ABSTRACT

BIG DATA ANALYTICS FOR GAS TURBINES

Srikanth Kashyap Jandhyala, MS
Department of Mechanical Engineering
Northern Illinois University, 2015
Dr.Abhijit Gupta, Director

Gas turbines have many important variables such as the load, turbine speed, fuel gas flow, and inlet and outlet pressures. The volume, velocity, variability and complexity of the data from various sensors are huge. Monitoring of gas turbines consequently needs big data analytics which is the process of collecting, organizing and analyzing large data sets. The need for big data analytics stems from the need to increase the efficiency, improve operations and predict various trends and comment on the performance. Big data implies that the data sets are too large to be analyzed or even viewed by conventional methods and software. The gas turbine data set considered in this thesis is a time series data.

In this thesis, a large data set from gas turbines is first made readable by converting it into the CSV format, as it is beyond the dimensional capability of Microsoft excel. The data is then analyzed using various statistical tools such as R-software. Combustion instabilities have been observed in the units and units with high dynamics have been determined. Data quality issues and missing data were observed. The limits of the blade path temperature spreads have been determined and the correlation of operational parameters were determined. Principle

Component analysis was performed to reduce the dimensionality of the data and observe health of the gas turbine operation in terms of dynamic behavior.

NORTHERN ILLINOIS UNIVERSITY
DEKALB, ILLINOIS

AUGUST 2015

BIG DATA ANALYTICS FOR GAS TURBINES

BY

SRIKANTH KASHYAP JANDHYALA

A THESIS SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

DEPARTMENT OF MECHANICAL ENGINEERING

Thesis Director:

Abhijit Gupta

ACKNOWLEDGEMENTS

It gives me immense pleasure to thank my advisor **Dr.Abhijit Gupta** who guided me in each and every step to complete this thesis work. He supported me in regards to research and scholarship. His great attitude towards work made me work comfortably without any hassle. I would also like to thank my committee members Dr. Sanjib Basu and Dr. Jenn Tern Gau who actively took time and gave some valuable comments to improve my thesis work.

A special grace to **Praveen Uppaluri** who encouraged me to complete my work and boosted me to overcome the faults in developing my thesis and guiding me by being my mentor in completing my master’s program.

.

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

1. INTRODUCTION 1

1.1 Big Data........ 1

1.1.1 Dimensions of Big Data 1

1.2 R-Software 2

1.2 Gas Turbines 2

1.3 Literature Review 2

1.4 Types of Sensors Used in Gas Turbines........ 4

1.5 Approach........6

2. DATA EXPLORATION 7

2.1 System Requirements to Run Big Data........ 7

2.2 Splitting the Data into Chunks 8

2.3 Exploring the Data 10

2.4 Determining the range of operational variables 14

2.5 Time Series Analysis........ 15

2.6 Data Quality 19

3. DYNAMICS AND BLADE PATH TEMPRRATURE SPREAD 23

3.1 Dynamics.. 23

3.2 Acoustic oscillations 24

3.3 Range of amplitudes for different bins... 25

3.4Counts of Dynamics 27

3.3 Blade Path Temperature Spread... 32

4. STATISTICAL ANALYSES 37

4.1 Correlation Analysis... 37

4.2 Principle Component Analysis.. 43

5. CONCLUSION.. 48

REFERENCES 49

LIST OF TABLES

Table 1: Distribution of the Data 9

Table 2: Operational Data 11

Table 3: Dynamics Data 12

Table 4: Logic tag Data 13

Table 5: Thermocouple Data 13

Table 6: Quantile range of generator load (DM2) 14

Table 7: Amplitude and frequency range 25

Table 8: Observations of Bin-4 28

Table 9: Observations of can-2 30

Table 10: Counts of blade path temperature spread 34

Table 11: Correlation matrix considering one file 39

Table 12: Correlation matrix of case-3 42

Table 13: Importance of components 45

Table 14: PCA analysis 46

LIST OF FIGURES

FIGURE PAGE

Figure 1: Working of gas turbine 3

Figure 2: Sensor placement diagram 5

Figure 3: 3D surface plot of generator load 15

Figure 4: Time Vs Generator Load 17

Figure 5: Time Vs Turbine speed 17

Figure 6: Time Vs Load 18

Figure 7: Time Vs Turbine speed 19

Figure 8: Missing data 20

Figure 9: Missing dynamics 21

Figure 10: Damaged burner assembly 23

Figure 11: Classification of bins 26

Figure 12: 3D histogram showing dynamics of Bin-4 29

Figure 13: 3D histogram showing dynamics of unit-3 29

Figure 14: 3D histogram of can-4 31

Figure 15:3D histogram of unit-6 32

Figure 16: 3D histogram of blade path temperature spread 35

Figure 17: Correlation plot 38

Figure 18: Pair plot of case-1 40

Figure 19: Pair plot of case 2 41

FIGURE PAGE

Figure 20: Pair plot of case-3 42

Figure 21: Scree plot 44

CHAPTER.1- INTRODUCTION

1.1 Big Data

Terabytes and terabytes of data at multiple locations in various formats whose scale, diversity and complexity require new algorithms and techniques to analyze and make value of it. Hence, this is termed as big data [1]. These data can be obtained from sensors used for various applications such as for climate information, posts to social media sites, cell phone applications and GPS etc. [1]

1.1.1 Dimensions of Big Data

1. Volume: It refers to the amount of data collected over a period of time, usually the data is gigabytes or terabytes.[2]
2. Velocity: It refers to the data which is streaming and has minutes to respond with analysis, has high volumes of data continuously collected in the cloud [2].
3. Variety: It is the type of data being collected or analyzed, the data can be in various formats such as media, logic tags, text , numeric etc.[2]

Data analysis is gaining importance in current world, the sensors are getting inexpensive day by day and automation is increasing, hence ,a lot of data is getting stored which fills up the server capacity. An effective utilization of this data using the statistical and software techniques is today's world challenge. This data can be used to predict the behavior of the system in an accelerated period of time.

1.2 R-Software

Bell labs developed R-software. R was initially written by Robert Gentleman and Ross Ihaka also known as R&R of the statistics department of university of Auckland [3]. This software is used for statistical computing and graphics. This software is available in open source. It is used for building the data analytics algorithms. R software is in built with manipulation, calculations and graphical display. [3]. Programs can be executed and stored and can be loaded in into the R-package database for future use.

1.3 Gas Turbines

Gas turbine is the engine at the heart of the power plant which produces electricity [4]. The electricity produced must be continuously utilized as it cannot be stored in any other form. The process of burning of fuel to produce power is called as combustion. During the process of combustion the natural gas is converted into mechanical energy, further used to rotate a generator which is used to generate electricity [4] .This electricity moves around power grids to supply to home and offices. The operation of the gas turbine is shown. (See Figure: 1)

1.4 Literature Review

Ann P. Dowling *et al* [5] modeled acoustic analysis of combustion dynamics to investigate dynamics when lean premixed pre vaporized combustion is used of NOx. Also determined that acoustics of the gas turbine from compressor exit to turbine entry play an important role in combustion instabilities.

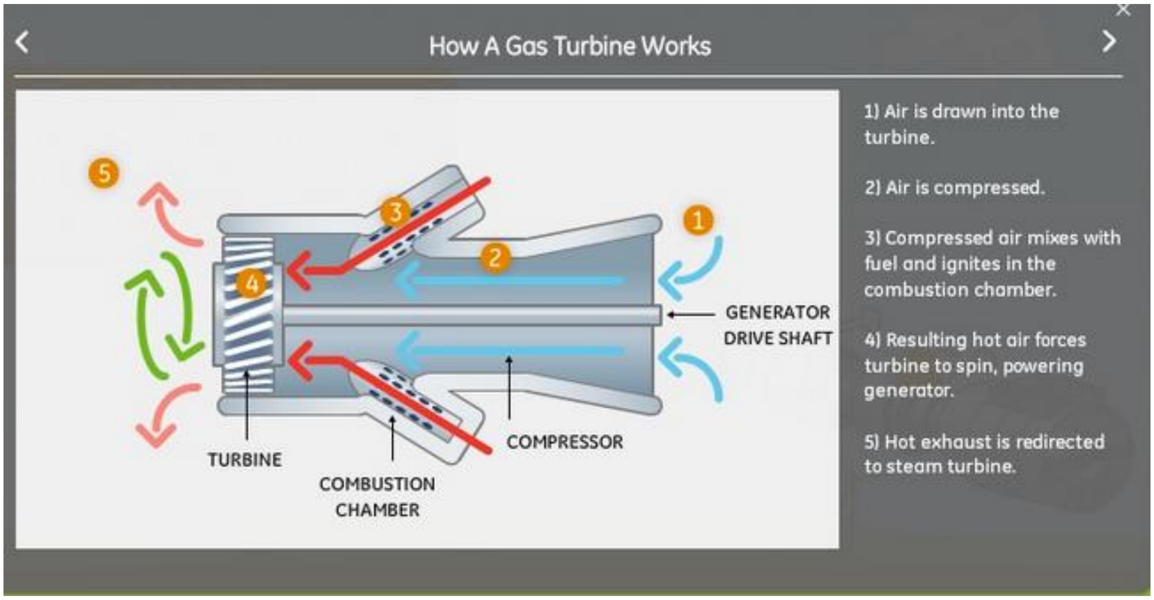

Figure 1: Working of gas turbine [4]

Robert E Dundas *et al* [6] monitored the performance of the gas turbine to prevent failure, monitoring of blade path temperature spread is described and examples of failure are discussed.

Barna Saha *et al* [7] described the data quality issues and explains the use of fours V's in big data which are volume, velocity, variety and veracity.

David E Hobson *et al* [8] measured vibration in the combustion chamber and observed the bandwidth, which is a measure of damping of resonant mode of the combustion chamber's acoustic resonance. It decreases towards zero as the machines approach their combustion stability limits.

Keith Mc Manus *et al* [9] determined the thermos acoustic response of an industrial scale gas turbine combustor. A model is created to observe the dynamic behavior of the combustor over a variety of operating conditions.

Mauro Venturini *et al* [10] used statistical techiniques for gas turbine analysis and obtained probability density function for prediction of dynamics and turbine degradation.

1.5 Types of Sensors Used in Gas Turbines

A sensor is a device which records a change in the physical property of the system [11]. It records and responds to the change in the form of a signal. The sensors in the gas turbines are placed in different locations. Figure: 2 shows the placement of sensor on a gas turbine. Pressure sensors and temperature sensors are the important sensors placed in the gas turbine, other sensors used are fuel sensors, control sensors etc. [11]

Pressure sensors are used to measure pressure oscillations, amplitude and pulsations in gas turbines, and also to determine the combustion instabilities in the gas turbine [11]. The pressure sensors are of main focus in this thesis as it determines the dynamics in the gas turbines. The types of pressure sensors used are.

1. Remote sensors

   These sensors are portable, these sensors are placed in separate housing and these are connected to the sensing lines connected to the combustor. These sensors are simple to setup and are cost effective [11].

2. Closed coupled sensors

   These sensors are mounted on the gas turbine, they operate at a higher frequency than the remote sensors. Due to high operating temperatures of the gas turbine, these sensors have high thermal capability to sense precisely at high temperatures [11].

