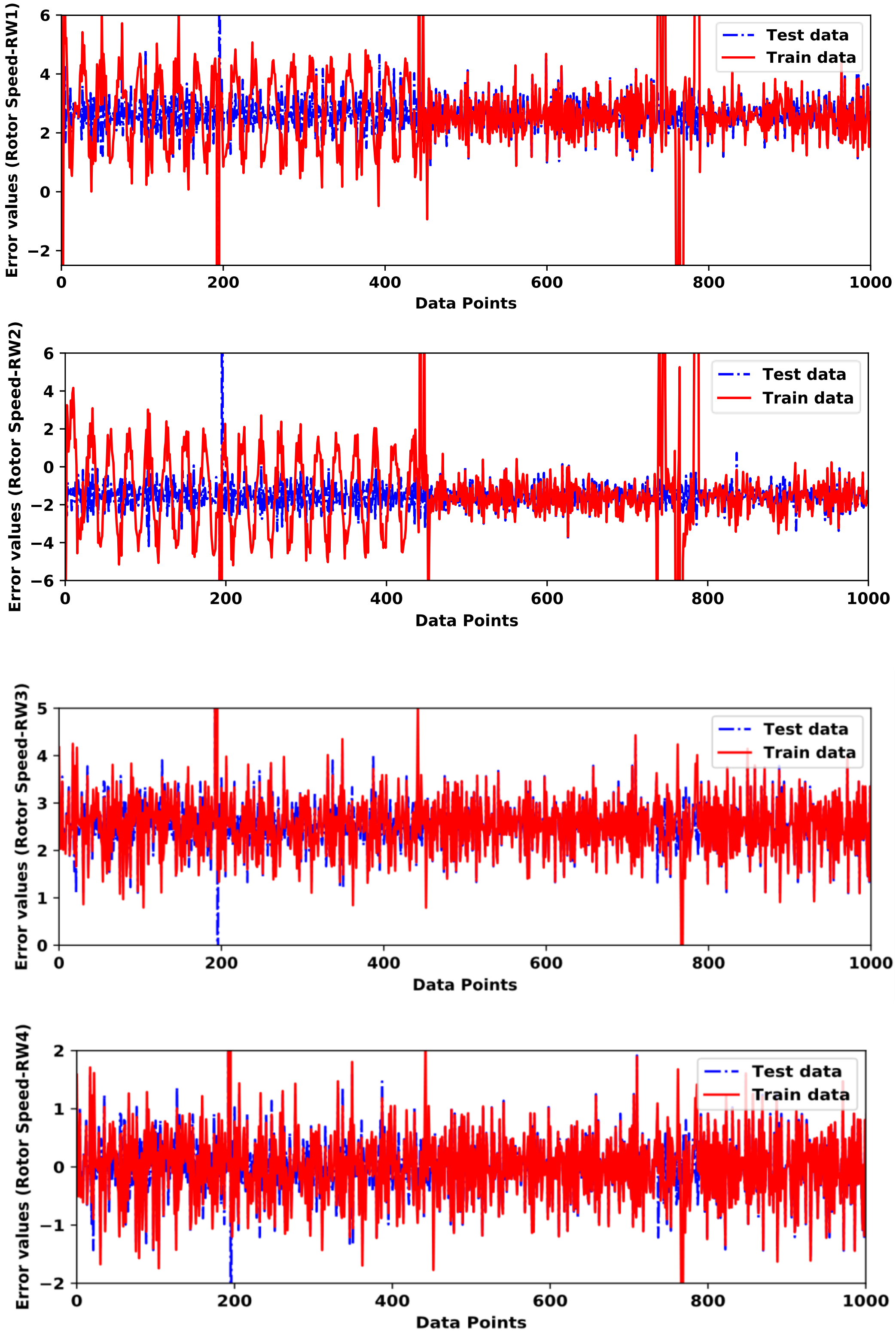

Figure 3.18 Error Plots of Rotor Speed for all Four Reaction Wheels

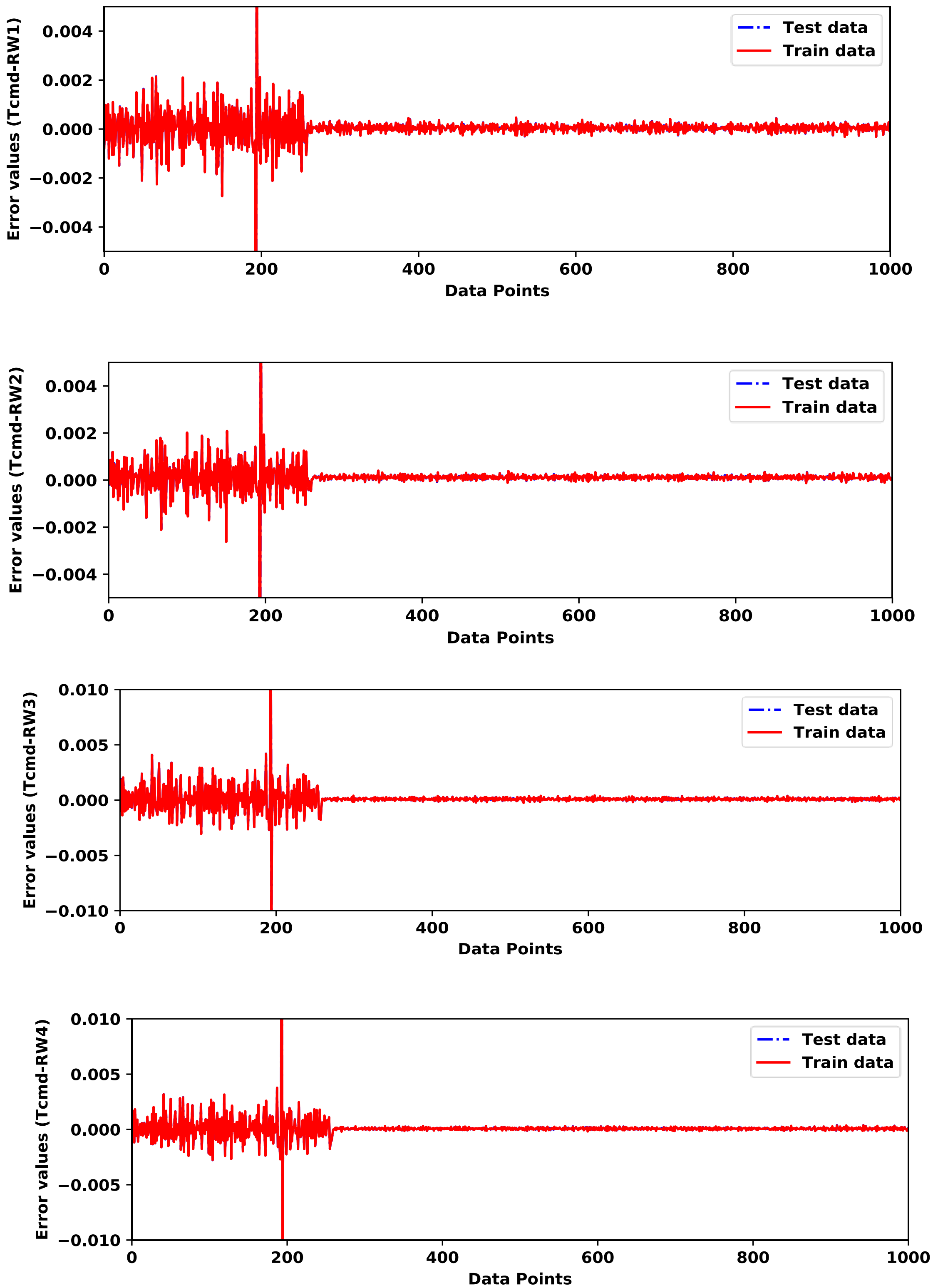

Figure 3.19 Error Plots of Torque Command for all Four Reaction Wheels

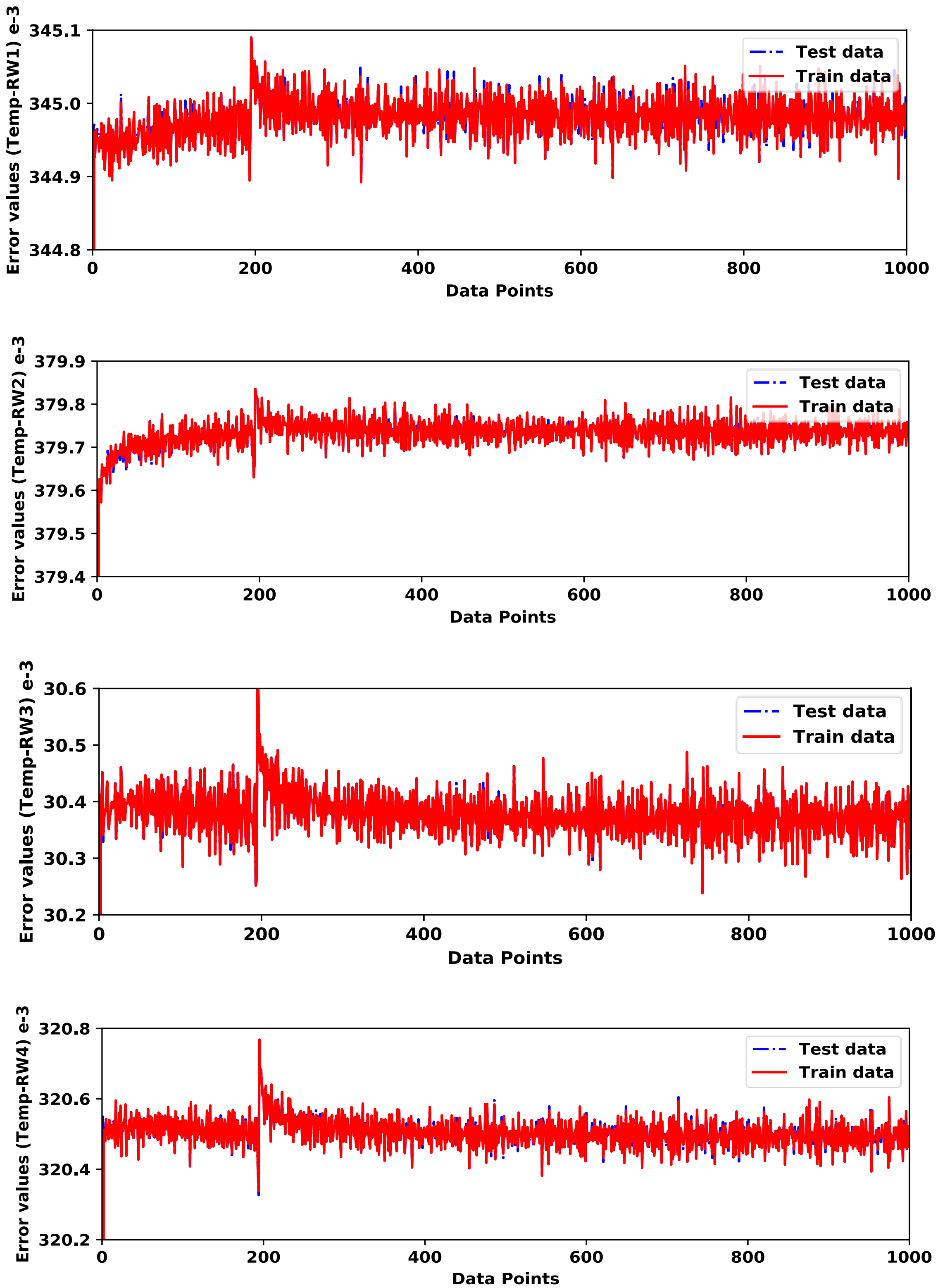

Figure 3.20 Error Plots of Temperature for all Four Reaction Wheels

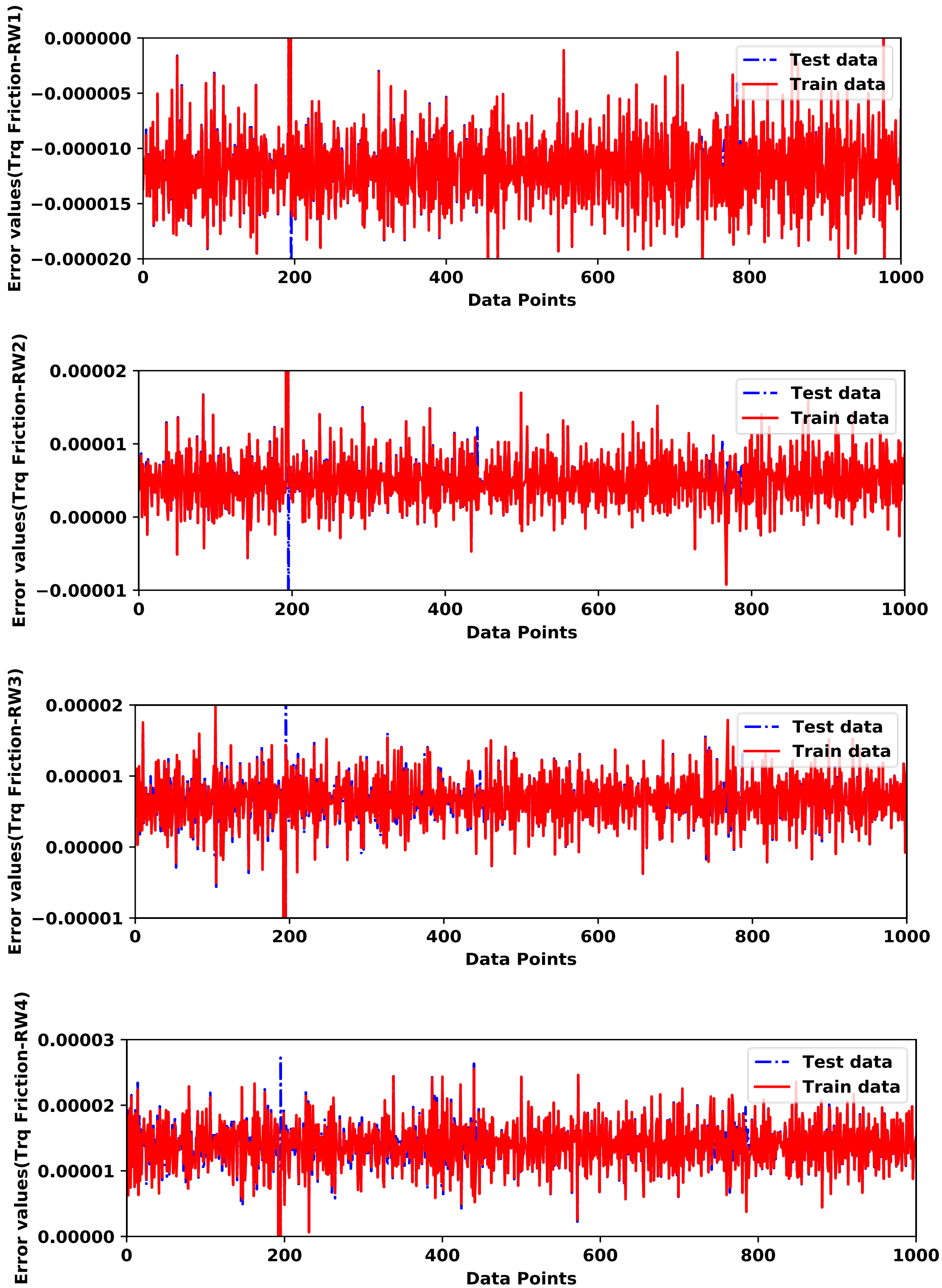

Figure 3.21 Error Plots of Torque Friction for all Four Reaction Wheels

3.5.2 Fault Detection - Actual Data

An LSTM model was developed which would input speed, temperature, torque command and torque friction values from the previous time steps to predict the respective features in the following time step using sliding window technique. The model was trained and validated using data of the year 2009 as shown in the previous section. The same model is utilized to predict the sensor values of the year 2012. The differences in the predicted output from the model and actual faulty output are calculated and the sudden change is captured to identify the fault in the RW2 and RW4 as shown below.

Figure 3.22 and Figure 3.23 show the error signals of temperature profile and torque command detecting the abrupt fault of RW2 at the data point 205, which is an equivalent time frame on the exact failure date (July 14, 2012).

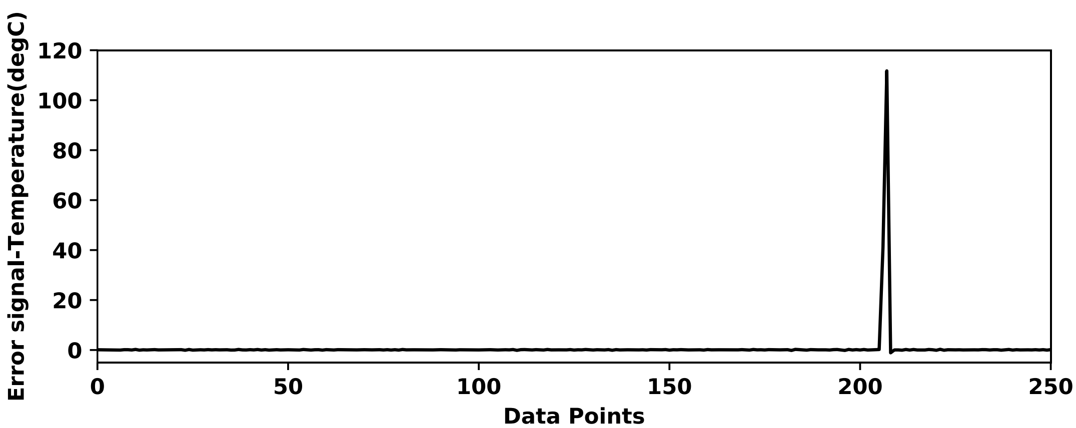

Figure 3.22 Error Signal: Abrupt Fault - Temperature of RW2

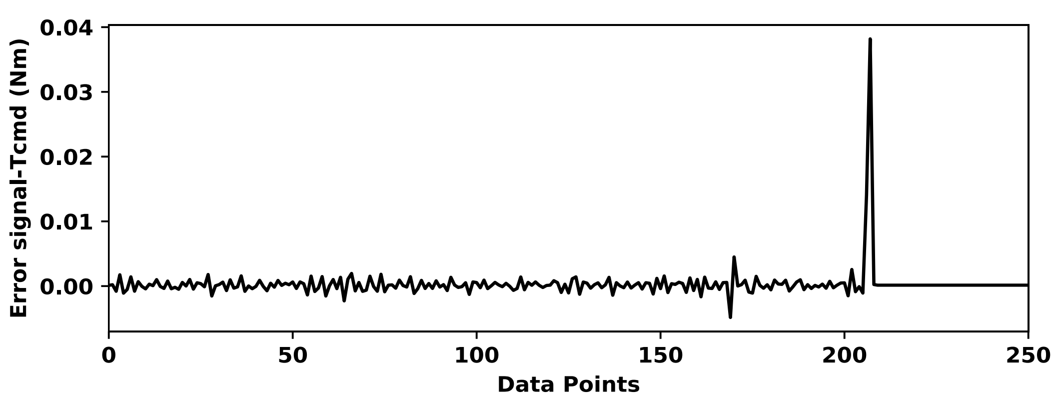

Figure 3.23 Error Signal: Abrupt Fault - Torque command of RW2

Similarly, Figure 3.24 to 3.27 represents the error signals of rotor speed, temperature profile, torque command and torque friction detecting the abrupt fault for RW4. The timeline equivalent of May 11, 2013, in the corresponding plots is at the 610th datapoint.

