

**Centrifugal Compressor Performance Degradation
Monitoring and Prediction Using Hybrid Model:
Thermodynamic-Based and Artificial Intelligence**

BY

Murtadha AlabdrabAlredha

A Thesis Presented to the

COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

MECHANICAL ENGINEERING

December, 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis is written by Murtadha Alabdrab Alredha under the direction of his thesis advisor and approved by his thesis committee. It has been presented and accepted by the Dean of Graduate Studies, in partial fulfilment of requirements for the degree of **Master of Science in Mechanical Engineering**



Dr. Mohamed A. Antar

Department of Chairman



Dr. Suliman Al-Homidan

Dean of Graduate Studies

Date



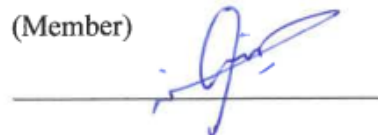
Dr. Fadi A Al-Badour

(Advisor)



Dr. Abdelsalam Mohammad Alsarkhi

(Member)



Dr. Haitham Hassan Saleh

(Member)



Dr. Ammar Ayad Alzaydi

(Member)



Dr. Rami Khalid Mohammad Suleiman

(Member)



© Murtadha AlabdrabAlredha

2024

Dedication

I dedicate this thesis work to my wife, Maryam, who has been a pillar of support and encouragement throughout graduate school and life. I sincerely appreciate you being in my life. Also, I dedicate this thesis work to the global Equipment Condition Monitoring team at Air Products, who maintain a positive and inspiring atmosphere and share their knowledge and experience, it is a privilege to work with you.

ACKNOWLEDGMENTS

I want to take a moment to express my sincere gratitude to my advisor, Dr. Fadi A Al-Badour, for his relentless support, guidance, and expertise throughout my research journey. His thoughtful feedback and encouragement have played a crucial role in shaping this work.

I also deeply appreciate the members of my research committee, Dr. Abdelsalam Mohammad Alsarkhi, Dr. Haitham Hassan Saleh, Dr. Ammar Ayad Alzaydi and Dr. Rami Khalid Mohammad Suleiman for their constructive criticism and valuable suggestions. Their diverse viewpoints have significantly enhanced the quality of this thesis.

I am thankful to King Fahd University of Petroleum and Minerals for their financial support, which made this research possible. Their belief in the significance of my work has been a strong source of motivation.

A heartfelt thank you goes to my colleagues and peers, who created a stimulating academic environment and generously shared their knowledge. The collaborative spirit within the Mechanical Engineering Department has been vital in enriching the depth and scope of my research.

I am also grateful to my family for their unwavering support and understanding during this challenging journey. Their love and encouragement have laid the foundation for my academic endeavors.

Finally, to my friends who have stood by me, providing support, laughter, and a much-needed distraction at times your friendship has been a treasured part of my academic experience. This work reflects the collective efforts of everyone who believed in and contributed to its realization.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	V
TABLE OF CONTENTS.....	VI
LIST OF TABLES.....	X
LIST OF FIGURES.....	XI
LIST OF ABBREVIATIONS AND NOMENCLATURE.....	XIV
ABSTRACT.....	XX
ملخص الرسالة.....	XXI
CHAPTER 1 INTRODUCTION	1
1.1 Background.....	1
1.2 Aim	2
1.3 Outline of the Thesis	4
CHAPTER 2 LITERATURE REVIEW	6
2.1 Monitoring of Centrifugal Compressors	7
2.1.1. Monitoring Methods for Mechanical Deterioration	8
2.1.2. Monitoring Methods for Performance Deterioration.....	10

2.2 Mathematical-Base Modelling.....	13
2.3 Combined mathematical and Artificial intelligence modelling.....	14
2.4 Statistical-base Modelling	15
2.5 Modelling Methods Comparison	16
2.6 Summary	19
 CHAPTER 3 BASIC PRINCIPLES USED IN A TURBOMACHINERY.....	21
3.1 Continuity Equation.....	21
3.2 Momentum Equation	22
3.3 First Law of Thermodynamics	24
3.4 Second Law of Thermodynamics	26
3.5 Gibbs Equation Entropy and Enthalpy Relation	27
3.6 Stagnation Temperature and Pressure	29
 CHAPTER 4 DATA EXPLORATIONS	32
4.1 Data Generation Goal.....	32
4.2 Data source, Type and Range	33
4.3 Simulated Case-Study.....	35
 CHAPTER 5 THEORETICAL PERFORMANCE ANALYSIS AND MODELLING.....	38

5.1 Velocity Triangle	38
5.2 Rothalpy preservation in Compressors.....	40
5.3 Zone Flow Analysis.....	41
5.4 Effect of Circulatory Flow (Slip)	43
5.5 Compressors Isentropic and Polytropic efficiencies.....	45
5.6 Performance Models of Centrifugal Compressors	48
5.7 Result and Discussion of Theoretical Models	53
 CHAPTER 6 ARTIFICIAL INTELLIGENCE; MACHINE LEARNING PERFORMANCE	
MODELLING..	58
 6.1 Modeling methodology.....	58
6.1.1. Data cleaning	59
6.1.2. Data categorizing and selection	60
6.1.3. Model Training and Validation.....	61
6.1.4. Model testing.....	62
 6.2 Machine learning used algorithms	62
6.2.1. Linear Regression Model.....	63
6.2.2. Decision Tree Model	64
6.2.3. Random Forest Model	65
6.2.4. Neural Network Model	66
 6.3 Result and Discussion of Artificial Intelligence models	67

CHAPTER 7 HYBRID THEORETICAL AND MACHINE LEARNING MODELLING	80
7.1 Hybrid Modelling Methodology.....	82
7.2 Result and discussion of hybrid models.....	83
7.3 The Proposed Hybrid Modelling Compared To Latest Advance AI Modelling.....	95
CHAPTER 8 DEGRADATION CONDITION MONITORING	98
8.1 Degradation Detection workflow	98
8.1.1. Outlier removing.....	99
8.1.2. Performance parameters calculation	100
8.2 Degradation detection Result and discussion	100
CHAPTER 9 CONCLUSION AND PERSPECTIVES	104
9.1 Conclusion	104
9.2 Perspectives	106
REFERENCES	108
APPENDIX	117
VITAE	129

LIST OF TABLES

Table 2.1: Modelling Methods In Literature With Maximum RMSE Score	18
Table 2.2: Modelling Methods In Literature With Maximum ME Score	19
Table 4.1: Data Input, Range And Unit.....	34
Table 4.2: Data Output And Unit.....	35
Table 4.3: Case-Study Data Input, Range And Unit	37
Table 5.1: Zone 1 (0-1) Inlet Nozzle (Duct) tag mapping	50
Table 5.2: Zone 2: (1-2) Impeller (Rotor) Tag Mapping.....	51
Table 5.3: Zone 3: (2-3) Diffuser (Stator) Tag Mapping.....	52
Table 5.4: theoretical model scores	54
Table 6.1: Linear Regression model scores	68
Table 6.2: Decision Tree model scores	71
Table 6.3: Random Forest model scores	73
Table 6.4: Neural Network model scores.....	76
Table 7.1: P3 Theoretical Model And Random Forest Scores.....	81
Table 7.2: T3 Theoretical Model And Random Forest Scores	81
Table 7.3: PR Theoretical Model And Random Forest Scores.....	81
Table 7.4: Hybrid Linear Regression Model Scores.....	83
Table 7.5: Hybrid Decision Tree Model Scores.....	86
Table 7.6: Hybrid Random Forest Model Scores	88
Table 7.7: Hybrid Neural Network Model Scores	91
Table 7.8: Most Recent Advance Performance Modelling Techniques with MAE score.....	96

LIST OF FIGURES

Figure 2.1 [5]: Integrally Geared Centrifugal Compressor Arrangement.....	8
Figure 3.1: Continuity Equation.....	21
Figure 3.2: [40] Rotor Rotation And Tangential Speed For Centrifugal Compressor	22
Figure 4.1 [43]: Efficiency Trends With Maintenance Shutdown 18 Months.....	36
Figure 5.1 [39]: Inlet And Outlet Impeller Cross Section With Coordinates.....	39
Figure 5.2: Inlet And Outlet Velocity Triangle.....	39
Figure 5.3: Compressors Modelling Zones.....	42
Figure 5.4: Impeller Outlet Velocity Triangle With Slip	44
Figure 5.5 [39]: Compressors Isentropic And Polytropic Efficiencies.....	46
Figure 5.6: Centrifugal Compressors Modelling: Output And Input.....	49
Figure 5.7: Inlet Velocity Triangle	51
Figure 5.8: Outlet Velocity Triangle	52
Figure 5.9: Actual Outlet Temperature T_3 Versus Calculated.....	55
Figure 5.10: Actual Outlet Pressure P_3 Versus Calculated	56
Figure 5.11: Actual Outlet Pressure P_R Versus Calculated.....	57
Figure 6.1: MATLAB Cross-Validation Function With 5 Folds And 5 Group	62
Figure 6.2 [37]: Linear Regression Model	63
Figure 6.3 [47]: Decision Tree: Decision Nodes And Leaf Nodes	64
Figure 6.4 [38]: Random Forest Model Illustration	65
Figure 6.5 [38]: Neural Network Model Layers Connection	66
Figure 6.6: T_3 Linear Model Predicted Versus Actual And Residual.....	68

Figure 6.7: P3 Linear Model Predicted Versus Actual And Residual	69
Figure 6.8: PR Linear Model Predicted Versus Actual And Residual	70
Figure 6.9: T3 Decision Tree Model Predicted Versus Actual And Residual.....	71
Figure 6.10: P3 Decision Tree Model Predicted Versus Actual And Residual	72
Figure 6.11: PR Decision Tree Model Predicted Versus Actual And Residual	73
Figure 6.12: T3 Random Forest Model Predicted Versus Actual And Residual.....	74
Figure 6.13: P3 Random Forest Model Predicted Versus Actual And Residual	75
Figure 6.14: PR Random Forest Model Predicted Versus Actual And Residual	76
Figure 6.15: T3 Neural Network Model Predicted Versus Actual And Residual.....	77
Figure 6.16: P3 Neural Network Model Predicted Versus Actual And Residual.....	78
Figure 6.17: PR Neural Network Model Predicted Versus Actual And Residual	79
Figure 7.1: T3 Hybrid Linear Model Predicted Versus Actual And Residual	84
Figure 7.2 [38]: P3 Hybrid Linear Model Predicted Versus Actual And Residual..	85
Figure 7.3 [38]: PR Hybrid Linear Model Predicted Versus Actual And Residual .	86
Figure 7.4: T3 Hybrid Decision Tree Model Predicted Versus Actual And Residual	87
Figure 7.5: P3 Hybrid Decision Tree Model Predicted Versus Actual And Residual	87
Figure 7.6: PR Hybrid Decision Tree Model Predicted Versus Actual And Residual	88
Figure 7.7: T3 Hybrid Random Forest Model Predicted Versus Actual And Residual	89

Figure 7.8 P3 Hybrid Random Forest Model Predicted Versus Actual And Residual	90
Figure 7.9: PR Hybrid Random Forest Model Predicted Versus Actual And Residual	91
Figure 7.10: T3 Hybrid Neural Network Model Predicted Versus Actual And Residual.....	92
Figure 7.11 P3 Hybrid Neural Network Model Predicted Versus Actual And Residual	93
Figure 7.12: PR Hybrid Neural Network Model Predicted Versus Actual And Residual.....	94
Figure 8.1: Moving-Window For Outlier Remover	100
Figure 8.2: Outlet Temperature T3 Undegraded Predicted And Degraded Actual	102
Figure 8.3: Outlet Pressure P3 Undegraded Predicted And Degraded Actual.....	102
Figure 8.4: Pressure Ratio PR Undegraded Predicted And Degraded Actual	102
Figure 8.5: Efficiency η Undegraded Predicted And Degraded Actual	103
Figure 8.6: Efficiency η Degradation Deviation Percentage Indicator.....	103
Figure 8.7: Pressure Ratio PR Degradation Deviation Percentage Indicator.....	103

LIST OF ABBREVIATIONS AND NOMENCLATURE

ASU	Air Separation Unit
MAE	Mean Absolute Error
SSE	Sum Of Squared Errors
SST	Sum Of Squared Total
ME	Mean Error
RMSE	Root Mean Square Error
KPLS	Kernel Partial Least Squares
GPA	Gas Path Analysis
AI	Artifactual Intelligence
EKF	Extended Kalman Filtering
PF	Particle Filter
LSTM	Long Short-Term Memory
ANNs	Artificial Neural Networks
Res-BPNN	Residual-Back Propagation Neural Network
MK-MMD	Multi-Kernel Maximum Mean Discrepancy
ML	Machine Learning
TL	Transfer Learning

TSABC	Artificial bee colony algorithm with tabu search
EMD	Empirical Mode Decomposition
TDA	Topological Data Analysis
UD	Un-Degraded
c	Speed Of Sound
M	Mach Number
R^2	Degree Of Fitness. The Coefficient Of Determination
n_s	Sampling Rate
y_i	Actual Measure
y'_i	Predicted Measure
$\dot{m}$	Mass Flow Rate
ρ	Fluid Density
A	Cross Section Area
V	Fluid Related Velocity
τ	Torque
$V_{\theta j}$	Related Tangential Velocity At Location j
r_j	Rotation Radius At Location j
ω	Rotor Angular Speed
P	Power

U	Blade Linear Speed
u	Internal Energy Per Unit Of Mass
g	Acceleration Of Gravity
z	Elevation Distance
e	Specific Energy
h	Enthalpy
P	Absolute Pressure
v	Volume
$\dot{Q}$	Rate Of Heat Transfer
$\dot{W}$	Rate Of Work
q	Heat Transfer Per Unit Mass
w	Work Done Per Unit Mass
h_0	Stagnation Enthalpy
T_0	Stagnation Temperature
P_0	Stagnation Pressure
T	Absolute Temperature
S	Entropy
η	Efficiency
η_p	Polytropic Efficiency
Z_c	Compressibility Factor

Y	Axial Coordinate Axis
r	Radial Coordinate Axis
θ	Cylindrical Coordinate Axis
β	Blade Flow Angle
N	Rotational Speed in RPM rotation per minutes
r0:	inlet pipe or nozzle
r1	r1h-r1s: Inducer inlet radius
r1h	impeller hub radius
r1s	impeller shroud radius
beta1	mean-line impeller inlet blade angle,
r2	impeller outlet radius
beta2	mean-line impeller outlet blade angle,
b2	impeller outlet blade height
r3	diffuser outlet radius
b3	diffuser channel width
Zb	number of blade
PR	pressure ratio

eff	compressor isentropic efficiently
T0	inlet temperature
P0	inlet pressure
m	mass flow rate
P3	outlet Pressure
T3	outlet temperature
P3 Linear	outlet Pressure (kPa) Linear Tree regression prediction
T3 Linear	outlet temperature (K) Linear Tree regression prediction
PR Linear	Pressure ratio linear Linear Tree prediction
P3 Tree	outlet Pressure (kPa) Decision Tree regression prediction
T3 Tree	outlet temperature (K) Decision Tree regression prediction
PR Tree	Pressure ratio Decision regression prediction
P3 NN	outlet Pressure (kPa) Neural Network regression prediction
T3 NN	outlet temperature (K) Neural Network regression prediction
PR NN	Pressure ratio Neural Network regression prediction

P3 LinearH	outlet Pressure (kPa) Hybrid Linear Tree regression prediction
T3 LinearH	outlet temperature (K) Hybrid Linear Tree regression prediction
PR LinearH	Pressure ratio linear Hybrid Linear Tree prediction
P3 TreeH	outlet Pressure (kPa) Hybrid Decision Tree regression prediction
T3 TreeH	outlet temperature (K) Hybrid Decision Tree regression prediction
PR TreeH	Pressure ratio Hybrid Decision regression prediction
P3 NNH	outlet Pressure (kPa) Hybrid Neural Network regression prediction
T3 NNH	outlet temperature (K) Hybrid Neural Network regression prediction
PR NNH	Pressure ratio Neural Hybrid Network regression prediction

ABSTRACT

Full Name : Murtadha AlabdrabAlredha

Thesis Title : CENTRIFUGAL COMPRESSOR PERFORMANCE
DEGRADATION MONITORING AND PREDICTION USING
HYBRID MODEL: THERMODYNAMIC-BASED AND
ARTIFICIAL INTELLIGENCE

Major Field : Mechanical Engineering

Date of Degree : December 2024

This thesis introduces an online degradation rate estimation approach utilizing a condition monitoring technique that integrates Artificial Intelligence models and Thermodynamic models. It offers a solution for operation to assess the condition and stop only when it is needed. The thesis reviews two main approaches; the thermodynamic ideal design approach and supervised machine learning approach. A review of theoretical turbomachinery concepts for centrifugal compressors was conducted to formulate ideal equations. The performance output of these ideal equations differs from real on operation machines due to environmental condition and other losses. Also, the thesis covers performance estimation using supervised machine learning algorithms, considering on real life historical data training. Four main algorithms were implemented: linear regression, Decision Tree, Random Forest, and Neural Network. Although machine learning algorithms consider real data to include all losses and environmental condition, they need huge data which leads to high computational cost and time to incorporate all operational modes and variable, which in most cases are not available in open literature. To overcome these two problems; a hybrid approach is proposed which combines both approaches of design turbomachinery models and supervised machine learning. Four primary comparison indicators were used to prove the effectiveness of proposed method and to evaluate error scores : Mean Absolute Error (MAE), Sum of Squared Errors (SSE), Sum of Squared Total (SST). The most effective model based on performance was chosen and applied to a simulated degradation scenario to assess its efficiency. The thesis concludes based on findings that even with limited information about compressor design and limited historical data, proposed hybrid models can be built with acceptable errors to estimate degradation.