3. On Turbine instability sensors

These are high temperature sensors which are mounted directly on the sensors. These sensors are highly reliable as they produce consistent results. Due to the placement of the sensors right on top of the can the combustion analysis is highly accurate [11]

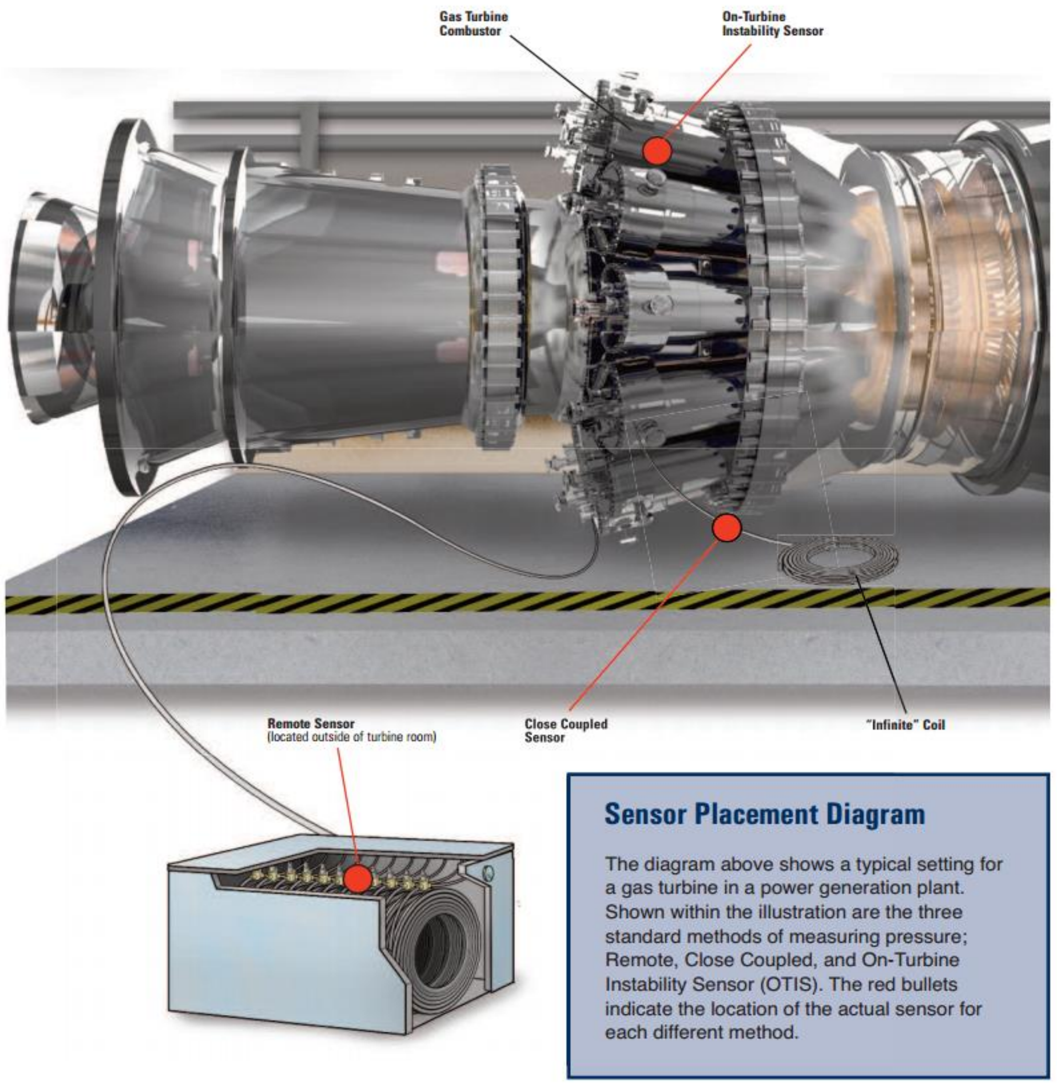

Figure 2: Sensor placement diagram [11]

Temperature sensors

Temperature sensors are connected to the turbine sensing lines. These sensors also enable vibration measurement in high temperatures which may go up to 1300F [11]. They are used to measure the temperature and respond to the signal if the temperature exceeds the threshold limit.

1.6 Approach

Dataset was obtained in a comma separated value (CSV) format. It is the actual data of a running gas turbine which is collected over a period of time. The gas turbine consists of 18 cans around the periphery of the turbine, each can is a separate combustion unit by itself. During the process of combustion each of the can undergoes the process of burning fuel and produces power which is collectively used to rotate the shaft of the engine to produce power. There are six operating ranges of amplitudes and frequencies, they are called as bins. Each bin has a specified range of amplitude and frequency. Therefore, every can operates with six ranges of bins. Data consists of values of operating frequencies of every can and their respective bins.

Data management techniques such as splitting the data into chunks and cleaning the unwanted data is performed to analyze the data using statistical tools and techniques. Combustion instabilities are the main cause of damage to the turbine. Occurrence of combustion instabilities has been identified in the obtained dataset using techniques such as correlation and Principal component analysis. All the techniques are further discussed with observations in the following chapters.

CHAPTER.2- DATA EXPLORATION

2.1 System Requirements to Run Big Data

R-software is used in this thesis to analyze big data obtained from gas turbines, the system requirements to run the big data are very important. A calculation on the memory size required by the data is calculated below. The data obtained was of 8.3 million rows and 277 columns which is about 10.6 Gigabytes in size.

If all the data of size 10.6GB is imported in the software the amount of memory required by the computer is determined.

8,291,143 rows and 277 columns

8,291,143 X 277= 2296646611

= 2296646611/2^20bytes/MB

= 17.11GB of RAM memory required.

As per the calculation 17.11Gigabytes of RAM is required to import the file into R-software at once. RAM should be at least 2.5 times the obtained value as there are other programs also running simultaneously in the computer. Therefore the actual RAM memory required is 17.11 X 2.5 = 43GB of RAM.

43GB of RAM is not available on regular computers hence the file needs to be split into chunks

2.2 Splitting the Data into Chunks

The data of these gas turbines was obtained in a CSV (comma separated value) format. This data was obtained in a single CSV file which has the size of 10.6 GB (Giga Bytes). All the obtained data was collected with respect to time and is running over five years.

It has about eight million rows and two hundred and seventy seven columns. This large file is beyond the dimensional capability of Microsoft excel to handle. The latest version of MS excel can handle data up to 350,000 rows only. But the functionality of MS excel becomes very slow after 100,000 rows hence leads to crashing of the file. Hence, the need for a software to analyze large data sets stems into Big Data Analytics R-software is used in the analysis as R is a free ware, requires no licensing and is also very efficient in handling larger data sets.

Since the file is too big, the data is split into chunks to exactly know what variables are present in the data. The splitting of the data was done using a software tool called CSV splitter. CSV splitter is a tool used to split large comma separated files into smaller files. CSV splitter splits the files based on either number of rows or number of files to be split.

This large data was first split into 83 files of which 82 files had 100,000 rows in them and the last file had 38,000 rows. Hence, has a total of 83 files as chunks which is the data over 14 operating gas turbines. Every turbine has about 400,000 rows each running into four or five CSV files. Each of these CSV files can easily be opened in MS excel. Hence, being able to view the columns or the variables in the gas turbines. Now these variables can be analyzed in R software to predict patterns and dynamics.

Observations

The larger data set is split into chunks which split up to 83 files. Table: 1 shows the observations of units of gas turbines in different files.

Table 1: Distribution of the Data

Unit number	**Data file number**
Unit-1	Data-1 to Data-4
Unit-2	Data-5 to Data-11
Unit-3	Data-12 to Data-17
Unit-4	Data-18 to Data-20
Unit-5	Data-21 to Data-27
Unit-6	Data-28 to Data-36
Unit-7	Data-37 to Data-41
Unit-8	Data-42 to Data-47
Unit-9	Data-48 to Data-52
Unit-10	Data-53 to Data-57
Unit-11	Data-58 to Data-63
Unit-12	Data-64 to Data-70
Unit-13	Data-71 to Data-76
Unit-14	Data-77 to Data-83

2.3 Exploring the Data

The obtained data was explored in R-software, the data consists of 277 columns, each column is considered as a variable in operating the gas turbine. The data obtained in every column is through a sensor mounted on the turbine which records the observation over a period of time.

The obtained data was found to be of 14 different units. The summary of the variables was obtained in R-software which showed the relevance of each variable which can be categorized into four types, they are Operational data, Logic tags, Dynamics data and thermocouple data. The dynamics Data is the main subject of interest in this thesis, there are 18 cans around the gas turbine each of which is a combustion unit by itself, Dynamic pressure sensors are placed all over the 18 cans which measure both amplitude and frequency.

•**Operational Data**: These are the variables which really tell how the system is working, some of the kinds of operational data are generator load, which gives the load applied on the gas turbine, fuel gas flow which is measured by the fuel sensor, Compressor discharge pressure which measures the pressure through the pressure sensor. Some of the important variables Load, fuel stroke reference, Compressor discharge pressure, and compressor discharge temperature etc.

•**Logic Tags**: Logic tags are the control commands used in operation of a gas turbine , these logic tags are of less relevance in this thesis as they are usually binary either ‘0’ or ‘1’ . Hence they cannot be analyzed in R –software efficiently.

•**Dynamics data**: Dynamics is the measure of instability of the turbine, it basically has two main observations the high cycle fatigue and the low cycle fatigue. The high cycle fatigue is proven to be dangerous for the functionality of the turbine. The dynamics data consists of two kinds of observations the amplitude and the frequency, later in this thesis the effects of high amplitude and high frequency are discussed.

•**Thermocouple data**: This data gives the observations of the temperature in the turbine, the temperature is usually measured at the exhaust of the turbine, the units used are Fahrenheit, and usually the temperature at the exhaust is 1200F.

Observations of exploring the data

Operational data

A brief set of variables are shown for the operational data (see Table: 2)

Table 2: Operational Data

Description	**Tag name**
Generator Load in Mega Watts	DM2
Fuel Stroke reference	FI9
Compressor Discharge pressure	P5
Compressor discharge Temperature	T1
Fuel Gas Flow	MF5

Dynamics data

The dynamics consist of two parts one is the amplitude and other is the frequency. There are six bins numbered from 1to 6, for each bin there are observations of both amplitude and frequency. In total there are eighteen cans around the turbine, so each can has six bins of amplitude and frequency. Nomenclature example: "OA1a", "OA"- Overall amplitude, "1" – Bin No. "a"- 1st can

Listing out all the amplitudes of Bin "1" over all the eighteen cans are shown (see Table 3)

Table 3: Dynamics Data

Description	**Tag names**
Dynamics Amplitude bin-1 can-1	OA1a
Dynamics Amplitude bin-1 can -2	OA1b
Dynamics Amplitude bin-1 can -3	OA1c
Dynamics Amplitude bin-1 can-4	OA1d
Dynamics Amplitude bin-1 can-5	OA1e
Dynamics Amplitude bin-1 can-6	OA1f
Dynamics Amplitude bin- 1 can -7	OA1g
Dynamics Amplitude bin-1 can-8	OA1h
Dynamics Amplitude bin-1 can-9	OA1i
Dynamics Amplitude bin-1 can-10	OA1j
Dynamics Amplitude bin-1 can-11	OA1k
Dynamics Amplitude bin-1 can-12	OA1l
Dynamics Amplitude bin-1 can-13	OA1m
Dynamics Amplitude bin-1 can-14	OA1n
Dynamics Amplitude bin-1 can-15	OA1o
Dynamics Amplitude bin-1 can-16	OA1p
Dynamics Amplitude bin-1 can-17	OA1q
Dynamics Amplitude bin-1 can-18	OA1r

Similarly, the frequencies are listed over the 18 cans for each bin. Hence there are 72 variables of amplitude and 72 variables of frequencies in the data.