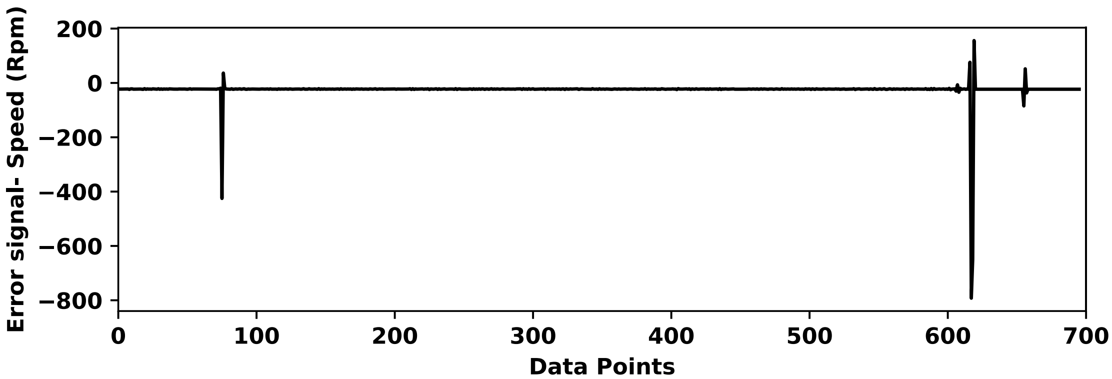

Figure 3.24 Error Signal: Abrupt Fault - Rotor Speed of RW4

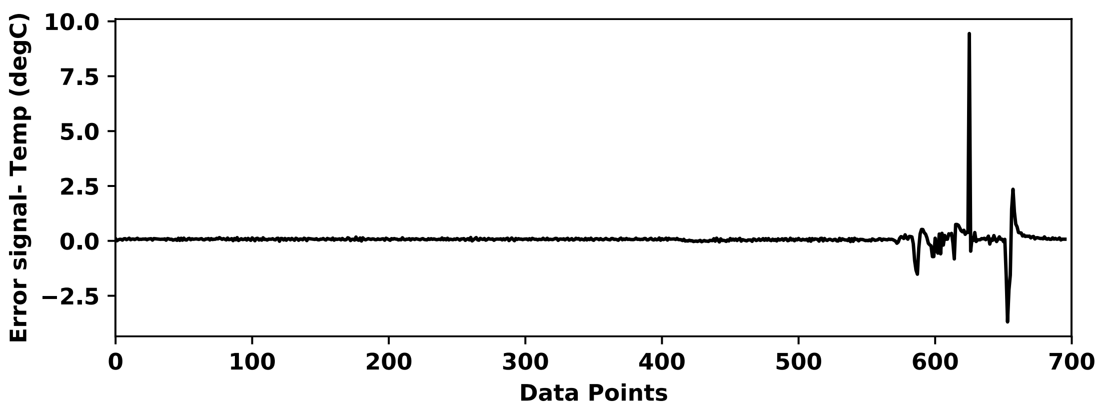

Figure 3.25 Error Signal: Abrupt Fault - Temperature of RW4

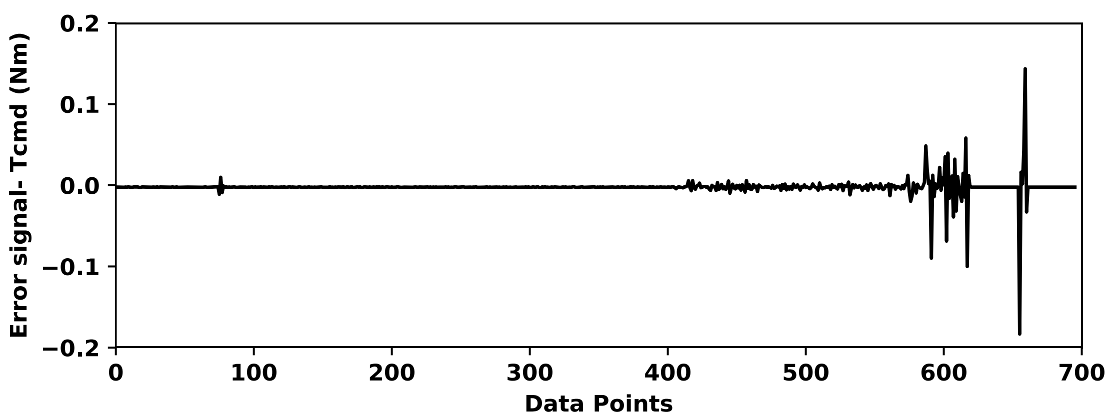

Figure 3.26 Error Signal: Abrupt Fault - Torque Command of RW4

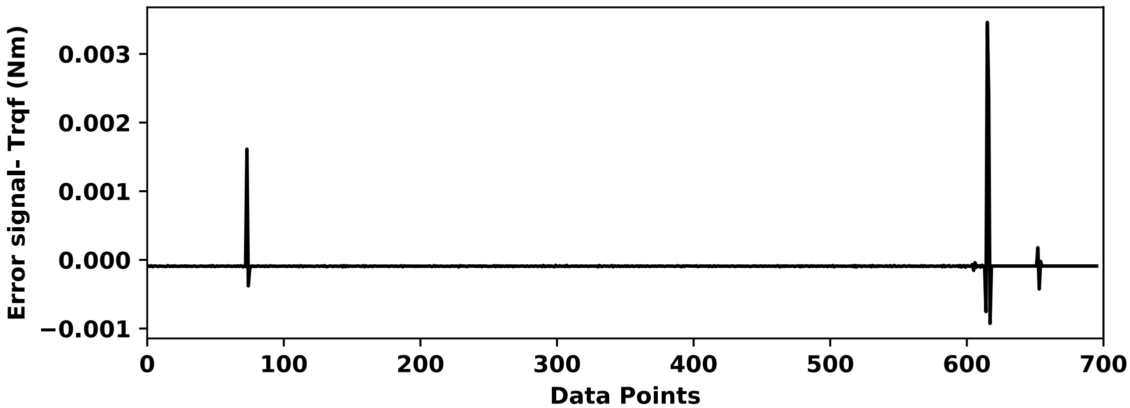

Figure 3.27 Error Signal: Abrupt Fault - Torque Friction of RW4

In addition to the fault detection, the plots of speed and corresponding torque frictional data showed a disruption signal. It is confirmed later that the rotor speed was changed on at that particular time to reorient the spacecraft and is hence neglected.

Also, the sample plots for the healthy reaction wheels RW1 and RW3 show no change in the error signal for the following years 2012 and 2013 respectively as shown in Figure 3.28 to Figure 3.33.

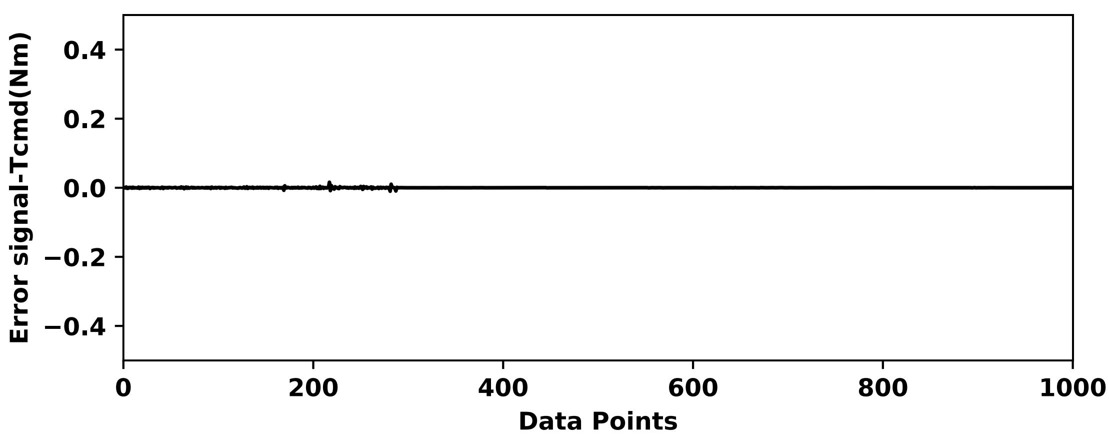

Figure 3.28 Error Signal: Healthy Wheel - Torque Command of RW1

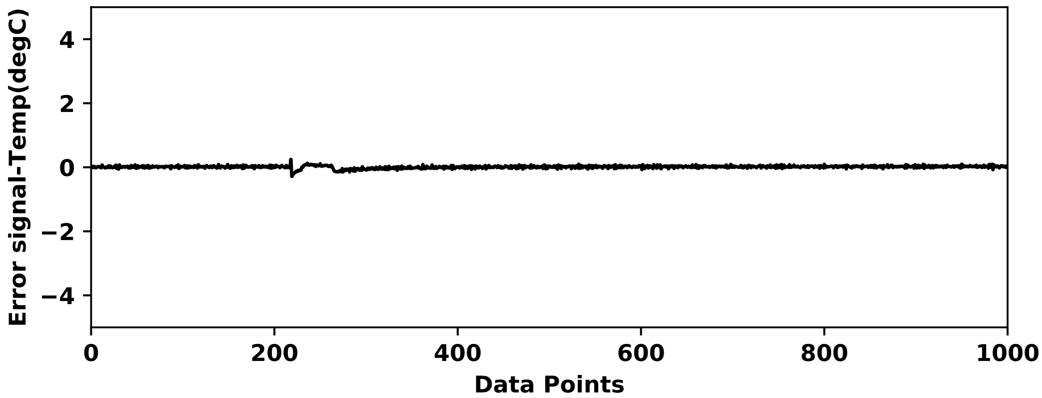

Figure 3.29 Error Signal: Healthy Wheel - Rotor Temperature of RW1

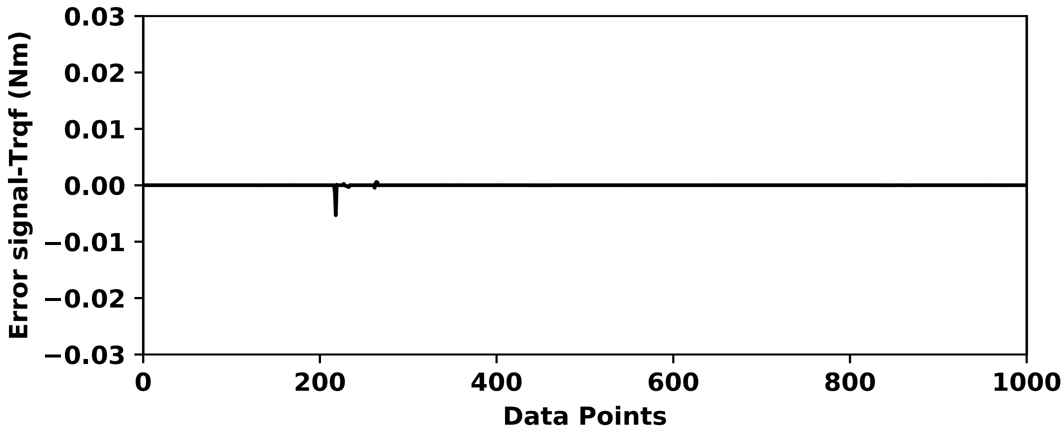

Figure 3.30 Error Signal: Healthy Wheel - Torque Friction of RW1

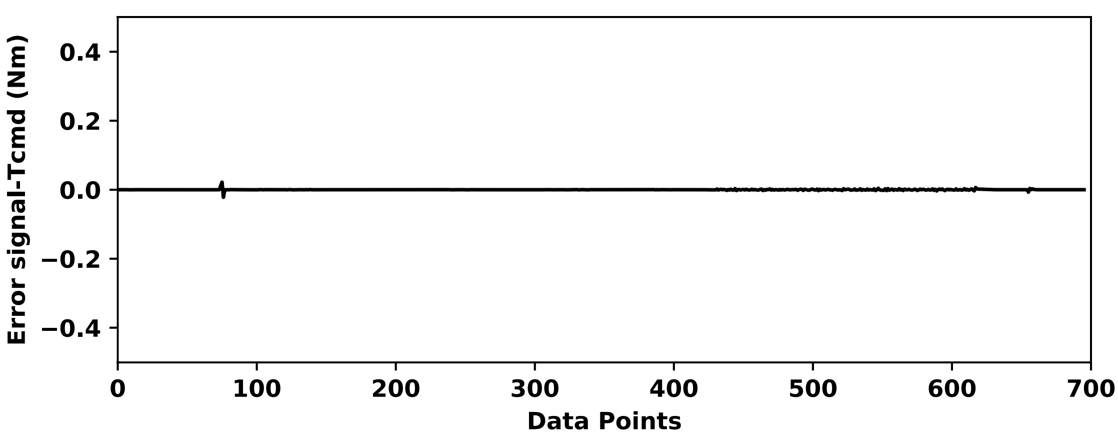

Figure 3.31 Error Signal: Healthy Wheel - Torque Command of RW3

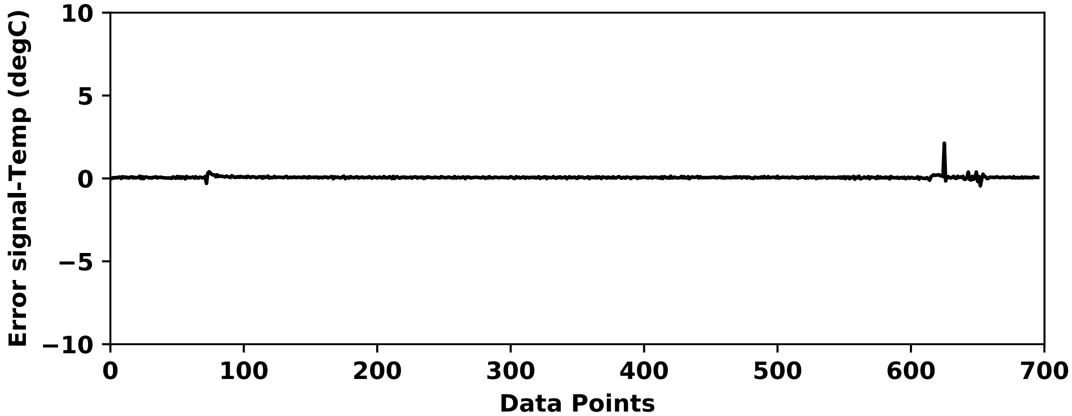

Figure 3.32 Error Signal: Healthy Wheel - Rotor Temperature of RW3

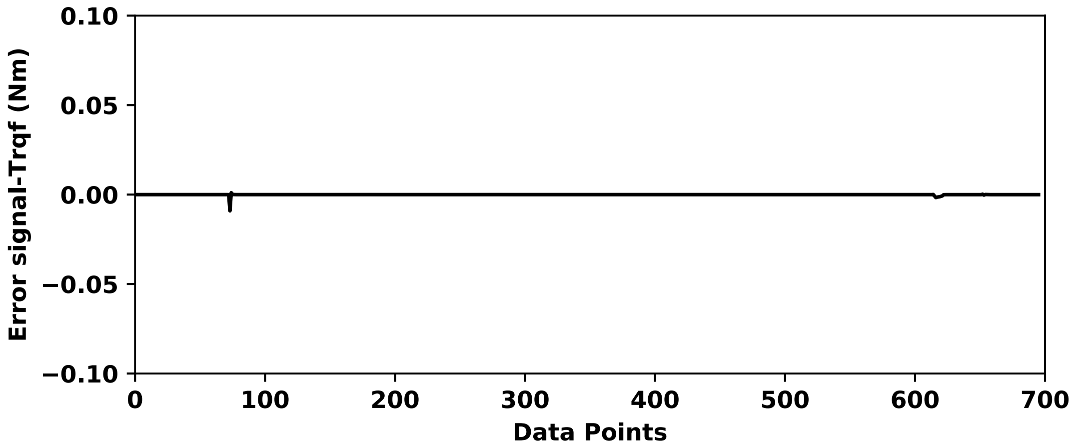

Figure 3.33 Error Signal: Healthy Wheel - Torque Friction of RW3

3.5.3 Fault Detection - Weibull Data

In this segment, the correlated Weibull parameters of torque command is used as the input to the model for predicting the abnormality in the reaction wheels 2 and 4. The results of all the four wheels are as shown in Figure 3.34 to Figure 3.37 respectively and are clearly predicting the fault in accordance with the classical methods.