ملخص الرسالة

الاسم الكامل: مرتضى حسن العبدرب الرضا

عنوان الرسالة: مراقبة وتنبؤ تدهور أداء ضاغط الطرد المركزي باستخدام نموذج هجين: قائم على الديناميكا الحرارية والذكاء الاصطناعي

التخصص: الهندسة الميكانيكية

تاريخ الدرجة العلمية: ديسمبر 2024

تقدم هذه الرسالة منهجاً لتقدير معدل التدهور أثناء التشغيل باستخدام تقنية مراقبة الحالة التي تدمج بين نماذج الذكاء الاصطناعي ونماذج الديناميكية الحرارية. وتوفر حلاً للمشغلين لتقييم حالة المعدات وإيقاف التشغيل فقط عند الحاجة. تستعرض الرسالة منهجين رئيسيين: منهج التصميم المثالي الديناميكي الحراري، ومنهج التعلم الآلي المُراقب. تم إجراء مراجعة للمفاهيم النظرية للألات التوربينية الخاصة بضاغط الطرد المركزي من أجل صياغة معادلات مثالية. تختلف مخرجات الأداء لهذه المعادلات المثالية عن الأداء الحقيقي للألات أثناء التشغيل بسبب الظروف البيئية وفقدان الطاقة. كما تغطي الرسالة تقدير الأداء باستخدام خوارزميات التعلم الآلي المُراقب، استناداً إلى التدريب على بيانات تاريخية واقعية. تم تنفيذ أربع خوارزميات رئيسية: الانحدار الخطي، شجرة القرار، الغابة العشوائية، والشبكة العصبية. ورغم أن خوارزميات التعلم الآلي تأخذ في الاعتبار البيانات الواقعية بما يشمل جميع الخسائر والظروف البيئية، إلا أنها تتطلب كميات ضخمة من البيانات، مما يؤدي إلى تكلفة حسابية وزمنية مرتفعة لتغطية جميع أوضاع التشغيل والمتغيرات، وهي بيانات غير متوفرة غالباً في الأدبيات العلمية المفتوحة. لتجاوز هاتين المشكلتين، تقترح الرسالة نهجاً هجيناً يجمع بين منهجي نماذج تصميم الألات التوربينية والتعلم الآلي المُراقب. تم استخدام أربعة مؤشرات مقارنة رئيسية لإثبات فعالية المنهج المقترح وتقييم درجات الخطأ: متوسط الخطأ المطلق (MAE)، مجموع مربعات الأخطاء (SSE)، ومجموع المربعات الكلية (SST). تم اختيار النموذج الأكثر كفاءة من حيث الأداء وتطبيقه على سيناريو تدهور مُحاكى لتقييم فعاليته. تخلص الرسالة إلى أنه حتى مع وجود معلومات محدودة عن تصميم الضاغط وبيانات تاريخية محدودة، يمكن بناء نماذج هجينة مقترحة بدقة مقبولة لتقدير التدهور.

CHAPTER 1

INTRODUCTION

1.1 Background

Centrifugal compressors are one of the most important pieces of equipment in different industries. They increase gas pressure by passing it through high-speed impellers. Pressurized gas from compressors is essential in the plant process. For that reason, the reliability of the compressors plays a vital role in the reliability of the whole plant. Plants without sufficient pressurized gas from the compressor may not be able to deliver the required products. Ensuring compressors operate with desirable efficiency is a must to keep operation.

To cost-effectively keep compressors, deliver the required function, maintenance is performed. Common centrifugal compressor maintenance can be classified into two main strategies: preventive scheduled-base and condition-base maintenance. Preventive maintenance depends on manufacturer instructions to stop the machine at fixed intervals for maintenance. Condition-based, on the other hand, is scheduled based on operator feedback from their routine inspection findings. The degree of deterioration can be assessed by regular inspections and diagnosis. After a specific long period of operation, the engine is overhauled. Overhauling is considered as scheduled and condition-based maintenance. Basically, the engine is dismantled, cleaned, inspected, repaired if required, and finally tested. The work done in the engine overhauling is based on both findings during regular inspections and manufacturing interval-based recommendations.

Condition monitoring methods are another widely used technique to assess machine health. It can be used while the machine is under operation. Data is gathered from different locations on machine and gas paths such as: vibration, temperature, flow, pressure, etc. to help assess the condition of machine. Mechanical faults can be predicted based on data trend such as: misalignment, unbalancing, and bearing faults.

These strategies, preventive, condition-based maintenance, and condition monitoring methods are capable of estimate the overall condition. However, they are not enough to evaluate the condition at component level. Actual condition per component can be estimated using more details analysis of engine gas path. As the gas passes through the internal machine part, it changes its physical properties. Monitoring and analyzing these changes in each stage of the machine can lead to visualizing the condition of each component.

1.2 Aim

The aim of this research is to determine whether physical deterioration on centrifugal compressors can be practically effectively detected during operation or not. Nowadays, industries have been evolving in data-driven decisions[ref]. There are a lot of built in instrumentation to measure upstream and downstream process data. However, the data alone are meaningless unless it is analyzed probably. There is still room to do much more to analyze these data to extract useful information. For instance, Tagging and linking data to physical design models provide comparison indicator in how far system from design and how it trends within time. This thesis tests the possibilities to have more visibility to the internal physical parts by utilizing data without stopping for inspection. The research attempts to close the gap between theoretical thermodynamic and aerodynamic balance and

actual real-time running compressors. It suggests a method to calibrate the ideal mass conservation law, momentum conversation law, and energy conversation to a real field example by using actual historical data that has hash environmental conditions and other losses.

Without a doubt, performance monitoring methods are valuable to improve maintenance and operation planning. Physical deterioration could lead to high shutdown and maintenance costs if not discovered early. However, there are many factors that influence the rate of performance reduction which makes it difficult to assess current conditions without considering historical trends. The approach that is taken in this thesis is to use reduced-order physical modelling and mathematical optimization methods to analyze the measurements collected by the sensors. Then, validate and tune using artificial intelligence, neural network, methods to improve prediction.

The main objectives for this thesis can be summarized as follows:

1. To build a reduced-order theoretical model for centrifugal compressors based on the thermodynamic concept of turbomachinery to estimate performance parameters.
2. To adopt and tune existing predicting models using synthetic data and machine learning algorithms.
3. To discuss, and compare different machine learning algorithms for effectiveness and accuracy.
4. To design deterioration detection and quantification methodology based on predicted and actual parameters in a simulated real degradation case.

1.3 Outline of the Thesis

Chapter 1 is an introduction chapter which gives a summary background about condition monitoring in general and then presents gas path monitoring. Centrifugal compressors are under mechanical stresses which can be indirectly measured by built sensors on the machine. However, these sensors are not enough to evaluate the performance of the machine. The gas path needs to be monitored to predict issues related to performance reduction. Models are built to link process parameters to performance indicators.

Chapter 2 is an intensive literature review for The last 10 years of research in gas path monitoring. Three main modelling methods were reviewed: Empirical mathematical, artificial intelligence, and statistical modelling. Models' accuracies and effectiveness were compared with different methods using Mean Error (ME) and Root Mean Square Error (RMSE).

Chapter 3 review the basic principle used in turbomachinery. A general review of common physical rules used in turbomachinery to be used in theoretical base modelling in chapter 5.

Chapter 4 introduces the source and goal for generated data. After that, it shows the type and range of data focused on this study. Then, describe the simulated degradation case study.

Chapter 5 is for Theoretical Performance Analysis and Modelling. Theoretical models are built based on compressor thermodynamic, energy conversation laws to link performance parameters with process parameters. Then it compares these theoretical models' results on

sample data with actual ones. Possible causes of differences between theoretical and actual are introduced and propose adaptation methods.

Chapter 6 uses different approach for modeling using Artificial Intelligence, Machine Learning Performance Modelling. The approach assumes no prior knowledge of thermodynamic relationships. The errors were compared to theoretical calculations. Different machine learning algorithms are presented and tested for their effectiveness in this section; linear, decision tree, forest tree, and neural network.

Chapter 7 presents Hybrid Theoretical and Machine Learning Modelling To improve the accuracy of theoretical models, combination of both model methods is suggested. Detailed steps and description of these techniques are presented and then tested.

Chapter 8 is about Degradation Detection. After selecting the best hybrid model with minimal error. The model was used in a case study to predict degradation. This section suggests a method to detect degradation, remove outliers, and quantify the performance degradation in real time by comparing predictive and actual trends. Finally, results from the implementation of the method are discussed.

Chapter 9 concludes the thesis report by summarizing the pros and cons of each methods and then suggests an area of improvement in future research.

CHAPTER 2

LITERATURE REVIEW

Condition-based monitoring is a technique used to reduce shutdown maintenance, visual inspection, and cost by relying on data to evaluate the machine's condition. Condition-based monitoring can be divided into two main categories. First: diagnostics which includes fault detection, isolation, and identification. [1] Fault detection includes abnormalities or unusual events in data trend capturing, fault isolation includes determining the location of faulty components, and fault identification is the root cause of the fault. Second: prognostics which is basically an intensification evaluation of the fault. It includes forecasting failure impending and estimating remaining useful life. Causes of rotating machinery deterioration can be classified into two main groups: mechanical nature and aerodynamic performance causes. [1] Mechanical nature is non-aerodynamic related such as misalignment, unbalance, loose components, bearing defects, and lack of lubrication. These faults can be monitored by vibration, oil and wear debris analysis, acoustics, thermography, lubrication flow parameters, load analysis, and metal temperature. On the other hand, aerodynamic or performance causes which are not mechanical nature related such as fouling, corrosion, and erosion can be detected by gas path analysis (GPA). [2] Gas path analysis methods, which is also called performance-based health monitoring, is one of the condition-based monitoring used to detect degradation faults. It attracts researchers and investors in the gas turbine and aircraft field. Different methods and approaches were attempted and tested to diagnose faults during operation before failure happened. Unlike mechanical faults monitoring health parameters are not directly measured. They are thermodynamically correlated to measured parameters. [2] from the following 10 years of

published papers on this subject, it can be concluded that there are mainly three main steps: engine model development, model adaptation, performance parameter estimation, and experimental or real data verification. The following literature review focuses on research studies that have been performed in the area of diagnostics and prognostics of gas turbine gas turbine engines in the field of power production and aviation but not much in compressors. This research applies the most common techniques that have been utilized on gas turbines, on centrifugal compressors. It compares different compressors' performance diagnostics methods and suggests the most effective approach based on real-time field data. There are three main modeling and adaptation methods; mathematical-base, Combined mathematical and Artificial intelligence, and/or statistical-base modeling.

2.1 Monitoring of Centrifugal Compressors

Increased operational demand leads to minimized shutdown for inspection and maintenance. Also, some failures develop in fast manor which could not be detected during regular inspection. Stopping for inspection and maintenance is not a preferable option. To achieve these requirements safely, condition monitoring methods are the solution for most industries. Investment is made to upgrade compressors with enough instrumentation to continuously measure real-time data to assess the condition without disturbing operation. These instruments are built inside the machine to capture more accurate data with less environmental noise. Temperature, differential pressure, vibration, flow rate, and guide vane position are the main indicators. The baseline for these indicators is created based on normal data range for different operational modes [3]. Any deviation from the baseline can indicate a potential failure. The degree of deviation refers to the severity of the machine's condition.

2.1.1. Monitoring Methods for Mechanical Deterioration

The centrifugal compressor is a mechanical design that increases the pressure of gas flow by rotating, or centrifugal force. They consist of a rotor that holds impellers and rotates on bearings. In figure (1.1) shows an example of an Integrally Geared Centrifugal Compressor design which is connected to a bull gearbox [4]. It reduces the rotational speed coming from an electrical motor.

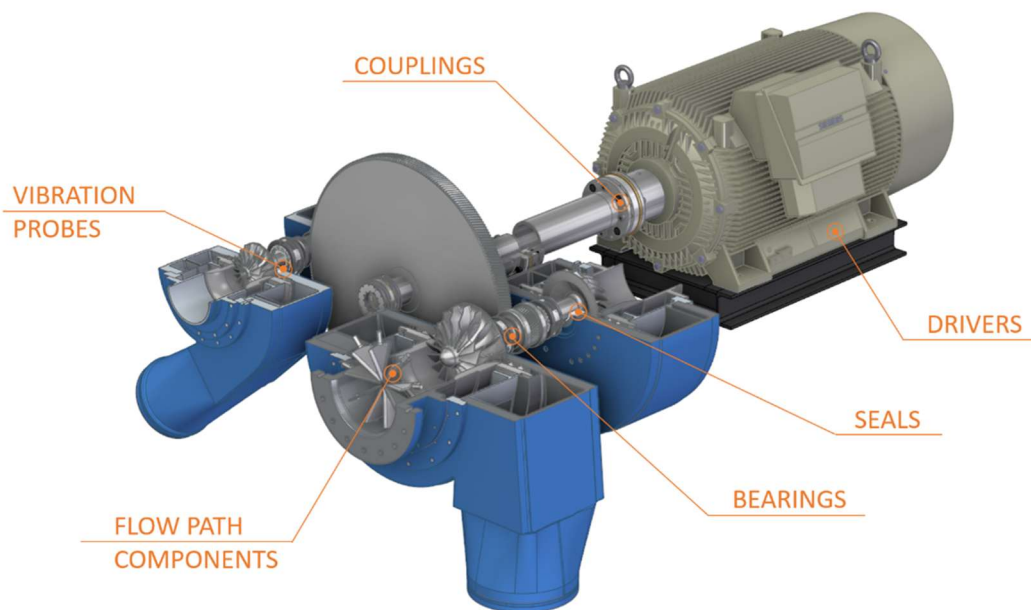


Figure 2.1 [5]: Integrally Geared Centrifugal Compressor Arrangement

The electrical motor transfers energy from electricity to mechanical rotational energy. The unique design of impellers, which include nozzles and diffuser allows gas to flow through them and increase pressure. This mechanism involves mechanical stresses. Forces from the weight of the rotor and impellers, gas impact and centrifugal force are handled by bearings. Bearings hold and restrict the rotor movement in all directions and allow only rotation in one coordinate about the rotor axis. For that reason, Bearings are also under high frictional force due to high-speed rotation and normal force from other components. To alleviate

stresses on bearings and allow smooth rotation as possible, lubrication oil from an auxiliary system is used. This lube oil system could be a source of stress on bearings if it does not supply enough lubricant flow, temperature, and pressure. Impellers, on the other hand, are exposed to high gas flow which contains impurities like dust, water, and other chemical components. Also, the impeller handle increased or decreased gas temperature due to compression. All these stresses cause wear out in the mechanical components of the compressor and eventually lead to failure.

Compressor failure does not necessarily mean it has stopped working. A compressor failure includes any deviation from the main function which is to increase gas pressure, flow, and temperature to a certain desirable value [6]. Wear out of mechanical parts is one of the causes of function deviation and leading to degradation in performance. Stresses on mechanical components within time generate a gradual increase in temperature and vibration. By monitoring these signals on the machine, mechanical deterioration can be predicted. Mechanical deterioration involves rotor misalignments, rotor imbalances, bearings faults, gearbox faults, blade or impellers defects or fracture, coupling issues, looseness, and others. Also, lube oil auxiliary system problems can be detected such as insufficient lubricant flow, high lubricant temperature or low lubricant pressure, and oil or water leakages. For example, fractured impeller could cause imbalance due to a change of weight distribution which increases vibration in rotating speed. Also, they may generate an increase in temperature over time due to friction. Another example is deformation on bearings due to high frictional force from poor lubrication could lead to excessive clearance which increases vibration and temperature.

To sum up, mechanical deterioration monitoring is a signal-based monitoring that basically indirectly measures stress intensiveness on mechanical components [3]. Temperature and vibration are the main signals that need to be measured in different locations to monitor mechanical parts. As per design internal components are exposed to different stresses. Within time these stresses will lead to component wearing out which increases vibration and temperature. Also, gas impurities may affect the impellers' geometry and surface which leads to an increase in vibration and temperature. Mechanical deterioration in the late stages will eventually affect compressor function leading to shutdown.

2.1.2. Monitoring Methods for Performance Deterioration

One of the main objectives of monitoring is to keep the machine performing within the acceptable range. A decrease in performance is not only caused by mechanical deterioration. Moreover, the machine could still perform perfectly even with mechanical deterioration signs [7]. Performance only has a considerable decrease in the late mechanical deterioration stage. For that reason, signal-based monitoring is not enough for performance monitoring. To assess compressor performance, the feed gas needs to be evaluated directly. This type of performance monitoring is called gas-path monitoring since it can predict faults that are in the gas path. It is a model-based monitoring that considers physical and thermodynamic laws to evaluate the degree of performance degradation. Unlike mechanical fault monitoring, health parameters are not directly measured. They are thermodynamically correlated to measured parameters.

Models are built for the machine from the conservation of mass, momentum, and energy laws to evaluate machine performance. The model input, or the dependent variable, is the controllable measurable parameters; guide vanes position which refers to gas flow rate,

inlet gas pressure and temperature, ambient temperature and relative humidity, rotation speed and motor power, and cooling water temperature. The model outputs, or independent variables, are the performance rate indicators; pressure ratio, polytropic efficiency, polytropic power, stage outlet pressure and temperature, and flow rate. Each indicator will have a separate model. The deviation between calculated and measured performance rate indicators means a drop in performance. However, it is almost impossible to have actual performance parameters as same as calculated ones. Compressors in real life are running under many uncontrollable harsh environments which lead to losses. For that reason, calculated models usually validate or adapt to an acceptable data range. Different adaptation approaches were implemented. Empirical approaches such as scaling, stage stacking, and regression model. Artificial intelligence approaches such as neural networks. Faults that can be predicted by performance monitoring are fouling, corrosion, erosion, clearance change, etc. These faults change the internal parts' geometry which leads to changes in performance parameters such as pressure ratio, flow rate, and efficiency.

2.1.2.1. Fouling

Fouling is the formation of unwanted material deposits on internal compressor parts. The air that the compressor takes from its surroundings is not clean. [8] Usually, it is polluted by other industrial processes or environments such as sand, dust, soil, insects, fly ash or hydrocarbons. Even with good filtration prior compressor inlet still, some of these foulants escape with the air. Built-up foulant on the gas path could modify aerofoil shape and increase surface roughness which causes changes in inlet flow angle and a rise in frictional losses. Moreover, these deposit materials could cause a grow in blade thickness which reduces flow capacity or even over time leads to flow blockage.

2.1.2.2. Corrosion

Corrosion, oxidation, is the destructive and continued chemical reaction between the metallic surfaces and certain flow contaminants, such as: salts, mineral acids or reactive gases. [9] Compressors are built with metal, and their surface is reactive with air. Due to this reason, a special coating is applied to isolate metal surfaces from direct air exposure. However, within time and harsh environments, this coating is weakened and allows metal and air direct contact. Air usually has contaminations that are reactive with Metal. This reaction causes metal surfaces to rust or corrode. Thickness reduction may occur due to high-speed velocities of the flow which make corroded material break away from the surface leaving small holes in the metal (pits). Pits could cause local stresses which diminish the mechanical strength of the blades therefore reducing the fatigue life of the blades.

2.1.2.3. Erosion

Abrasive material removal due to particle impact in the gas path. These particles are larger than 10 μm which at high speed could cause damage to the internal surface. Surface roughness increases which leads to flow separation, changes in incident angle, and finally pressure losses. [10] Good filtration before entering the compressor could minimize or solve this issue but within a long period of operation, compressors' internal parts, or piping might produce these abrasive materials due to wear out and affect other parts. Also, filters themselves, due to accumulated dirt could allow some particles to compressor.