Logic Tag Data

The logic tag data are the controls used in operating the turbine, they are binary hence 0 or 1. A sample of the logic tag variables in the data are shown (see Table 4)

Table 4: Logic tag Data

Description	**Tag name**
Logic Tag-1	L1
Logic Tag-2	L2
Logic Tag-3	L3
Logic Tag-4	L4
Logic Tag-5	L5
Logic Tag-6	L6

Thermocouple Data

This data has observations of the temperature inside the turbine at various locations measured by the temperature sensors .Some of the variables of the thermocouple data are shown (see Table5)

Table 5: Thermocouple Data

Description	**Tag names**
Combustion monitor actual spread	T10
Combustion monitor actual spread-1	T7
Combustion monitor actual spread-2	T8

2.4 Determining the range of operational variables

The range of the operational variables i.e. the maximum and the minimum values is associated with each of the operational variable. Therefore in R software, a loop is run to determine the quantile values of each of the operational variables. Now the actual range of each of the operational variables is known. This quantile determines the percentile values of each of the operational parameter over 14 units. They are split into 0%, 5%, 25%, 50%, 75%, 100% and the mean is also calculated in this analysis. Considering the generator load (DM2) and finding the quantile range of the operational variables (see Table 6)

Table 6: Quantile range of generator load in Megawatts (DM2)

DM-2	**0%**	**5%**	**25%**	**50%**	**75%**	**95%**	**100%**	**mean**
unit-1	-199.76	81.04	250.84	380.92	401.89	412.34	427.50	324.59
unit-2	-14.14	216.59	236.12	277.01	346.64	410.01	437.99	292.12
unit-3	-9.28	122.11	223.03	258.93	273.27	282.62	292.50	236.86
unit-4	-3201.68	200.58	342.70	390.84	400.73	416.23	430.15	356.92
unit-5	-145.29	96.58	146.27	203.10	237.02	280.44	304.80	192.20
unit-6	-73.92	193.88	241.17	269.84	278.58	287.24	295.31	254.30
unit-7	-7.95	51.46	193.49	225.70	249.54	265.25	287.49	208.41
unit-8	-8.30	84.69	191.70	225.75	239.98	259.06	285.67	206.86
unit-9	-8.05	101.58	142.72	191.87	233.82	266.30	291.64	187.41
unit-10	-8.60	108.00	153.44	194.34	232.84	262.00	287.33	190.12
unit-11	-9.55	90.02	106.59	179.54	243.13	264.90	280.41	175.04
unit-12	-148.21	80.94	102.81	168.62	243.87	270.35	285.46	171.79
unit-13	-5.10	255.55	303.20	358.84	403.70	424.21	434.72	346.24
unit-14	-8.29	305.85	334.63	359.37	403.88	425.06	437.09	360.98

After the table is obtained a 3D surface plot is also plotted to know the range of the variables over all the 14 different gas turbine units as shown (see Figure 3)

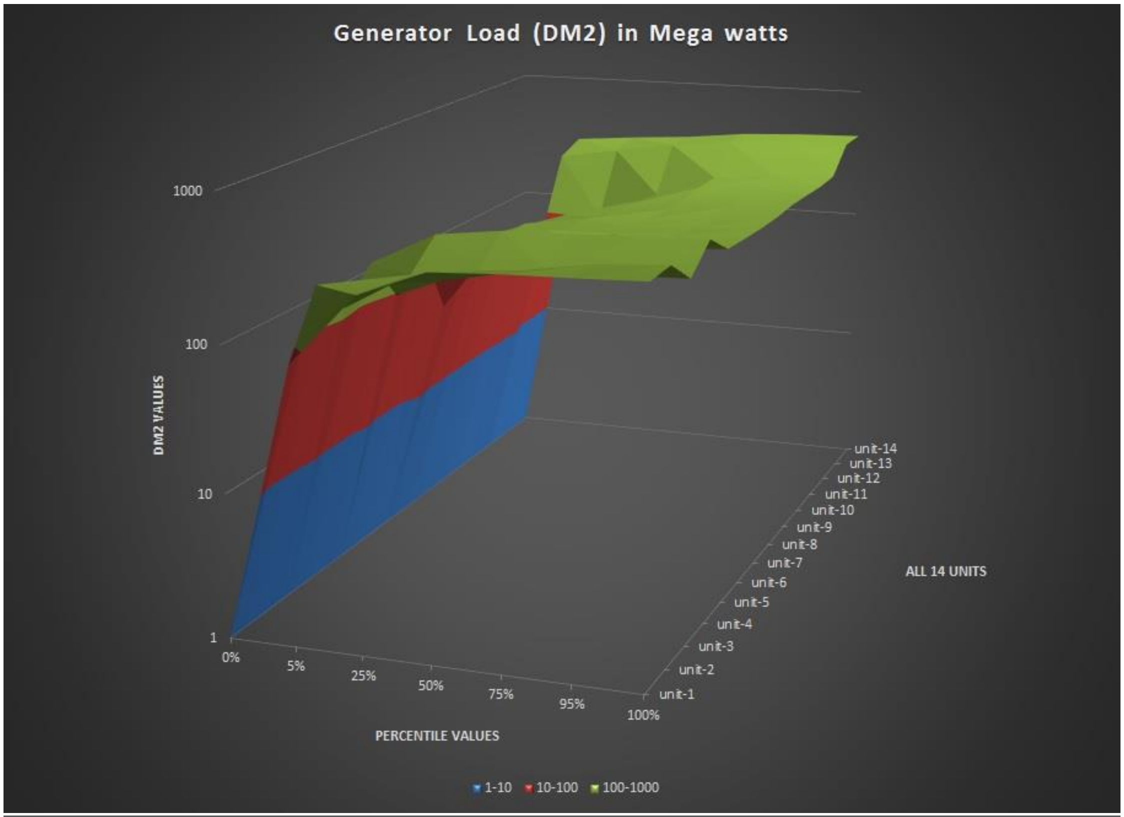

Figure 3: 3D surface plot of generator load

2.5 Time Series Analysis

After exploring the data and finding the types and categories of variables, the next step to find whether the data is continuous with respect to time, the data is collected every minute so at every minute one data point is generated , hence we plot time series plots to know this. Some of the operational variable are taken and plotted against time and the continuous patterns are observed.

For our convenience and to reduce the character size the time stamp data is modified into numbers in the plots, the general form of time stamp is MM/DD/YYY/HH/MM/SS i.e. Month/Day/Year/Hour/Min/Sec. Each interval in the plot should be technically showing the entire time stamp. But when Plotting in R-software, Software reads everything in character size, hence it was too big to be analyzed in R-software, it had to be eventually modified and given numbers.

For Example: 01/01/2015 10:00:00 is numbered as 1

01/01/2015 10:01:00 is numbered as 2

Since the data is too large and is split into 83 files , now we check using time series plots whether the data is continuous or not. Hence, two sets of plots have been plotted for some operational variables, the two sets are named as SET-A and SET-B plots.

SET-A

Plots of data collected over one operational day of the Gas Turbine. In this set, the key variables are plotted against time to observe the continuous patterns and dependability of variables.

One of the file was opened and one operational day was chosen say any random day when the turbine is running from 9am to 6pm. So, when at every minute the data points are generated, we should have 540 points as time for a given day. These time points are plotted with some operational variable. Plotting a variable load with respect to time as shown in the figure, the load increases stepwise as the time advances (see figure 4)

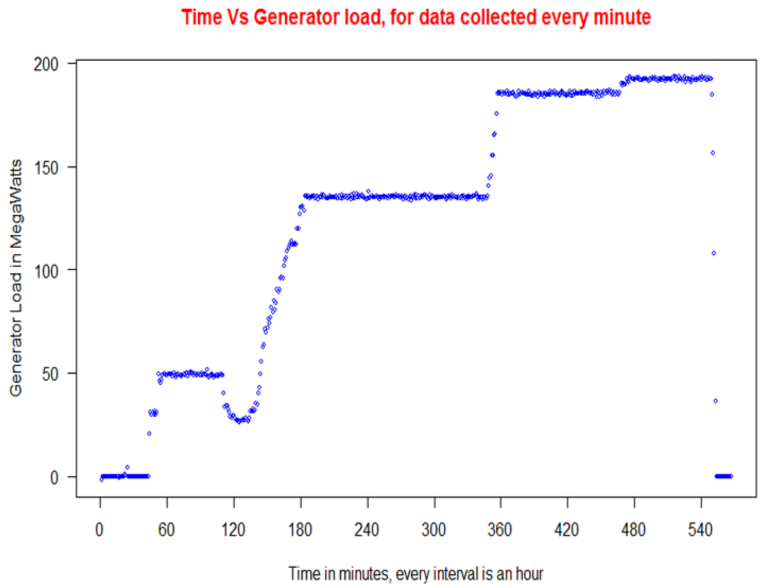

Figure 4: Time Vs Generator Load

As long as the unit is generating load the turbine speed is at 100% (see figure 5)

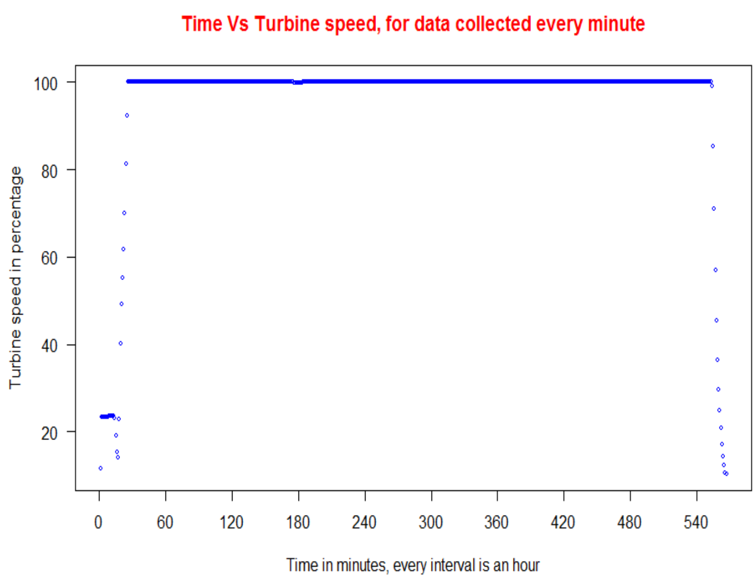

Figure 5: Time Vs Turbine speed

SET- B Plots

Plots of data collected over one year, considering one data point at particular time each day for 365 days. (i.e. data points at 10am for every day.)

Every day one point is collected over a span of one year for one unit of the data. For filtering and separation of one point everyday MS excel was used. This gives us a wider picture whether the data is continuous with respect to time or not. Hence, for time 365 points are generated over one year and are plotted against some of the operational variables. The two plots shown in figures show that unit is not generating load, but the turbine is still running .This phenomenon is called_idling. (See figure 6 &7)

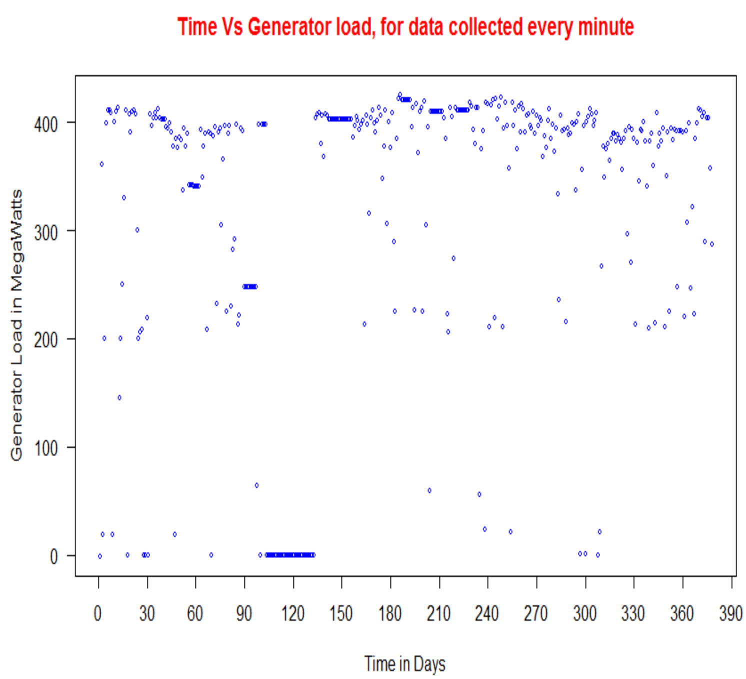

Figure 6: Time Vs Load

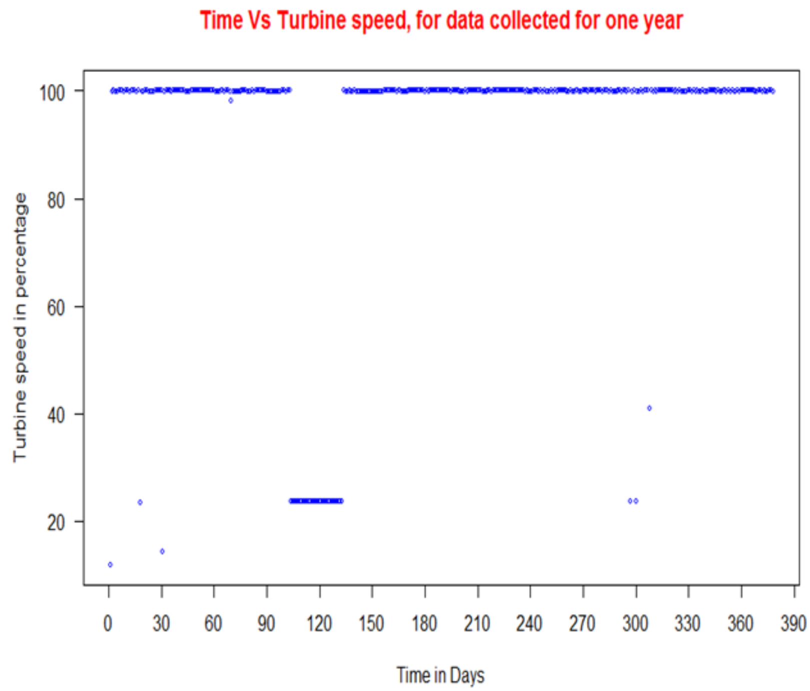

Figure 7: Time Vs Turbine speed

2.6 Data Quality

In this data also there is a quality issue was observed while measuring dynamics. There is an issue of incompleteness in this data. During the process of writing the codes in R and running the results for a variable turned out to be zero in all the files. A code was written to count the points above 2Psi in dynamics and the variable named OA1o to OA6o showed no values in all 83 files. It is possible for a sensors to be defective of OA1o to OA6o for one unit i.e. in a few files, but these six variables are showing no values in all the 83 files. This is an issue with quality of the data. The sensor could have not been functional at one unit, but as per the results in all the files these six variables turn out to be zeros or with no values of amplitudes in them. This could be an entry quality which doubts whether the data entered the system correctly at the origin? It is the most easiest to identify but most difficult to correct such kind of errors. There could be other errors in this data, but this error in quality was found in my observation. (See figure 8)