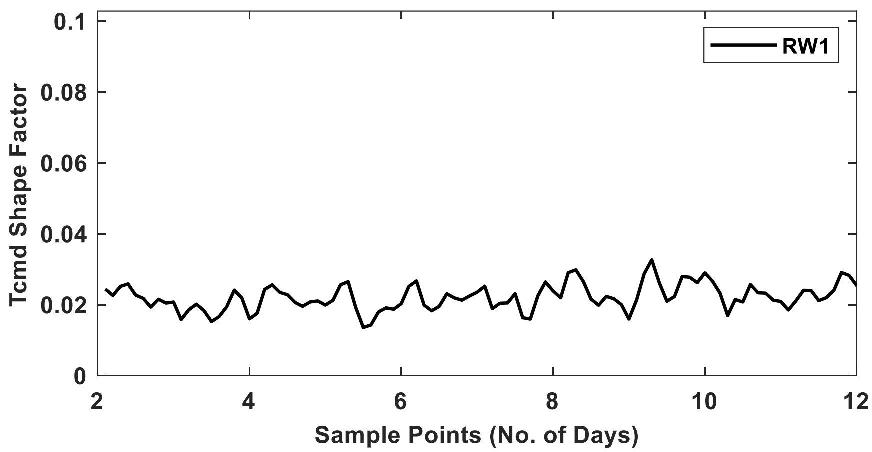

Figure 3.34 RW1: LSTM - Time Series Plot of Tcmd - Shape Factor (Healthy)

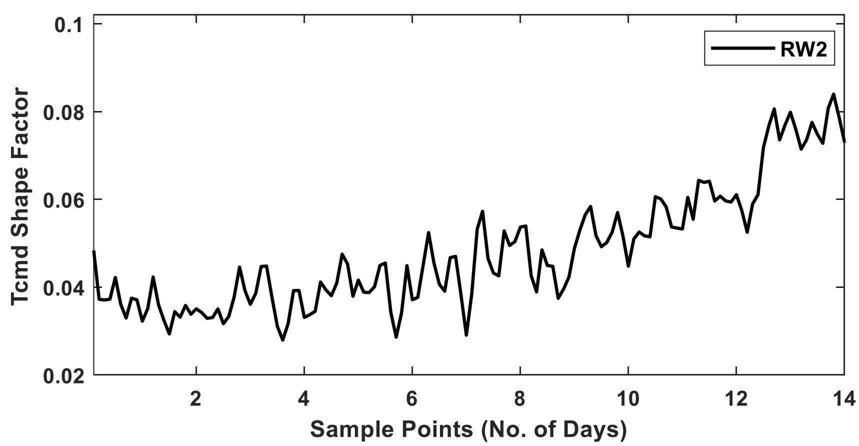

Figure 3.35 RW2: LSTM - Time Series Plot of Tcmd - Shape Factor (Faulty)

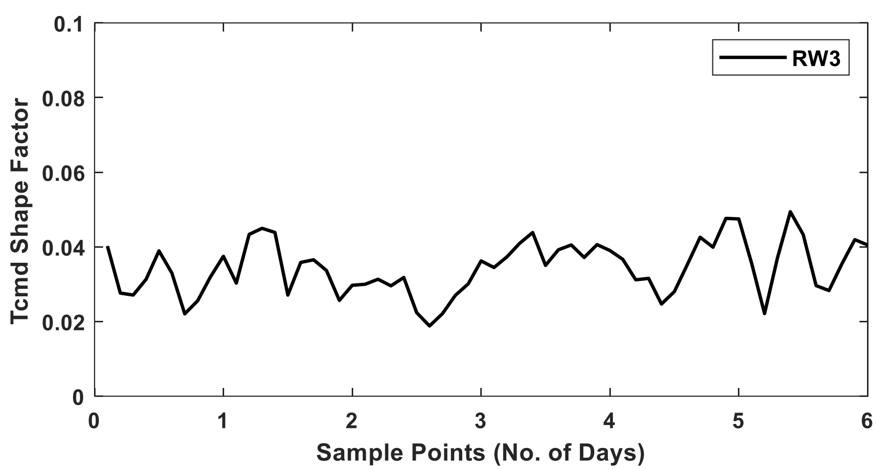

Figure 3.36 RW3: LSTM - Time Series Plot of Tcmd - Shape Factor (Healthy)

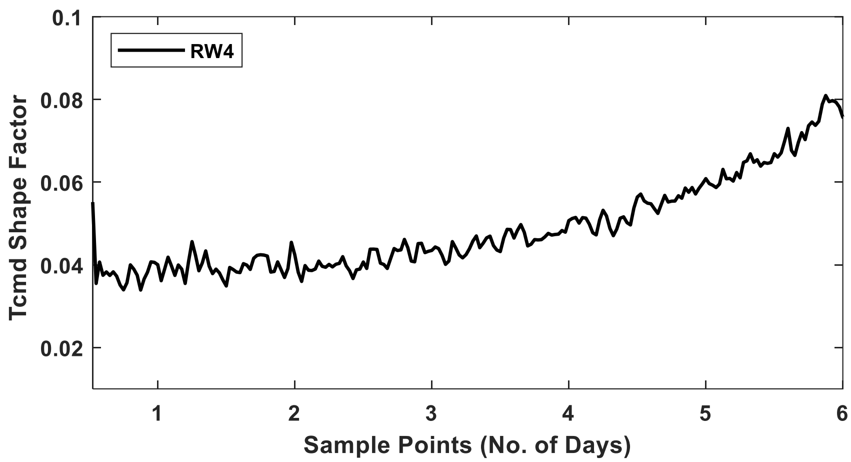

Figure 3.37 RW4: LSTM - Time Series Plot of Tcmd - Shape Factor (Faulty)

3.5.4 Fault Detection – Intercept Values

Similarly, in this section, the intercept values calculated using the torque friction are used as the input for predicting the fault in advance. The results of all the four wheels are as shown in Figure 3.38 to Figure 3.41 respectively and the numerical values in comparison with the statistical results are given in Table 3.5.

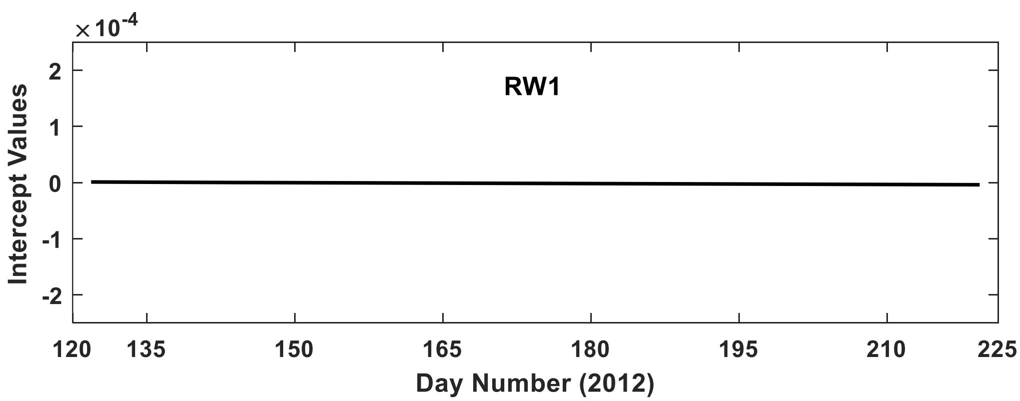

Figure 3.38 RW1: Torque Friction Intercept Values of 2012 - No-fault

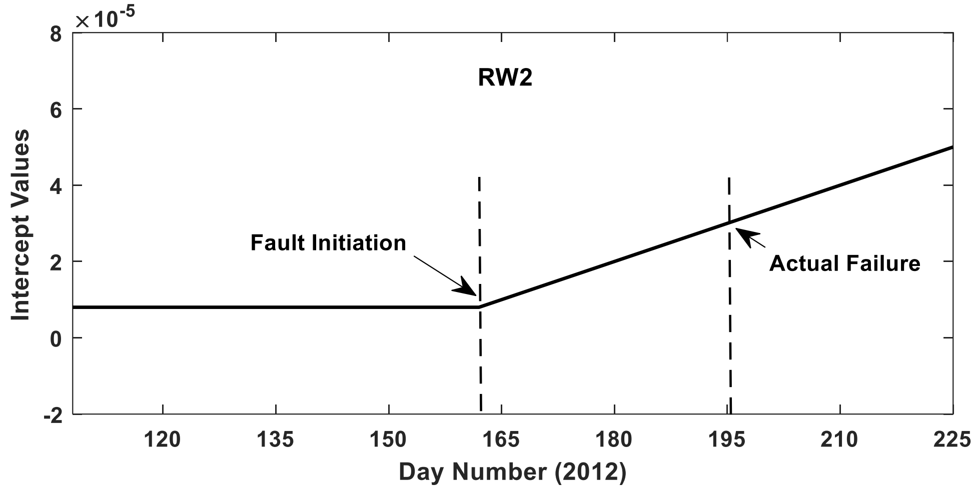

Figure 3.39 RW2: Torque Friction Intercept Values of 2012 - Fault Detection

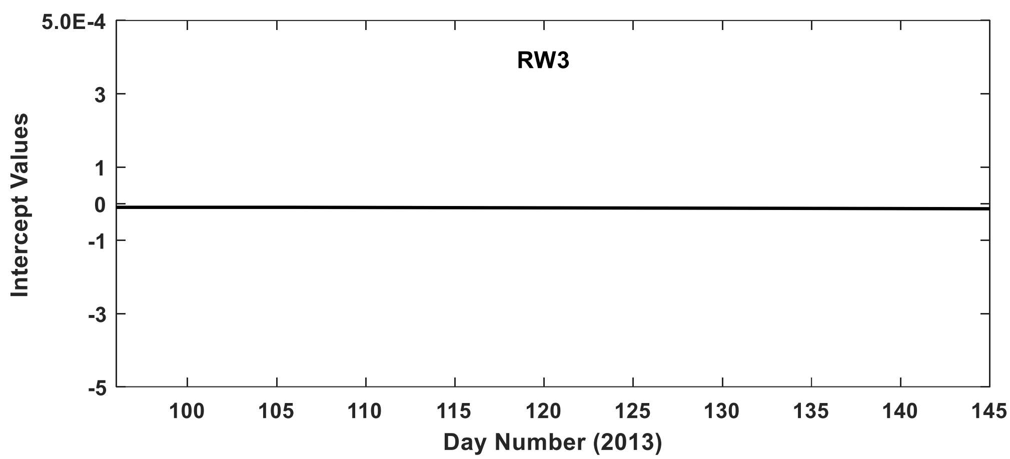

Figure 3.40 RW3: Torque Friction Intercept Values of 2013 – No-Fault

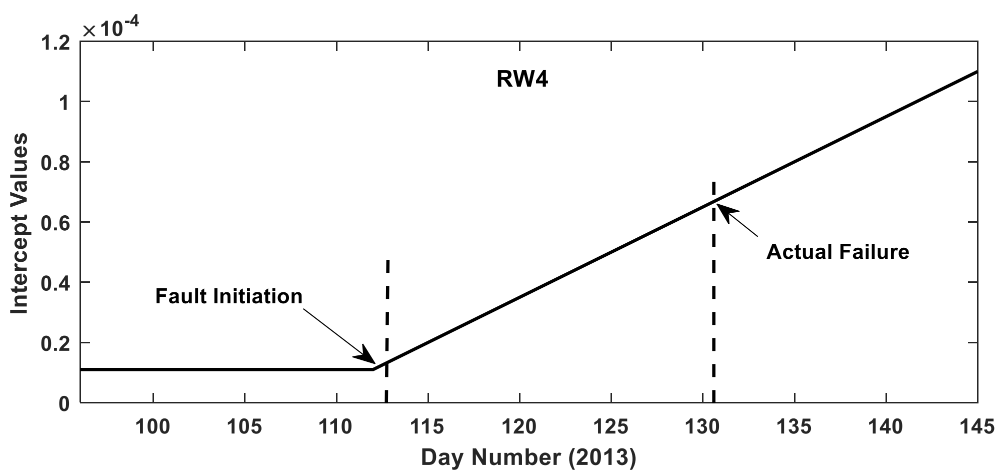

Figure 3.41 RW4: Torque Friction Intercept Values of 2013 - Fault Detection

Table 3.5 Remaining Useful Life After Fault Detection: Intercept (Data-Driven)

Reaction Wheel Assembly	True RUL (days)	Estimated RUL (days)	Percentage Error (%)	Actual Date of Failure	Predicted Date of Failure
RW1	N/A	N/A	N/A	Operational	Beyond 5 years
RW2	56	40	28.57	July 14, 2012 (196th day)	June 4, 2012 (156th day)
RW3	N/A	N/A	N/A	Operational	Beyond 5 years
RW4	22	15	31.81	May 11, 2013 (131st day)	April 26, 2013 (116th day)

3.5.5 RUL - Prediction with Weather Data

To train the model in order to predict the remaining useful life with the given sensor values are insufficient. From the obtained data, the total number of faulty events is only two (RW2 and RW4) out of four and both are differently induced. Many unidentified bearing failures and friction anomalies of the reaction wheels on the spacecraft lead to an analysis of the space charging environment by United technologies corporation [105]. In this section, the space weather data of Coronal Mass Ejection (CME) is obtained in terms of the Kp Index [106]. The 3-hours Kp values are converted to Ap Index and are averaged to each day which is plotted for each year from 2009 to 2013 as shown in Figure 3.42.

From the 2012 plot, it is clearly noted the Ap index reached the highest value around 80 ahead of 100 days to failure of RW2 (196th day). On the exact failure date, the value of the Ap index is 90 (41st largest geomagnetic storm since 1994) and this is used as an indication. Similarly, in the year 2013, the Ap index (63) was high on 78th day that is approx. 1.5 months before the RW3 failure (131st day). From this it can be clearly stated there is a strong correction between space weather and bearing issues. Hence the weather data is used for predicting the remaining useful life of the reaction wheel 2 and reaction wheel 4 with the adaptive threshold assigned for all the five years based on the data using the prediction interval technique.

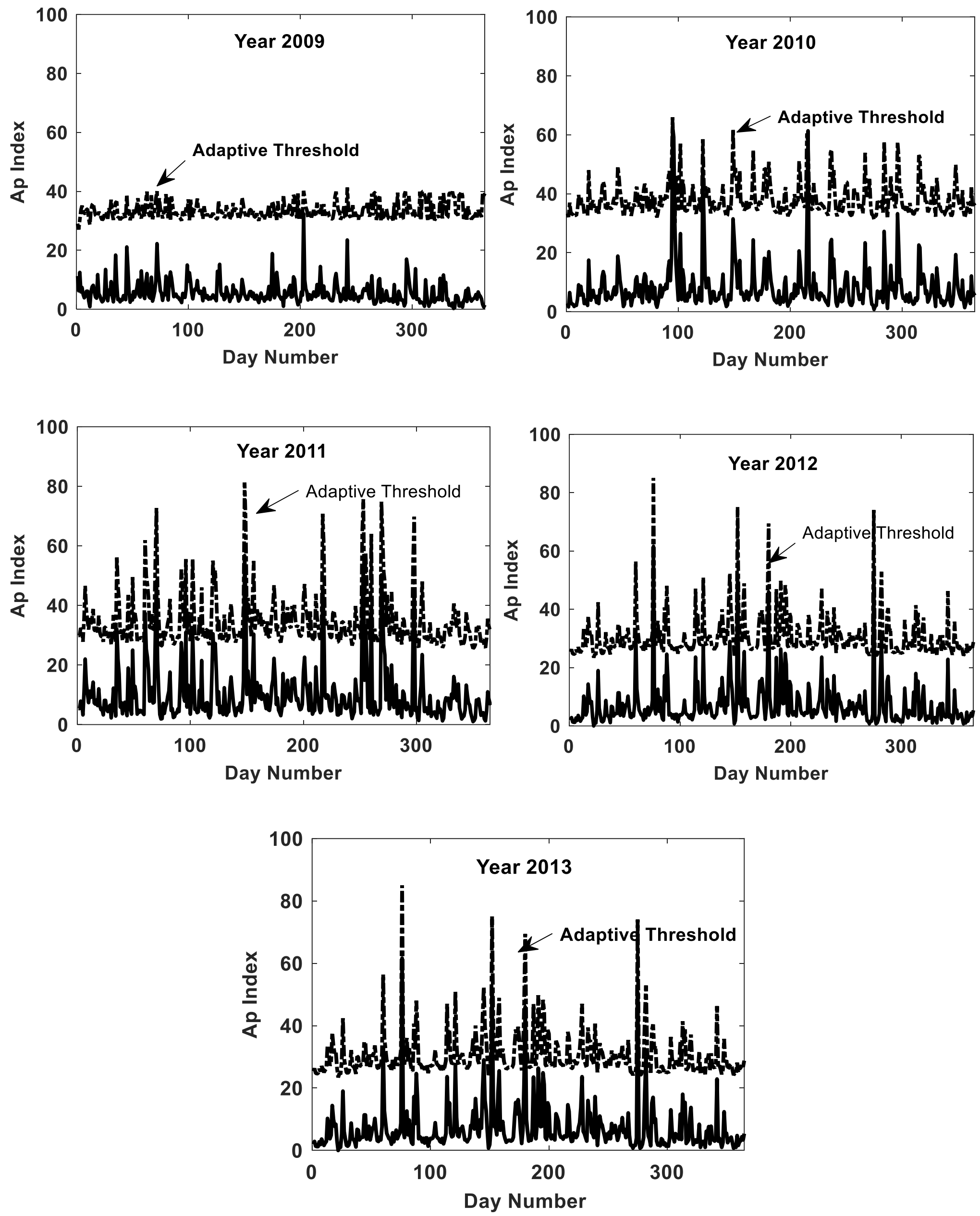

Figure 3.42 Plots of Averaged Ap Index for the Year 2009 to 2013

The LSTM model is simulated for predicting the remaining useful life of the two respective reaction wheels 2 and 4 using the Ap index values. For RW2, the daily averaged data for the year 2012 is plotted for the actual and predicted values. From Figure 3.43, it is observed the predicted value is 15 days ahead of the actual failure event. Similarly, the model predicted the RW4 incident 14 days prior to the malfunction (Figure 3.44). Figures

3.45 and 3.46 is a supportive plot for the other 2 healthy reaction wheels for 2013 until the given data points.

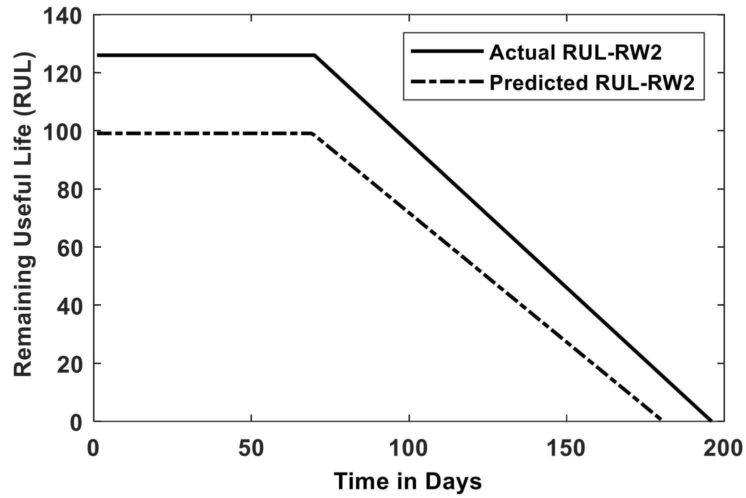

Figure 3.43 Comparison Plot of Actual and Predicted RUL (RW2) - 2012

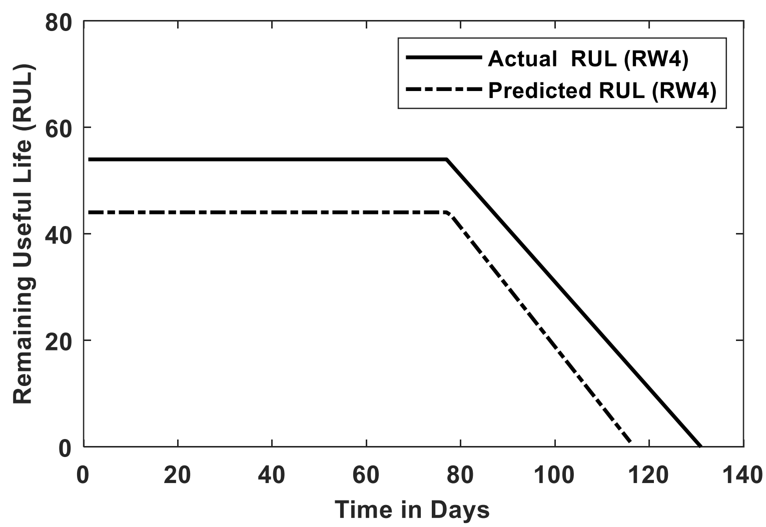

Figure 3.44 Comparison Plot of Actual and Predicted RUL (RW4) - 2013

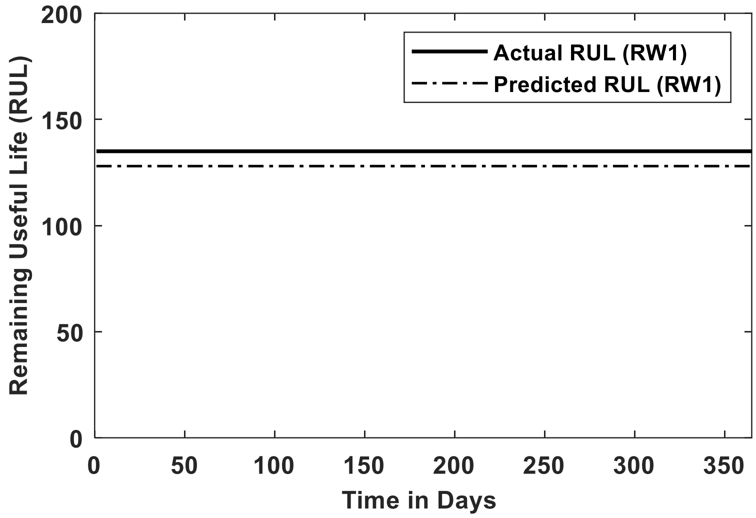

Figure 3.45 Comparison Plot of Actual and Predicted RUL (RW1) - 2013

Table 3.6 shows the actual failure date from the given data and the predicted failure date from the LSTM model.