2.2 Mathematical-Base Modelling

The mathematical model is to build the relationship between measured parameters; inlet and outlet temperature and pressure, motor power and speed, environmental condition, and health parameters: pressure ratio, polytropic efficiency, polytropic power, stage outlet pressure and temperature, by applying theoretical thermodynamic and aerodynamic rules. These modeling methods include Scaling, stage stacking or regression model. [1] they can be expressed in linear and non-linear relationship. However, linear relationships are not effective due to the non-linearity and complexity of gas turbines. There are attempts to modify linear equations to fit non-linear correlations. One of the linear approximation ways is Newton Rapson method. This method can be found in gas path monitoring [11] [12]. The partial least-squares technique is another way for mathematical modelling. [13] [14] least-square is a method to find the best fit by minimizing the sum of the offsets or residuals of points from the plotted curve. Kernel partial least squares (KPLS) which is the modified version of least square to deal with non-linearity, was used by Fei Chu et al. [13] to estimate operational parameters for centrifugal compressors. His result was significantly improved comparing to linear method as error result for both efficiency and pressure ratio within 2.5 %. Models need to be adopted using collected data to optimize their coefficients. Adaptation or tuning is basically a way to fit the mathematical model with target machine data with minimum residual. Li, X et al [14] use stacking calculation to develop a compressor map by stage. They combined different station performances to get the generalized performance curve. They also uses novel component map adaptation to optimize model coefficients. Moreover, partial least square regression is used to minimize error. Novel component map adaptation assumes speed lines, the pressure ratio is an elliptic

function of the corrected mass flow rate. In addition to Xu Li et al paper, novel component map adaptation can be found in a number of recent papers. A published paper by Dengji Zhou et al [15] uses the same novel model in periodic analysis. Novel regression is used by Elias Tsoutsanis et al [16] to refine existing engine performance models to different modes of gas turbine operation. Other researchers combine it with other methods to improve accuracy. Peng Jin [17] combines novel approach with extended Kalman filtering (EKF) and adopts the model by using simulated data from MATLAB. EKF is the nonlinear version of the Kalman filter which linearizes about an estimate of the current mean and covariance. Lü, F.; Huang, J.; & Lv, Y. [18] combine EKF with particle filter (PF) and use it to diagnose aircraft engines. EKF as a method alone proves its effectiveness as well. Feng Lu et al [19] applies EKF on an experimental two-spool mixing-exhaust turbofan engine and simulated data. Kalman filter is used by Jingchao Lia and Ying, Y. [11] to compare it with the Newton-Raphson algorithm in gas turbines.

2.3 Combined mathematical and Artificial intelligence modelling.

Hybrid methods are widely used approaches as well which have proven their capabilities to overcome sensor noises, complexities, and nonlinearities. They basically combine artificial intelligence methods with mathematical models. Lin et al. [20] combine a novel approach with a long short-term memory (LSTM) network. Real-life operation data of aero-engines are used in his research, twenty CFM56-5B aero-engines, which is a kind of twin-shaft Turbofan aero-engine and provides propulsion and power for Airbus 320 civil aircraft. LSTM is one of the Artificial Neural Networks (ANNs) techniques that are based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. An artificial neural network (ANN) with a

thermodynamic model was utilized by M. Talaat et al. [21] on in-field gas turbine data. Residual-Back Propagation Neural Network (Res-BPNN) combined with stacks of multi-layer Multi-Kernel Maximum Mean Discrepancy (MK-MMD) is also one of the ANN methods published by Xingjie ZHOU et al. [22] which used it on different civil aero-engine models. In addition, Kun Wang and Zepeng Wang use ANN in their research separately to model the gas path of turbines. Kun Wang et al. [23] utilizes simulated data for two -spool turbine engine. Zepeng Wang et al. [24] compares six different neural network methods on simulated data. Fuzzy logic (FL) is one of the artificial intelligence methods which is used in gas path monitoring. It is based on fuzzy rules, set of if-then statements. A paper published by Odokwo [25] is one of example which utilized fuzzy logic in MATLAB and uses experimental data PG9171 model manufactured by General Electric. Transfer learning with novel model is used by Junqiang Liu [26] on experimental data from China Eastern Airlines and by Shanxuan Tang et al. [27] on experimental data from a two-spool split flow turbofan engine. Transfer learning (TL) is a technique in machine learning (ML) in which knowledge learned from a task is re-used in order to boost performance on a related task. T.V.Breikin et al. [28] uses Genetic algorithms, with dynamic modelling on gas turbines to estimate operational parameters. Also, Genetic algorithms were utilized by Hu Yu et al. [29] combined with Artificial bee colony algorithm with tabu search (TSABC) on experimental data from a turbofan engine model.

2.4 Statistical-base Modelling

Other recent effective methods are a statistical base. The hierarchical evolutionary model [30] [31] [32] is inspired by a ubiquitous organizing principle in biology and a key reason evolution produces complex, evolvable organisms. Empirical Mode Decomposition [33]

[34] is a technique, known as data-adaptive multiresolution, that can decompose a signal into components that have physical meaning. EMD can be used to analyze non-linear and non-stationary signals by separating them into components at different resolutions. Stochastic dynamics [35] [36] is the dynamics of information accrual in models involving noisy, time-varying stimulus information in continuous time are mathematically represented by stochastic differential equations. vector-based deterioration [37] is a method also called "non-local means denoising" inspired by image-reducing noise. This involves replacing each pixel in the noisy image with the weighted average of other pixels in its surrounding neighborhood. The weight of each pixel is determined based on how similar its neighborhood is to the one around the pixel being replaced. Bee colony algorithm [29] is a probabilistic technique, inspired by the behavior of real bees, used to solve computational problems that can be reduced to finding optimal paths through graphs. Topological data analysis [38] is an approach to the analysis of datasets that uses techniques from topology to extract valuable information. Dealing with high-dimensional, incomplete, and noisy datasets can be challenging, but TDA provides a general framework that can analyze such data in a manner that is insensitive to the metric chosen. Additionally, TDA can provide dimensionality reduction and robustness to noise, making it a useful tool for data analysis.

2.5 Modelling Methods Comparison

Among the presented methods, the most accurate and effective in recognizing faults is needed to be found. Researchers evaluate their models by calculating Root Mean Square Error (RMSE) Which measures the mean predicted and actual difference.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad \text{Eq 2-1}$$

Other researchers use Mean Error to evaluate the fitness of the model to actual data. It is the ratio between predicted and actual differences with actual.

$$RME = \frac{y_i - y'_i}{y_i} \quad \text{Eq 2-2}$$

For the sake of comparison, the maximum RMSE or ME was considered in each paper in the literature. The following Tables (Tables 2.1 and 2.2) show a comparison between the number of papers found in the literature. The minimum RMSE is 0.012 which was found in [16] using the sparse deep stacking network method on the gas turbine model developed under the MATLAB2020 environment. The minimum ME is 0.04 was obtained using the transfer learning method [11] on experimental data; a two-spool split flow turbofan developed by Beihang University.

Table 2.1: Modelling Methods In Literature With Maximum RMSE Score

#	Modelling method	citation number	application	RMSE
1	Genetic algorithms	[28]	experimental data from turbofan engine model.	3.02
		[29]	on-field data aero gas turbine	1.90
2	kernel partial least square	[13]	centrifugal compressor power plant of a steelwork	0.0631
		[14]	three shaft marine gas turbine	
3	Kalman particle filter	[18]	Turbofan Engine	0.525
		[19]	two-spool mixing-exhaust turbofan engine	0.0546
		[11]	single-shaft power generation gas turbine with natural gas as fuel	0.356
		[17]	simulated data using MATLAB	0.59
4	novel Model	[26]	euro-engine experimental data from China Eastern Airlines	0.5
5	hierarchical evolutionary model	[30]	experimental durability test on centrifugal compressor	1.51
6	neural network	[21]	real gas turbine data	0.0331
		[22]	cruise data from different civil aero-engine models and compared with other transfer learning algorithms	1.65
		[24]	gas turbine model, the engine model was developed under the MATLAB2020 environment. A complete thermodynamic engine model	0.012
7	long short-term memory network	[20]	real-life operation data of aero-engines, twenty CFM56-5B aero-engines, which is a kind of twin-shaft turban aero-engine and provides propulsion and power for Airbus 320 civil aircraft.	0.48
8	empirical mode decomposition	[34]	a single-stage centrifugal compressor.	1.18
		[33]	practical application cases and experimental from aero-engine	0.44

Table 2.2: Modelling Methods In Literature With Maximum ME Score

#	Modelling method	citation number	application	ME
1	novel Model	[16]	MATLAB Simulink environment, gas turbine power plants.	0.15
2	transfer learning	[38]	experimental data a two-spool split flow turbofan developed by Beihang University	0.04
3	fuzzy logic	[25]	experimental data on PG9171 model manufactured by General Electric.	0.9
4	Newton–Raphson method	[12]	A single shaft gas turbine engine, which is constructed based on the core part of a turbofan engine.	1.82
5	stochastic dynamics responses	[35]	aerothermodynamic model is implemented in MATLAB Simulink.	2.5

2.6 Summary

To sum up, based on the above review of different methods, model-based methods are accurate methods that link the physical nature of degradation to measured parameters. However, it needs details geometrical information about equipment and the surrounding environment which in most cases is not available. Also, for a complex system that includes many stages and sub-components, requires a complex model that requires time and effort. On the other hand, data-based models using artificial intelligence is nowadays handy and easy to compute. regardless of how complex the system and even non-linearity, AI can provide sufficient accuracy. However, AI is a black box which blindly considers only input data to compute output without evaluating if the data is reasonable or not. Usually, data collected from instruments involve noises and malfunctions which leads to fake fault

detection. Also, AI needs huge historical data to be optimized to capture all operational and environmental variation which in most cases is not available. Different researchers recognized these challenges. To help overcome these challenges, over the past 10 years combining methods model and data-driven method are the favorite approach to handle noises, uncertainty, and lost data and to avoid fake faults. while Model-base limit model probabilities outcome by linking design limitation. Data-driven approach tune and calibrate model-base to be more close to real system which include energy losses. This study introduces a comparison and effectiveness evaluation of different gas path hybrid modeling methods, model-based and data-driven based, on industrial centrifugal compressors. After searching the literature, it was concluded these methods have been studied and tested on aero-engine and gas turbines, but they are yet to be developed for industrial centrifugal compressors. The outcome of this study will contribute to optimizing maintenance costs, and increasing availability and reliability of equipment. By developing a performance model for compressors, the maintenance strategy will be more condition, database monitoring than preventive maintenance or reactive maintenance. As a result, equipment will have more operational time than shutdown time for inspection or maintenance. Also, by continuously monitoring live stream data will have more visuality of potential failure and have early warning to plan for maintenance.

CHAPTER 3

Basic Principles Used in a Turbomachinery

Compressors are turbomachinery machines that transfer energy between a rotor and a fluid. A Shaft contains rotors that are attached to blades. Blades are fixed with an angle and a shape on the shaft to rotate and transfer energy to the fluid when going through it. Turbomachinery converts rotor velocity, or mechanical energy, fluid static, and dynamic pressure continuously. This chapter is a review for basic physical law used in designing turbomachinery equipment.

3.1 Continuity Equation

As shown in Figure (3.1), consider a pipe that allows mass flow from an inlet (1) to an outlet (2) while the pipe's surface's sides are walls that do not allow any mass through them. The continuity equation, or principle of mass conservation, states that the inlet mass flow rate is always equal to the outlet mass flow rate [39], $\dot{m}$

$$\dot{m}_1 = \dot{m}_2$$

Eq 3-1

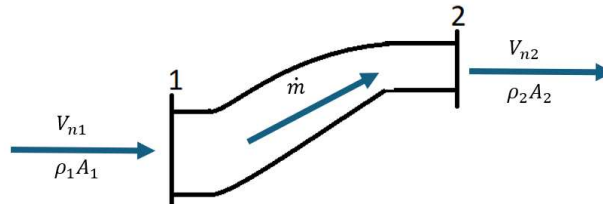


Figure 3.1: Continuity Equation

The mass flow rate in a uniform flow can be related to the density and velocity of the fluid and the cross-sectional area of the flow. So, equation 3-1 can be written as:

$$\rho_1 A_1 V_{n1} = \rho_2 A_2 V_{n2}$$

Eq 3-2

ρ Fluid Density

A Cross Section Area

V_n Fluid absolute Normal Velocity

Density can be constant in incompressible fluid like water. In that case, it can be said that increasing cross-sectional area leads to a decrease in normal velocity. However, this is not necessarily true for compressible fluid like air when density is not constant. Even with the fact that variables are not steady in the real process, it is safe to assume as elementary analysis that all variables have steady values.

3.2 Momentum Equation

Fluid flow with a certain pressure and speed in turbomachinery is generated from an applied rotating force, or torque. Considering compressors, for example, torque is generated by a coupled electrical motor which transfers energy through a rotating shaft to

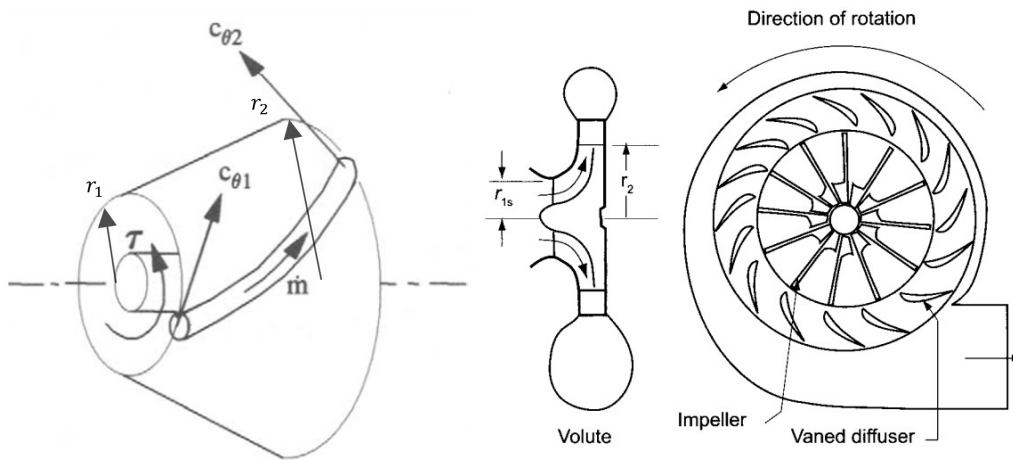


Figure 3.2: [40] Rotor Rotation And Tangential Speed For Centrifugal Compressor

impellers which changes the tangential momentum of fluid mass. This required torque can be found by applying Newton second law [40] as follows:

$$\textbf{Applied torque} = \textbf{rate of change of moment of momentum} \quad \text{Eq 3-3}$$

Linear momentum is the quantity of motion of a moving mass which can be measured as mass multiplied by velocity. Since fluid motion in turbomachinery is rotational, angular momentum, or moment of momentum, needs to be measured by multiplying moment by a radius of rotation. So, equation 3-3 can be written as:

$$\tau = m \frac{d}{dt} (rV_{\theta}) \quad \text{Eq 3-4}$$

Considering the flow has one inlet 1 and one exit 2 and flow is steady.

$$\tau = \dot{m}(r_2V_{\theta 2} - r_1V_{\theta 1}) \quad \text{Eq 3-5}$$

τ Torque

$\dot{m}$ Mass flow rate

$V_{\theta j}$ Fluid tangential Velocity at θ axis at location j

r_j Rotation radius at location j

The required power from motor can be found as

$$P = \tau\omega \quad \text{Eq 3-6}$$

ω rotor angular speed

P Power

Substituting torque from equation 3-5 to equation 3-6

$$P = \dot{m}(\omega r_2 V_{\theta 2} - \omega r_1 V_{\theta 1}) \quad \text{Eq 3-7}$$

$$P = \dot{m}(U_2 V_{\theta 2} - U_1 V_{\theta 1}) \quad \text{Eq 3-8}$$

Where $U = \omega r$ is blade linear speed. Dividing by $\dot{m}$ both sides, we can get specific work

$$w = \frac{P}{\dot{m}}$$

$$w = U_2 V_{\theta 2} - U_1 V_{\theta 1} \quad \text{Eq 3-9}$$

3.3 First Law of Thermodynamics

The first law of thermodynamics is the principle of energy conservation [41]. Consider Figure 3.1 which is a pipe with a control volume and has single inlet and single outlet. The uniform steady flow energy equation is as follows:

$$\dot{m} \left(u_1 + P_1 v_1 + \frac{1}{2} V_1^2 + g z_1 \right) + \dot{Q} = \dot{m} \left(u_2 + P_2 v_2 + \frac{1}{2} V_2^2 + g z_2 \right) + \dot{W} \quad \text{Eq 3-10}$$

The sum of internal energy u , Kinetic energy $\frac{1}{2} V^2$, and potential energy gz is called fluid specific energy $e = u + \frac{1}{2} V^2 + gz$. The sum of internal energy u and the flow work pv is called enthalpy $h = u + Pv$ which P is fluid absolute pressure and v is volume. $\dot{Q}$ is the

heat transfer rate and $\dot{W}$ is the delivered work rate. by including enthalpy h in the previous equation and dividing by mass flow rate $\dot{m}$ the energy equation can be shortened as follows:

$$h_1 + \frac{1}{2}V_1^2 + gz_1 + q = h_2 + \frac{1}{2}V_2^2 + gz_2 + w \quad \text{Eq 3-11}$$

In which q and w are heat transfer and work done per unit mass. The sum of enthalpy, kinetic energy and potential is called the stagnation enthalpy h_0

$$h_0 = h + \frac{1}{2}V^2 + gz \quad \text{Eq 3-12}$$

In this study, the flow of gases is considered in compressors. Potential energy due to elevation difference can be neglected because it is small. Also, turbomachinery flows are nearly adiabatic which means heat transfer from surrounding to system q can be removed from the equation. Then, energy equation for adiabatic, controlled volume and steady state flow for gas flow in compressors can be reduced to $h_{01} = h_{02} + w$ Re-arrange the equation, work done on the fluid in compressor is.

$$w = h_{02} - h_{01} \quad \text{Eq 3-13}$$

For ideal gas enthalpy delta can be found using Isobaric specific heat C_p and change in temperature.

$$h_{02} - h_{01} = C_p(T_{02} - T_{01}) \quad \text{Eq 3-14}$$

3.4 Second Law of Thermodynamics

The second law of thermodynamics can be described in two simple statements; heat always flows spontaneously from hotter to colder regions of matter, not all heat can be converted into work in a cyclic process [39]. To express these two-statements, a thermodynamic variable was introduced called entropy S which is basically measure the difference of heat transfer divided by temperature

$$\Delta S = \frac{\Delta Q}{T} \quad \text{Eq 3-15}$$

There are mainly two types of heat transfer in turbomachinery system, heat losses from system to machine's walls or surroundings and heat loss from fluid molecular friction. Equation 3-15 then can be written as

$$\Delta S = \frac{\Delta Q}{T} + \frac{\Delta Q_f}{T} \quad \text{Eq 3-16}$$

Entropy of system due to heat loss to surroundings stays constant if process is reversable which means there is no heat transfer from state 1 to state 2 to surroundings. Even with changes in temperature, pressure, and velocity, they would return to their original values downstream. In other hand, entropy would increase from state to another when there is a change in heat transfer between two objects in system which makes heat flows from hot to cold surface. This process is called irreversible since part of energy has been lost and cannot be recovered.

$$S_2 = S_1, \text{ (Reversible Process)}$$

$$S_2 > S_1 \text{ (Irreversible Process)}$$

In reality, there is no reversible process. Losses always exist. However, with temperature difference between the walls and the fluid is low and the temperature gradient in the flow direction is small, it can be assumed a reservable process with acceptable error. This can be achieved with high fluid flow rate so the energy transfer by bulk motion far exceeds the transfer by conduction and radiation in the flow direction. The second term of equation is energy losses which can happen from fluid itself. When fluid pressure and temperature increases, internal frictional heat is produced due to an increase of molecules interaction. Fluid can be considered internally reversable when heat losses is low. In both cases, there are differences between calculated values assuming constant entropy and real values. Losses are unavoidable. Heat cannot totally be converted to work. For this reason, efficiency needs to be calculated which is basically measure of system disorder, or deviation from calculated with constant entropy.