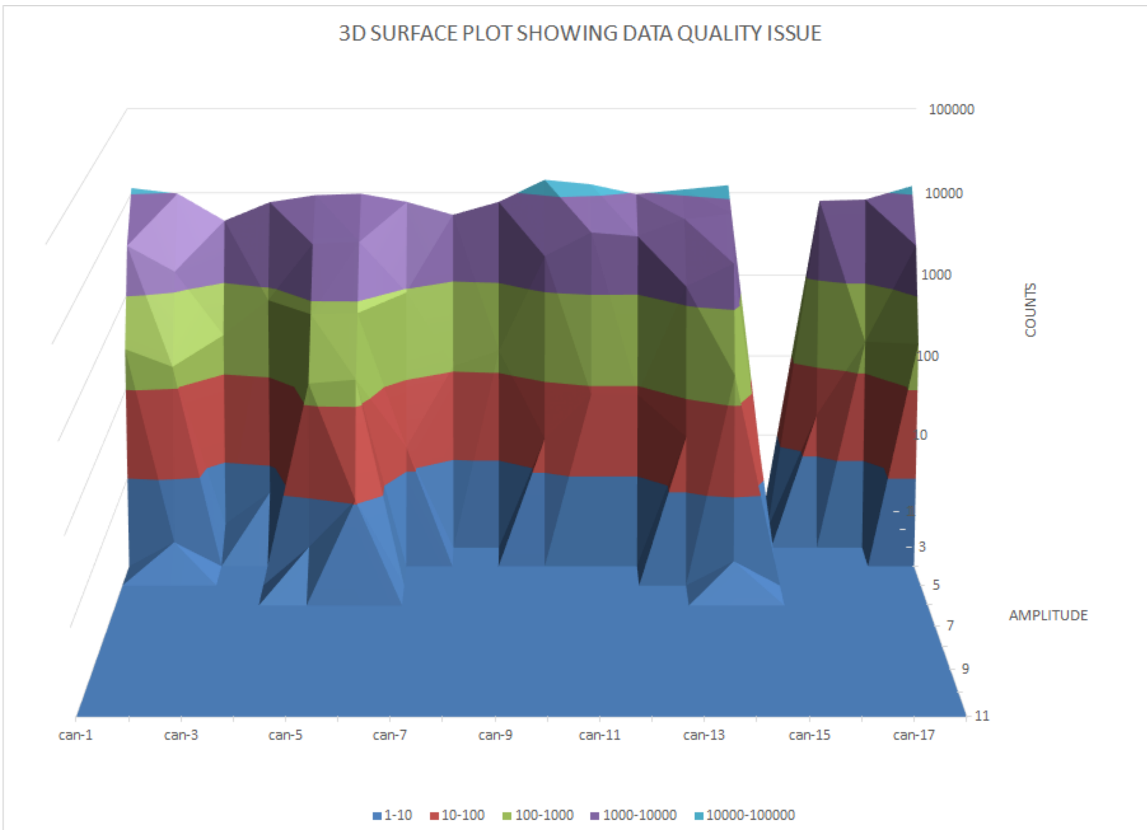

Figure 8: Missing data

Another data missing quality was observed in the data. Unit no 14 and Unit 15 did not have any data for dynamics over the 18 cans. (See figure 9)

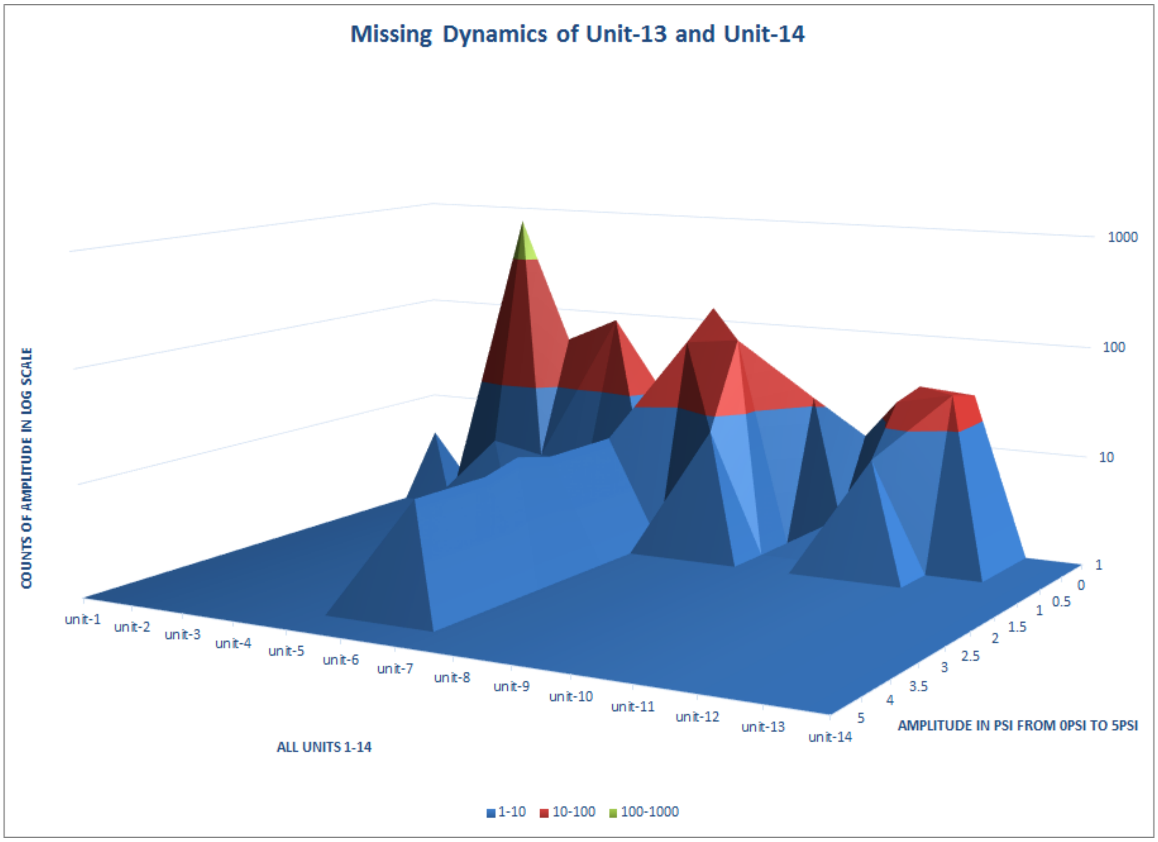

Figure 9: Missing dynamics

CHAPTER.3- DYNAMICS AND BLADE PATH TEMPRRATURE SPREAD

3.1 Dynamics

Dynamics are the interactions between combustion and acoustics in the combustor unit of the gas turbine [12]. These interactions are caused due to the disturbances of the air flow into the combustion unit [12]. These disturbances in the combustion effect the operation of the turbine causing stress and thermal loading which with effect the material properties and degrade the life of the turbine by failing it.

Combustion systems are prone to dynamics. The pressure oscillations elevate mechanical load in the chamber, which results in high amplitude vibrations. Figure 10 illustrates a comparison between an intact burner assembly and a damaged burner assembly due to dynamics [13].

Figure 10: Damaged burner assembly [13]

3.2 Acoustic oscillations

The sound emitted from the combustor is described in various ways. Some companies term them as "rumble" which describes all the acoustic oscillations at all ranges of frequency [14]. As otherwise "rumble" or "growl" is for the noise in the low frequency range of 50-180 Hz. For higher frequencies the term "howl" or "humming" is used. Prediction of accurate dynamics with amplitudes and frequencies is not possible as there is no accurate universal model. The dynamics is usually carried out by case to case basis [14].

Arthur H. Lefebvre [14] provided the name "GROWL" to describe acoustic oscillations in this paragraph, and here is a quote about "GROWL":

> The characteristics of growl vary from engine to engine and between engine startups, but its onset may occur soon after ignition is accomplished and it may persist at engine speeds up to idle .Growl is considered to be undesirable because it lengthens engine startup time and reduces compressor stall margins. According to Seto some compressors are quiet tolerant to grow. Whereas others appear to have growl- related stall problems.
>
> Certain engine parameters affect growl. Increase in combustor inlet air temperature decreases the speed range and intensity of growl, whereas an increase in combustion pressure has the opposite effect of promoting growl.
>
> Methods of alleviating growl include improvements in primary zone flow patterns and structures and modifications to the fuel injections system to raise the fuel delivery pressure. As growl is normally most prevalent when the primary –zone fuel air ratio is near the weak extinction limit , any change to the acceleration schedule that raises the fuel/air ratio in the primary zone ,or any change in fuel spray characteristics which lowers the lean blowout limit , such as reduction in spray cone angle , will tend to suppress growl. (pg. 263)

Arthur H. Lefebvre [14] provided the name “HOWL” to describe acoustic oscillations in this paragraph, and here is a quote about “HOWL”:

> The phenomenon of howl or humming is closely related to growl but it occurs at higher engine speeds. Its frequency is usually in the range from 200 to 500 Hz. As with growl, its intensity is dependent on ambient air temperature and falls of rapidly as engine inlet air temperatures rise above normal atmospheric values. It is sensitive to fuel type and diminishes in severity with an increase in fuel volatility.
>
> With growl the engine compressor instabilities play an important role and may even be the trigger for growl, but with the howl the compressor is much less aerodynamically involved. The primary cause of howl appears to be from fuel pressure perturbations. Isolating the feedback mechanism tends to eliminate howl. (pg.263)

3.3 Range of amplitudes for different bins

As per the classification in section 3.2 the acoustic noise have been divided into two broad categories, growl and howl. One of the 83 files was considered and the plot of amplitude and frequency was plotted in R-software. This plot is a scatter plot of amplitude plotted against frequency. This data has six bins numbered from 1 to 6, each bin is associated with a frequency and amplitude range. As per the classification of growl and howl, we can divide the bins into two broad categories. All the six bins ate labelled with colors and each bin has specific amplitude and frequency range. (See Table 7)

Table 7: Amplitude and frequency range

Bin	**Color**	**Amplitude range**	**Frequency range**
Bin-1	Red	Up to 4psi	0-30 Hz
Bin-3	green	Up to 0.5psi	30- 100 Hz
Bin-4	orange	Up to 2.5psi	100-200 Hz
Bin-2	yellow	Up to 0.5psi	200-400 Hz
Bin-6	pink	Up to 0.3psi	400-1000Hz
Bin-5	purple	Up to 0.1psi	1000-1500 Hz

The range of amplitude and frequency has been determined for different bins from the scatter plot. Bin-1, Bin-3 and Bin-4 are categorized under “growl” and Bins-4, 2, 6&5 are categorized into “howl”. (See figure 11)