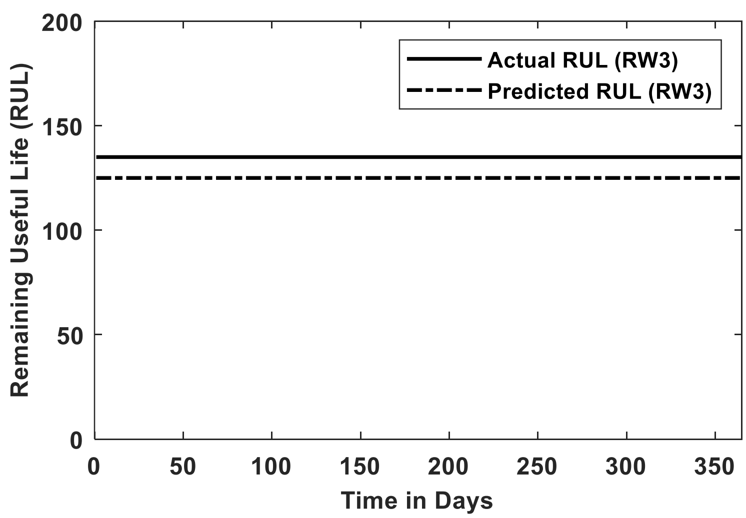

Figure 3.46 Comparison Plot of Actual and Predicted RUL (RW3) - 2013

Table 3.6 LSTM - Remaining Useful Life Prediction Results for Faulty Wheels

Reaction Wheel Assembly	Actual Date of Failure	Predicted Date of Failure	Predictions Ahead
RW1	Operational	N/A	N/A
RW2	July 14, 2012 (196th day)	June 29, 2012 (181st day)	15 days
RW3	Operational	N/A	N/A
RW4	May 11, 2013 (131st day)	April 27, 2013 (117st day)	14 days

3.6 Conclusions

A detailed analysis of the failure of reaction wheels onboard the Kepler spacecraft based on telemetry data is studied. Several methodologies are applied; these include raw data analysis, correlation, Weibull and friction model. The results of the statistical analysis show that Reaction wheel 2/ Reaction wheel 4 possesses a signature of failure two to three months prior to its complete failure. Thus, it is possible to preplan, develop and upload new attitude control algorithms to extend the life of the spacecraft in case of onboard reaction wheel failures.

The data-driven LSTM model is applied successfully for detecting the faults of Reaction wheel 2/ Reaction wheel 4 using given data, Weibull parameters and intercept values. Finally, from the space weather data, the remaining useful life close to the actual value is estimated and it can be concluded that the reaction wheels made up of an insulating material such as ceramics, will not react with the electrical charges and thus failure can be prevented. In the future, a detailed investigation of the effects of space weather will be performed.

CHAPTER 4

4. Data-Driven Prognostics

4.1 Introduction

In this Chapter, the aftereffects of the abnormal condition in the system are projected with remaining useful life prediction. In an engineering field, the prognosis is the method of predicting the operative time after which the plant will no longer perform its envisioned operation. Any mechanical component or system tends to degrade from its normal operation either due to wear/tear or aftermath the initiation of fault. The predictive maintenance framework of the health management system should perform all necessary corrective actions to maintain the normal operation throughout the lifetime of the mission. Based on the literature survey, this section for prognosis is developed with the data-driven technique (LSTM network) with some amendments using filtering, merging and optimization techniques.

The Diagnosis and Prognosis Health Management (DPHM) module should effectively possess [107],

a. complete control over the configuration of overall components of the structure to safety and reliability.
b. managing the ability to record all the random failures to assist in identifying recurring issues.
c. life predicting capability with reconditioning or replacement of the parent component after the onset of a failure event.

Maintenance activities cover both panned and unplanned actions for sustaining the process in its nominal condition.

Data Source: National Aeronautics and Space Administration

https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan

4.2 Dataset

The sample used in prognostic management is from the collection of datasets developed by Modular Aero-Propulsion System Simulation (MAPPS) of NASA. These specimens consist of multiple multivariant time-series sensor values of several same configuration turbofan engines simulated with 40,800 kg of thrust at about 36,000 ft, of lower subsonic regime (0.7-0.9 Mach) with temperature ranging from -51 ^{O}C to 40 ^{O}C. All the sensor values extracted are from the nominal state until the point of failure.

There are two different sets of datasets available in the repository that is used in this Chapter for simulation, they are

1. Turbofan Engine Degradation simulation dataset (C-MAPPS)
2. PHM08 Challenge Dataset

4.2.1 C-MAPPS Dataset

This dataset was an engine degradation simulation provided by the prognostics center of excellence at NASA Ames. It has four sets (FD001 to FD004) of train and test data along the with RUL values simulated under different fault modes and operational settings as shown in Table 4.1.

Table 4.1 C-MAPPS Turbofan Dataset [108]

Dataset	Train Trajectories	Test Trajectories	Conditions	Fault Modes
FD001	100	100	(one) Sea level	HPC degradation
FD002	260	259	Six	HPC Degradation
FD003	100	100	(one) Sea level	HPC and Fan degradation
FD004	248	249	Six	HPC and Fan degradation

The given file has 26 columns with a unit number, time (cycles), operational settings and sensor values contaminated with noise [108]. The description of all the sensor values of turbofan engines (Table 4.2) and its run to failure representation for one case (FD001) of the training dataset is as shown in Figure 4.1.

Table 4.2 C-MAPPS Sensor Values of Turbofan Engine [108]

Sensor Data	Category	Unit
1	Total Temperature at Fan Inlet	ºR
2	Total Temperature at LPC Outlet	ºR
3	Total Temperature at HPC Outlet	ºR
4	Total Temperature at LPT Outlet	ºR
5	Pressure at Fan Inlet	psia
6	Total Pressure in the Bypass Duct	psia
7	Total Pressure at HPC Outlet	psia
8	Physical Fan Speed	rpm
9	Physical Core Speed	rpm
10	Engine Pressure Ratio (P50/P2)	--
11	Static Pressure at HPC Outlet	psia
12	Ratio of Fuel Flow to Ps30	pps/psia
13	Corrected Fan Speed	rpm
14	Corrected Core Speed	rpm
15	Bypass Ratio	--
16	Burner Fuel-Air Ratio	--
17	Bleed Enthalpy	--
18	Demanded Fan Speed	rpm
19	Demanded Corrected Speed	rpm
20	HPT Coolant Bleed	lbm/s
21	LPT Coolant Bleed	lbm/s

Each time series dataset has 21 sensor values from a different engine of the same type with three different operational settings that substantially affects the performance of the engine. Each engine has a different manufacturing variation and initial wear which is not considered as a fault. Fault is developed at some point during the operation, the provided repository is divided into two sets until system failure.

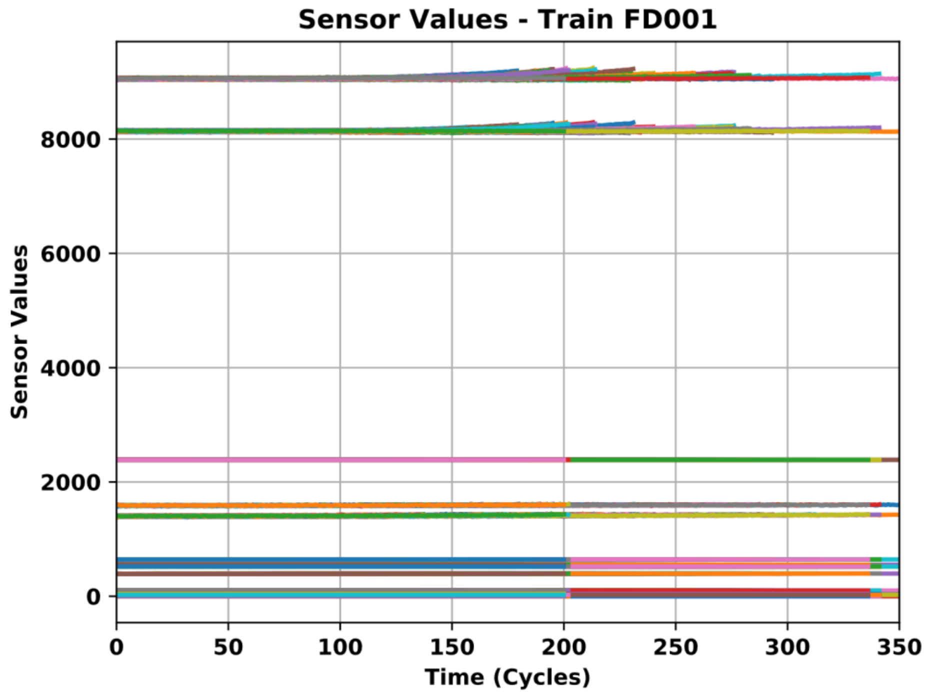

Figure 4.1 Run to Failure Sensor Values of First Set Fault Mode (CMAPPS)

4.2.2 PHM08 Dataset

The second dataset is a similar turbofan engine degradation simulation model with train and test sensor values except for the true RUL information [109] as shown in Table 4.3.

Table 4.3 PHM08 Turbofan Dataset [109]

Dataset	Train Trajectories	Test Trajectories	Final Test Trajectories
1	218	218	435

This data is made available to the public for the challenge competition held at the 1st International Conference on Prognostics and Health Management (PHM08) in 2008. Similarly, the file has 26 columns with a unit number, time (cycles), operational settings and noisy sensor values as same as categories described in Table 4.2. The run to failure representation of the test dataset is given in Figure 4.2.

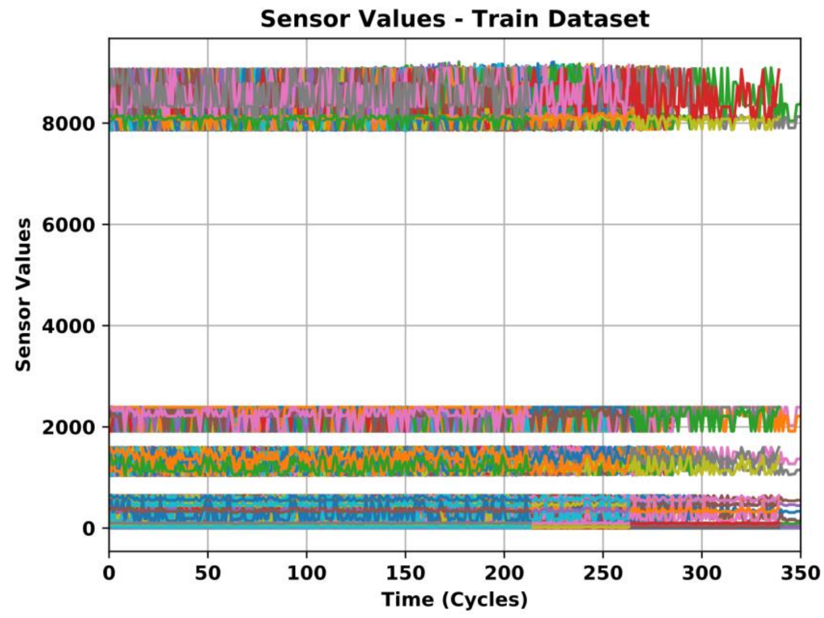

Figure 4.2 Run to Failure Sensor Values (PHM08)

4.3 Pre-Processing

Both the C-MAPPS and PHM08 sensor data are adulterated with noise which will affect the estimation of the remaining useful life based on the training given to the proposed model. Noise in the data is defined as a meaningless attribute attached to the parametric value which adversely affects the actual feature. The file has a total number of 21 sensor values (which will be given as the input to the model), due to unwanted and repeated features the model may overfit the results. Thus, the following steps are followed as a cleaning process before handling the input/output characteristics with the model.

1. Stationary, unrelated, non-correlative sensor values are detached from the database.
2. All the sensor values with different magnitudes are normalized with a unified interval range.
3. Noise contamination is removed with an efficient digital filter for smoothing the oscillating data.
4. The similar functionality sensor values are grouped together and are combined to a single value using merging techniques.

4.3.1 Data Reduction

The primary step is to remove the irrelevant sensor values form based on the visualization. The sensor values, S1 (Total temperature at fan inlet), S5 (Pressure at fan inlet), S6 (Total pressure in bypass duct), S10 (Engine pressure ratio (P50/P2)), S16 (Burner fuel-air ratio), S18 (Demanded fan speed), S19 (Demanded corrected speed) remains constant throughout the cycle for all the operating settings and hence is removed from the analysis.

4.3.2 Normalization

The process of changing the range of different scaled variables to a common distribution is known as normalization. In the present case, the above-mentioned Figures 4.1 and 4.2 clearly shows the difference in the range of sensor values. These variations cause problems in the learning rate and highly effects the process of updating weights at every iteration in training the model. This issue has to avoided and thus a mathematical rescaling is applied to the data to convert all the 21 sensors to a standard form (Figure 4.3).

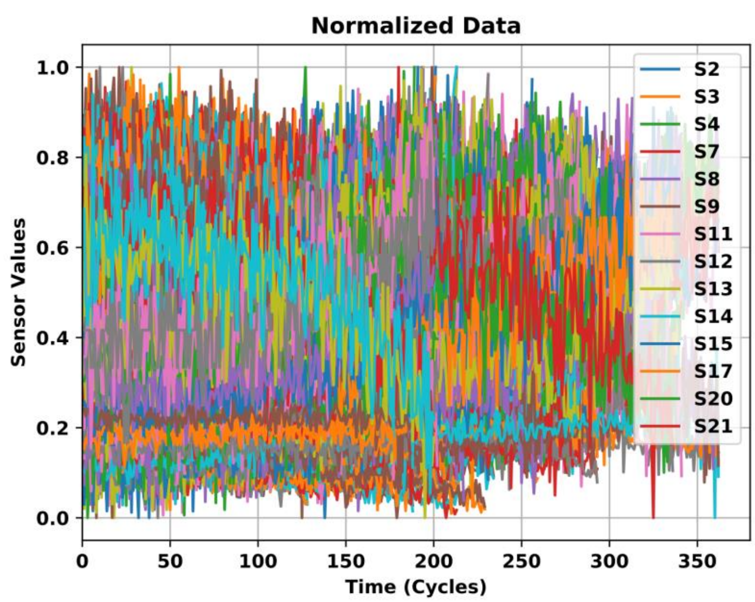

Figure 4.3 Normalized Data of the Correlative Sensor Values

In this approach, all the features are transformed between 0 and 1 using minmax_scaler function in pandas and the equivalent form is represented in the equation as,

$$S_{nor} = \frac{S_i - \min(S)}{\max(S) - min(S)} \tag{4.1}$$

where S = ($S_1, S_2, \ldots\ldots\ldots, S_n$) and S_{nor} is the ith normalized data.

4.3.3 Filtering - Savitzky - Golay (S-G) Filter

A Savitzky-Golay filter is a low-pass, well-adapted data-based method used for smoothing the data without distorting the actual feature. This was popularized by Abraham Savitzky and Marcel J.E. Golay [110] and is thereafter the most widely cited work in data cleaning. The S-G filters optimally curve fits the data points to an N-ordered polynomial using least-square methods but are applicable only for the odd-numbered datasets. There are studies [111] that extend the even number of data smoothing with feasibility validation.

The properties of the S-G filter [111] are summarized below:

- Property 1: The odd-indexed coefficients of the impulse response polynomial are zero and S-G filters are identical for even integers.
- Property 2: S-G smoothing with zero-order polynomial is identical to moving average filters.
- Property 3: The impulse response is symmetric, and the frequency response is purely real.
- Property 4: S-G filters are either on the unit circle of z-plane or complex conjugate on reciprocal groups.
- Property 5: These filters have a very flat frequency response in their passbands.
- Property 6: The normalized cutoff frequency depends on the order of the polynomial and length of the impulse response.
- Property 7: They have minimum attenuation characteristics in the stopband regions.

They perform better than standard averaging filters and are effective in preserving high-frequency signal components but are less effective in reducing the noisy signals. An example of filtered total temperature (LPC outlet) value is as shown in Figure 4.4.

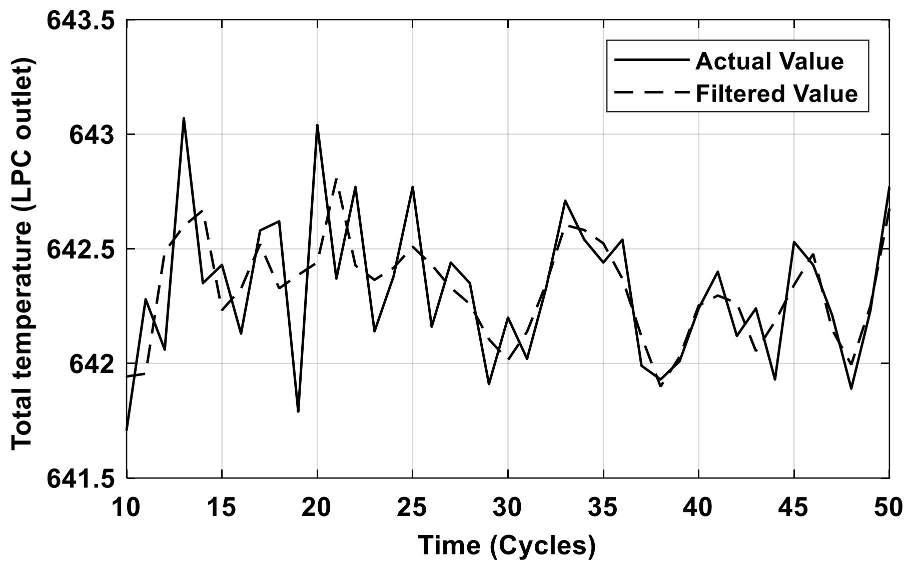

Figure 4.4 Savitzky-Golay Filtered S2 Sensor Value

4.3.4 Parameter Merging

The next step in preprocessing is to convert the complex multivariate time series input data into a simplified form. The multiple values have to be merged together and are performed by several approaches. Jamie Baalies Coble [109] has performed a prognostic study by merging the data sources using three predefined metrics such as monotonicity, prognosability, and trendibility. The scores from each of them are obtained for all the inputs and are merged based on their correlative strength. In the current work, a mathematical procedure called principal component analysis is applied to reduce the dimensions using orthogonal transformation.

4.3.4.1 Principal Component Analysis (PCA)

PCA is a data compression method based on the correlation that reduces the attribute to a lower-dimensional space [112]. This dimension-reduction tool solidifies the whole data columns into a small set that contains all the important information.