3.5 Gibbs Equation Entropy and Enthalpy Relation

For closed systems and with negligible potential and kinetic energy, interactions happen in work, heat, and internal energy. From the first law of thermodynamics, it can be expressed mathematically

$$dU = \delta Q - \delta W \quad \text{Eq 3-17}$$

Work can be calculated due to compression or expansion as pressure multiply by change of volume

$$\delta W_s = PdV \quad \text{Eq 3-18}$$

For internally reversible process, heat transfer change can be linked to entropy change as presented in equation 3-15 and it can be written as

$$\delta Q_s = TdS \quad \text{Eq 3-19}$$

Substituting these two equations 3-18 and 3-19 into 3-17

$$dU = TdS - PdV \quad \text{Eq 3-20}$$

Or by dividing by mass, it can be written as

$$du = Tds - Pdv \quad \text{Eq 3-21}$$

Re-arranging the equation to be with respect to entropy, the equation called first Gibbs equation [41]

$$Tds = du + pdv \quad \text{Eq 3-22}$$

Knowing that $u = h - pv$ and differentiating give $du = dh - pdv - vdp$. Substituting this into the first Gibbs equation gives

$$Tds = dh - vdp \quad \text{Eq 3-23}$$

Which is the second Gibbs equation. This equation is useful to graph entropy S and enthalpy h diagram to estimate efficiencies and compare it with actual. For ideal gas $dh = C_p dt$ and $v = \frac{RT}{P}$ substitute in equation 3-23 and dividing both side by T we will have

$$ds = C_p \frac{dT}{T} - R \frac{dP}{P} \quad \text{taking integral} \quad \text{Eq 3-24}$$

$$\Delta s = C_p \log \frac{T_2}{T_1} - R \log \frac{P_2}{P_1} \quad \text{Eq 3-25}$$

3.6 Stagnation Temperature and Pressure

Due to the expansion and compression process that occurred inside turbomachines, pressure is changing which leads to fluid velocity changes across stages. For this reason, it is more convenient to combine kinetic energy in each point in machine with enthalpy as presented before in equation 3-12 which is called stagnation enthalpy h_0 . Based on the first law of thermodynamics, if there is no work nor heat transfer, stagnation enthalpy is constant.

For ideal gas cases, the relation between pressure and temperature can be derived from enthalpy and stagnation enthalpy. If fluid is ideal, then $h = C_p T$ substituting this into $h_0 = h + \frac{1}{2} V^2$ it will be as $C_p T_0 = C_p T + \frac{1}{2} V^2$. Dividing by C_p in both side stagnation temperature is defined as:

$$T_0 = T + \frac{1}{2} \frac{V^2}{C_p} \quad \text{Eq 3-26}$$

Dividing by T in both side and knowing $C_p = \frac{\gamma R}{\gamma - 1}$ it gives

$$\frac{T_0}{T} = 1 + \frac{1}{2} (\gamma - 1) \frac{V^2}{\gamma R T} = 1 + \frac{1}{2} (\gamma - 1) M^2 \quad \text{Eq 3-27}$$

Where M is Mach number, defined as the ratio of fluid speed V to speed of sound c ; $M = \frac{V}{c}$. Speed of sound in air as function of absolute temperature is $c = \sqrt{\gamma R T}$ substitute in

Mach number formula; $M = \frac{V}{c} = \frac{V}{\sqrt{\gamma R T}}$. Stagnation pressure can be derived using Gibb's

relation equation 3-25; $Tds = dh - vdp$ for adiabatic and isentropic $ds = 0$ and for ideal

gas $v = \frac{RT}{p}$ then $dh = \frac{RT}{p} dp$ Substituting $dh = C_p dT$ and re-arranging gives $\frac{dp}{p} = \frac{C_p}{R} \frac{dT}{T}$.

Knowing $C_p = \frac{\gamma R}{\gamma - 1}$ we obtain:

$$\frac{dp}{p} = \frac{\gamma}{\gamma - 1} \frac{dT}{T} \quad \text{Eq 3-28}$$

By integrating from 1 – 2:

$$[\ln P]_{p_1}^{p_2} = \frac{\gamma}{\gamma - 1} [\ln T]_{T_1}^{T_2} \quad \text{Eq 3-29}$$

$$\ln P_2 - \ln P_1 = \frac{\gamma}{\gamma - 1} (\ln T_2 - \ln T_1) \quad \text{where } \ln(a) - \ln(b) = \ln\left(\frac{a}{b}\right)$$

Eq 3-30

$$\ln\left(\frac{P_2}{P_1}\right) = \frac{\gamma}{\gamma - 1} \ln\left(\frac{T_2}{T_1}\right) \quad \text{Raise to power } e \text{ both side}$$

Eq 3-31

$$e^{\ln\left(\frac{P_2}{P_1}\right)} = e^{\ln\left(\frac{T_2}{T_1}\right)^{\frac{\gamma}{\gamma - 1}}} \quad \text{Eq 3-32}$$

$$\frac{P_2}{P_1} = \left(\frac{T_2}{T_1}\right)^{\frac{\gamma}{\gamma - 1}} \quad \text{Eq 3-33}$$

Considering fluid move from static to stagnation the relation for stagnation temperature and pressure can be written as:

$$\frac{P_0}{P_1} = \left(\frac{T_0}{T_1}\right)^{\frac{\gamma}{\gamma - 1}} \quad \text{Eq 3-34}$$

Substituting Eq 3-26 into 3-34:

$$\frac{p_0}{p_1} = \left(1 + \frac{1}{2}(\gamma - 1)M^2\right)^{\frac{\gamma}{\gamma-1}} \quad \text{Eq 3-35}$$

From gas law density $\rho = \frac{p}{RT}$, then $\frac{\rho_0}{\rho} = \frac{\frac{p_0}{RT_0}}{\frac{p}{RT}}$, so $\frac{\rho_0}{\rho} = \left(\frac{p_0}{p}\right)\left(\frac{T}{T_0}\right)$

$$\frac{\rho_0}{\rho} = \left(\frac{T_0}{T_1}\right)^{\frac{1}{\gamma-1}} \quad \text{Eq 3-36}$$

CHAPTER 4

DATA EXPLORATIONS

The thesis study is targeting industrial multistage centrifugal compressors in an Air Separation Unit (ASU). Air from the surroundings is the fluid that needs to be compressed. The components along the path of the gas are the focus of this study which mainly are Rotors, Ducts, Stators. This chapter describes type, range and source of data that has been used to prove and validate thesis proposal.

4.1 Data Generation Goal

Most plants depend mainly on the provided performance map by the manufacturer to evaluate the condition of their machine. However, these maps are experimental which are limited to specific geometrical details and narrow operational condition range. In addition, Machine after overhauling would not be same performance as newly installed-one. Every time there is overhauling performance map changes. For that reason, an adoptable theoretical and database performance map is needed which will cover wider geometrical details and operational condition range and can be calibrated on real data from a running machine.

Due to data privacy and scarcity restriction, simulated data was considered which is very close and similar to real data. While using real data would be more realistic and accurate, synthetic data would be enough to prove the concept of this research. To make this study more practical, the considered types of data are the ones which can easily get from the most standard compressors. These data are measured by built-in instruments in the gas path in different locations which include mainly temperature, pressure, valve position, differential

pressure, and flow rate. Also, data from the motor is considered in modelling as Power and speed.

4.2 Data source, Type and Range

Cyril Picard, Massachusetts Institute of Technology, Cambridge, MA, [42] in May 2023 published a large synthetic dataset for centrifugal compressors. They are tabular design data; each row represents a design variant and each column is feature. Features include geometrical details, fluid properties, input, and output operational conditions. Each feature has bounded upper and lower value which cover real-world compressors design range. Some of the bounds are independent while others depend on other bounds to match design limitation and consideration. Data out of these boundaries for each feature has been sampled randomly from distribution by different methods to reduce dataset size without biasing or losing important data attribute. 19 million design variants were generated and validated base on 14 real-world compressors' geometries with actual performance parameters. Normal noise with 1% of the design space was applied. Geometrical parameters and operation conditions with their range as listed in table (4.1) are input to model, which is built and trained on actual data. The Model then calculates inlet and outlet velocity triangles at the impeller and inducer and corrected by accounting for losses using established correlations. After that, outlet and performance parameters can be found, isentropic efficiency, pressure ratio, outlet temperature.

The dataset was reduced by filtering out other non-interest working fluid such as: ammonia, isobutane, pentane, propane. Also, dataset was reduced by filtering not working condition rows such as choking and surge. After filtering, dataset size focus in this study become 290000 rows.

Table 4.1: Data Input, Range And Unit

independent geometrical parameters			
r_2	impeller outlet radius	[5,250]	mm
β_1	mean-line impeller inlet blade angle	[-60,0]	degree
β_2	mean-line impeller outlet blade angle	[-70,-35]	degree
Z_b	Number of Blades	[5,20]	-
Dependent geometrical parameters			
r_0	inlet nozzle	$[r_{1s}, r_2]$	mm
r_{1s}	impeller shroud radius	$[0.1r_2, 0.5r_2]$	mm
r_{1h}	impeller hub radius	$[1.2r_{1h}, 0.8r_2]$	mm
r_3	diffuser outlet radius	$[r_2, 4r_2]$	mm
b_2	impeller outlet blade height	$[0.015r_2, 0.3r_2]$	mm
b_3	diffuser channel width	$[0.5b_2, 1.5b_2]$	mm
input operational parameters			
m	mass flow rate	[0.00006120, 40]	Kg/s
T_0	inlet temperature	[170, 400]	K
P_0	inlet pressure	[14, 1260]	KPa
ω	rotational speed	[382, 53100]	rad
$Power$	actual Power	[1.7, 3500000]	Watt
fluid properties: Air			
R	Specific Gas constant	287	J/kg.K
C_p	Isobaric Specific heat	1007	J/(kg*K)
γ	specific heat ratio Gamma	1.4	-

Table 4.2: Data Output And Unit

output performance parameters				
T_3	outlet temperature	172 - 430	K	
P_3	outlet pressure	16 - 1710	KPa	
η	isentropic efficiency	0.65 – 0.88	-	
PR	Pressure ratio	1 – 1.56	-	

4.3 Simulated Case-Study

To evaluate models' effectiveness to detect performance degradation. Another dataset is needed to simulate degradation trend. While the presented dataset focuses on normal performance from varying geometrical details and operation condition, case-study data set should have fixed geometrical details and less variation in operational condition. Looking at ASU plants centrifugal compressors, inlet temperature and pressure are ambient temperature and atmospheric pressure. Pressure is almost constant and temperature are varying due to seasonal changes, Low in winter and high in summer. As per Marta Zagorowska et al. [43] performance degradation rate starts to behave linearly then with long operational hour it changes to exponential. Figure (4.1) shows the efficiency reduction over 18 months with 4 stops for cleaning.

Dataset are generated using degradation model develop by Marta Zagorowska et al. [43] on real data. This model has a combination linear and exponential function which switch between them based on running hours. Inputs to the model are normal operation inlet and

outlet condition and running hours. Outputs are degraded efficiency, pressure ratio. Table (4.3) shows selected geometry details and inlet operation condition range.

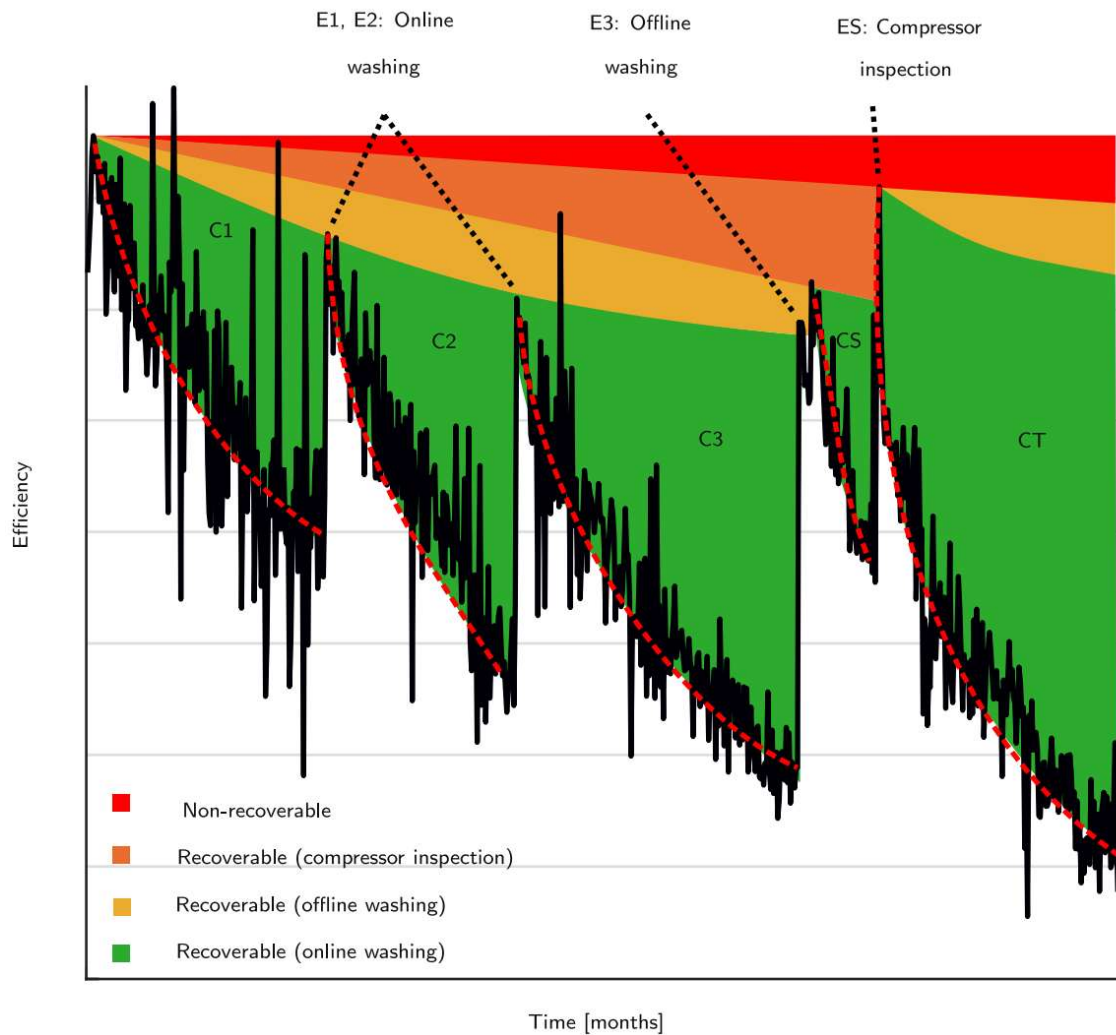


Figure 4.1 [43]: Efficiency Trends With Maintenance Shutdown 18 Months

Table 4.3: Case-Study Data Input, Range And Unit

Fixed Geometrical Parameters			
r_0	inlet nozzle	44.97	mm
r_{1s}	impeller shroud radius	44.53	mm
r_{1h}	impeller hub radius	26.55	mm
r_2	impeller outlet radius	68.19	mm
r_3	diffuser outlet radius	71.43	mm
β_1	mean-line impeller inlet blade angle	-9.16	degree
β_2	mean-line impeller outlet blade angle	-54.02	degree
b_2	impeller outlet blade height	6.29	mm
b_3	diffuser channel width	6.92	mm
Z_b	Number of Blades	16	-
Input Operational Parameters			
m	mass flow rate	0.42	Kg/s
T_0	inlet temperature	[283, 318]	K
P_0	inlet pressure	[98, 102]	KPa
ω	rotational speed	3545	rad
$Power$	actual Power	19667	Watt
fluid properties: Air			
R	Specific Gas constant	287	J/kg.K
C_p	Isobaric Specific heat	1007	J/(kg*K)
γ	specific heat ratio Gamma	1.4	-

CHAPTER 5

Theoretical Performance Analysis and Modelling

This chapter includes equations governing and variable reducing to build theoretical-base models to evaluate compressor performance. From previously explained laws in chapter 3, models variables are narrows to fit specifically centrifugal compressor by considering geometrical design and gas path direction.

5.1 Velocity Triangle

Let $\mathbf{Y}$, $\mathbf{r}$, $\boldsymbol{\theta}$ to be 3d coordinates which $\mathbf{Y}$ axis of rotation, $\mathbf{r}$ is parallel to impeller plane perpendicular to $\boldsymbol{\theta}$. $\boldsymbol{\theta}$ is tangent to circumference of rotation. In the compressor inlet, Fluid flow velocity u_1 enters to impeller at direction of axis of rotation $\mathbf{Y}$ with 90 degrees from tangential $\boldsymbol{\theta}$ axis. At impeller outlet, another axis is considered perpendicular to axis of rotation $\mathbf{r}$. Fluid flow exit with a velocity u_2 with an angle α_2 from tangential axis $\boldsymbol{\theta}$. U_1 , U_2 are the blade velocities at inlet 1 and outlet 2. Blade velocities are tangential to blade circle, so it is always on $\boldsymbol{\theta}$ coordinate. They can be calculated as rotation speed multiply by radius of blade. While speed of shaft is constant, U_1 has different value than U_2 due to different blades radius r_1 and r_2 . By velocity triangle related velocity of fluid can be found which is the resultant velocity from blade and fluid vectors in inlet and outlet V_1 and V_2 with angle β_1 and β_2 measured from tangential axis $\boldsymbol{\theta}$.