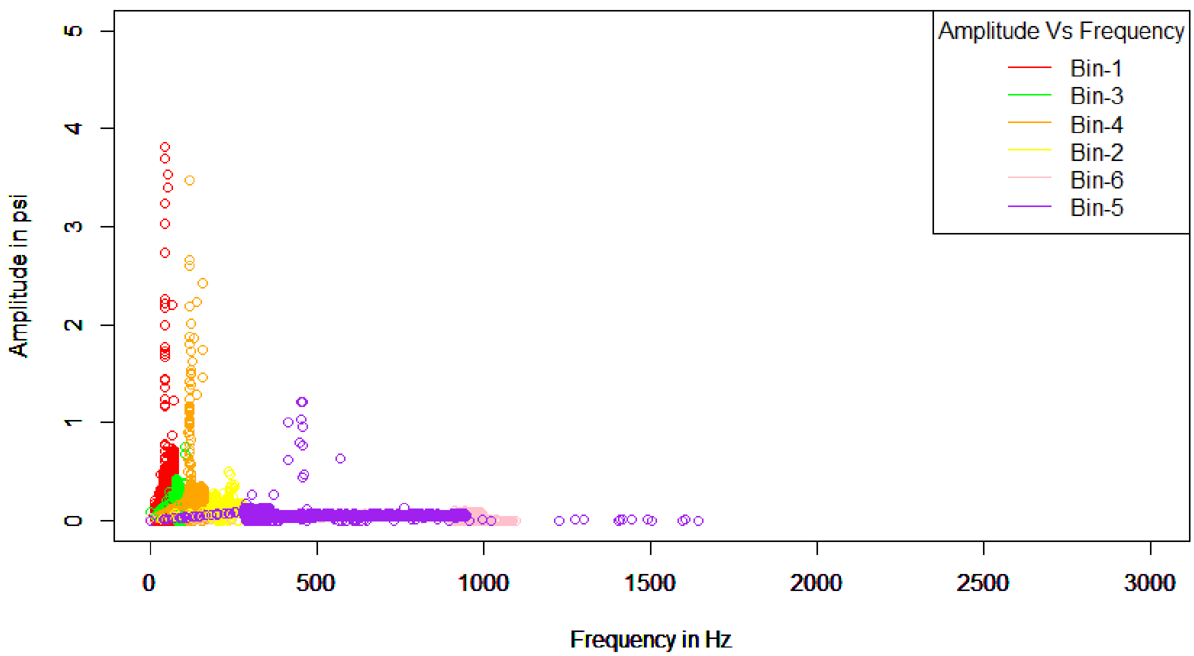

Figure 11: Classification of bins

In this data of gas turbines dynamics has a lot of importance as major part of data has the columns or variables focusing on dynamics. As mentioned in section 1.6 we have 18 cans around the turbine and each can provides us the data of dynamics in six different bins which means we have 18 times 6 columns that is 108 columns or variables.

3.4Counts of Dynamics

R-software was used to give the counts of variables. In R we need to import the file and run the analysis, but this process was not suited for this data as the files are too big to import at once. A code was is written in R to loop the files can perform operations of all files in a single loop by importing one file after another.

The data has six bins. Each Bin is considered at a time and the range amplitude is specified for each of the bin splitting it into 10 different sizes. For example in Bin-1 the Range of the amplitude is from 0psi to 10 psi and is split into intervals 0,1,2,3,4,5,6,7,8,9,10.

The results of the code yield the counts of amplitude in psi from every interval specified. For example: in the first interval 0to1 it counts the points in this interval in every file and Provides the summary.

Observations of “Growl”

Arthur H. Lefebvre [14] determined the range of amplitude for growl is from 0-200Hz, Hence, in this range the there are three bins in this data, Bin-1, Bin-3 and Bin-4 have their amplitude corresponding to the growl. Considering “Growl”.

The code of counting the amplitudes in a loop was run through all the 83 files and the results are yielded over all the 18 cans present in that bin. The observations of the Bin-4 are shown in a table (see Table 8). All the columns in the table indicate the amplitude in different intervals and the rows indicate the units in the data. The table is the result of can -4 of Bin-4.

Table 8: Observations of Bin-4

Range of Amplitude in “psi”	1	2	3	4	5	6	7	8	9	10
unit-1	144	4	1	0	0	0	0	0	0	0
unit-2	0	0	0	0	0	0	0	0	0	0
unit-3	7734	737	5	0	0	0	0	0	0	0
unit-4	183	50	4	1	0	0	0	0	0	0
unit-5	409	85	4	0	0	0	0	0	0	0
unit-6	22	17	8	9	9	9	11	9	9	9
unit-7	438	132	71	0	0	0	0	0	0	0
unit-8	219	33	81	13	0	0	0	0	0	0
unit-9	47	21	0	0	0	0	0	0	0	0
unit-10	10	10	2	0	0	0	0	0	0	0
unit-11	101	33	28	10	0	0	0	0	0	0
unit-12	101	30	36	0	0	0	0	0	0	0
unit-13	0	0	0	0	0	0	0	0	0	0
unit-14	0	0	0	0	0	0	0	0	0	0

In the 3D Histogram of can -4 of Bin-4 over 14 different units (see Figure 12), this histogram shows the amplitudes of each unit. All the units of can-4 are in the low amplitude range, this range is the “growl” range. As we can see all the units are having amplitudes within the range of 0-4 psi, but unit-6 has an amplitude of up to 10psi. This is called as dynamics as unit-6 is at high risk of failure.

Now, considering the only single unit and observing the dynamics over all the 18 cans. Unit-3 is considered and the histogram of uni-3 of all the 18 cans is shown (see Figure 13). In this histogram we can see that most of the cans are having amplitudes in the range 0-4psi, but few cans have amplitudes above 4psi which is a probably showing dynamics in specific cans.

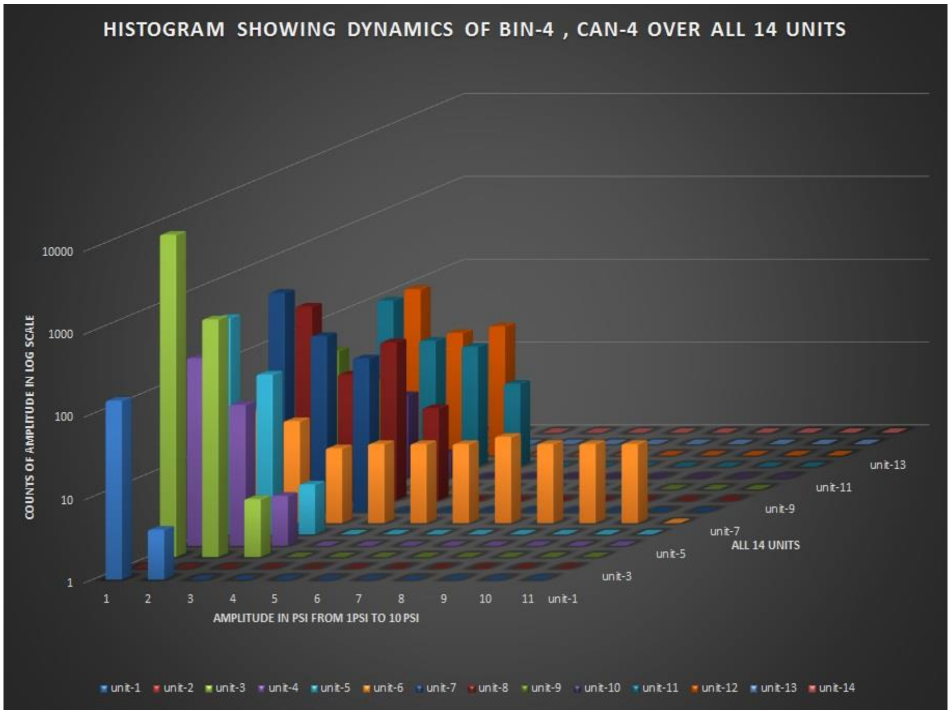

Figure 12: 3D histogram showing dynamics of Bin-4

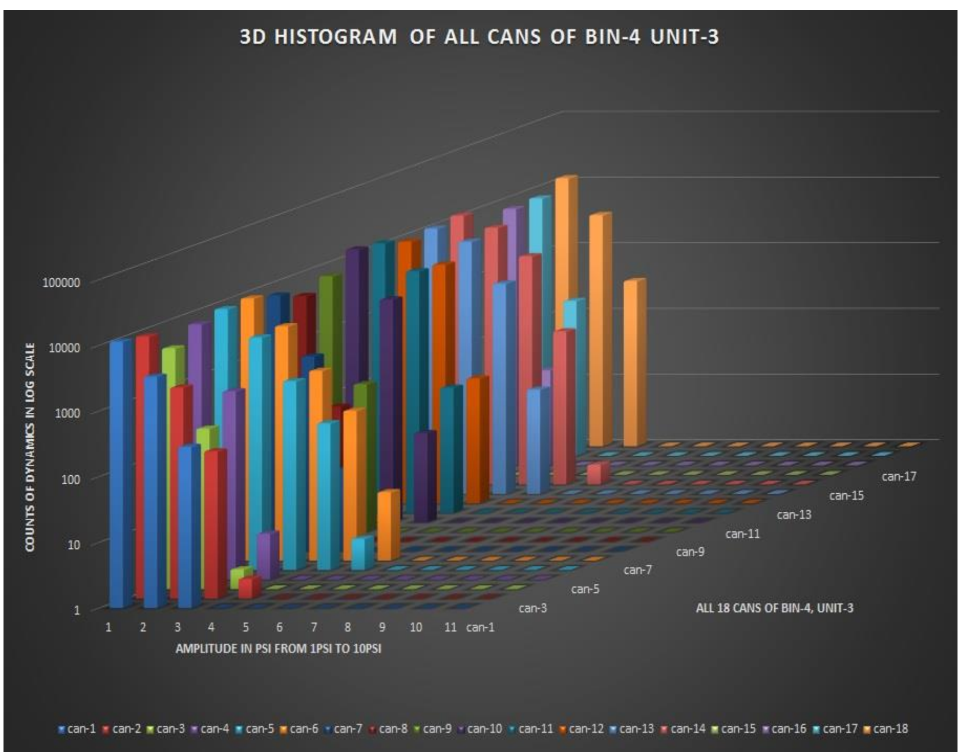

Figure 13: 3D histogram showing dynamics of unit-3

Observations of "Howl"

Arthur H. Lefebvre [14] determined the range of amplitude for Howl is from 200Hz-500Hz, Hence in this range there are three bins in this data, Bin-2, Bin-5 and Bin-6 have their amplitude corresponding to the Howl. The code of counting the amplitudes in a loop was run through all the 83 files and the results are yielded over all the 18 cans present in that bin. The observations of the Bin-2 are shown in a table 9. All the columns in the table indicate the amplitude in different intervals ranging from 0psi to 5psi and the rows indicate the units in the data. The above table is the result of can -4 of Bin2.

Table 9: Observations of can-2

Range of amplitude in "psi"	0	0.5	1	1.5	2	2.5	3	3.5	4	5
unit-1	4	1	0	0	0	0	0	0	0	0
unit-2	0	0	0	0	0	0	0	0	0	0
unit-3	737	5	0	0	0	0	0	0	0	0
unit-4	50	4	1	0	0	0	0	0	0	0
unit-5	85	4	0	0	0	0	0	0	0	0
unit-6	17	8	9	9	9	11	9	9	9	0
unit-7	132	71	0	0	0	0	0	0	0	0
unit-8	33	81	13	0	0	0	0	0	0	0
unit-9	21	0	0	0	0	0	0	0	0	0
unit-10	10	2	0	0	0	0	0	0	0	0
unit-11	33	28	10	0	0	0	0	0	0	0
unit-12	30	36	0	0	0	0	0	0	0	0
unit-13	0	0	0	0	0	0	0	0	0	0
unit-14	0	0	0	0	0	0	0	0	0	0

3D Histogram of can -4 of Bin-2 over 14 different units (see Figure 14). This histogram shows the amplitudes of each unit. As per Howl, Unit-6 is having amplitudes above 1.5 psi and all other units have amplitude within 1.5psi. This shows that unit -6 is prone to dynamics.