The segments involved in the process of principal component analysis is as given below,

1. The values have to be standardized prior to the reduction process and the deviation of data points from its mean with respect to the other values is computed.

$$Standardized\ value = \frac{Value - Mean}{Std\ deviation} \tag{4.2}$$

2. The eccentric correlative behavior is identified using the covariance matrix (C). The covariance matrix is a symmetric matrix (n x n) that contains covariances associated with all the possible pairs of the given variable set.

$$\mathrm{C} = \begin{pmatrix} cov\ (x,x) & cov(x,y) \\ cov\ (y,x) & cov(y,y) \end{pmatrix} \tag{4.3}$$

3. The third step is to compute the eigenvalues and eigenvectors of the covariance matrix to find the principal components of the data.
4. All these component values are selected based on its significance, and a feature vector with the high factor eigenvalues is created.
5. In the final step, the featured values are reoriented by multiplying the transpose of the feature vector with the transpose of the original dataset to represent the principal components.

$$Featured\ Values\ = (Feature\ Vector)^T * (Actual\ data)^T \tag{4.4}$$

Finally, PCA is an orthogonal linear transformation of data to a new coordinate system with the projection of the greatest variance with significant components. In the turbofan data, the sensor values are grouped based on the directional pattern and merged to a single value for effective prediction. An example of a principal component plot of a single sensor value in the training set of C-MAPPS is as shown in Figure 4.5.

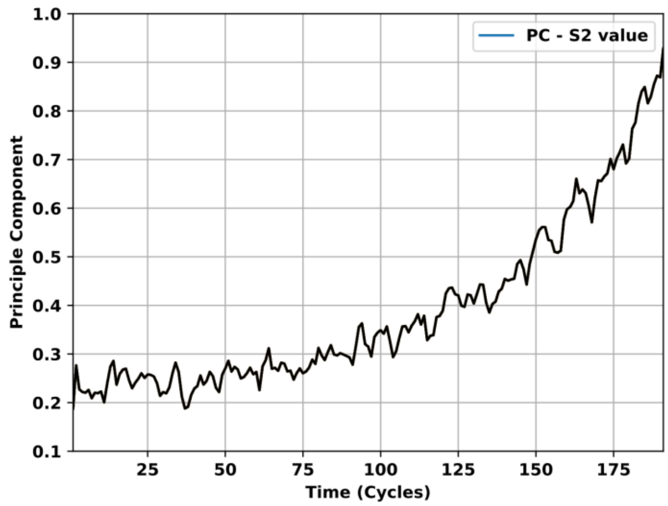

Figure 4.5 Orthogonal Transformation of Total Temperature (LPC Outlet)

The method based on kernel PCA is applied to predict the anomaly based on the cause-effect relationship of the abnormal telemetry data. This approach with a change in direction of the principal axes follows von Mises Fisher distribution for setting the threshold.

4.4 Modeling with Optimization Techniques

Optimization is the process of obtaining the best results from the complicated circumstances of any given problem. The reason for using optimization techniques is to minimize the effort and time in gaining efficient parameters and to maximize the desired outcome with reliability [80]. An objective function f(x) is defined in such a way to maximize or minimize certain parameters based on the set of constraints.

Constructing the LSTM model involves many parameters, in order to achieve better results, two best-known optimization algorithms (GA and PSO) are used to optimize these hyperparameters. Both the techniques are coded, applied and compared with the conventional model.

4.4.1 Genetic Algorithm (GA)

The genetic algorithm is a heuristic approach based on the theory of natural evolution for solving both constrained and unconstrained problems. This algorithm imitates the procedure of natural selection in which the appropriate population is selected to reproduce the fittest offspring of the future generation [81]. The optimal solution is achieved over the successive iterations (generations), for all problems even if the objective function is nonlinear, discontinuous, nondifferentiable or stochastic.

4.4.1.1 GA operators

The GA uses three types of operators every step to create the next generation from the current population [113].

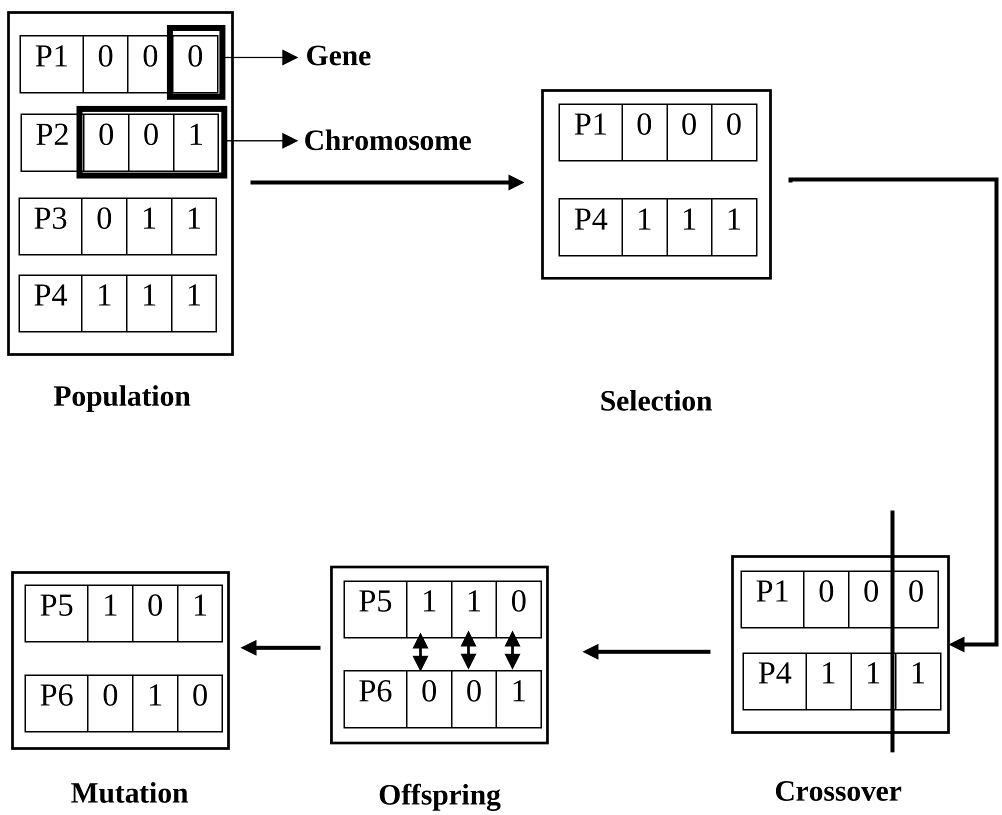

Figure 4.6 Selection Process of Genetic Algorithm [113]

a. Selection - collection of the fitter chromosome (parents) based on the fitness score to form the next generation.

b. Crossover - rules to exchange the locus of the chromosomes of two parents to form the children. Offspring are formed after reaching the crossover point.

c. Mutation - random changes within the population applied to the parents to form the offspring with maintaining diversity along with preventing premature convergence.

The fitness function is applied for each offspring and a fitness score is calculated to determine the accuracy of suitability for the future generation with optimal capability (Figure 4.6). The algorithm repeats and terminates until the convergence of significant offspring with a set of solutions. GA is applied is varied fields of science and engineering, automatic programming, machine learning, economics, natural immune systems, ecology studies, social systems, etc.

4.4.1.2 Steps of Genetic Algorithm

This algorithm is typically between 50 to 500 generations depending on the type of problem to obtain the highly fit chromosomes, the steps are:

1. Initialize the process with a randomly generated population of n-size with m-bit chromosomes.

2. The fitness function f(x) in population is calculated for each chromosome.

3. The following steps are repeated until the creation of n-offspring,

   a. A pair of parent chromosomes are selected with the highest probability of fitness function.

   b. With the probability of cross overrate the pair chosen at the random point is used to form new offspring. (If no crossover occurs - then exact copies of their parents are carried.

c. Transmute the two offspring at each point with the mutation rate to generate the new population.

4. If the number of offspring is odd, truncate the new population to an even number, otherwise, proceed.

5. Replace the current population with the new population and go back to step 2 and repeat the steps until the optimal solution.

4.4.2 Particle Swarm Optimization (PSO)

PSO was first introduced by James Kennedy and Russell Eberhart in 1995 [114] for nonlinear functions in ties with bird flocking, fish schooling and swarming theory. PSO is a swarm intelligence metaheuristic and nondeterministic optimization technique to find an optimal value of the target function. Five basic principles of Swarm intelligence are,

a. Proximity principle
b. Quality principle
c. Diverse response
d. Principle of Stability
e. Principle of Adaptability

The particle swarm optimization (Figure 4.7) paradigm adheres to all the above principles for manipulating the space calculations over a series of time steps. The bunch of particles called *the swarm* are allowed to move around and explore the given space. The motion of the particle direction is directed by,

1. Inertia of the particle's previous velocity.
2. Distance between the best positions of individual particles – cognitive force.
3. Social force from the best-known position of the swarms.

Each particle (hyperparameter of the model) is initialized randomly to a prescribed position in the given space [80].

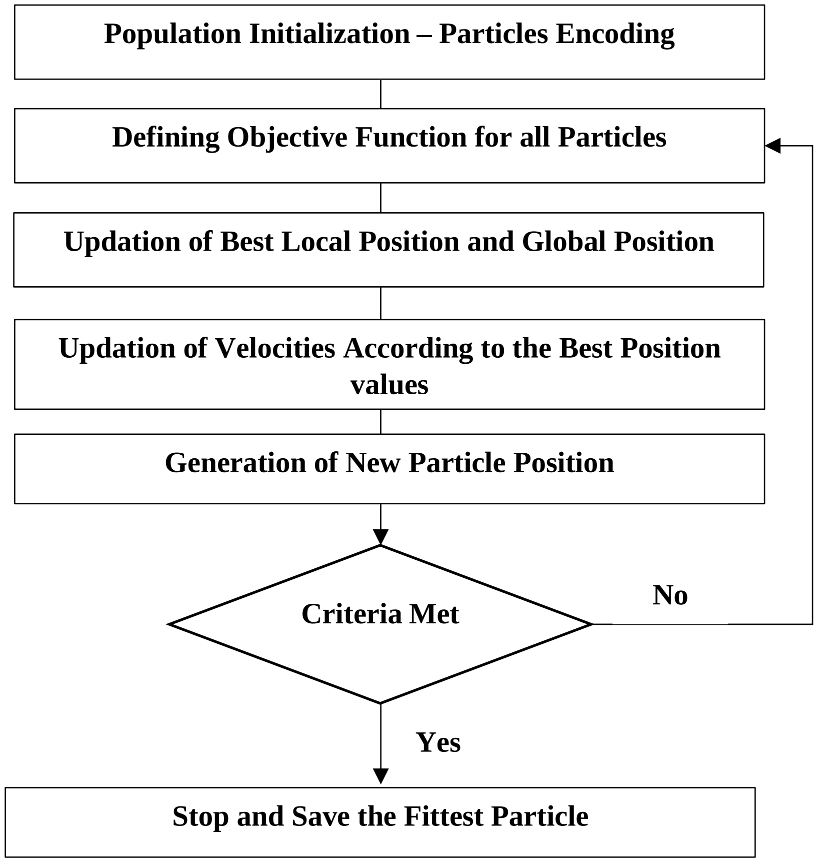

Figure 4.7 Steps in PSO Algorithm [114]

The Euler Integration for the physical movement of the particle's position is given by,

$$\vec{x_i}(t) = \vec{x_i}(t-1) + \vec{v_i}(t) \tag{4.5}$$

where,

$\vec{x_i}$ is position vector

$\vec{v_i}$ is velocity vector

i is the number of dimensions (number of components to optimize)

Similarly, the current velocity $\vec{v_i}$ of the particle is updated from the initial random value with the velocity of the best particle (with two stochastic variables).

$$\vec{v_i}(t) = \vec{v_i}(t-1) + \rho_1[\vec{p_i} - \vec{x_i}(t-1)] + \rho_2\left[\vec{p_g} - \vec{x_i}(t-1)\right] \tag{4.6}$$

where,

$\vec{p_i}$ is variable that take cares of the velocity vector corresponding to the previous velocity

$\vec{p_g}$ is the variable corresponding to the velocity of the best particle

ρ_1 and ρ_2 are the random constants for social and cognitive behavior of the particles

This method is known to be a better method for training the network in LSTM. PSO compares the fitness of each network and finds the best global value for an optimized solution. Here, the model parameter such as the number of neurons, optimizers contains a position (weight) and velocity (to update the weight closer to global best). The particles swarm around the hyperspace and update their position to reach the optimal condition of the neighborhood and continue to move towards the global optima.

In the training process of the LSTM model, the network fitness is determined by the mean square error for the entire set and is defined as,

$$Fittness = \sum (Actual\,Value - Predicted\,Value)^2 \tag{4.7}$$

4.5 Results

The LSTM model shaped in Chapter 2 with all relevance and tuning is used for prognosis in finding the remaining useful life of the turbofan engine for the two given datasets with all the 4 faulty conditions. Later the results with an optimized model using the above-mentioned techniques are obtained and compared with the existing literature.

4.5.1 Normal Model

In this section, the predictions of the remaining useful life of the engines for the datasets of CMAPSS (FD001, FD002, FD003, FD004) and PHM08 are presented. The evaluation score is given by the expert team and an error plot with prediction accuracy is shown with proper justification [108][115].

The finalized hyperparameters of the model for the faulty conditions based on the learning from pollution analysis along with trial and error are shown in Table 4.4.

Table 4.4 LSTM Model Hyperparameters

Parameters	**C-MAPPS FD001**	**C-MAPPS FD002**	**C-MAPPS FD003**	**C-MAPPS FD004**	**PHM08**
Number of Neurons	50,50	128,128	50,50	64,64	50,50
No. of Hidden Layers	2	2	2	2	2
Activation Function	Linear	ReLU	ReLU	ReLU	ReLU
Batch Size	64	200	64	200	200
Optimizer	RMSprop	Adam	RMSprop	Adam	Adam
Loss Function	MSE	MSE	MSE	MSE	MSE

4.5.1.1 C-MAPPS Dataset

This dataset has 4 subsets with different operating settings and fault conditions as shown in Table 4.1. All the engines operate normally and develop a fault at some point in time and this study estimates the remaining operating cycles before the failure.

(i) The subsets FD001 and FD003 exhibit similar patterns and are analyzed with the same hyperparameters (Table 4.4). Sensor 1,5,6,10,16,18,19 displays constant measurements thought the cycle and are no useful for the degradation pattern in modelling. The other 14 sensor values are grouped together based on their principle directional behavior (sensor 2,8,11,15,17 - showed a very slight increase (PCA combo 1) ; sensor 3,4 - displayed slight increase with a magnitude more than 40 units (PCA combo 2); sensor 7,12,20,21 - displayed decrease in pattern (PCA combo 3); sensor 9,14 - exhibited an increase in trend with magnitude more than 110 units (PCA combo 4)) and are merged into a single value using PCA technique.

Figures 4.8 and 4.9 represents a comparison plot of the actual and predicted remaining useful life of the simpler datasets FD001 and FD003.

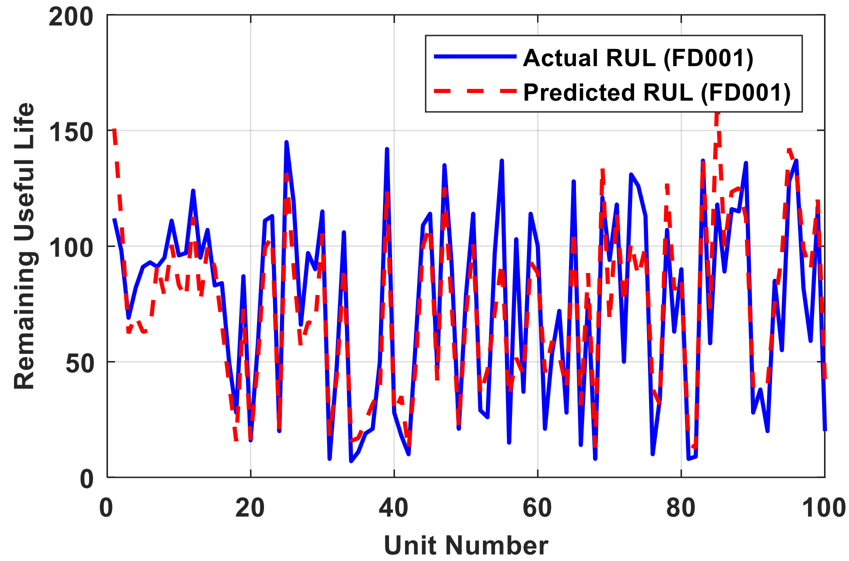

Figure 4.8 Actual and Predicted RUL [FD001]

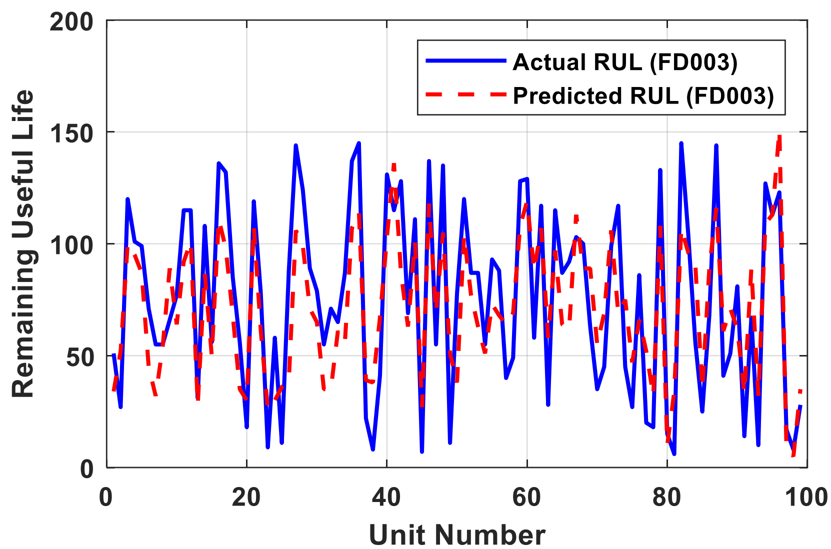

Figure 4.9 Actual and Predicted RUL [FD003]

(ii) The subsets FD002 has six operating conditions with one fault mode, on the other hand, FD004 has six operating conditions with 2 fault conditions which makes this data much complex and difficult to analyze. Both these sets are examined with small changes in the model (Table 4.4). In both these cases, other than corrected core speed (S14) none of the sensors show any convergence or divergence pattern. Hence sensor-14 is taken and is converted to the kernel space with principal components using the PCA technique.