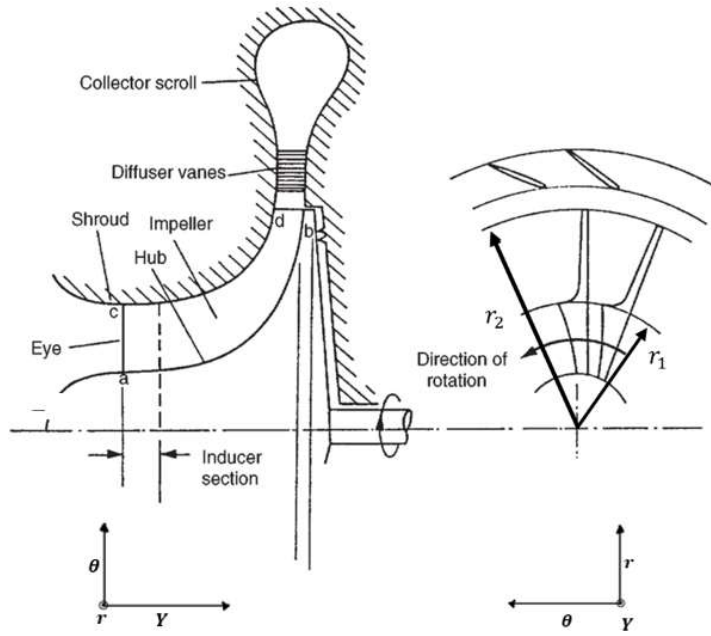


Figure 5.1 [39]: Inlet And Outlet Impeller Cross Section With Coordinates

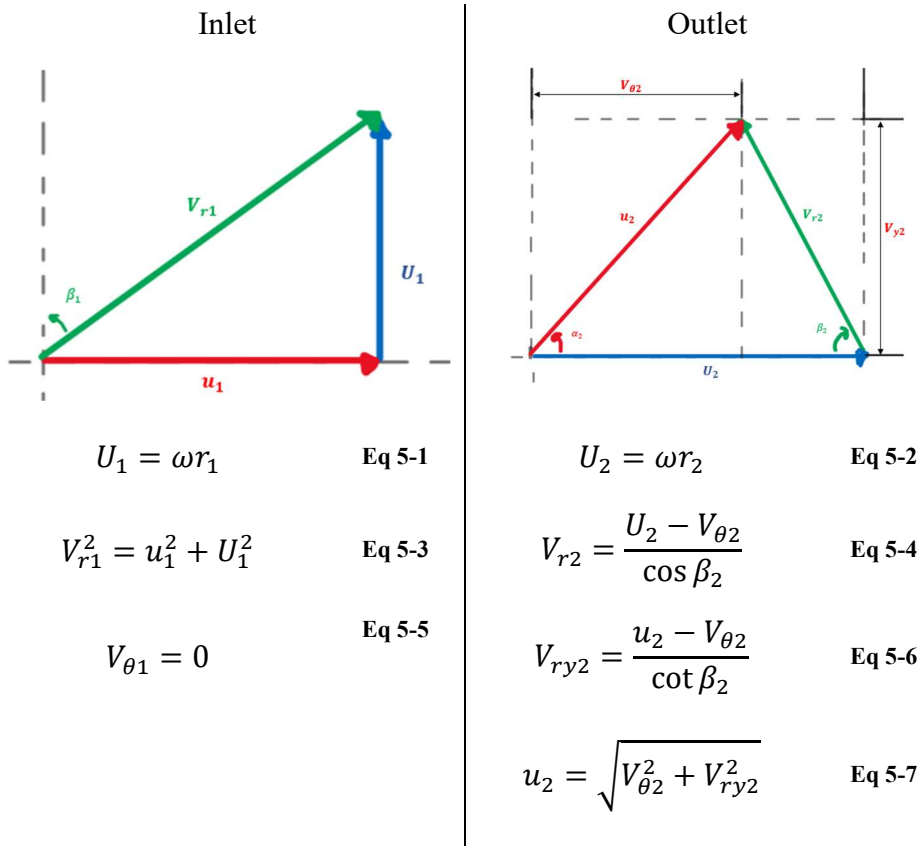


Figure 5.2: Inlet And Outlet Velocity Triangle

5.2 Rothalpy preservation in Compressors

Combining equation 3-9 specific work from momentum equation and equation 3-13 specific work done of fluid from rise in stagnation enthalpy, we have:

$$w = U_2 V_{\theta 2} - U_1 V_{\theta 1} = h_{02} - h_{01} \quad \text{Eq 5-8}$$

Re-arranging equation by knowing $h_0 = h + \frac{1}{2}u^2$ with no potential energy.

$$h_1 + \frac{1}{2}u_1^2 - U_1 V_{\theta 1} = h_2 + \frac{1}{2}u_2^2 - U_2 V_{\theta 2} = I \quad \text{Eq 5-9}$$

I is called Rothalpy. Rothalpy is fixed at inlet and outlet of each stage. It can be used to link fluid mechanical properties, enthalpies through stages. Equation can be simplified by considering velocity triangles as following:

$$u^2 = V_y^2 + V_\theta^2 \quad \text{Eq 5-10}$$

$$V_r^2 = V_y^2 + (U - V_\theta)^2 \quad \text{Eq 5-11}$$

$$V_\theta^2 - (U - V_\theta)^2 = u^2 - V_r^2 \quad (\text{Eq 5-10}) - (\text{Eq 5-11})$$

$$\text{Eq 5-12}$$

$$UV_\theta = \frac{1}{2}(u^2 + U^2 - V_r^2) \quad \text{Eq 5-13}$$

$$h_1 + \frac{V_{r1}^2}{2} - \frac{U_1^2}{2} = h_2 + \frac{V_{r2}^2}{2} - \frac{U_2^2}{2} = I \quad \text{Substitute Eq 5-13 into Eq 5-9}$$

$$\text{Eq 5-14}$$

5.3 Zone Flow Analysis

Since three-dimensional motion involves high complexity and higher order equations, simplified one-dimensional flow analysis approach is considered in this study which has enough reasonable approximation [39]. Other dimensional flow details can be covered later by comparing one-dimensional model with actual data collected from site to calibrate one-dimensional models and reduce uncertainties. The main assumption for one-dimensional approach is that fluid is uniform over flow cross-sections and can be presented with one vector at centre of cross-sections. Also, the study fluid in this research is air. According to Walsh P; Philip; Fletcher, Paul [44], air may be considered as an ideal gas without compromising calculation accuracy, except when the mass fraction of water vapour is more than 10%. The specific water vapor content at 100% humidity and 36 °C is 4%, which decreases as the temperature drops. The critical pressures and temperatures at which air ceases to act as an ideal gas are 37.7 bar and -140 °C, respectively. It is evident that these conditions will not be achieved by the compressor being studied, and the specific humidity will not exceed 10% in any situation (except during cleaning cycles). Therefore, although it is appropriate to apply the ideal gas equations for the purposes of this thesis, compressibility factors will also be utilized.

$$Z_c = \frac{P}{\rho RT} \quad \text{Eq 5-15}$$

Tracking fluid flow from compressor entrance to exit in each stage, the fluid passes through mainly three channels [39]; Inlet nozzle, impellers, diffuser. In each of these channels, fluid parameters changes. These channels can be classified into 2 main categories, as can be seen

in figure 3.5, based on the nature of thermodynamics changes in them: Ducts, Stators and Rotors.

- Ducts (nozzle, diffuser): vaneless zones, adiabatic, no work is transferred.
- Rotors (Impellers): vaned zone, adiabatic, work is transferred.

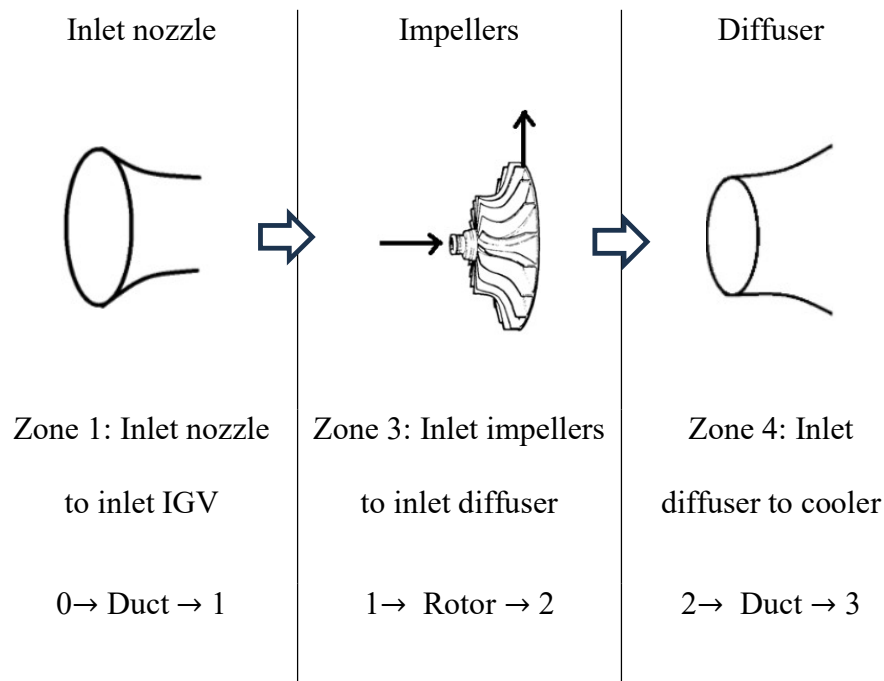


Figure 5.3: Compressors Modelling Zones

In duct, there is no relative velocity V_r since there is no applied work by rotating impeller.

Fluid flow will be considered, and stagnation enthalpy is constant in inlet and outlet.

$$h + \frac{1}{2}u^2 = \text{constant} \quad \text{Eq 5-16}$$

In rotor, since there is applied work by rotating impeller, there is a relative velocity.

Rothalpy preservation is applied

$$h + \frac{1}{2}u^2 - UV_\theta = \text{constant} \quad \text{Eq 5-17}$$

5.4 Effect of Circulatory Flow (Slip)

The fluid inside impeller has circulation flow opposite to impeller rotation. This happens because of inertia effect. As a result, actual fluid outlet velocity u'_2 and resultant flow V'_2 will not be same as calculated theoretical velocities due to reduce of flow outlet angle β'_2 . Let V'_2 be actual velocity and V_2 be calculated one then the slip velocity V_s can be found as:

$$V_s = V_2 - V'_2 \quad \text{Eq 5-18}$$

Slip factor is the ratio of actual by calculated velocity

$$S_f = \frac{V'_2}{V_2} \quad \text{Eq 5-19}$$

Outlet velocity triangle then can be modified by considering circulatory effect. figure (5.4) shows difference between ideal outlet velocity triangle (solid line) and actual one (dashed line).

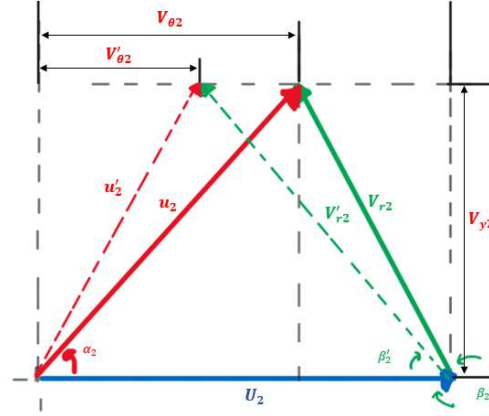


Figure 5.4: Impeller Outlet Velocity Triangle With Slip

A number of experiments were done by different researchers to estimate slip factor. Based on the outlet blade angle range three main estimations will be considered as following:

$$\text{Stodola [45]} \quad S_f = 1 - \frac{(\pi/Z_b) \sin \beta_2}{1 - \left(\frac{V_{r2}}{U_2}\right)} \quad \beta_2 = (20 - 30) \quad \text{Eq 5-20}$$

$$\text{Busemann [45]} \quad S_f = \frac{A - B \left(\frac{V_{r2}}{U_2}\right) \cot \beta_2}{1 - \left(\frac{V_{r2}}{U_2}\right) \cot \beta_2} \quad \beta_2 = (30 - 80) \quad \text{Eq 5-21}$$

$$\text{Staniz [45]} \quad S_f = 1 - \frac{0.63 \left(\frac{\pi}{Z_b}\right)}{1 - \left(\frac{V_{r2}}{U_2}\right) \cot \beta_2} \quad \beta_2 = (80 - 90) \quad \text{Eq 5-22}$$

Slip velocity can be estimated from Staniz (1952) [39]. He applied relaxation methods in calculating slip velocity. His experimental result was slip velocity is independent of outlet impeller angle and depends only on number of blades as:

$$V_s = 0.63U_2 \left(\frac{\pi}{Z_b} \right)$$

Eq 5-23

5.5 Compressors Isentropic and Polytropic efficiencies

Compressors are designed to convert mechanical energy to fluid which is delivered through coupling and shaft. So, efficiency is a question of whether the mechanical energy fully converted to fluid or not. The isentropic efficiency of compressors is defined as a ratio of useful energy input to fluid energy and power input to rotor. It is important to consider losses due to coupling, bearings, and other rotating parts. Electrical Power supplied to motor is not the same as power feed to shaft. Mechanical efficiency is a good indicator to measure these kinds of losses which is ratio of real power input to coupling of shaft and motor power from electrical source. From equation 3-13, Useful work done on the fluid in compressor can be ideally estimated with constant entropy and assumption of steady flow, reversable and adiabatic process as the difference of outlet isentropic stagnation enthalpy inlet stagnation enthalpy. Actual adiabatic work input to rotor in other hand can be calculated with same equation calculating the difference between the output and inlet actual stagnation.

$$\eta_c = \frac{h_{02s} - h_{01}}{h_{02} - h_{01}} = \frac{W_{ideal}}{W_{actual}}$$

Eq 5-24

Compressors with multi-stages have different pressure ratio in each stage. Isentropic efficiency in this case can be misleading if used to compare efficiencies of different stage. Consider a multi-stage compressor with same isentropic efficiency in each stage. Isentropic of whole machine will be different than the sum of stages' efficiencies. For that reason,

Polytropic efficiency is used to find total efficiency which is sum of each isentropic stage efficiencies.

$$\eta_p = \frac{\{(h_{0xs} - h_{01}) + (h_{0ys} - h_{0x}) + \dots\}}{h_{02} - h_{01}} \quad \text{Eq 5-25}$$

$$\eta_p > \eta_c$$

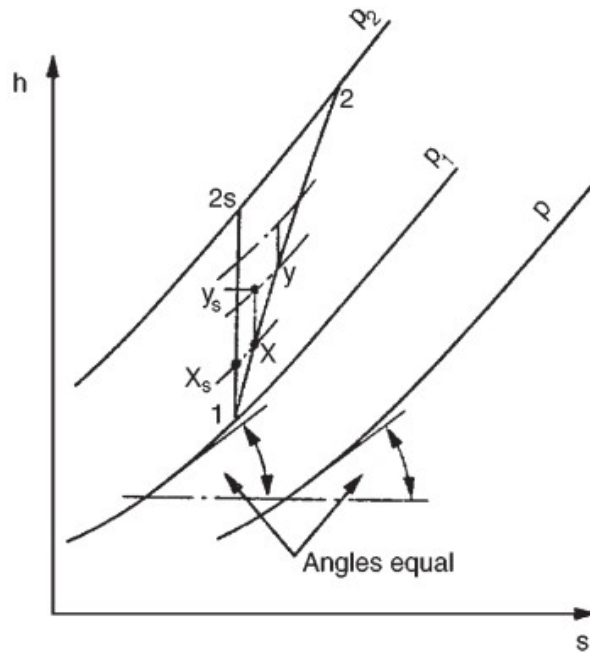


Figure 5.5 [39]: Compressors Isentropic And Polytropic Efficiencies

Efficiency equation can be rewritten in terms of stagnation pressure or temperature. As they are main contribution to enthalpy changes. Referring to equations 3-14 and 3-34, ideal work with no loss can be written as:

$$W_{ideal} = C_p(T_{02i} - T_{01}) \quad \text{Eq 5-26}$$

$$W_{ideal} = C_p T_{01} \left(\frac{T_{02i}}{T_{01}} - 1 \right)$$

Dividing by T_{01}

Eq 5-27

$$W_{ideal} = C_p T_{01} \left[\left(\frac{P_{02i}}{P_{01}} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right]$$

Eq 5-28

$$knowing \frac{P_2}{P_1} = \left(\frac{T_2}{T_1} \right)^{\frac{\gamma}{\gamma-1}}$$

Actual input work on the other hand will have higher value than ideal work because of slip and other friction losses and can be calculated knowing blade and fluid velocities as presented in equation 3-9. In most compressors and in normal operation fluid enters compressor axial in Y direction with minimal whirl angle. It has almost 90 degrees to blade velocity. In other words, $V_{\theta 1}$ is small and negligible. For that reason, equation 3-9 can be reduced to:

$$w = U_2 V_{\theta 2}$$

Eq 5-29

Also, the outlet flow velocity angle will not have the same outlet blade angle due to circulatory flow effect which is called slip. Actual angle is less than ideal one. Slip factor S_f need to be considered. Then actual work can be estimated as:

$$w_{actual} = S_f U_2 V_{\theta 2}$$

Eq 5-30

Putting these equations 5-28 and 5-30 together to find efficiency:

$$\eta_c = \frac{W_{ideal}}{W_{actual}} = \frac{C_p T_{01} \left[\left(\frac{P_{03}}{P_{01}} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right]}{S_f U_2 V_{\theta 2}}$$

Eq 5-31

For given efficiency stagnation pressure ratio can be found by rearranging this equation as:

$$\frac{P_{03}}{P_{01}} = \left(1 + \frac{\eta_c S_f U_2 V_{\theta 2}}{C_p T_{01}} \right)^{\frac{\gamma}{\gamma-1}} \quad \text{Eq 5-32}$$

5.6 Performance Models of Centrifugal Compressors

Centrifugal compressors in industry are used to raise fluid pressure and mass flow rate. So, the main parameters to assess performance for compressors are outlet pressure ***P*** and flow rate ***m***. Pressure and flow rate are raised due to centrifugal force from impellers' kinetic energy which is driven by rotating motor shaft. Geometrical details play a role in blade centrifugal force which affect fluid speed and pressure. Not to mention, fluid physical properties also contribute on performance. Since losses are unavoidable, work input from motor will not be the same as output. Initial or static thermodynamic parameters need to be considered as well. To put this in a few words, contribution factors for compressor performance are:

- Speed and Power of motor shaft ***ω, Power***
- Designed parameters: Impeller radii ***r*** and blade angles ***β*** .
- Fluid Parameters: density ***ρ*** and constant ***R, C_p***.
- Thermodynamic variables: enthalpies which include inlet temperature and pressure; ***h*** .

$$P_{out}, T_{out}, u_{out} = f \left(\underbrace{h_{in}, \dot{m}}_{\substack{\text{Thermodynamics} \\ \text{Variables}}} , \underbrace{\omega, r, \beta, Power}_{\text{Design Parameters}} , \underbrace{\rho, R, C_p}_{\text{Fluid Constant}} \right) \quad \text{Eq 5-33}$$

Choosing compressor as a closed system, Inlet pressure and temperature to compressor are considered as input and uncontrollable variable which usually come from surroundings or

from other processes. Impellers' geometrical details and motor shaft speed are both fixed design inputs. Outlet temperature, pressure and outlet flow rate are considered as output variable that should be calculated from the main given input parameters. Since design is dimensional and fixed for each machine, variable numbers are reduced by normalizing and creating dimensionless variable.