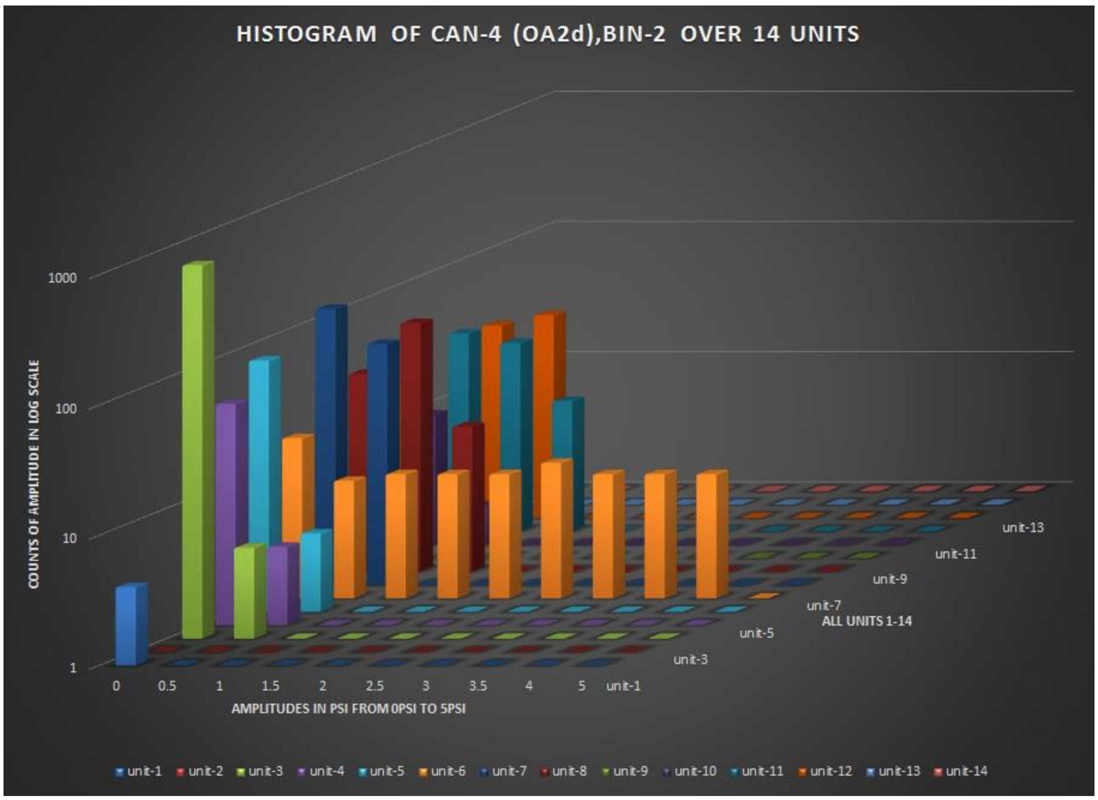

Figure 14: 3D histogram of can-4

This 3D histogram shows (see Figure 15) the dynamics of all cans of unit-6. All cans are showing amplitudes within the range of 0-1.5 PSI except for can-4 and can-14 which are showing dynamics more than 1.5 Psi, hence these cans are prone to high dynamics.

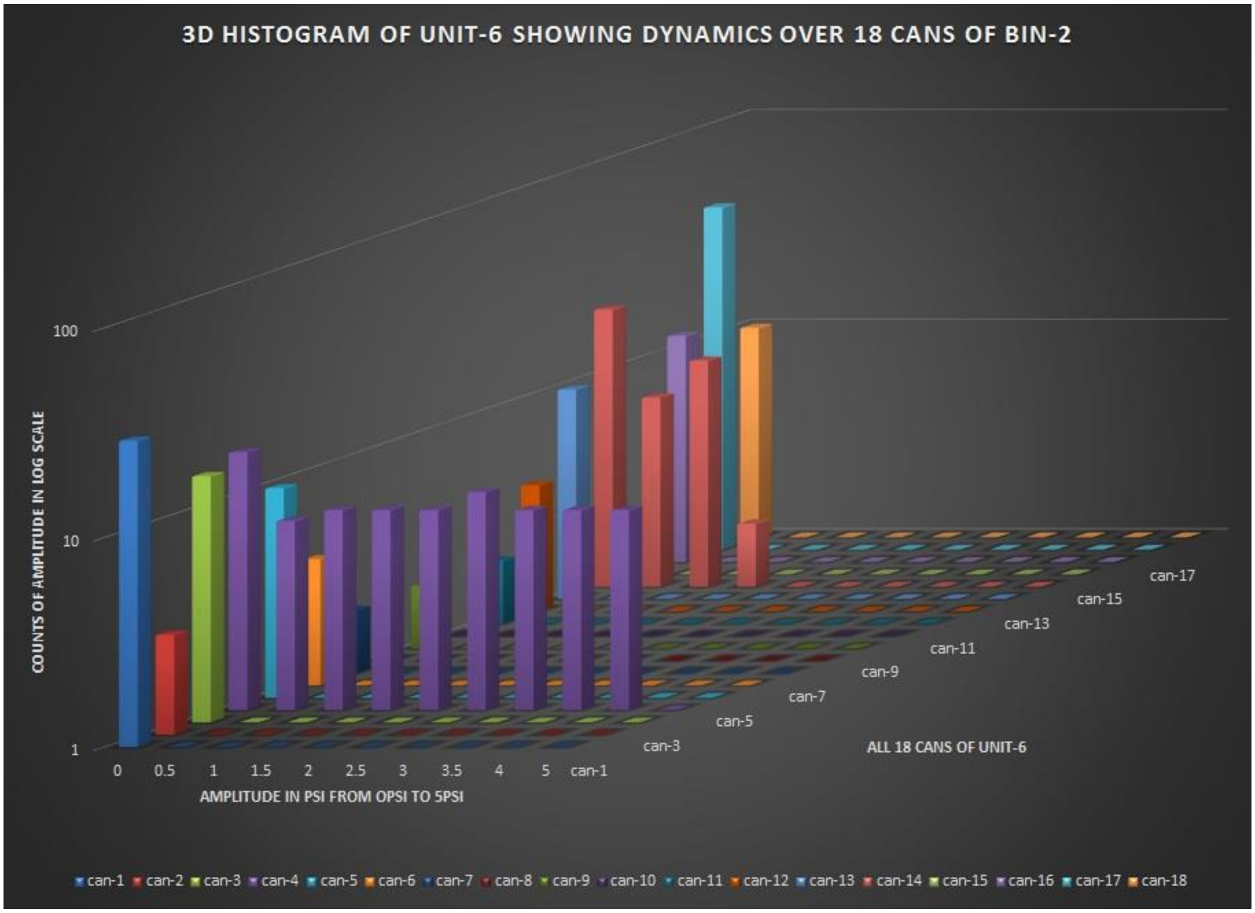

Figure 15:3D histogram of unit-6

3.3 Blade Path Temperature Spread

Monitoring the circumferential temperature spread is important for avoiding serious damage in the cans of the turbine. A thermocouple ring is placed in the exhaust, with thermocouple line of each of the cans. The difference between the maximum and the minimum temperatures within the cans is the actual blade path temperature spread. Usually about 16 to 18 thermocouples are used [6]. All thermocouples record the same temperature. Alert is produced via signal when the actual temperature difference exceeds certain limits. The degree of distortion may signify a problem with the fuel nozzles or with one or more of the combustors [6]. Manufacturers often set operating limits on the spread. Following are the typical set of such limits. Robert E. Dundas [6] explains the conditions:

1. If the spread is greater than 90F (50C) operator is set an alert.
2. If more than 5 thermocouples have failed shut down the engine.
3. If more than three adjacent thermocouples have failed, shut down the engine.

The use of blade path temperature spread helps in monitoring the defects in the gas turbine Such as fuel nozzle malfunction, turbine blade failure. etc. [6]

In the gas turbine data there are three variables pertaining to the temperature spreads.

T7- Combustion monitor actual spread-1

T8- Combustion monitor actual spread-2

T9- Combustion monitor actual spread-3

The data has three temperature spreads. R-software was used to count the ranges of the temperature spreads over all the 14 units. The temperature spread was split into intervals 0,10,20,30,40,50,90,120,200,300,400. The results of one of the spread is shown (see Table 10)

Table 10: Counts of blade path temperature spread

	0 to 10 F	10 to 20 F	20 to 30 F	30 to 40F	40 to 50 F	50to 80F	80 to 120F	120 to 200F	200 to 300F	300 to 400F
unit-1	64572	46760	19523	29790	149692	96374	1614	1247	192	38
unit-2	2859	26630	372088	243969	73733	2839	728	91	1304	310
unit-3	2647	411	8670	96349	105125	363027	376	100	31	4
unit-4	3842	34006	31058	219728	13674	14122	1309	3447	123	10
unit-5	6732	233122	187902	249744	29200	7227	1552	594	466	168
unit-6	10029	3015	45893	35429	348228	403587	299	297	70	3
unit-7	8391	66750	61333	100395	223481	33812	6149	2924	3295	153
unit-8	8354	46166	89008	234585	68998	67603	39248	161	1	1
unit-9	7678	107541	194829	126633	69983	4727	953	793	416	59
unit-10	5823	6462	198113	184429	50326	12188	2096	1320	870	508
unit-11	4425	260089	122836	172346	8946	3877	493	554	361	207
unit-12	4778	165670	240214	133965	81203	22150	13629	6290	3857	1941
unit-13	4099	1838	187902	324229	76426	25154	27	16	11	6
unit-14	4045	11402	189555	377859	86479	1824	101	50	52	7

Figure 16 shows the 3D histogram of temperature spreads and their respective counts in the ranges. T7 is the combustion thermocouple actual spread-1 which is the value of the difference in the temperature of the first and the last thermocouple. As per the author Robert E. Dundas [6] if the spread is greater than 90F, operator is set an alert and if the temperature spread is greater than 108F then unit is shut down. In the above plot Unit-3, Unit-6, unit-7, unit-12 are at risk of setting up alert and shutting down the unit.

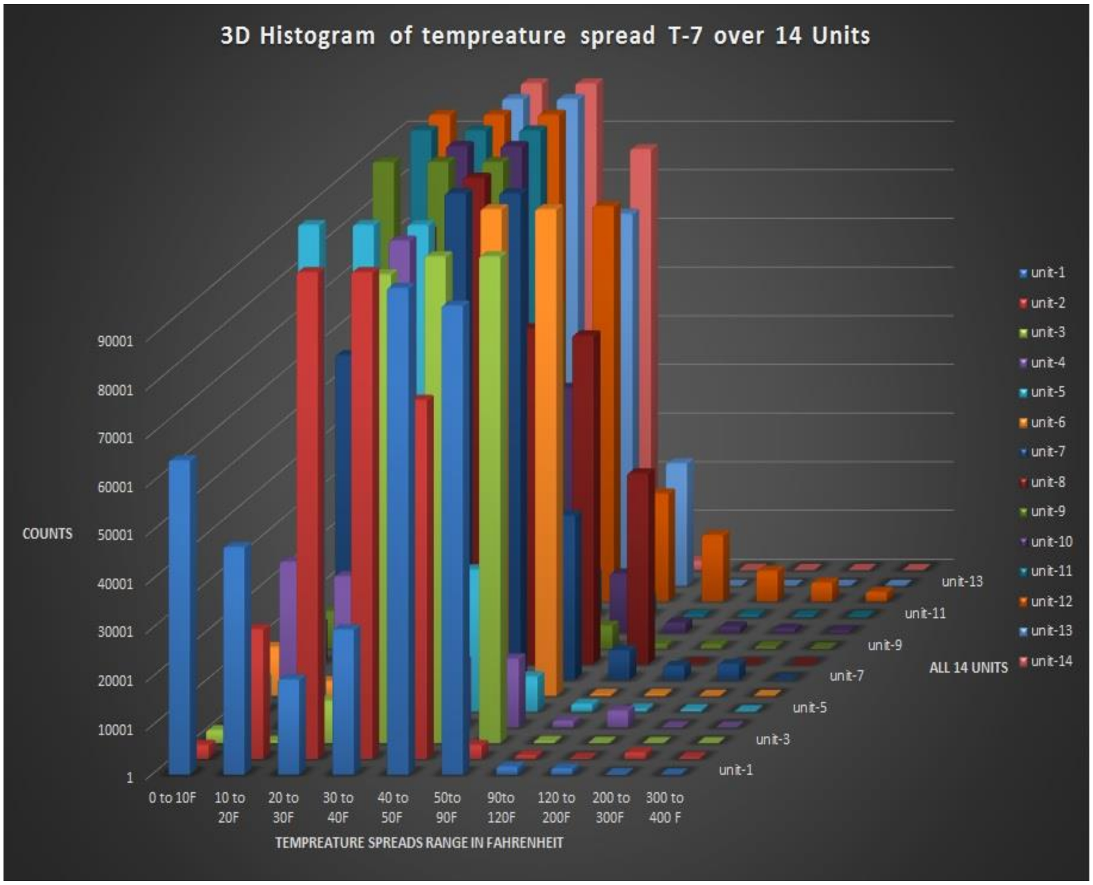

Figure 16: 3D histogram of blade path temperature spread

CHAPTER.4- STATISTICAL ANALYSES

4.1 Correlation Analysis

Correlation is a statistical measure which determines the strength of dependability of two variables [15]. In this analysis the correlation coefficient is estimated which predicts dependability of the operating variables. Coefficient is denoted as “r”. “r” ranges from -1 and 1 and measures the direction and strength and linear association between the two variables The correlation between two variables can be either positive , negative or zero [15].