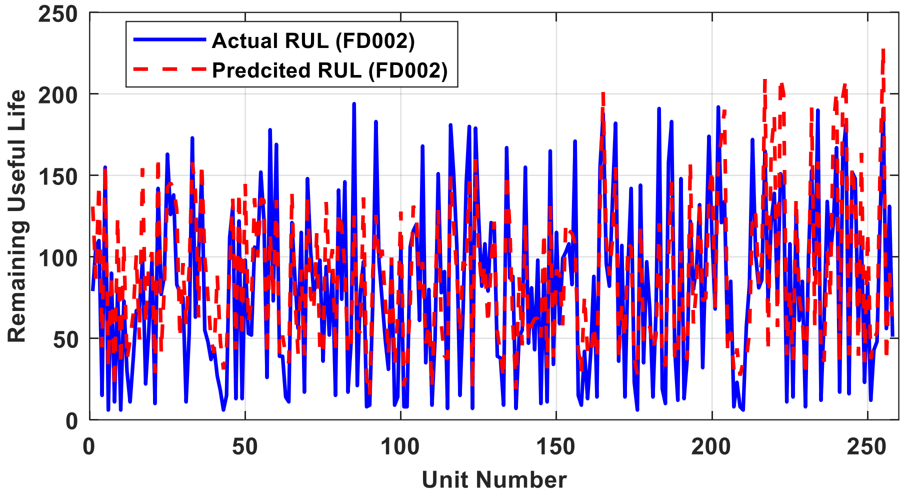

Figure 4.10 Actual and Predicted RUL [FD002]

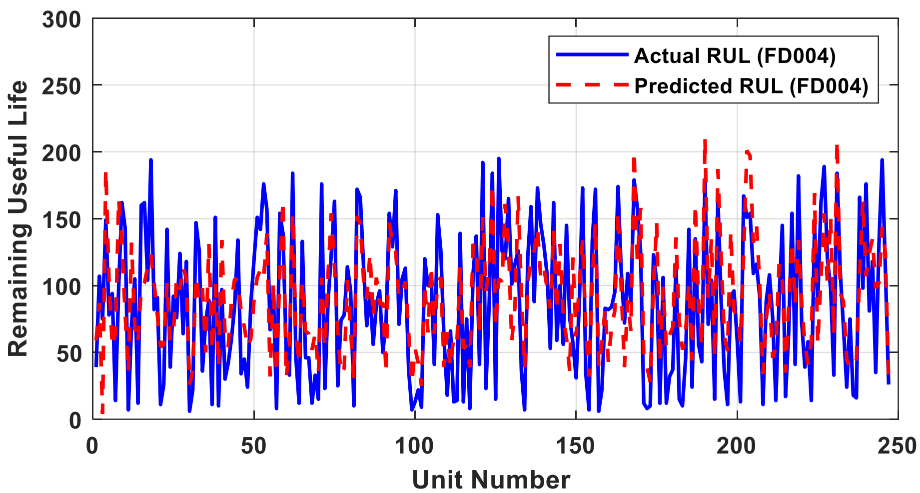

Figure 4.11 Actual and Predicted RUL [FD004]

Figures 4.10 and 4.11 represents the comparison plot of both actual and model-predicted plots for both the complex datasets FD002 and FD004.

4.5.1.2 PHM08

Prognostics and Health Management (PHM08) dataset is similar to the C-MAPPS data except for the true RUL values.

The sensor patterns are as same as the FD003 dataset (C-MAPPS) and hence the relevant hyperparameters are applied with data pre-processing and deep LSTM model. Figure 4.12 shows the predicted values of RUL for the test and final-test data sets (no reference is given for comparison).

In the prognosis framework, early prediction is generally desirable than late prediction. Evaluation metrics namely, root mean square error (RMSE) and scoring function (SF) [108] are employed to assess the performance of the model. These functions are calculated for every set of modeling the mathematical expressed as shown below,

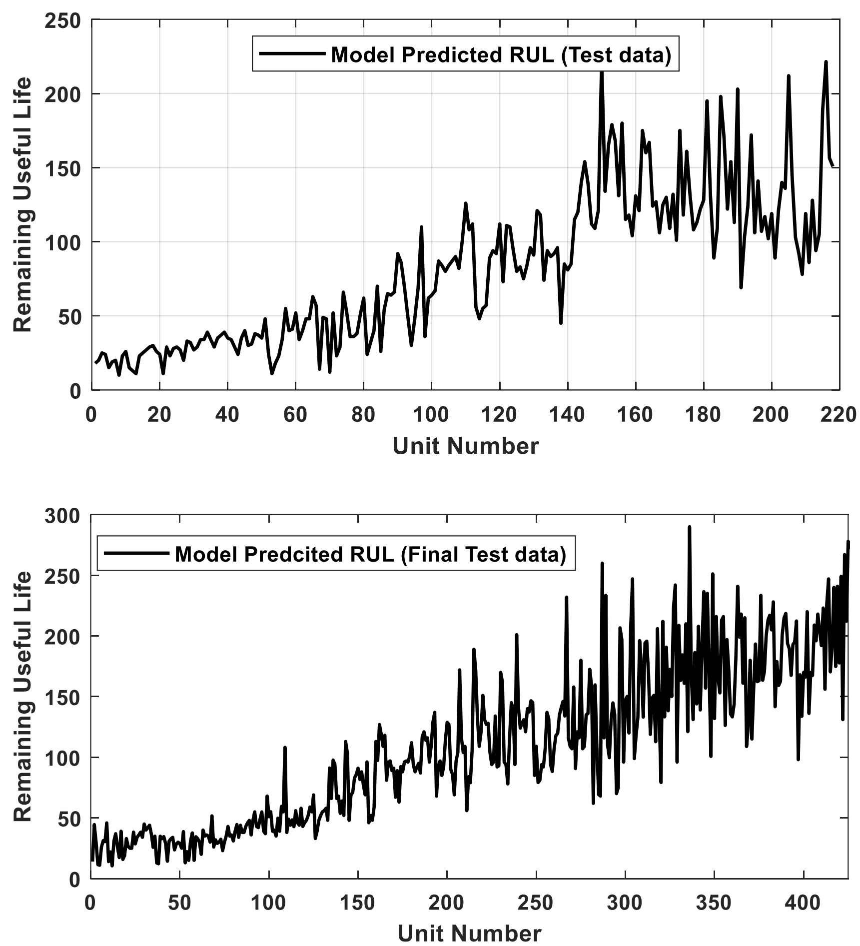

Figure 4.12 Predicted RUL for Test Dataset - PHM08

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}(b_i)^2 \tag{4.8}$$

$$SF = \sum_{i=1}^{N} s_i, \tag{4.9}$$

and

$$s_i = \begin{cases} e^{\frac{-b_i}{a1}} - 1, & for\ b_i < 0 \\ e^{\frac{b_i}{a2}} - 1, & for\ b_i \geq 0 \end{cases}$$

where,

N = total number of the samples

$b_i = RUL_{predicted} - RUL_{true}$

a1 = 13, a2 = 10 predetermined values

A scoring function is an exponential form and a bad prediction would affect the overall performance score. Figure 4.13 represents the plots of these two evaluation metrics for the given error values.

The summary of model results with all its variants including filters and PCA combos is presented in Table 4.5. The present model is ineffective in predicting the score values for the PHM08 challenge dataset.

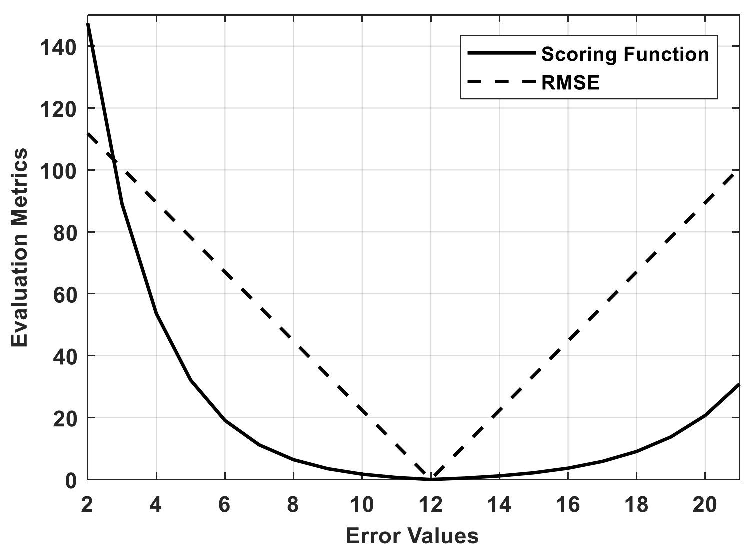

Figure 4.13 Scoring Function and RMSE Plots with Error Values

Table 4.5 Summary of Performance Metrics for the Given Dataset

CMAPPS-FD001				
PCA Combos	**Score**		**RMSE**	**Computational Time (s)**
	With Filter	**Without Filter**		
1,2	2136	2350	24.51	32
1,3	2149	2199	28.31	36
1,4	3628	3678	31.54	35
2,3	2655	2705	32.64	37
2,4	11913	12764	39.32	49
3,4	3024	3897	30.67	36
1	3848	3898	30.14	37
2	5222	5272	32.04	39
3	3008	3058	30.87	41
4	4974	5024	33.87	36
CMAPPS-FD002				
Sensor Value	**Score**		**RMSE**	**Computational Time (s)**
	With Filter	**Without Filter**		
14	11643	10758	38.78	108
CMAPPS-FD003				
PCA Combos	**Score**		**RMSE**	**Computational Time (s)**
	With Filter	**Without Filter**		
1,2	2456	2356	25.65	42
1,3	2821	2721	31.13	46
1,4	3268	3854	33.43	41
2,3	3515	3754	34.46	53
2,4	13133	11654	49.22	65
3,4	5024	5664	42.75	45
1	4897	4978	35.34	49
2	4131	4752	33.08	49
3	5118	5853	32.75	53
4	6378	6454	34.83	56
CMAPPS-FD004				
Sensor Value	**Score**		**RMSE**	**Computational Time (s)**
	With Filter	**Without Filter**		
14	13578	11051	43.44	115
PHM08				
Dataset	**Score**		**RMSE**	**Computational Time (s)**
Test set	993104.48		48.47	67
Final Test set	N/A		43.56	78

4.5.2 Model with GA

The results from the conventional LSTM states the necessity of improvements in the present model. The genetic algorithm is iterated for many generations until the evolution of improved individuals. The given dataset is split into 75 % and 25% for training and validation respectively [116]. The set of genes and parameters used in this approach is given in Table 4.6 and Table 4.7 respectively.

The following iterative algorithm is used for obtaining the optimal solution,

For i = 1, Population_max

Objective function f (x) = Accuracy Metrics (Value $_{t-1}$ - Value $_{t}$)

If f (x) ≤ Threshold, go to 10

End

10 Stop

Table 4.6 Genes Used in GA Approach (Hidden Layers = 2 and Batch Size = 64)

Genes	Hyper-Parameter	Choices and Range	Optimized Values
1	Number of Neurons (Unit 1)	1 – 500	32
2	Number of Neurons (Unit 2)	1 – 500	32
3	Activation Function - Input Layer	Sigmoid, ReLU, Softmax, Tanh, ELU, SELU, Linear	Sigmoid
4	Activation Function - Hidden Layer	Sigmoid, ReLU, Softmax, Tanh, ELU, SELU, Linear	Sigmoid
5	Activation Function - Output Layer	Sigmoid, ReLU, Softmax, Tanh, ELU, SELU, Linear	Tanh
6	Loss Functions	Categorical Crossentropy, Binary Crossentropy, Mean Squared Error, Mean Absolute Error, Sparse Categorical Crossentropy	Mean Squared Error
7	Optimizer	Stochastic Gradient Descent, RMSprop, Adagrad, Adadelta, Adam, Adammax, Nadam	Adam
8	Epoch	Early Stopping Criteria	N/A

Table 4.7 Parameters Used in GA Approach

Parameters	Value
Classes	1
Population_max	50
Generations	10
Threshold	0.0050

The optimized hyperparameters obtained from the algorithm is expedited in the LSTM model for all the four cases (FD001, FD002, FD003, FD004). The results for the best PCA combo (1,2) are presented in Table 4.8 and the comparison plot is shown from Figure 4.14 to Figure 4.17.

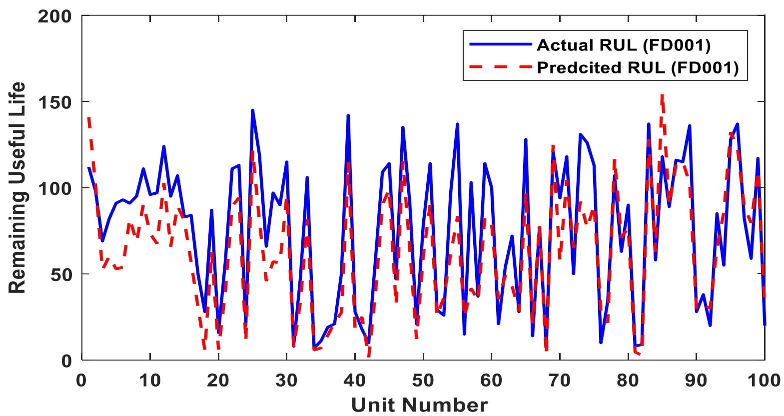

Figure 4.14 GA Optimized Actual and Predicted RUL [FD001]

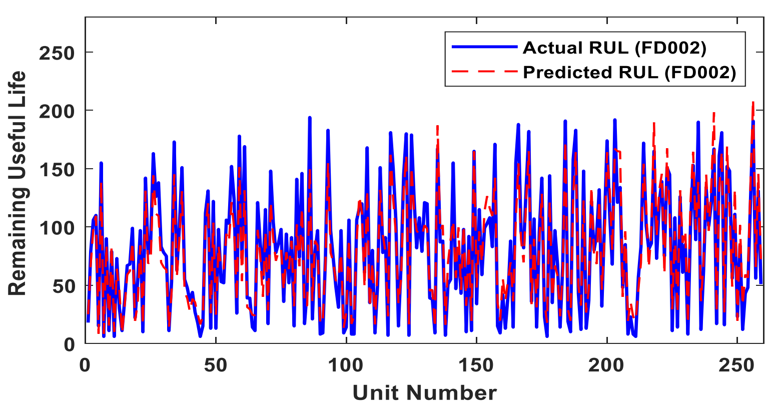

Figure 4.15 GA Optimized Actual and Predicted RUL [FD002]

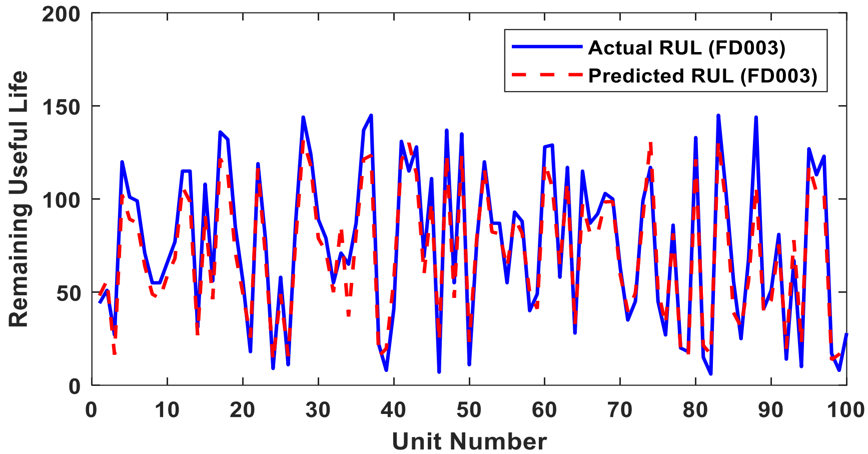

Figure 4.16 GA Optimized Actual and Predicted RUL [FD003]

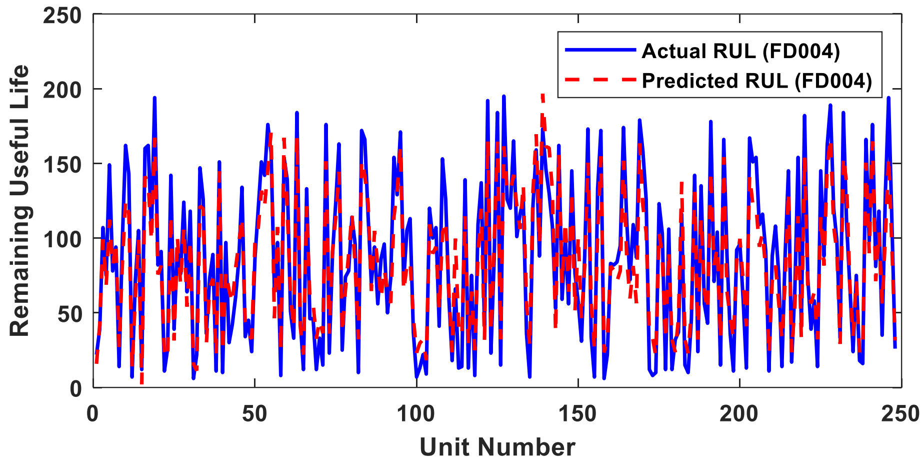

Figure 4.17 GA Optimized Actual and Predicted RUL [FD004]

Table 4.8 GA Optimized Performance Parametric Values

PCA Combos	Score		RMSE	Computational Time
	With Filter	Without Filter		
1,2	660	780	15.82	~ 48 hours
	3870	3845	21.14	
	840	823	13.77	
	4574	4698	23.69	

4.5.3 Model with PSO

The main advantage of this population-based optimization method is its fast convergence due to the association of particles in the swarm. The stochastic constants towards local best (pbest) and global best (gbest) are chosen to be 0.1 and 0.2 respectively for the training dataset. The inclusion of the validation dataset may have improved the results. The initial set of parameters specified to the process is given in Table 4.9.

The following iterative algorithm is used for obtaining the optimal solution,

10 i = 1

20 Objective function f_1 (i) = Training Loss [mse]

30 gbest= f_1 (i)

40 i=i+1

50 Objective function f_1 (i) = Training Loss [mse]

70 if f_1 (i) $<$ f_1 (i-1), then go to 30

80 Else if gbest > gbest_min, go to 40

90 Stop

Table 4.9 Parameters Given to the PSO Algorithm (Hidden Layers = 2 and Batch Size = 64)

S.No	Hyper-Parameter	Choices and Range	Optimized Values
1	Number of Neurons (Unit 1)	1 – 32	16
2	Number of Neurons (Unit 2)	1 – 64	32
3	Activation Function - Input Layer	Sigmoid, ReLU, Softmax, Tanh, ELU, SELU, Linear	Linear
4	Activation Function - Hidden Layer	Sigmoid, ReLU, Softmax, Tanh, ELU, SELU, Linear	Sigmoid
5	Activation Function - Output Layer	Sigmoid, ReLU, Softmax, Tanh, ELU, SELU, Linear	Linear
6	Optimizer	Stochastic Gradient Descent, RMSprop, Adagrad, Adadelta, Adam, Adammax, Nadam	RMSprop

These optimized values are utilized in the LSTM model and the results for the best PCA combo (1,2) are presented in Table 4.10 and the comparison plot for all four cases is shown from Figure 4.18 to Figure 4.21.