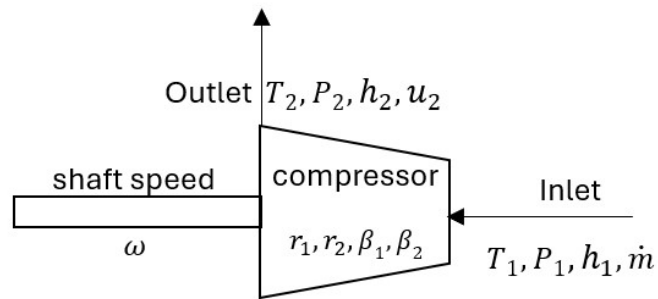
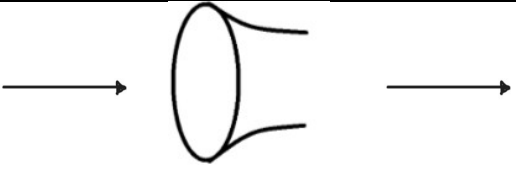


Figure 5.6: Centrifugal Compressors Modelling: Output And Input

Zone 1: (0-1) Inlet Nozzle (Duct)

As demonstrated in Figure 5.6, the fluid enters the compressor from the inlet casing to the impeller eye, or as sometimes called the inducer section. The inlet casing in most designs has a nozzle shape which helps to increase the velocity of the fluid by reducing the cross-sectional area. No work is done. The inlet measured, given, parameters at point 0 and outlet need to be calculated at point 1.

Table 5.1: Zone 1 (0-1) Inlet Nozzle (Duct) tag mapping

Input	0	Zone 1	1	Output
Measured: $T_0, P_0, h_0, \dot{m}, \rho_0$				Calculated: T_1, P_1, h_1, u_1, u_0
	Geometrical Dimension r_0, r_1			
	Constant R, C_p, γ			

$$\rho_0 = \frac{Z_0 P_0}{RT_0}$$

$$u_0 = \frac{\dot{m}}{\rho_0 \pi r_0^2}$$

$$u_1 = \frac{u_0 r_0}{r_1}$$

$$C_p(T_0 - T_{ref}) + \frac{1}{2}u_0^2 = C_p(T_1 - T_{ref}) + \frac{1}{2}u_1^2$$

using Eq 32, Find ρ_0

Eq 5-34

using Eq 4, Find u_0

velocity inter in Y direction

Eq 5-35

using Eq 7 Find u_1

Eq 5-36

using Eq 33, find T_1

knowing $h = C_p dt$

Eq 5-37

Final Combined equation for Temperature and pressure:

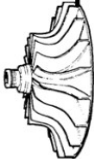
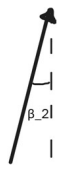
$$T_1 = T_0 + \frac{1}{2C_p}u_0^2 - \frac{1}{2C_p}u_1^2$$

Eq 5-38

Zone 2: (1-2) Impeller (Rotor)

Then, the fluid goes to the impeller eye which is located at the impeller center. In this channel, the fluid energy increases due to kinetic energy transfer through angular momentum of whirling impeller. As a result, both pressure and velocity increases. In this zone, there is applied work to fluid from rotating impeller which is rotated by electrical motor.

Table 5.2: Zone 2: (1-2) Impeller (Rotor) Tag Mapping

Input	1	Zone 3	2	Output
T_1, P_1, h_1, u_1				T_2, P_2, h_2, u_2
	Geometrical Dimension $\beta_1, r_1, r_{1h}, b_2, r_2, \beta_2$ Constant $R, C_p, \gamma,$			

$$V_{ry1} = u_1$$

Eq 5-39

$$V_{r\theta1} = U_1 = \omega r_1 = \frac{2\pi N r_1}{60}$$

Eq 5-40

$$V_{r1} = \sqrt{U_2^2 + u_2^2}$$

Eq 5-41

$$V_{\theta1} = 0$$

Eq 5-42

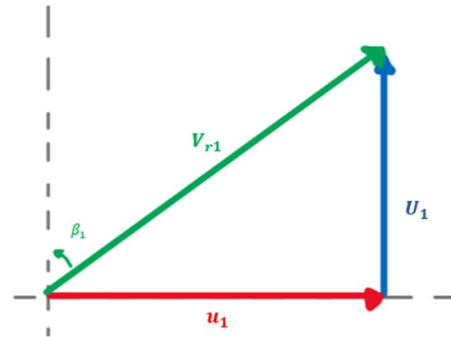


Figure 5.7: Inlet Velocity Triangle

$$U_2 = \omega r_2 = \frac{2\pi N r_2}{60} \quad \text{Eq 5-43}$$

$$V_{\theta 2} = \frac{\text{Power}}{\dot{m} U_2} \quad \text{Eq 5-44}$$

$$V_{r2} = \frac{U_2 - V_{\theta 2}}{\cos \beta_2} \quad \text{Eq 5-45}$$

$$V_{ry2} = \frac{u_2 - V_{\theta 2}}{\cot \beta_2} \quad \text{Eq 5-46}$$

$$u_2 = \sqrt{V_{\theta 2}^2 + V_{ry2}^2} \quad \text{Eq 5-47}$$

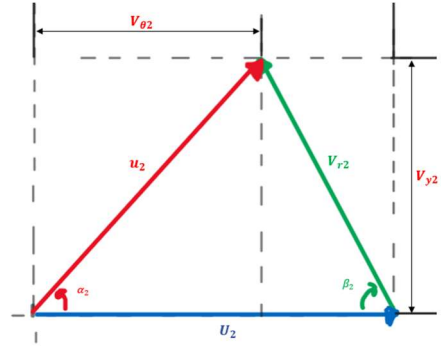


Figure 5.8: Outlet Velocity Triangle

using Eq 34, find T_2 knowing $h = C_p dt$

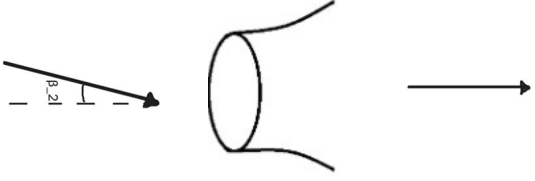
$$C_p(T_1 - T_{ref}) + \frac{1}{2}u_1^2 - U_1 V_{\theta 1} = C_p(T_2 - T_{ref}) + \frac{1}{2}u_2^2 - U_2 V_{\theta 2} \quad \text{Eq 5-48}$$

Final Combined equation for Temperature:

$$T_2 = T_0 + \frac{1}{2c_p} u_0^2 - \frac{1}{2c_p} u_2^2 + \frac{1}{2c_p} U_2 V_{\theta 2} \quad \text{Eq 5-49}$$

Zone 3: (2-3) Diffuser (Stator)

Table 5.3: Zone 3: (2-3) Diffuser (Stator) Tag Mapping

Input	2	Zone 3	3	Output
$T_2, P_2, h_2, u_2, \beta_2$				T_3, P_3, h_3, u_3
	Constant $R, C_p, \gamma, r_2, b_2, r_3, b_3$			

$$u_3 = \frac{u_2 b_2}{b_3}$$

Eq 5-50

using Eq 7 Find u_3

$$C_p(T_2 - T_{ref}) + \frac{1}{2}u_2^2 = C_p(T_3 - T_{ref}) + \frac{1}{2}u_3^2$$

Eq 5-51

using Eq 33, find T_1

knowing $h = C_p dt$

Final Combined equation for Temperature:

$$T_3 = T_0 + \frac{1}{2C_p}u_0^2 - \frac{1}{2C_p}u_3^2 + \frac{1}{2C_p}U_2V_{\theta 2}$$

Eq 5-52

outlet pressure P_3 is from compressor efficiency equation 38 knowing η_c

$$P_3 = \left(1 + \frac{\eta_c S_f U_2 V_{\theta 2}}{C_p T_{01}}\right)^{\frac{\gamma}{\gamma-1}} * P_0$$

Eq 5-53

5.7 Result and Discussion of Theoretical Models

Theoretical models are built based on ideal assumptions to simplify equations. They assume no heat loss during compression. So, comparing calculated and actual [42], errors are expected. In addition to RMSE and MSE equation 2-1, 2-2, error and effectiveness of these models are measured using following equations.

Mean Absolute Error (MAE)

$$\frac{1}{n} \sum_{i=1}^n |y_i - y'_i|$$

Eq 5-54

Goodness of fit (R^2)

$$R^2 = \frac{SST - SSE}{SST}$$

Eq 5-55

$$\text{Sum of Squared Errors (SSE)} \quad \sum_{i=1}^n (y_i - y'_i)^2 \quad \text{Eq 5-56}$$

$$\text{Sum of Squared Total (SST)} \quad \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{Eq 5-57}$$

y_i Actual reading [42]

y'_i Predicted or Calculated reading

$\bar{y}$ Average Calculated reading

Calculated output was found from 290000 input design variants [42]. The results are as follows:

Table 5.4: theoretical model scores

Model Score Calculated Theoretical	RMSE	MSE	MAE	R^2
T_3 (K)	12.202	148.905	10.594	0.9711
P_3 (kPa)	12.851	165.145	8.26	0.9970
PR	0.0509	0.0026	0.0520	0.7986

Theoretical calculation method has a maximum of RMSE 12.8 kPa errors. It estimates outlet temperature with an average difference of 10.594 K, outlet pressure with 8.26 kPa and Pressure ratio with 0.0520. This is clearer graphically in figures 5.9 to 5.11. Result value plotted verses actual one and then a line is fit. Residual shows how far calculated from actual. As can be seen, calculated model works better in low inlet temperature and pressure. Higher residual is clear in high temperature and pressure.

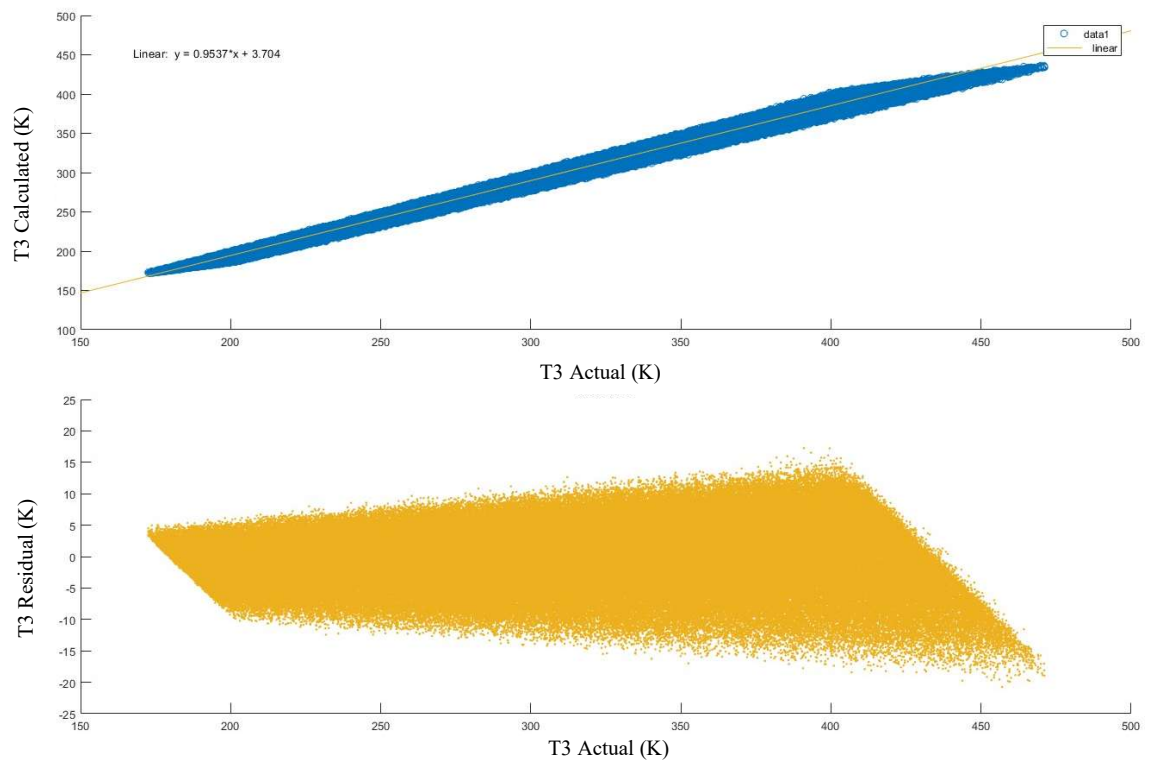


Figure 5.9: Actual Outlet Temperature T_3 Versus Calculated

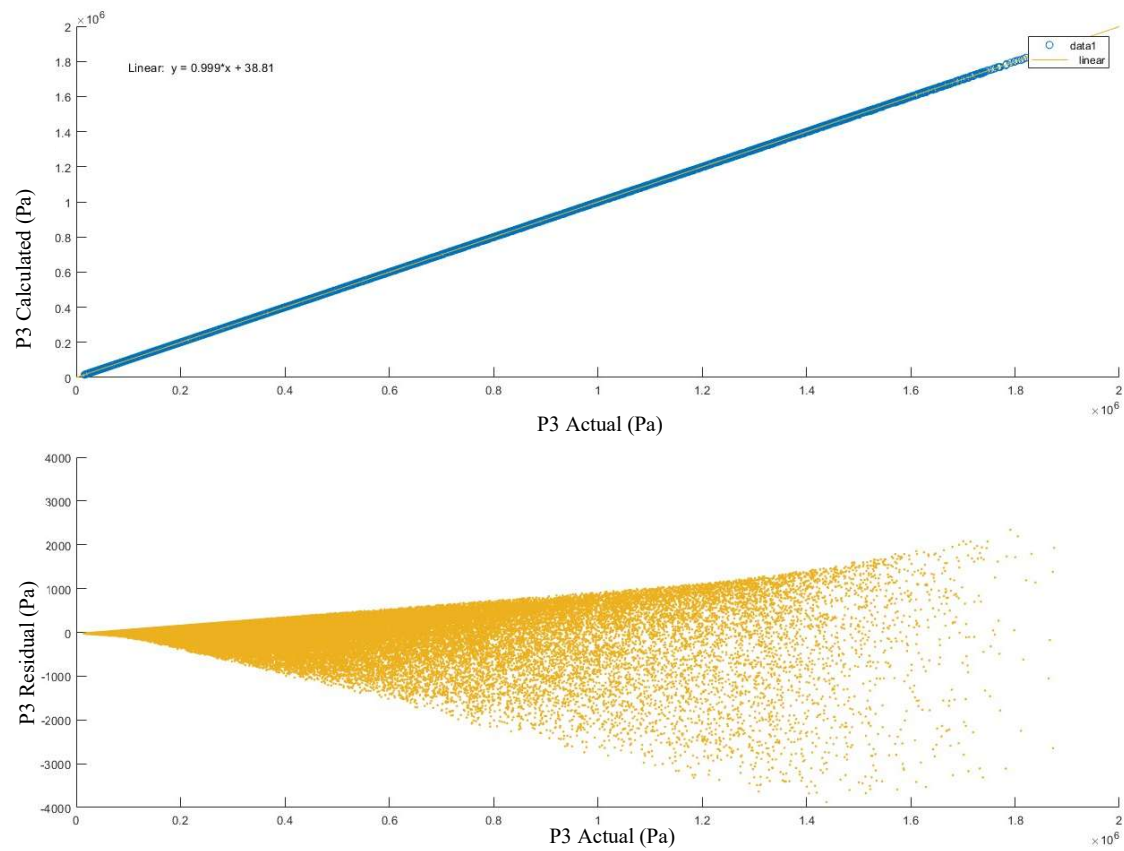


Figure 5.10: Actual Outlet Pressure P_3 Versus Calculated

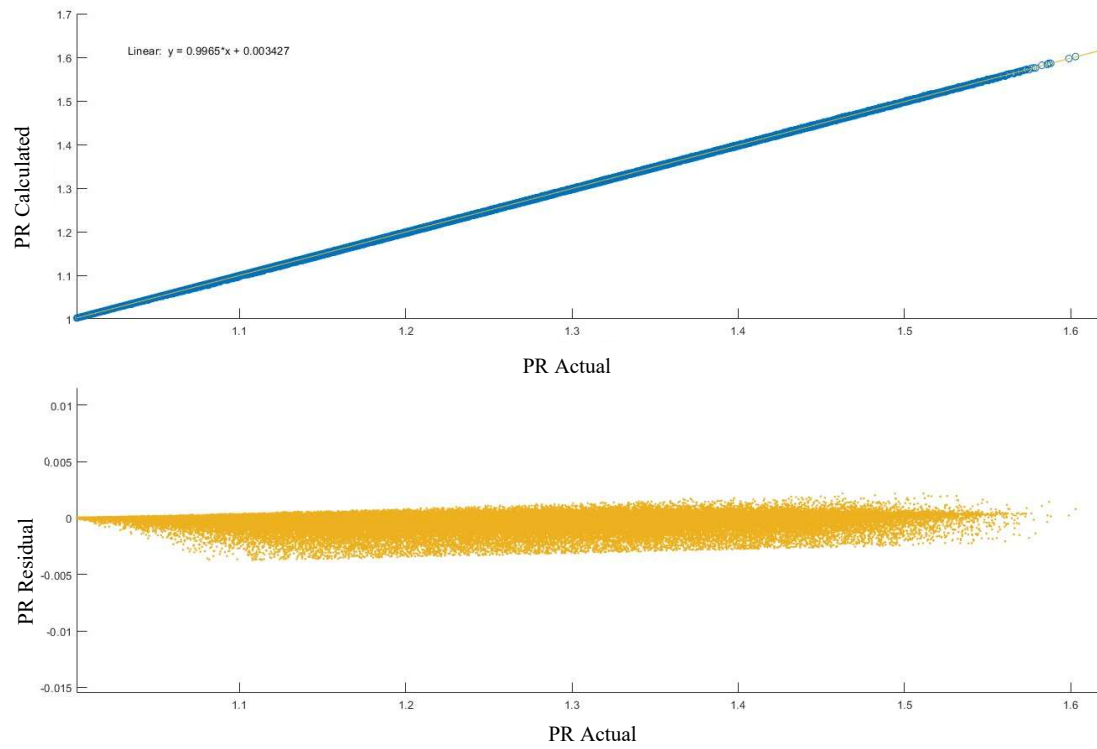


Figure 5.11: Actual Outlet Pressure *PR* Versus Calculated

CHAPTER 6

Artificial intelligence; Machine Learning Performance Modelling

Artificial intelligence is a tool that has been developed within time to make computer thinks, analyses, solves and learns as a human base on experiences. Nowadays with cost effective and increasing processing power, availability of data and cloud storing, artificial intelligence has become more important than ever in all fields. It helps to save time and cost to take decisions based on real data with minimal risk. Before, due to the high cost of processing and lack of data, people rely on conclusion base on limited data from experiments which not necessarily include all possible factors and could have a high number of errors. Artificial intelligence has its role in operation and maintenance decisions. With massive amounts of historical data collected within years, it helps take decisions whether to stop for maintenance or keep running. This chapter describes Artificial Intelligence modelling methodology, used regression models in the thesis and result testing and validation.