For example, a correlation of r = 0.95 suggests a strong, positive association between two variables, whereas a correlation of r = -0.5 suggest a weak, negative association. A correlation close to zero suggests no linear association [15].

Scenario of different values of “r” has been illustrated in the plot. (See Figure 17)

- Plot-1 depicts a strong positive association (r=0.95)
- Plot- 2 depicts a weaker association (r=0.2)
- Plot-3 depicts no correlation (r=0)
- Plot- 4 depicts strong negative association (r= -0.9)

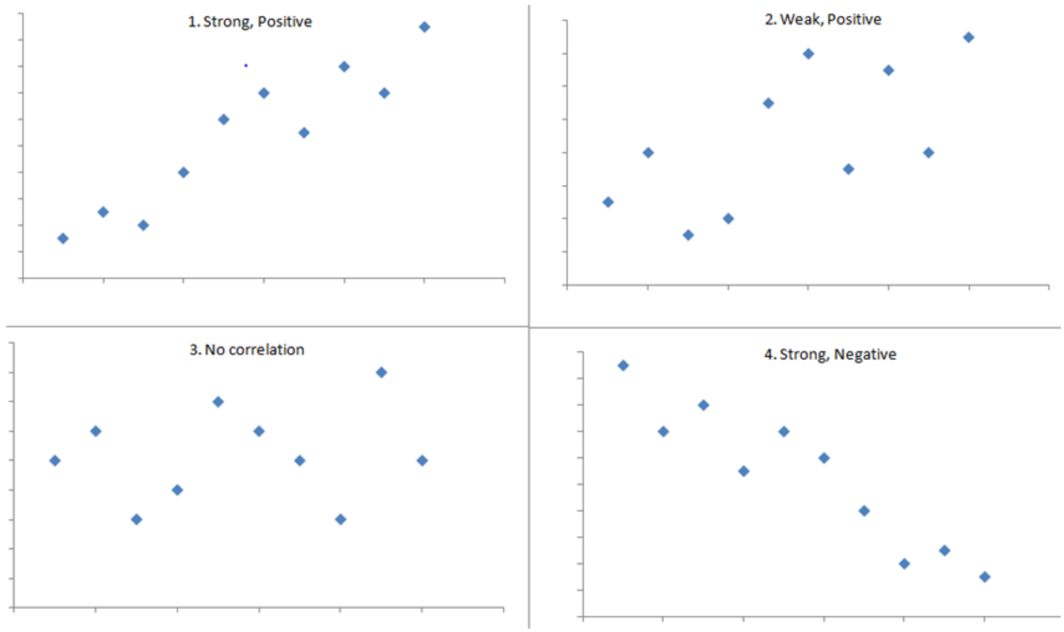

Figure 17: Correlation plot [15]

The correlation analysis was performed on the operational variables. There are about 40 operational parameters in this data. All these parameters were analyzed in R and the dependent and independent variables were resulted. Code is developed in R-software to determine the correlation coefficient and the pair plot.

The 100th percentile values of the operational variables were created a table and imported in R to correlate any relationship between them. Since the load (DM2) is the primary parameter for the gas turbine. Looking at the 100th percentile of the load over the 14 units. There are 9 units with operating load up to 300MW and there are 5 units with load up to 450 MW.

The variables considered for the correlation analysis are.

DM2- Generator Load, FI9- Fuel stroke reference, MF5- Fuel gas flow

T3- Fuel gas temperature, P5- Compressor discharge pressure

Three cases have been considered for representation of the correlation of operating parameters

Case-1: Operational variables considering only one file among all the operational variables.

Case-2: Operational variables considering the 100th percentile of all 14 units.

Case-2: Operational variables considering the 100th percentile of 9 units under 300MW of Load.

Case-1

One of the 83 files is considered and the pair plot (see Figure 18) and the correlation table (see Table 11) have been generated. From the matrix it is observed that there is a relationship between the two parameters.DM2 vs FI9 -----99.8%, DM2 Vs MF5 ---- 99.8%. We cannot really estimate the correlation with one file as the data is big. So the correlations should be associated with all 14 units.

Table 11: Correlation matrix considering one file

	DM2	FI9	MF5	T1	T3	P5	P8
DM2	1	0.998384	0.998742	0.773028	-0.60961	0.996883	-0.48123
FI9	0.998384	1	0.99922	0.784388	-0.59265	0.998288	-0.48011
MF5	0.998742	0.99922	1	0.786028	-0.59194	0.998613	-0.48017
T1	0.773028	0.784388	0.786028	1	-0.17786	0.791549	-0.24114
T3	-0.60961	-0.59265	-0.59194	-0.17786	1	-0.59024	0.428511
P5	0.996883	0.998288	0.998613	0.791549	-0.59024	1	-0.46661
P8	-0.48123	-0.48011	-0.48017	-0.24114	0.428511	-0.46661	1

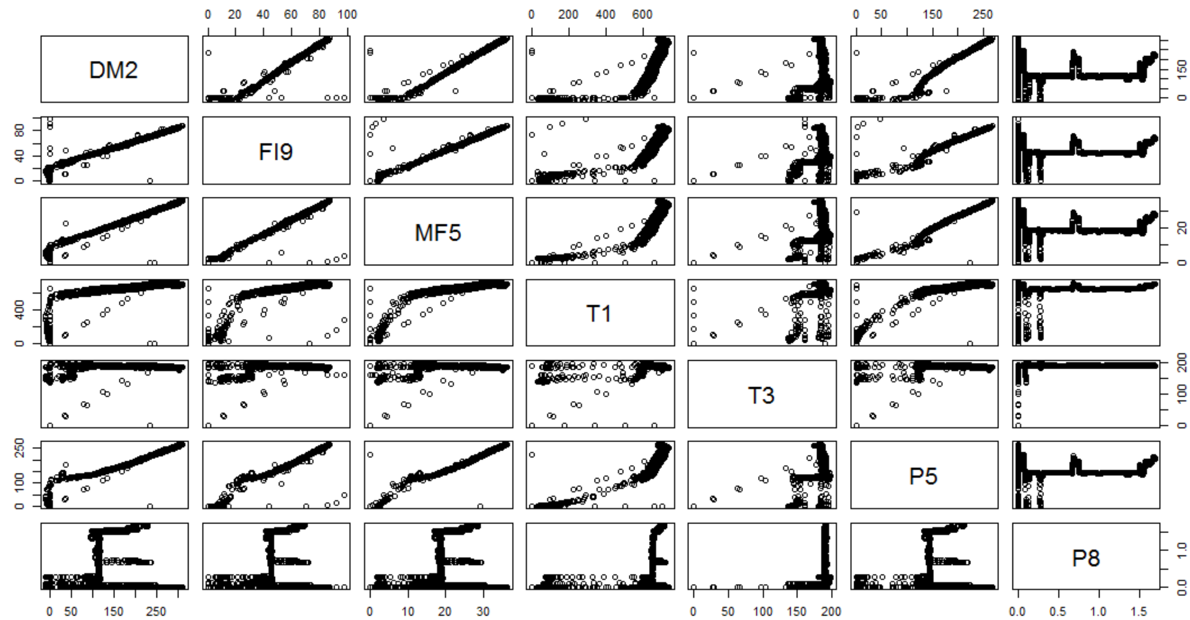

Figure 18: Pair plot of case-1

Case-2

In this case the 100th percentile values of the load DM2 are considered and correlation pair plot is plotted over all the 14 units (see Figure 19). Two sets of clusters created in each of the plot associated with DM2 which is the generator load of the turbine. Some clusters have a maximum value above 350 MW and some clusters have maximum value below 300 MW. So here from this plot we can conclude that there are some units operating close to 300MW and some units are operating 400 MW. In the next plot values for DM2 have been considered separately and pair plots have been generated.

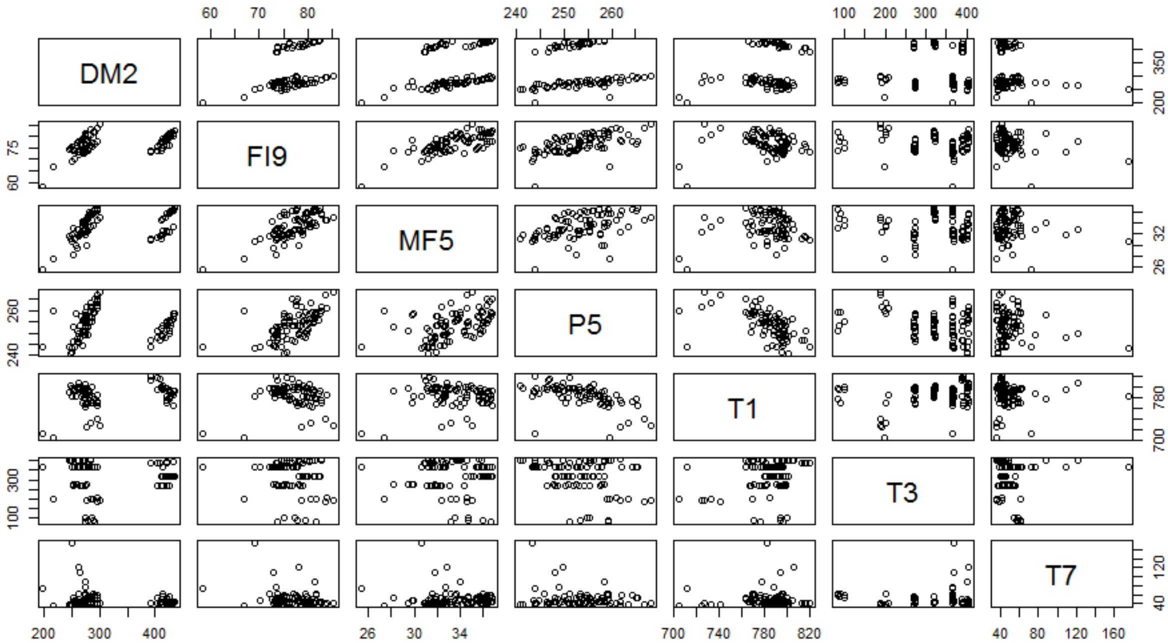

Figure 19: Pair plot of case 2

Case-3

In this case the nine units which are below 300 MW are considered to get consistency in the correlation results. This plot makes clear correlations as this is over all the 9 different units (see Figure 20). From this plot we can conclude the following MF5 is dependent on DM2. FI9 is dependent on DM2, P5 is dependent on DM2 as all these have positive correlation values.DM2 and T7 are independent as they have negative correlation and DM2 and T3 also have negative correlation so they are said to be independent variables. Table 12 shows the correlation table.