Table 4.10 PSO Optimized Performance Parametric Values

PCA combos	Score		RMSE	Computational Time
	With Filter	Without Filter		
1,2	876	651	15.23	~ 2 hours
,,	4124	4225	22.87	
	765	618	14.53	
	4897	4714	26.11	

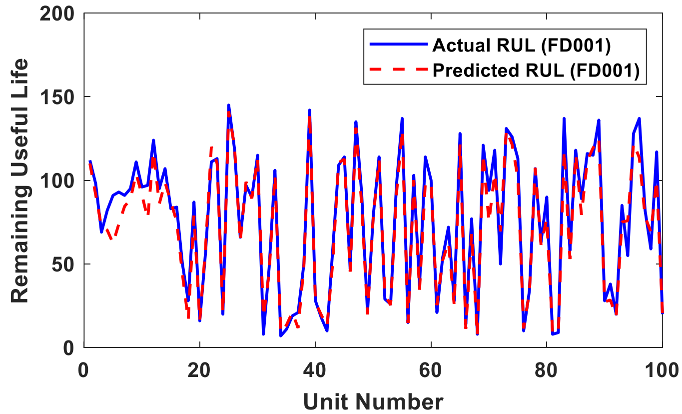

Figure 4.18 PSO Optimized Actual and Predicted RUL [FD001]

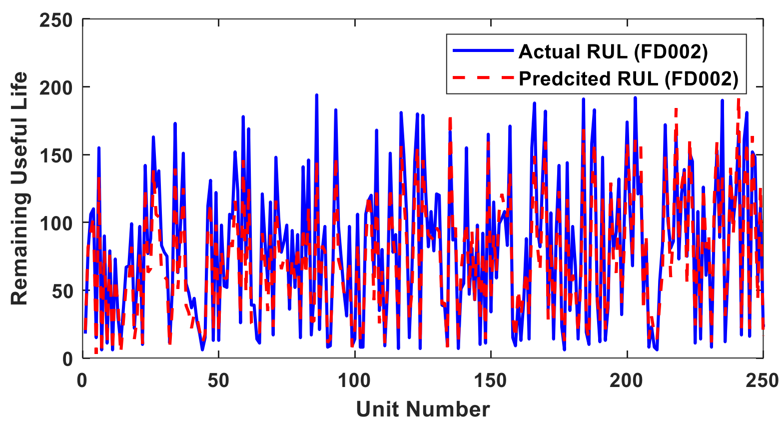

Figure 4.19 PSO Optimized Actual and Predicted RUL [FD002]

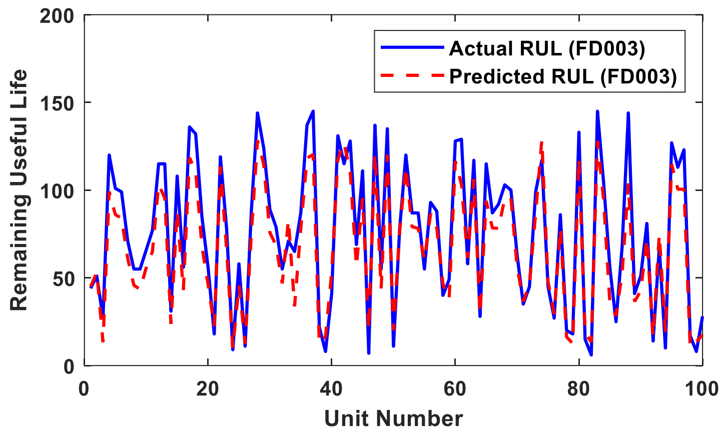

Figure 4.20 PSO Optimized Actual and Predicted RUL [FD003]

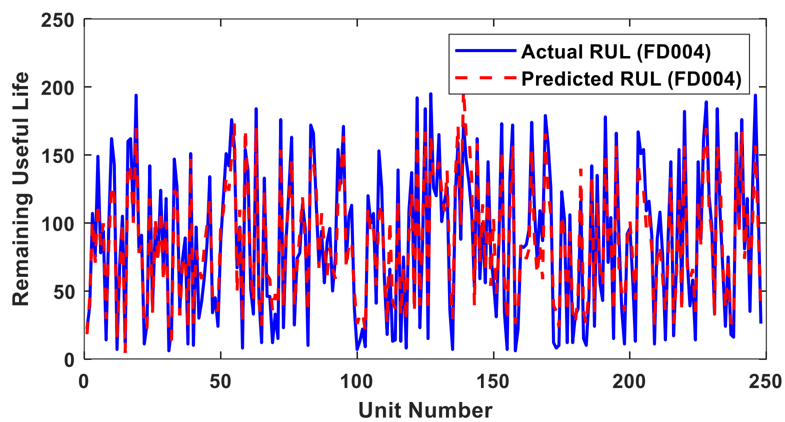

Figure 4.21 PSO Optimized Actual and Predicted RUL [FD004]

The best score from the above summary table is compared with earlier literature for both the metrics as shown in Table 4.11 and Table 4.12.

Table 4.11 Comparison of RMSE on C-MAPSS Dataset with Reference [115]

Dataset	**MLP**	**SVR**	**RVR**	**Deep LSTM**	**Proposed LSTM Model**	**Proposed LSTM (GA Optimized)**	**Proposed LSTM (PSO Optimized)**
CMAPPS FD001	37.56	20.96	23.80	16.14	24.51	**15.82**	**15.23**
CMAPPS FD002	80.03	42.00	31.30	24.49	38.78	**21.14**	**22.87**
CMAPPS FD003	37.39	21.05	22.37	16.18	25.65	**13.77**	**14.53**
CMAPPS FD004	77.37	45.35	34.34	28.17	43.44	**23.69**	**26.11**

Table 4.12 Comparison of Score Function on C-MAPSS Dataset with Reference [115]

Dataset	MLP	SVR	RVR	Deep LSTM	Proposed LSTM Model	Proposed LSTM (GA Optimized)	Proposed LSTM (PSO Optimized)
CMAPPS FD001	18000	1380	1500	338	2199	**660**	**651**
CMAPPS FD002	7800000	590000	17400	4450	10758	**3870**	**4124**
CMAPPS FD003	17400	1600	1430	852	2356	**823**	**618**
CMAPPS FD004	5620000	371000	26500	5550	11051	**4574**	**4714**

Remarks

4.6 Conclusions

In this Chapter, an improvised LSTM neural network with genetic algorithm and particle swarm techniques are employed to optimize the parameters of the model. Both the algorithms initiate from a random population and reckon a fitness value for each given parameter. PSO and GA are almost similar in updating the optimized offspring; unlike GA, PSO gets updated with internal velocity and does not have genetic operators like crossover and mutation. The information-sharing mechanism is different in PSO and GA. A most important advantage in PSO is the trend of convergence rate, the solution gets optimized quickly.

The variants in the conventional LSTM dictates the PCA combo 1,2 to be the best predicting solution with the least score value for the 4 cases of C-MAPSS as 2136, 10758, 2356, 11051 and 9560 for PHM08 (Table 4.5). It can be also noted that the usage of filters does not make much difference in the score value. The best values are compared with the earlier works and are seen to be with fewer improvements, thus two of the optimized techniques are applied.

The PSO algorithm and GA modeled in python framework are compared, it can be inferred that RMSE and score values are much improvised than the conventional neural network (Table 4.11 and Table 4.12). Among the two techniques, PSO comparatively produces better results with a faster computational converge time of 2 hours which is 2.75 days (72 hours) faster than the GA technique. The experimentation shows that the PSO-LSTM model has better accuracy and performance.

CHAPTER 5

5. Conclusions and Future Work

This thesis dealt with the fault diagnosis and prognosis of the aerospace systems that include datasets of an aircraft turbofan engine and Kepler satellite. The main modules of the AOCS framework and health monitoring systems have discoursed with fault detection, isolation/identification, and prognosis. The historical report of the failures was reviewed and major causes were revealed, utilized for the current methodology. The proposed module is capable of detecting the failure in the primary stage, followed by the isolation/ identification of severity/location of the fault and finally determining the remaining useful life in advance. The early RUL estimation helps in charging remedial actions and strategic planning of maintenance activities well in advance. In light of the severity and difficulty associated with fault analysis frameworks for the complex systems, three major modules were examined and enriched frameworks were proposed and evaluated with addressing some of the restrictions sensed in the early literature.

Data analysis is the advanced scientific approach with multi-disciplinary capability to handle complex systems without expert knowledge and experience. Artificial Intelligent techniques outperform human ability with high competence in solving complicated tasks. Furthermore, the proposed techniques with optimization tools can be extended to any similar aerospace systems with higher complications. Features of the data, types of faults and limitations of data methodology were scrutinized thoroughly.

Model development, fault diagnosis and prognosis to avoid downtime and mission failure, along with the optimizing tools available in the literature were studied in Chapter 1. The next chapter reveals the terminology of fault, malfunction, and failure with the desired requirements of the fault management system. A complete analysis of conventional and data-driven methods were explained with the applications on aircraft and spacecraft. An LSTM neural network was developed and tested with a simpler pollution dataset for evaluating the performance and reliability of the model. In Chapter 3, the Kepler telemetry dataset is studied for diagnosing the fault with the conventional approaches and data-driven techniques. Finally, in Chapter 4 the turbofan engine dataset is used in improvising the proposed model with optimization tools in successfully estimating the remaining useful life of the aircraft engines.

5.1 Highlights

This section explains the overall important structures of the proposed approach to the complex datasets (aircraft and satellite).

In Chapter 2, a modified and efficient LSTM model was proposed to address the agility and limitations of the available literature. The proposed model used the pollution dataset and studied the selection process in choosing the hyper-parameters. The common issues such as stationarity, the laziness of the model in predicting the previous data as the future step were addressed appropriately with inferences of the Dickey-Fuller test and sliding window- multistep predictions respectively.

In Chapter 3, the fault detection of a complex Kepler mission in its critical state (2 operating reaction wheels) was carried with the conventional and modern data-driven approaches. The most effective statistical methods were applied to the telemetry data and various new inferences were extracted from each of the methods and are merged together to detect the fault well in advance. Later the proposed LSTM model was varied to adapt this dataset for detecting the fault of the reaction wheels.

In Chapter 4, the performance of the recurrent neural network was reviewed, various preprocessing techniques such as parameter merging, principle component analysis, and digital filters were applied to improve the accuracy of prognosis. The projected model was verified and reviewed for all the different cases with its performance scores. Later, the two most effective and widely used optimization techniques (genetic algorithm and particle swarm optimization) were incorporated into the model separately. The prognostic ability of both the algorithms were compared with the metrics and reported in the tabular form.

5.2 Summary of Contributions

The main contributions in this thesis of developing the modern data-driven method for analyzing the complex aerospace systems with the capability of fault detection and prognosis are given as follows,

Model Development: The proposed data-driven model (1) considers the time-series environmental pollution dataset to predict and estimate the future trends with high-performance computing platform and Python language with TensorFlow backend [117], (2) the nonlinearities caused by the anthropogenic and environmental factors which adds further complexity were in considered in the analysis, (3) the network was trained using supervised learning to forecast the data in multi-steps by varying the free parameters, (4) the major issue of the data-driven approach such as overfitting and underfitting were resolved. The fine-tuned proposed model with a month ahead capability of predicting the future with superior performance compared to other available techniques in the literature was implemented.

Fault Detection: Based on the earlier literature, the conventional statistical methods were used for extracting the significant feature of the data. (1) correlational coefficients were studied initially for extracting the dependencies among the given input, (2) from the Weibull analysis the range of shape and scale factors were compared with the normal data to detect the faulty condition, (3) variations in the intercepts and slope values of the frictional data senses the abnormal behavior, (4) LSTM model was applied to the featured dataset for detecting the RW failure with the calculated weather data from the sun.

Fault Prognosis: Remaining useful life of the set of similar turbofan engines were successfully determined well ahead of the downtime with higher accuracy. The LSTM model was employed with (1) usage of digital filters in reducing the noise levels, (2) normalization and principal component analysis - for converting the sensor values to a standardized time variational elements, (3) parameter merging - to integrate the sensor values to a low-dimensional frame for simpler analysis.

Optimized Model: The proposed model was enhanced with two of the best available optimization techniques. (1) the model merged with genetic algorithm (GA) performs better in tuning the hyper-parameters with improved prognostic efficiency but takes a huge amount of time in converging (3 days approx.), (2) PSO-LSTM accomplishes the optimal solution with higher convergent rate. This optimized model is capable of estimating the RUL with a low error percentage and is applicable for all the scenarios of the fault with external disturbances and noise levels.

5.3 Future Work

The number of imminent works needs to be envisioned in the data-driven modeling considering the growth and demand of the present world. From this thesis, the applicability of the algorithm is limited only to certain types of time-series problems. Improvements in addition to the contributions of this work will enhance the performance of the model with universal capability. A few of the future works are outlined below,

Exploring the Model on Other Systems: The proposed algorithm of this thesis is substantially applied to the complex systems. It will be interesting to explore all the other relevant aerospace systems with different operating conditions. However, all the considered datasets are time-based and are only relevant to time-variant problems. The data-driven structure should also be proficient in solving the snags like classification and image-based problems.

Hybrid Approaches on Neural Network Models: The present framework focused on the major issues in data-driven approaches along with optimization tools. It is also feasible to merge the two different neural networks to generate an enhanced approach with greater accuracy and reliability. Combinations of the two different models like CNN + FNN, RNN + CNN, FNN + RNN, conventional + modern can compensate the inconsistencies, discrepancies, and inaccuracies from the existing methodology. It is also highly appreciated if the optimization tools are merged with this integrated model.

Integration of External Features and Disturbances: As discussed at the beginning of this work, the output of any high fidelity data-driven model depends on the early history of faults and failure trends. The predictions and known abnormal patterns can be easily estimated with proper training and evaluation of the designed model. The complexity arises with the sudden external interaction or a new disturbance that completely changes the behavior of the system. The severity of these aspects changes the model completely to an unfeasible form.

Investigation of Noisy Datasets: Almost every dataset obtained from the machinery or sensors are noisy. The irregularities and contaminations in the featured data will affect the performance of the modeling. There is only a minimal change in the error and score values with the usage of the Savitzky-Golay filter in the second dataset. The fitters could be ineffective or not required (neural network is already a fitting model). Even a better technique might improve the results. Thus, to examine the exact cause of the issue further investigation on filtering is required.

Performing Analysis for Incomplete Data: In this digital world, information about any system can be studied from its datasets. Big data analysis is a study of multi-variant collection of different sets with various attributes. The major issue is the availability of the data with all its dimensional features. In most of the cases, procurement of data is problematic and difficult due to ethical and privacy matters. Therefore, a framework that explores acquaintance about the system with the given incomplete data would be much competent.

References

[1] Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. DOI:10.1080/10618600.2017.1384734

[2] Hajirahimova, M. S., & Aliyeva, A. S. (2017). About Big Data Measurement Methodologies and Indicators. *Journal of Modern Education and Computer Science (IJMECS), 9*(10), 1-9. DOI:10.5815/ijmecs.2017.10.01

[3] Gao, Z., Cecati, C., & Ding, S. X. (2015, March). A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. *IEEE Transactions on Industrial Electronics*, *62*(6), 3757-3767. DOI:10.1109/TIE.2015.2417501

[4] Godard. (2010). Fault-Tolerant Control of Spacecraft. Aerospace Department, Ryerson University, Toronto, Canada

[5] Kumar, K. (2008, June). Attitude Dynamics and Control of Satellites Orbiting Rotating Asteroids. *Acta Mechanica*, *198*(1-2), 99-118. DOI:10.1007/s00707-007-0508-y

[6] Kumar, D. K. (2012). *Fundamentals of Dynamics and Control of Space Systems* (2nd ed.). California, CreateSpace Publisher

[7] Allied Market Research. (2017). Small Satellite Market by Type (Nanosatellite, Microsatellite, Minisatellite), End User (Civil, Defense, Commercial), Application (Earth Observation & Meteorology, Communication, Scientific Research & Exploration), and Region. Retrieved July 13, 2019, from https://www.alliedmarketresearch.com

[8] Muthusamy, V. (2017). *Fault Detection, Isolation, and Prognosis for Control Moment Gyros Onboard Satellites* (Unpublished master's thesis). Aerospace Engineering, Ryerson University, Toronto,Canada

[9] Ruiter, A. H., Damaren, C. J., & R.Forbes, J. (2013). *Spacecraft Dynamics and Control: An Introduction*. United Kingdom, John Wiley & Sons, Ltd

[10] National Aeronautics and Space Adminstration. (1997). Reaction/Momentum Wheel. Retrieved June 13, 2019, from https://spinoff.nasa.gov/spinoff1997/t3.html

[11] Rahimi, A. (2017). *Fault Diagnosis and Prognosis of Satellite Attitude Control System with Reaction Wheels and Control Moment Gyros* (Unpublished doctoral dissertation). Aerospace Department, Ryerson University, Toronto, Canada

[12] Hyungjo, Y., & Panagiotis, T. (2002). Spacecraft Adaptive Attitude and Power Tracking with Variable Speed Control Moment Gyroscopes. *Journal of Guidance, Control, and Dynamics, 25*(6), 1081-1090. DOI:10.2514/2.4987

[13] Chen, R. H., Ng, H. K., Speyer, J. L., Guntur, L. S., & Carpenter, R. (2006). Health Monitoring of a Satellite System. *Journal of Guidance, Control, and Dynamics*, *29*(3), 593-605. DOI:10.2514/1.15012

[14] Lee, A. Y., & Wang, E. K. (2013, August). In-Flight Performance of Cassini Reaction Wheel Bearing Drag in 1997–2013. *Journal of Spacecraft and Rockets, 52(*2*)*, 470-480. DOI:10.2514/6.2013-4631

[15] Landau, Elizabeth (2017). Dawn Observing Ceres; 3rd Reaction Wheel Malfunctions. Retrieved June 15, 2019, from https://www.nasa.gov/feature/jpl/dawn-observing-ceres-3rd-reaction-wheel-malfunctions

[16] Putnam, D. (2018). Kepler's Second Mission: K2. Retrieved April 5, 2019, from https://www.nasa.gov/mission_pages/kepler/overview/index.html

[17] Tafazoli, M. (2009). A Study of on-orbit Spacecraft Failures. *Acta Astronautica*, *64*(2-3), 195-205. DOI:10.1016/j.actaastro.2008.07.019

[18] Japan Aerospace Exploration Agency (2014, May). Design Standard Attitude Control System. Retrieved June 01, 2019, from http://sma.jaxa.jp/en/TechDoc/Docs/E_JAXA-JERG-2-510A.pdf

[19] D.Mattingly, J. (2005). *Elements of Gas Propulsion*. New Delhi, Tata McGraw Hill Education Private Limited

[20] Hanachi, H., Mechefske, C., Liu, J., Banerjee, A., & Chen, Y. (2018, September). Performance-Based Gas Turbine Health Monitoring, Diagnostics, and Prognostics: A Survey. *IEEE Transactions On Reliability*, *67*(3), 1340-1363. DOI: 10.1109/TR.2018.2822702

[21] Transducer, S. (1987). The Vital Role of Sensors in Aircraft Engine Monitoring and Control. *Aircraft Engineering and Aerospace Technology*, *59*(7), 5-28. DOI:10.1108/eb036469

[22] Tumer, I. Y., & Bajwa, A. (1999). A Survey of Aircraft Engine Health Monitoring Systems. 35th American Institute of Aeronautics & Astronautics Joint Propulsion conference and Exhibit (20-24), Los Angeles, California. DOI:10.2514/6.1999-2528

[23] Zaidan, M. A., R.Relan, A.R.Mills, & Fleming, P. (2016). Gas Turbine Engine Prognostics Using Bayesian Hierarchical Models: A Variational Approach. *Mechanical System Signal Processing*, *70-71*(1), 120-140. DOI: 10.1016/j.ymssp.2015.09.014

[24] Ontario, Ministry of the Environment, Conservation, and Parks. (2019, March). Retrieved April 05, 2019, from http://www.airqualityontraio.com/science/pollutatn/paticulates.php

[25] Boston Health Institute (2018). State of Global Air/2018: A Special Report on Global Exposure to Air Pollution and Its Disease Burden. Retrieved June 15, 2019, from https://www.stateofglobalair.org/sites/default/files/soga-2018-report.pdf