6.1 Modeling methodology

This research suggests supervised machine learning methods to predict rate of degradation based on collected data from installed sensors on machine. As proved in the last section, the ideal theoretical models are not accurate enough to predict degradation rate. Also, building theoretical real models which include all possible losses is complicated, require experts and often needs a lot of detailed information which usually is not available and includes high cost. To overcome these issues, machine learning could find relation with available upstream sensors data (input) to downstream data (output) even with limited

details and knowledge. No expertise is required, from data computer derive model to help predict output.

Supervised machine learning techniques map input to output based on provided example of input-output [46]. There are two types of supervised learning regression and classification. Regression models provide continues outcome while classification discreate outcome. Since we are dealing with continue numbers and expecting continuous result regression is the right choice. There are mainly 4 steps to build model [46]; data cleaning, data categorizing and selection, model training, model validation and testing.

6.1.1. Data cleaning

Data collected from sensors includes unwanted readings such as shutdown periods, unstable operation, noises, sensor malfunction. Before building the model on this data it needs to be removed because it could lead to misleading results. From Prior knowledge from operation or from trending data itself, shutdown period can be filtered out since during shutdown there is no flow and outlet pressure and temperature will be close to inlet once. The common way to filter out Noises and instability is by plotting frequency distribution histogram. Any readings away from normal 99% distribution are considered as outlier and need to be removed.

In this research, there are no noises nor sensor malfunction, but data need to be inspected to see if it is making sense within boundaries and not contradict with design limitation. Since data is generated and synthetic, output data needs to be verified as well to be within range and close to realistic.

6.1.2. Data categorizing and selection

Data Categorizing means identifying names, types and units of each reading from a data set. Naming each variable should easily lead to location of collection point and type of reading. For example, T3 reads temperature at location 3, outlet. Knowing location and type of reading helps in the next step selecting contributing variables. Also, the type of reading needs to be imported to software even if the name of variable is clear enough. There are four main data types that can be used in model: numbers, Boolean, categorical and datetime. Numbers include all integers and decimal numbers which can have mathematical operation. Boolean has two options true or false, as valve open or not, machine on or off etc. Categorical similar to Boolean but has more than 2 options like; operational mode A, mode B or mode C. datetime as its name includes day, month, year, hour, minute and second of collection. It can be added or subtracted to find number of days, minutes, hours etc.

After identifying type, selection is the next step. It is Simply choosing which group of data are input variables X (predictors), which are output variables (outcome) Y, and which data are irrelevant. It is easy to know what we need to predict Y, but it might be challenging to know which predictor should be selected. Not to worry, first guess should be taken then when evaluating the model, it will be clear which variable has more contribution.

In this research, data was categorized during data generation. However, each variable needs to be re-named to match the naming convention as described in the theoretical section. With the help of some basic thermodynamic knowledge, the first guess of predictors can be made. All variables can be selected as predictors, but this will make model sensitive and lead to misleading prediction. Output variables are outlet temperature, outlet pressure and

pressure ratio. Main contribution input variables to these as reviewed in previous chapters are Power, Speed, mass flow rate, inlet and outlet radius and inlet and outlet blade angle.

6.1.3. Model Training and Validation

Model training means finding correlations between selected predictors and outcome. Predictors and outcome are known as collected from sensors should be fed to software to find relation. Training should not be done on all datasets. The dataset should be divided into three groups: data for training, validation and testing. Validation is re-training the trained model in a new set of data to avoid overfitting. Overfitting means model is trained in specific data which fit very well but fail to predict accurately in any other data. Validation could prevent this issue by Multiple data training.

In this research, an undegraded dataset was divided randomly equally into two groups 1450000 each. One group is used for training and validation while the other group is used for testing. Training and validation group of data then divided as well using built-in MATLAB cross-validation function with 5 folds. It is splitting data into 5 groups and each group trained and tested to other ones. The resulted 5 outcomes models are averaged to one model. Figure (6.1) shows an illustration example for 5 folds and 5 group data.

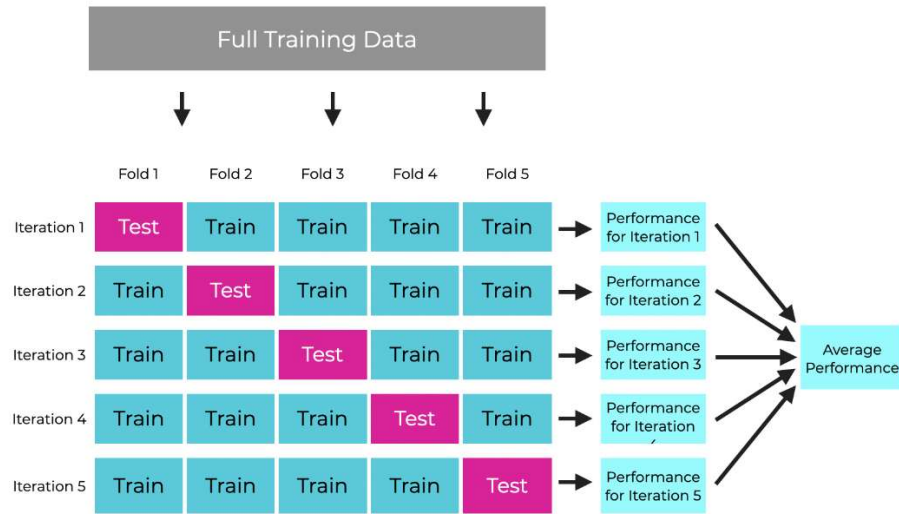


Figure 6.1: MATLAB Cross-Validation Function With 5 Folds And 5 Group

6.1.4. Model testing

Model testing is to evaluate prediction accuracy in new data, not training nor validation data. Unlike training and validation, outcome should not be included in processing. The model that was trained and validated should generate an outcome list from given predictors. Then predicted outcome is compared to actual to calculate residual and errors. If errors within acceptable limit model is published and used for degradation predictions. As presented in theoretical section RMSE, MSE, MAE and R^2 are the main indicators to evaluate model prediction accuracy.

6.2 Machine learning used algorithms

four mains commonly used supervised machine learning algorithms are investigated in this thesis. They are Built-in MATLAB regression learner app: linear regression, decision tree, random forest tree and neural network.

6.2.1. Linear Regression Model

Linear regression model fit a straight line among input or dependent variables [47]. This straight shows increasing or decreasing constant slope. If there is one predictor, the model is called linear, more than one is called multi-linear. Since linear regression is a mathematical base, it can be expressed as mathematical relation using constant coefficients β_n for each dependent variable where n is number of input dependent variables. y is output independent variable and X_n is input dependent variable.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad \text{Eq 6-1}$$

While linear relationship is easy to build and interpret mathematically, it fails to predict non-linear relationships

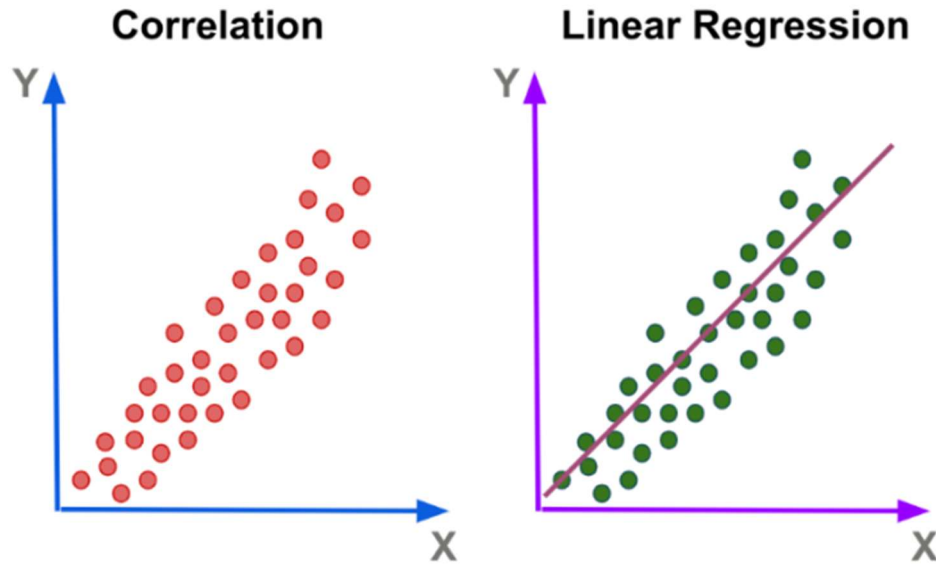


Figure 6.2 [37]: Linear Regression Model

6.2.2. Decision Tree Model

The basic algorithm of decision Tree model is separating data by asking questions [48]. Each question narrows possible outcomes until model get confident single prediction. These questions are limited to two answers, true or false. The Model is called tree because it is built in the form of tree with decision nodes and leaf nodes (Figure 6.3). Decision nodes have 2 or more branches. Each branch represents value for attribute test. Leaf nodes represent decision on numerical target. The coarse tree built-in algorithm has been used which has 36 leaf size.

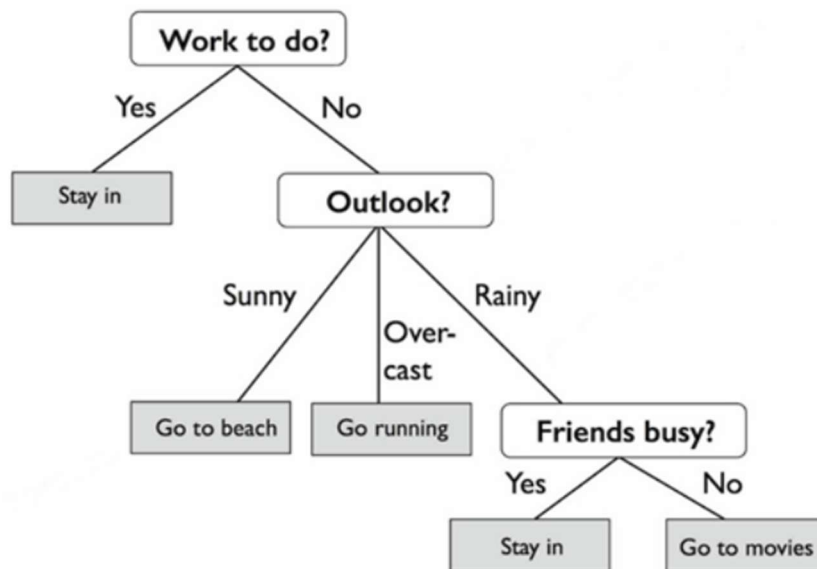


Figure 6.3 [47]: Decision Tree: Decision Nodes And Leaf Nodes

6.2.3. Random Forest Model

Random forest ensemble model type which uses decision trees algorithm but employs several decision trees to forecast the result rather than just one [48]. There are several output results when there are several decision trees. The average of the decision tree outputs in the random forest represents the actual predicted value. By generating random subsets of the features and using those selections to construct smaller trees, random forests avoid the tendency of decision trees as an algorithm to overfit the model. The structure of Random Forest is given in figure 6.4. Bagged trees with 8 leaf size and 30 learner trees are used to predict the outcome.

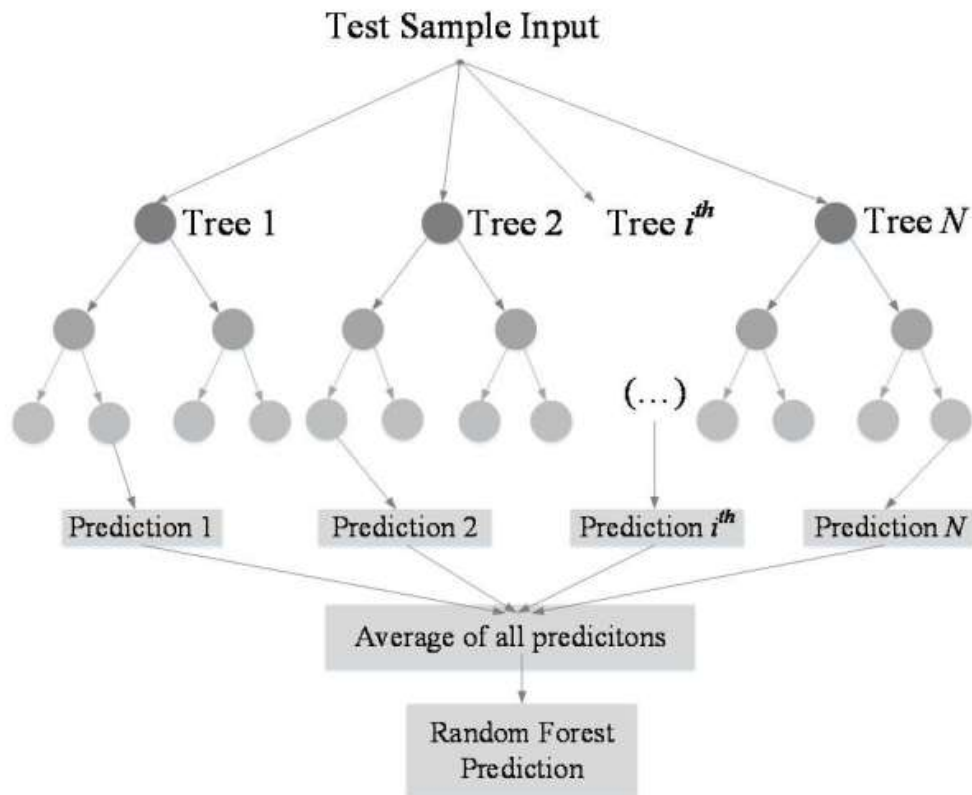


Figure 6.4 [38]: Random Forest Model Illustration

6.2.4. Neural Network Model

Neural network models mimic how humans are connected to neurons which are the fundamental structural tissues of the brain. Neurons in a mathematical model are connected to form a network known as a neural network. The fundamental tasks of neurons include accepting information, processing computations, producing information, and creating new neurons by joining several neurons together. Depending on their function, all of the neurons in the neural network model are separated into input, hidden, and output layers; each layer is functionally coupled. The input layer is sent in units of hidden layers based on the input unit and connects to the external input mode. In this case, the output layer is employed to create the output mode, while the hidden layer is the neural network's inner processing unit layer. The Bilayered Neural network model has been used with 2 fully connected layers. Both the first and second layer are 10 size. Iteration limit was set at 1000. More than these numbers can be considered as well however they need more computation time and difference in result accuracy is negligible Figure 6.5 demonstrates type of layer and how they are connected.

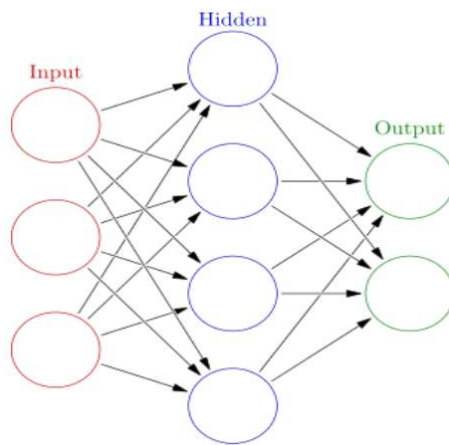


Figure 6.5 [38]: Neural Network Model Layers Connection

6.3 Result and Discussion of Artificial Intelligence models

Tables from 6.1 to 6.4 shows each model score, RMSE, MSE, MAE and R^2 . Figures 6.5 to 6.16 show the difference between the multiple regression model's predicted and actual values; outlet pressure, pressure ratio, outlet temperature. The Y-axis plots the predicted values produced by the regression model on the test set, while the X-axis plots the actual values from the test set. The model has never seen the data in the test set because it is not used for training. Thus, it can be said the best result with minimal error is random forest model which can predict outlet temperature with 2.36 K absolute mean difference, outlet pressure with 2.7 kPa means difference and pressure ratio of 0.0195 mean difference. While the score seems acceptable as an average value. Graphically it can be seen that there is a good amount of points out of mean value which reach to 40 K for outlet temperature, 6 kPa for outlet pressure and 0.3 for pressure ratio. These differences for random forest best fit model are still not acceptable and could lead to false conclusion.

While accuracy has been improved compared to thermodynamic calculation model, machine learning technique still not enough to capture degradation rate. According to Marta Zagorowska et al. [43] normalized degradation deviation from efficiency until maintenance shutdown is between 0.03 to 0.09. Similarly, the pressure ratio has a small residual range within 0.05. Errors from both theoretical thermodynamic calculation models and machine learning have a larger error range than degradation range. For that reason, more development is needed. The next section suggests hybrid method to improve accuracy.