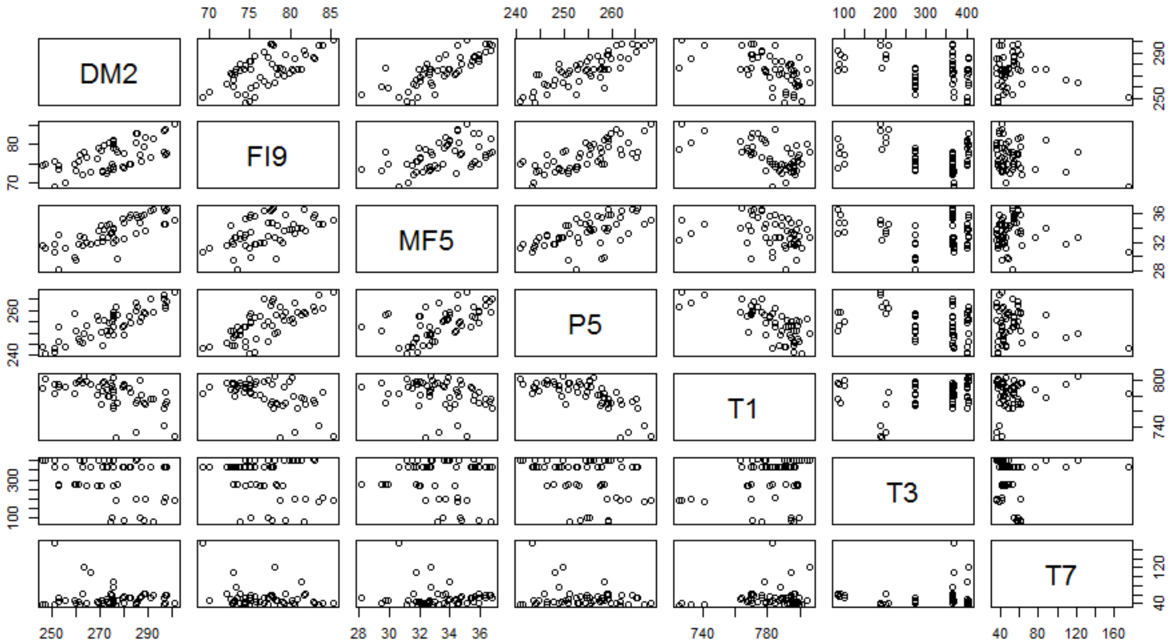

Figure 20: Pair plot of case-3

Table 12: Correlation matrix of case-3

	DM2	FI9	MF5	P5	T1	T3	T7
DM2	1	0.617025	0.785232	0.826978	-0.54926	-0.306	-0.13213
FI9	0.617025	1	0.504778	0.731953	-0.54384	-0.22835	-0.28118
MF5	0.785232	0.504778	1	0.573103	-0.28199	-0.07139	-0.07951
P5	0.826978	0.731953	0.573103	1	-0.69307	-0.30242	-0.1785
T1	-0.54926	-0.54384	-0.28199	-0.69307	1	0.26486	0.155036
T3	-0.306	-0.22835	-0.07139	-0.30242	0.26486	1	0.072064
T7	-0.13213	-0.28118	-0.07951	-0.1785	0.155036	0.072064	1

4.2 Principle Component Analysis

Principal component analysis (PCA) is used to identify the patters to reduce the dimensions of the data [16]. With large number of variables in the dataset the correlation matrix becomes too large to interpret. There would be many correlations to interpret in the pair plot. Graphical display of the plot is not possible as the dataset is very large [16]. This method is used to predict the bin or can causing actual dynamics to the turbine. Principle component analysis for this data is performed on the results of the counts of dynamics and counts of operational variables.

Considering the counts of Dynamics

PCA analysis is performed in R-software. The input file for PCA analysis in R is the table created with counts which consists of all the bins as columns with their intervals and the rows are all the 14 units with 18 cans in each. PCA analysis in R generates 60 principle components. The no of principle components are dependent on the no of columns in the table.

Figure 21 describes the importance of variables always the first component will have the highest cumulative proportion, the first component has highest variances among all the components. The no of components to be selected depends upon the cumulative proportion it covers the data. The next table shows the sample cumulative proportion of the data and the components to be selected.

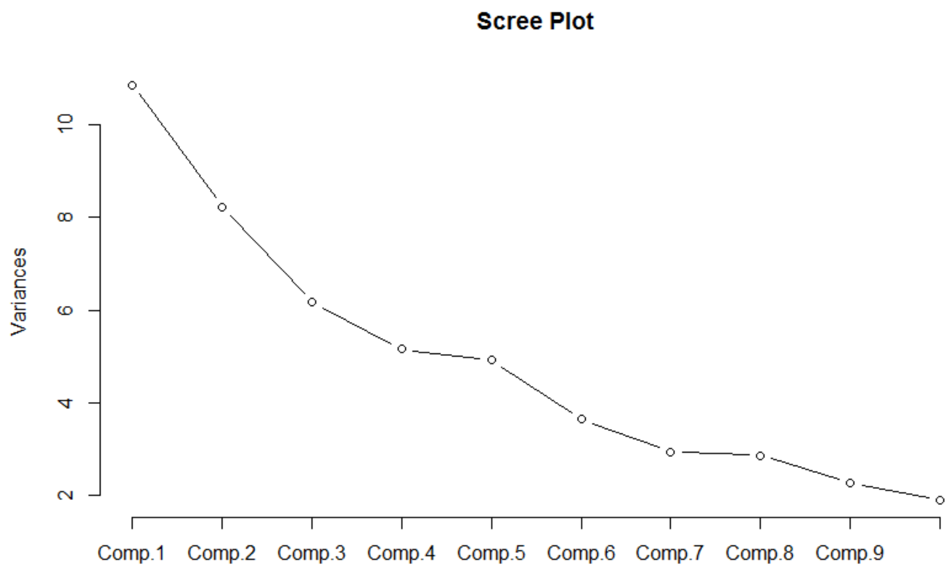

Figure 21: Scree plot

From Table 13 we consider the cumulative proportion of each component. To select the no of components we consider the components which covers the 80% of the data. In this case we consider components 1 to comp 10.

The results of the loadings of each of the components are displayed in the table (see Table13). There are absolute numbers generated for each of the bins. The highest number of in the absolute values in the bins represents high dynamics in the turbine. PCA analysis provides us the information of the Bins with maximum dynamics and high risk of failure.

Table 13: Importance of components

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	3.295136	2.869269	2.483251	2.270861	2.218674	1.908667	1.71592
Proportion of Variance	0.184033	0.139537	0.104518	0.087404	0.083432	0.061746	0.049905
Cumulative Proportion	0.184033	0.32357	0.428087	0.515491	0.598923	0.660669	0.710574

Components to be considered are components comp1 to comp10 which covers 80% of the data Component 1 is the highest with 18.4% proportion of variance and component 2 is the next highest with 13.9% of the variance .There are total 60 components. After 18 components the proportion of the variance reduces and is less than 1%. (See Table 14) for the results.

Principle component -1

Covers 18% of the data, it is showing equal importance for dynamics in Bin-2 and Bin-4. For Bin-4 it’s giving higher importance for dynamics greater than 5psi, For Bin-2 it is showing highest importance for values greater than 1.5 psi.

Principle component-2

PC-2 is showing high values (>0.2) for Bin-1 to Bin-5, In PC-1 we already know that Bin-2 and Bin-4 are important , as by rule of PCA it will ignore the same exact values in PC-2. But still in PC-2 the maximum value of high dynamics impact are of Bin-4 and Bin-2.

Principle component-3

PC-3, Bin-6 and Bin-5 have high absolute values showing these bins have high dynamics but have only 10% of the cumulative proportion.

Table 14: PCA analysis

	comp-1	comp-2	comp-3	comp-4	comp-5	comp-6	comp-7	comp-8
B1.0			-0.105	-0.126	0.137			
B1.0.5				-0.107			-0.146	-0.18
B1.1		0.12		-0.129		-0.139	-0.14	-0.121
B1.1.5		0.125		-0.139	-0.121		-0.122	-0.113
B1.2		0.193		-0.16	-0.176			-0.108
B1.2.5		0.213	-0.101	-0.181	-0.268			
B1.3		0.166	-0.115	-0.176	-0.304			
B1.3.5		0.153	-0.11	-0.169	-0.303			
B1.4		0.144	-0.103	-0.171	-0.305			
B1.5		0.131		-0.171	-0.307			
B3.0			-0.134	-0.155	0.113			
B3.0.5		0.21		0.157				
B3.1		0.17	0.113			0.348	0.137	-0.115
B3.1.5						0.384	0.215	-0.237
B3.2		0.147	0.113			0.363	0.196	-0.191
B3.2.5		0.189	0.158			0.148	-0.286	
B3.3			0.102	0.123		0.198	-0.44	0.125
B3.3.5				0.123		0.201	-0.441	0.125
B3.4		0.114	0.13	0.112		0.278	-0.357	
B3.5						0.38	0.209	-0.243
B4.1		0.253	0.134					
B4.2		0.243	0.178		0.119	-0.114		
B4.3		0.225	0.178		0.132	-0.147		
B4.4		0.224	0.159		0.132	-0.18		
B4.5	-0.212	0.109				-0.119		
B4.6	-0.299							
B4.7	-0.299							
B4.8	-0.299							
B4.9	-0.299							
B4.10	-0.299							
B2.0		0.243	0.178		0.119	-0.114		
B2.0.5		0.225	0.178		0.132	-0.147		
B2.1		0.224	0.159		0.132	-0.18		

	Comp-1	Comp-2	Comp-3	Comp-4	Comp-5	Comp-6	Comp-7	Comp-8
B2.1.5	-0.212	0.109				-0.119		
B2.2	-0.299							
B2.2.5	-0.299							
B2.3	-0.299							
B2.3.5	-0.299							
B2.4	-0.299							
B6.0.1		0.243	-0.139	0.151			0.11	0.193
B6.0.2				0.21	-0.117			
B6.0.3			0.119	0.242	-0.193			
B6.0.4			0.103	0.218	-0.174			
B6.0.5			0.123	0.244	-0.206			
B6.0.6			0.133	0.255	-0.23			
B6.0.7			0.121	0.238	-0.226			
B6.0.8				0.201	-0.211		0.139	
B6.1					-0.19		0.13	
B6.2		0.149	-0.135					0.194
B5.0.1		0.214		0.178				0.119
B5.0.2		0.107	-0.242	0.166				0.21
B5.0.3		0.108	-0.254	0.154				0.236
B5.0.4		0.12	-0.283	0.185				0.157
B5.0.5		0.117	-0.282	0.196				
B5.0.6		0.105	-0.268	0.189				-0.212
B5.0.7			-0.232	0.16			-0.126	-0.342
B5.0.8			-0.195	0.127			-0.147	-0.393
B5.1			-0.175	0.101			-0.144	-0.375
B5.2								

CHAPTER.5- CONCLUSION

Large Data Set of the gas turbine was mined to analyze the behavior of the gas turbines at various operating conditions. High Dynamics was observed in Bin-4 and Bin-2 by the counts of dynamics. The same bins were also observed for high dynamics by Principal component analysis. Blade path temperature spread was observed in unit7 and unit8, the difference in the actual temperature spreads was above 90F which is noted as an alert given to the operator of the turbine. Data Quality issues were observed such missing data and Arrhenius peaks. Hence, big data requires cleaning of the file before the analysis.

In future, further study on dynamics can be well understood if the failure data is also obtained. Techniques such as regression, Bootstrapping and clustering can also be used to predict the dynamics. Should work on Logic tags to predict the behavior in a better way.

REFERENCES

[1] Elena Geanina ULARU, F. C. P., Anca APOSTU, Manole VELICANU, 2012, "Perspectives on Big Data and Big Data Analytics " Database Systems Journals, III.

[2] Laney, D., 2001, "Big data defined," http://www.sas.com/en_us/insights/big-data/what-is-big- data.html

[3] W. N. Venables, D. M. S., 2015, " An Introduction to R," R- environment p. Pg2.

[4] "What is gas turbine and how does it work?," https://powergen.gepower.com/plan-build/tools-resources/power-generation-basics/how-gas-turbines-work.html

[5] Stow, A. P. D. S. R., 2003, "Acoustic Analysis of Gas Turbine combustors," Journal of propulsion and power, pp. 751-764.

[6] ABEGG, R. E. D. D. A. S. F., 1992, "Performace monitoring of gas turbines for failure prevention," ASME, pg 6

[7] Barna Saha, D. S., "Data Quality : The other Face of Big Data," AT&T Labs-Research.Pg1

[8] Hewitt, D. E. H. J. E. F. G., 2000, "Combustion Instabilities in Industrial Gas Turbines-Measurements.

[9]Shah, M. M. K. F. H. D. W. C. B. M., 2004, "Modelling and control of combustion Dynamics," Proceedings of the ASME TURBO EXPO 2004, 1, pp. 567-575.

[10] Venturini, M., and Therkorn, D., 2013, "Application of a Statistical Methodology for Gas Turbine Degradation Prognostics to Alstom Field Data," Journal of Engineering for Gas Turbines and Power, 135(9), pp. 091603-091603.

[11] "Combustion Dynamics Instrumentation," IMI Sensors, A. P. P. DIV, ed.

[12]] C.Lieuwen, B. T. Z. a. T., 2005, "Combustion instabilities: Basic concepts," American institute of aeronautics and astronautics, p. 3.

[13] Goy, C. J., James, S. R., and Rea, S., 2005, "Monitoring combustion instabilities: E.ON UK's experience," Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling, T. C. Lieuwen, and V. Yang, eds., American Institute of Aeronautics and Astronautics, pp. 163-175

[14] Lefebvre, A. H., 1998, "Combustion noise," Gas Turbine combustion,2,Pg263

[15] 2013, "Introduction to Correlation and Regression Analysis," http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Multivariable/BS704_Multivariable5.html.

[16] Jolliffe, I. 2014. Principal Component Analysis. Wiley StatsRef: Statistics Reference Online. .