[26] Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM*, *56*(12), 64-73. DOI:10.1145/2500499

[27] Laney, Dough. (2001, February). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Application Delivery Strategies, Meta Group. Retrieved August 30, 2019, from https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management- Controlling-Data-Volume-Velocity-and-Variety.pdf

[28] Oracle. (2013). Information Management and Big Data: A Reference Architecture, Retrieved August 15, 2019 from https://www.oracle.com/technetwork/topics/entarch /articles/info-mgmt-big-data-ref-arch-1902853.pdf

[29] Borne, K. (2014, April). Data Science Central. Retrieved August 10, 2019, from www.datasciencecentral.com: https://www.datasciencecentral.com/profiles/blogs/top-10-list-the-v-s-of-big-data

[30] Panimalar, A., Shree, V., & Kathrine, V. (2017, September). The 17 V's of Big Data. *International Research Journal of Engineering and Technology*, *4*(9), 329-333

[31] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambirdge, Massachusetts, London, The MIT Press

[32] Samuel, A. L. (2000). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *44*(1.2), 206–226. DOI:10.1147/rd.441.0206

[33] Boanccorso, G. (2017). *Machine Learning Algorithms*. Brimingham, United Kingdom, Packt Publishing

[34] Russell, S., & Norvig, P. (2010). *Artificial Intelligence - A Modern Approach*. Upper Saddle River, New Jersey, Prentice Hall

[35] Tuning, A. (1950). *Computing machinery and intelligence.* United Kingdom, Mind

[36] Gertler, J. (1988, December). Survey of Model-Based Failure Detection and Isolation in Complex Plants. *IEEE Control Systems Magazine*, *8*(6), 3-11. DOI:10.1109/37.9163

[37] Isermann, R. (1984, July). Process Fault Detection Based on Modeling and Estimation Methods - A survey. *Automatica*, *20*(4), 387-404. DOI:10.1016/0005-1098(84)90098-0

[38] Venkatasubramanian, V., Rengaswamy, R., Yin, K., & N.Kavuri, S. (2003, March). A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-based Methods. *Computers & Chemical Engineering*, *27*(3), 293-311. DOI:10.1016/S0098-1354(02)00160-6

[39] Simani, S., Fantuzzi, C., & Patton, R. J. (2003). Model-Based Fault Diagnosis Techniques. *Advances in Industrial Control*, 19-60. DOI:10.1007/978-1-4471-3829-7_2

[40] Marzat, J., Piet-Lahanier, H., Damongeot, F., & Walter, E. (2012, January). Model-Based Fault diagnosis for Aerospace Systems: A Survey. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 226*(10), 1329-1360. DOI:10.1177/0954410011421717

[41] Goebel, M., & Gruenwald, L. (1999). A Survey of Data Mining and Knowledge Discovery Software Tools. *SIGKDD Explorer*, *1*(1), 20-33

[42] Ahmad, H., Anwar, Z., & Shah, M. A. (2017). Data Mining Techniques and Applications - A Decade Review. *IEEE 23rd International Conference on Automation and Computing (ICAC),* (7-8 September), Huddersfield, UK, DOI:10.23919/IConAC.2017.8082090

[43] Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. (2012). Data Mining Techniques and Applications - A Decade Review From 2000 to 2011. *Expert Systems with Applications, 39*(12), 11303-11311. DOI:10.1016/j.eswa.2012.02.063

[44] Lee, I. N., Liao, S. C., & Embrechts, M. (2000). Data Mining Techniques Applied to Medical Information. *Medical Informatics and the Internet in Medicine, 25*(2), 81-102. DOI:10.1080/14639230050058275

[45] Srivastava, J., Cooley, R., Deshpande, M., & Tan, P.-N. (2000, January). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *ACM SIGKDD Explorations Newsletter*, *1*(2), 12-23. DOI:10.1145/846183.846188

[46] Sui, L., & Meng, L. (2001, July). Adding The value of The NAVTECH Road Database: an Implementation of Spatial Data Mining Techniques. *Journal of Geographical Sciences*, *11*(1), 69-73. DOI:10.1007/BF02837446

[47] Adderley, R., Townsley, M., & Bond, J. (2007). Use of Data Mining Techniques to Model Crime Scene Investigator Performance. *International Conference on Innovative Techniques and Applications of Artificial Intelligence (19-32)*, (11-13 December), Cambridge,United Kingdom. DOI:10.1016/j.knosys.2006.11.007

[48] Bagui, S., Mink, D., & Cash, P. (2007, April). Data Mining Techniques to Study Voting Patterns in the US. *Data Science Journal*, *6*(1), 46-63. DOI: 10.2481/dsj.6.46

[49] Parhizi, S., Shahrabi, J., & Pariazar, M. (2009). A New Accident Investigation Approach Based on Data Mining Techniques. *Journal of Applied Sciences*, *9*(4), 731-737. DOI:10.3923/jas.2009.731.737

[50] Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an

Academic Review of the Literature. *Decision Support Systems*, *50*(3), 559-569. DOI:10.1016/j.dss.2010.08.006

[51] Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153-1176. DOI:10.1109/COMST.2015.2494502

[52] Tahmasebian, S., Ghazisaeedi, M., Langarizadeh, M., Mokhtaran, M., Mahdavi-Mazdeh, M., & Javadian, P. (2017). Applying Data Mining Techniques to Determine Important Parameters in Chronic Kidney Disease and the Relations of These Parameters to Each Other. *Journal of Renal Injury Prevention*, *6*(2), 83-87. DOI:10.15171/jrip.2017.16

[53] Traore, B. B., Kamsu-Foguem, B., & Tangara, F. (2017, April). Data Mining Techniques on Satellite Images for Discovery of Risk areas. *Expert Systems with Applications*, *72*(1), 443-456. DOI:10.1016/j.eswa.2016.10.010

[54] Park, D., El-Sharkawi, M., Marks, R., Atlas, L., & Damborg, M. (1991). Electric Load Forecasting Using an Artificial Neural Network. *IEEE Transactions on Power Systems*, *6*(2), 442-449

[55] Trafalisa, T., Richman, M., White, A., & Santosa, B. (2002). Data Mining Techniques for Improved WSR-88D Rainfall Estimation. *Computers & Industrial Engineering*, *43*(4), 775-786. DOI:10.1016/S0360-8352(02)00139-0

[56] Park, D.-C., & Jeong, T.-K. J. (2002, May). Complex-Bilinear Recurrent Neural Network for Equalization of a Digital Satellite Channel. *IEEE Transactions on Neural Networks*, *13*(3), 711-725. DOI:10.1109/TNN.2002.1000135

[57] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780

[58] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A survey. *ACM Computing Surveys (CSUR)*, *41*(3), 15:1-15:58. DOI:10.1145/1541880.1541882

[59] Isermann, R. (2005). Model-Based Fault-Detection and Diagnosis - Status and Applications. *Annual Reviews in Control*, *29*(1), 71-85. DOI:10.1016/j.arcontrol.2004.12.0024

[60] Rahimi, A., Kumar, K. D., & Alighanbari, H. (2017). Fault Prognosis for Satellite Reaction Wheels Using Covariance Based Adaptive Unscented Kalman Filter and Particle Filter. *Acta Astronautica*, *134*(10), 159-169. DOI: 10.1016/j.actaastro.2017.02.003

[61] Castillo-Toledo, B., & Anzurez-Marin, J. (2005). Model-Based Fault Diagnosis Using Sliding Mode Observers to Takagi-Sugeno Fuzzy Model. *Proceedings of the 2005 IEEE InteMediterrean Conference on Control and Automation Intelligent Control (652-657)*, (27-29 June), Limassol, Cyprus. DOI:10.1109/.2005.1467091

[62] Hacker, J. M., Ying, J., & Lai, P. C. (2015). Reaction Wheel Friction Telemetry Data Processing Methodology and On-Orbit Experience. *Journal of Astronautical Sciences, 62*(3), 254-269. DOI:10.1007/s40295-015-0076-7

[63] Yairi, T., Nakatsugawa, M., Hori, K., Nakasuka, S., Machida, K., & Ishihama, N. (2017). Adaptive Limit Checking for Spacecraft Telemetry Data Using Regression Tree Learning. *IEEE International Conference on Systems, Man and Cybernetics (10-13)*, (10-13 October), The Hague, Netherlands. DOI:10.1109/ICSMC.2004.1401008

[64] Patton, R., Lopez-Toribio, C., & Uppal, F. (1999). Artificial Intelligence Approaches to Fault Diagnosis for Dynamic Systems. *International Journal of Applied Mathematics and Computer Science, 9*(3), 471-518

[65] R.Rengaswamy, Mylaraswamy, D., K.-E.Årzén, & V.Venkatasubramanian. (2001, December). A Comparison of Model-Based and Neural Network-Based Diagnostic Methods. *Engineering Applications of Artificial Intelligence, 14*(6), 805-818. DOI:10.1016/S0952-1976(02)00010-6

[66] Al-zyoud, I., & Khorasani, K. (2006). Neural Network-Based Actuator Fault Diagnosis for Attitude Control Subsystem of a Satellite. *IEEE Automation Congress, 2006 World Automation Congress*, (24-26 July), Budapest, Hungary

[67] Joshi, A., Gavriloiu, V., Barua, A., Garabedian, A., Sinha, P., & Khorasani, K. (2007). Intelligent and Learning-Based Approaches for Health Monitoring and Fault Diagnosis of RADARSAT-1 Attitude Control System. *IEEE International Conference on Systems, Man and Cybernetics (3177-3183)*, (7-10 October), Montreal, Canada. DOI:10.1109/ICSMC.2007.4414081

[68] Taylor, A., Leblanc, S., & Japkowicz, N. (2016). Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks. *IEEE International Conference on Data Science and Advanced Analytics (DSAA) (130-139)*, (7-10 October), Montreal, Canada. DOI:10.1109/DSAA.2016.20

[69] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (387-395)*, (19-23 August), London, United Kingdom. DOI:10.1145/3219819.3219845

[70] Byington, C., Roemer, M., & Galie, T. (2002). Prognostic Enhancements to Diagnostic Systems for Improved Condition-Based Maintenance. *IEEE Aerospace Conference Proceedings (2815-2824)*, (9-16 March), Big Sky, MT, USA. DOI: 10.1109/AERO.2002.1036120

[71] Qiu, J., B.Seth, B., Y.Liang, S., & Zhang, C. (2002, September). Damage Mechanics Approach For Bearing Lifetime Prognostics. *Mechanical Systems and Signal Processing, 16*(5), 817-829. DOI:10.1006/mssp.2002.1483

[72] Li, H., Pan, D., & Chen, C. L. (2015, May). Reliability Modeling and Life Estimation Using an Expectation-Maximization Based Wiener Degradation Model for Momentum Wheels. *IEEE Transactions on Cybernetics, 45*(5), 969-977. DOI:10.1109/TCYB.2014.2341113

[73] Mosallam, A., Medjaher, K., & Zerhouni, N. (2014, June). Data-Driven Prognostic Method Based on Bayesian Approaches for Direct Remaining Useful Life Prediction. *Journal of Intelligent Manufacturing*, *27*(5), 1037-1048.DOI:10.1007/s10845-014-0933-4

[74] Loutas, T. H., Roulias, D., & Georgoulas, G. (2013, October). Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression. *IEEE Transactions on Reliability, 62*(4), 821-832. DOI:10.1109/TR.2013.2285318

[75] Liao, L., & Köttig, F. (2014, March). Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction. *IEEE Transactions on Reliability, 63*(1), 191-207. DOI:10.1109/TR.2014.2299152

[76] H.Skima, K.Medjaher, C.Varniera, E.Dedua, & J.Bourgeois. (2016, October). A Hybrid Prognostics Approach for MEMS: From Real Measurements to Remaining Useful Life Estimation. *Microelectronics Reliability, 65,* 79-88. DOI:10.1016/j.microrel.2016.07.142

[77] Li, X., Ding, Q., & Sun, J.-Q. (2018, April). Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks. *Reliability Engineering and System Safety, 172*(1), 1-11. DOI:10.1016/j.ress.2017.11.021

[78] Ellefsen, A. L., Bjorlykhaung, E., Aksoy, V., Ushakov, s., & Zhang, H. (2019). Remaining Useful Life Predictions for Turbofan Engine Degradation Using Semi-Supervised Deep Architecture. *Reliability Engineering and System Safety, 183*(1), 240-251. DOI:10.1016/j.ress.2018.11.027

[79] Rahimi, A., Kumar, K. D., & Alighanbari, H. (2015, April). Enhanced Adaptive Unscented Kalman Filter for Reaction Wheels. *IEEE Transactions on Aerospace and Electronic Systems, 51*(2), 1568-1575. DOI:10.1109/TAES.2014.130766

[80] Das, G., Pattnaik, P. K., & Padhy, S. K. (2014). Artificial Neural Network Trained by Particle Swarm Optimization for Non-Linear Channel Equalization. *Expert Systems with Applications, 41*(7), 3491-3496

[81] Goldberg, D. E., & Holland, J. H. (1988, October). Genetic Algorithms and Machine Learning. *Machine Learning, 3*(2-3), 95-99. DOI:10.1023/A:1022602019183

[82] Isermann, R. (2006). *Fault-Diagnosis Systems*. Berlin, Springer

[83] Sobhani-Tehrani, E., & Khorasani, K. (2009). *Fault Diagnosis of Nonlinear Systems Using a Hybrid Approach*. London, Springer

[84] Frank, P. M. (1990). Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-Based Redundancy: A Survey and Some New Results. *Automatica, 26*(3), 459-474. DOI:10.1016/0005-1098(90)90018-D

[85] Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault Detection and Diagnosis in Industrial Systems.* London, Springer

[86] Wu, Q., & Saif, M. (2009, October). Model-Based Robust Fault Diagnosis for Satellite Control Systems Using Learning and Sliding Mode Approaches. *Journal of Computers, 4*(10), 1022-1032. DOI: 10.4304/jcp.4.10.1022-1032

[87] Sneider, H., & Frank, P. M. (1996, May). Observer-Based Supervision and Fault Detection in Robots Using Nonlinear and Fuzzy Logic Residual Evaluation. *IEEE Transactions on Control Systems Technology, 4*(3), 274-282. DOI:10.1109/87.491201

[88] Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000). *Data-Driven Techniques for Fault Detection and Diagnosis in Chemical Processes*. London, Springer

[89] Scholkopf, B., Smola, A., & Muller, K. R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation, 10*, 1299-1319

[90] Nwobi, F. N., & Ugomma, C. A. (2014). A Comparison of Methods for the Estimation of Weibull Distribution Parameters. *Metodološki zvezki, 11*, 65-78

[91] Fine, S., Singer, Y., & Tishby, N. (1998). The Hierarchical Hidden Markov Model: Analysis and Applications. *Machine Learning, 32*(1), 41-62

[92] Box, G. E., & Pierce, D. P. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association, 65*(332), 1509-1526

[93] Widodo, A., & Bo-SukYang. (2007, August). Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis. *Mechanical Systems and Signal Processing, 21*(6), 2560-2574. DOI:10.1016/j.ymssp.2006.12.007

[94] Chang, H. s., Hu, J., Fu, M. C., & Marcus, S. I. (2013). *Simulation-Based Algorithms for Markov Decision Processes.* London, Springer

[95] Diepen, G., Everlo, T. S., & Bouazzaoui, H. E. (2017, April). Deloitte, Retrieved August 21, 2019, from https://www2.deloitte.com/nl/nl.html,

https://www2.deloitte.com/nl/nl/pages/data-analytics/articles/part-2-artificial-intelligence-techniques-explained.html

[96] Shrivastava, G., Karmakar, S., & Kowar, M. K. (2012). Application of Artificial Neural Networks in Weather Forecasting. *International Journal of Computer Applications, 5*(18), 17-29

[97] Basheer, I. A., & M, H. (2000). Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *Journal of Microbiological Methods, 43*(1), 3-31. DOI: 10.1016/S0167-7012(00)00201-3

[98] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997, January). Face Recognition: A Convolutional Neural-Network Approach. *IEEE Transactions on Neural Networks, 8*(1), 98-113. DOI: 10.1109/72.554195

[99] Su, X., & Khoshgoftaar, T. M. (2009, August). A Survey of Collaborative Filtering Techniques. (J. Hong, Ed.) *Advances in Artificial Intelligence*, Florida, USA. DOI:10.1155/2009/421425

[100] Greff, Klaus, Srivastava, R., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE transactions on neural networks and learning systems, 28*(10), 2222-2232. DOI: 10.1109/TNNLS.2016.2582924

[101] D.A.Dickey, & A.Fuller, W. (n.d.). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association (427-431)*. DOI:10.2307/2286348

[102] Larson, K. A., McCalmont, K. M., Peterson, C. A., & Ross, S. E. (2014). Kepler Mission Operations Response to Wheel Anomalies. *2014 SpaceOpsConference*, (AIAA 2014-1882). DOI:10.2514/6.2014-1882

[103] (2013). *System Engineering Report- Kepler ADCS General Information.* NASA.

[104] MATLAB.(2018). *The MathWorks, Natick*. Massachusetts, United States.

[105] Bialke, W., & Hansell, E. (2017). Newly Discovered Branch of the Fault Tree Explaining Systemic Reaction Wheel Failures and Anomalies. *Proceedings - ESMATS 2017*, University of Hertfordshire, Hatfield, UK

[106] GFZ German Research Centre for Geosciences Helmholtz Center Potsdam. Retrieved March 08, 2019, from https://www.gfz-potsdam.de/en/home/: ftp://ftp.gfz-potsdam.de/pub/home/obs/kp-nowcast-archive/dat/

[107] Niu, G. (2016). *Data-Driven Technology for Engineering Systems Health Management.* Beijing, Springer

[108] Saxena, A., & Goebel, K. (2008). *Turbofan Engine Degradation Simulation Data Set*, *NASA Ames Prognostics Data Repository*. NASA Ames, Moffett Field, California, USA

[109] Coble, J. B. (2010). Merging Data Sources to Predict Remaining Useful. University of Tennessee, Knoxville

[110] Savitzky, A., & Golay, M. J. (1964). Smoothing and Differentiation of Data by Simplified Least-Squares Procedures. *Analytical Chemistry, 36*(8), 1627-1639. DOI:10.1021/ac60214a047

[111] Luo, J., Ying, K., & Bai, J. (2005, July). Savitzky–Golay Smoothing and Differentiation Filter for Even Number Data. *Signal Processing, 85*(7), 1429-1434. DOI:10.1016/j.sigpro.2005.02

[112] Jolliffe, I. (2002). *Principal Component Analysis, Second Edition.* Aberdeen, Springer

[113] Mitchell, M. (1996). *An Introduction to Genetic Algorithms.* Massachusetts, MIT Press Paperback Edition

[114] Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *IEEE International Conference*, (1942-1948). Washington, USA

[115] Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017). Long Short-Term Memory Network for Remaining. *IEEE International Conference on Prognostics and Health Management (ICPHM),* Texas, USA

[116] Grant, P (2016, August). Retrived June 27, 2018, from https://towardsdatascience.com/splitting-data-sets-cac104a05386

[117] Python Software Foundation. (n.d.). *Python Language Reference Version 3.3, TensorFlow with Spyder Interactive Environment*. http://www.python.org