1- Linear regression:

Table 6.1: Linear Regression model scores

Model Score	RMSE	MSE	MAE	R^2
Linear Regression				
T_3 (K)	9.673	93.585	7.605	0.9818
P_3 (KPa)	29.467	868.344	12.739	0.9843
PR	0.1030	0.0106	0.0829	0.1771

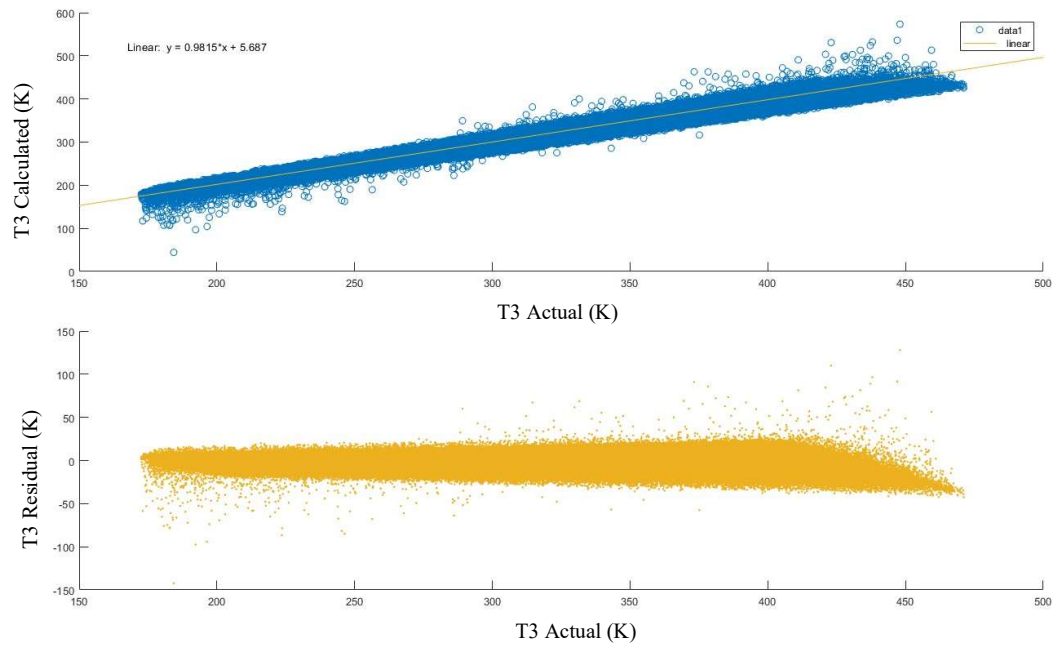


Figure 6.6: T3 Linear Model Predicted Versus Actual And Residual

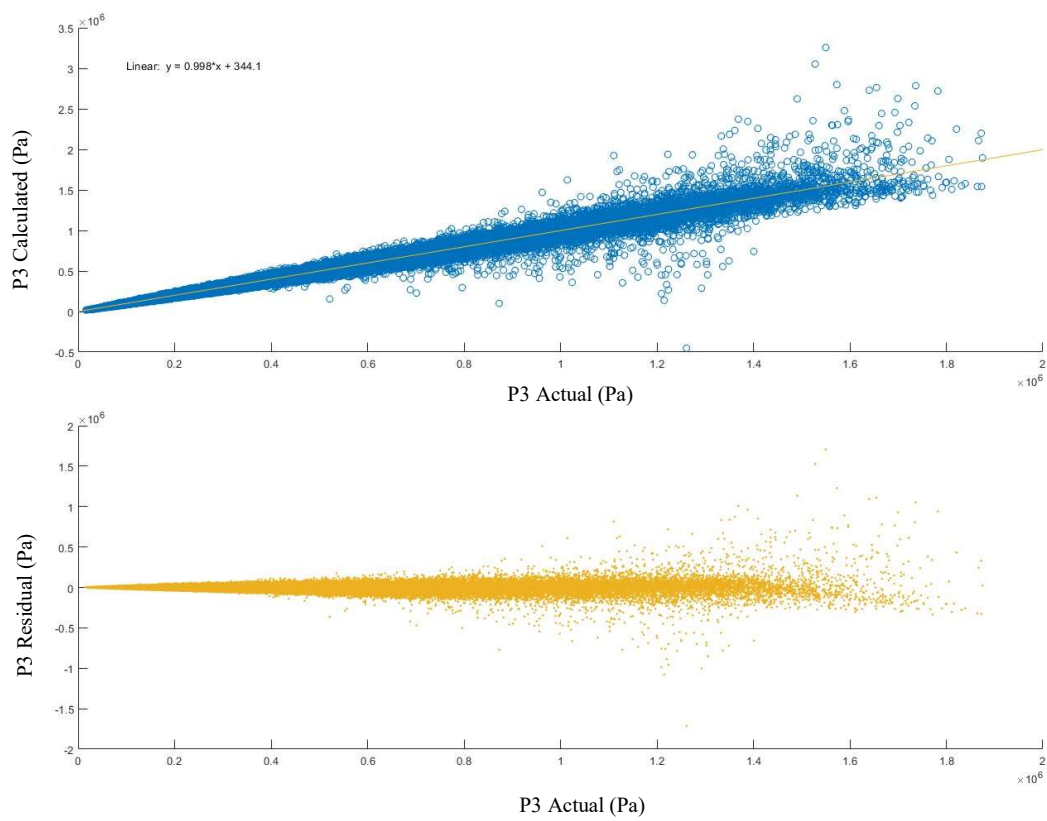


Figure 6.7: P3 Linear Model Predicted Versus Actual And Residual

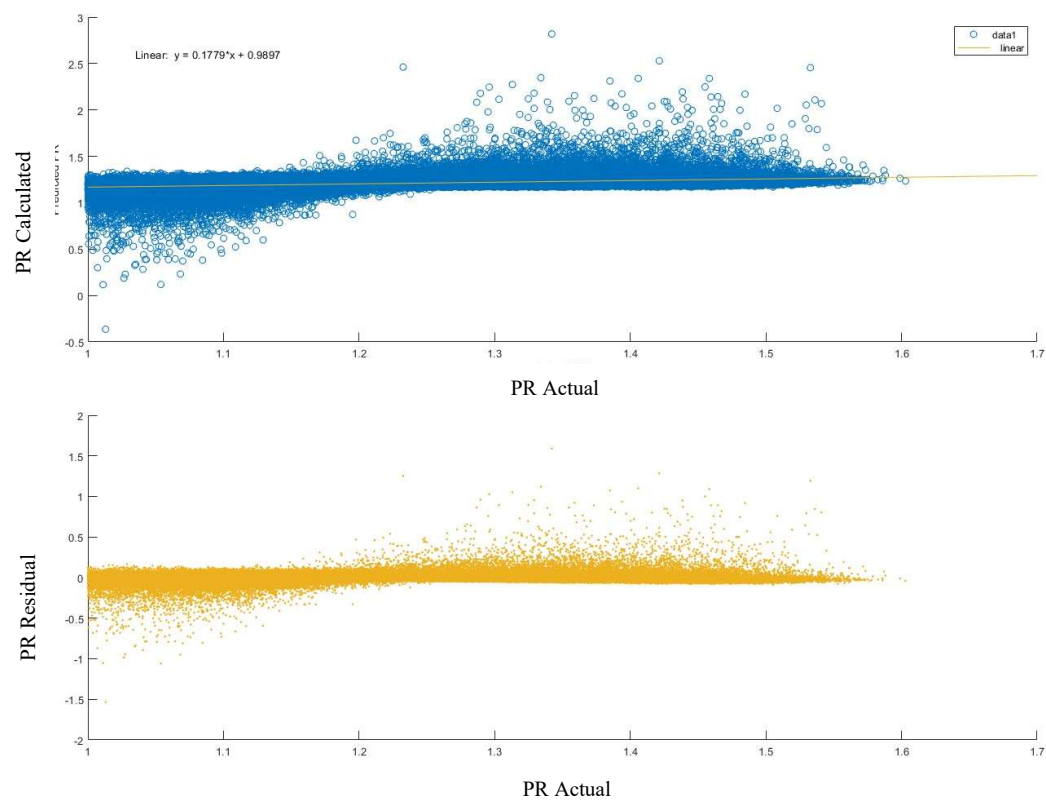


Figure 6.8: PR Linear Model Predicted Versus Actual And Residual

2- Decision Tree

Table 6.2: Decision Tree model scores

Model Score Decision Tree	RMSE	MSE	MAE	R^2
T_3 (K)	5.701	32.511	4.281	0.9937
P_3 (KPa)	9.799	96.034	4.402	0.9982
PR	0.0390	0.00152	0.0283	0.8815

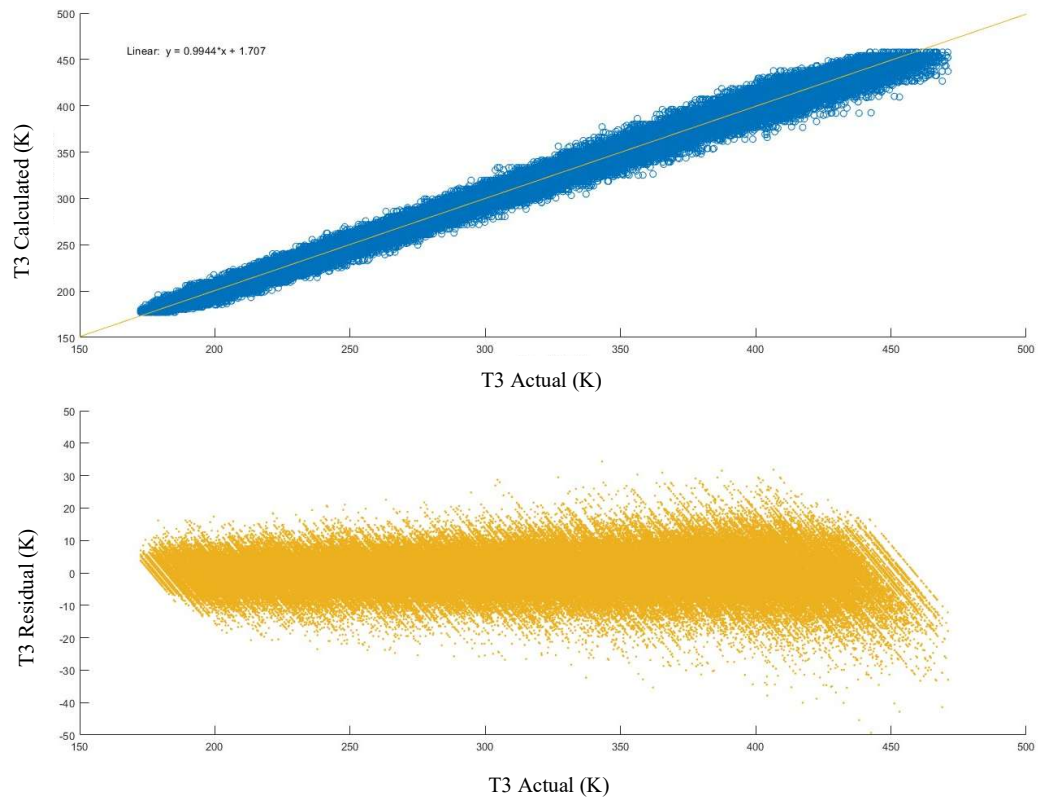


Figure 6.9: T3 Decision Tree Model Predicted Versus Actual And Residual

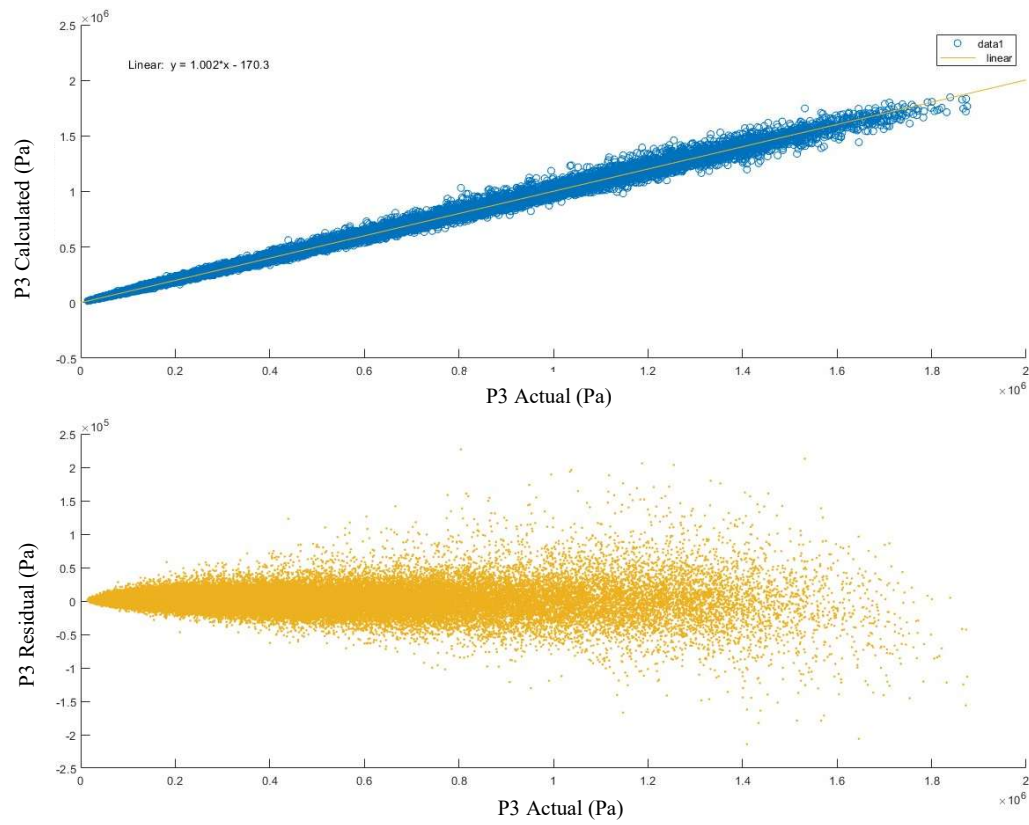


Figure 6.10: P3 Decision Tree Model Predicted Versus Actual And Residual

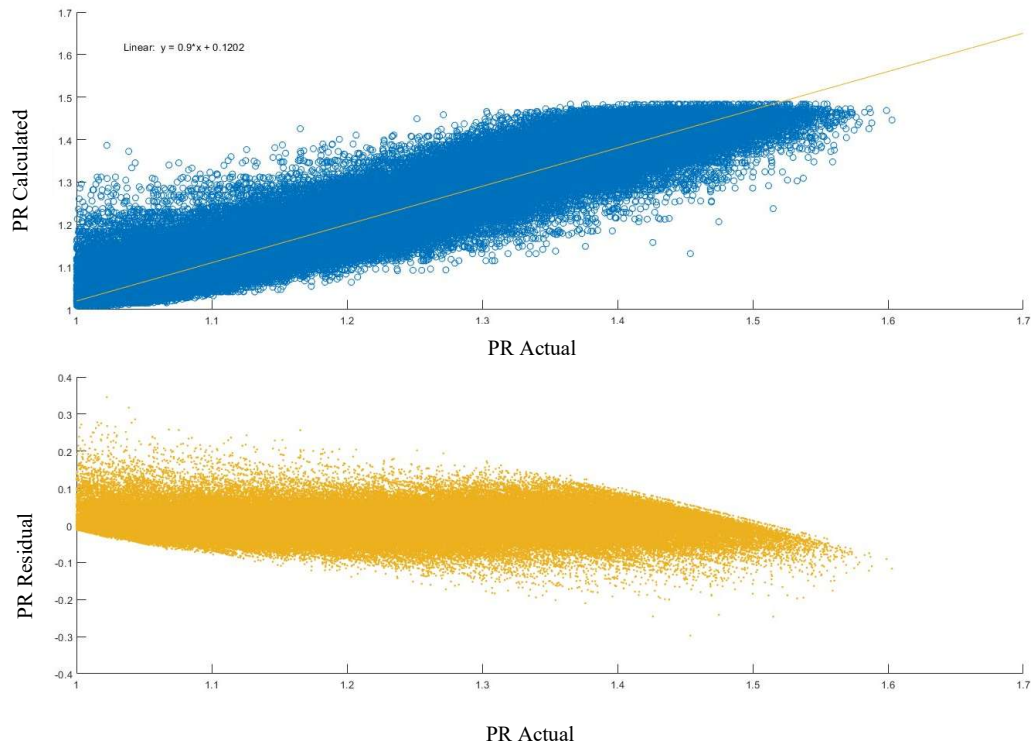


Figure 6.11: PR Decision Tree Model Predicted Versus Actual And Residual

3- Random Forest

Table 6.3: Random Forest model scores

Model Score Random Forest	RMSE	MSE	MAE	R^2
T_3 (K)	3.397	11.541	2.362	0.9977
P_3 (KPa)	6.427	41.312	2.701	0.9992
PR	0.0278	0.000776	0.0195	0.9398

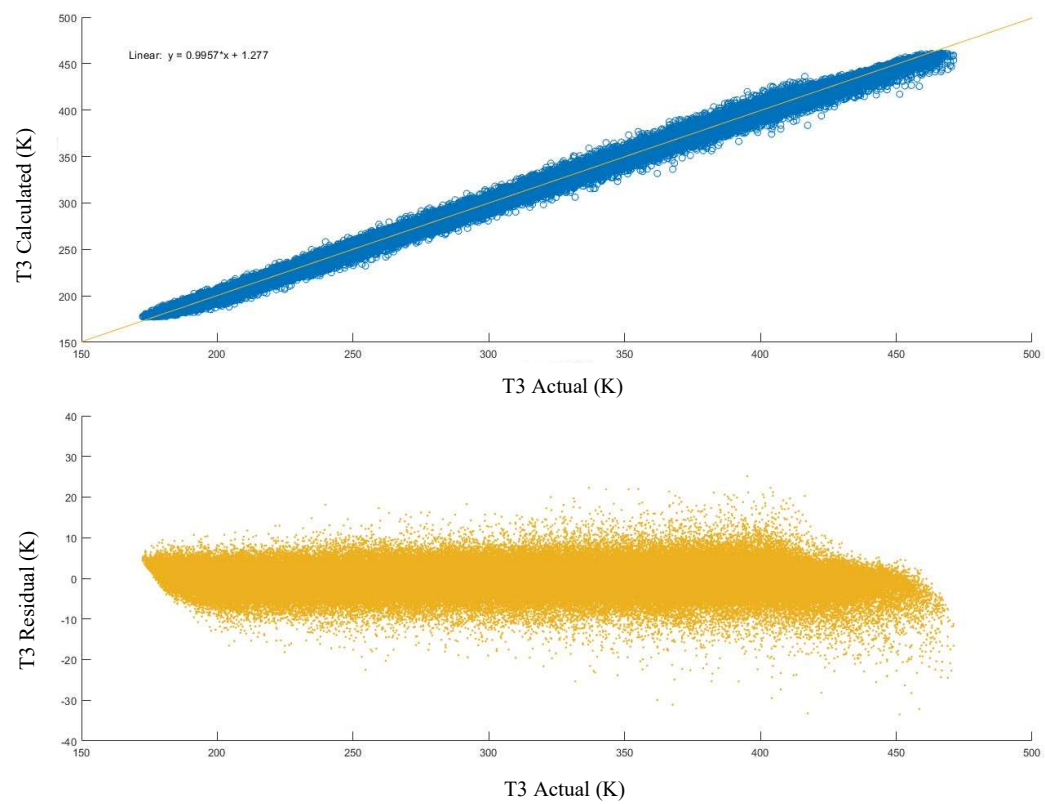


Figure 6.12: T3 Random Forest Model Predicted Versus Actual And Residual

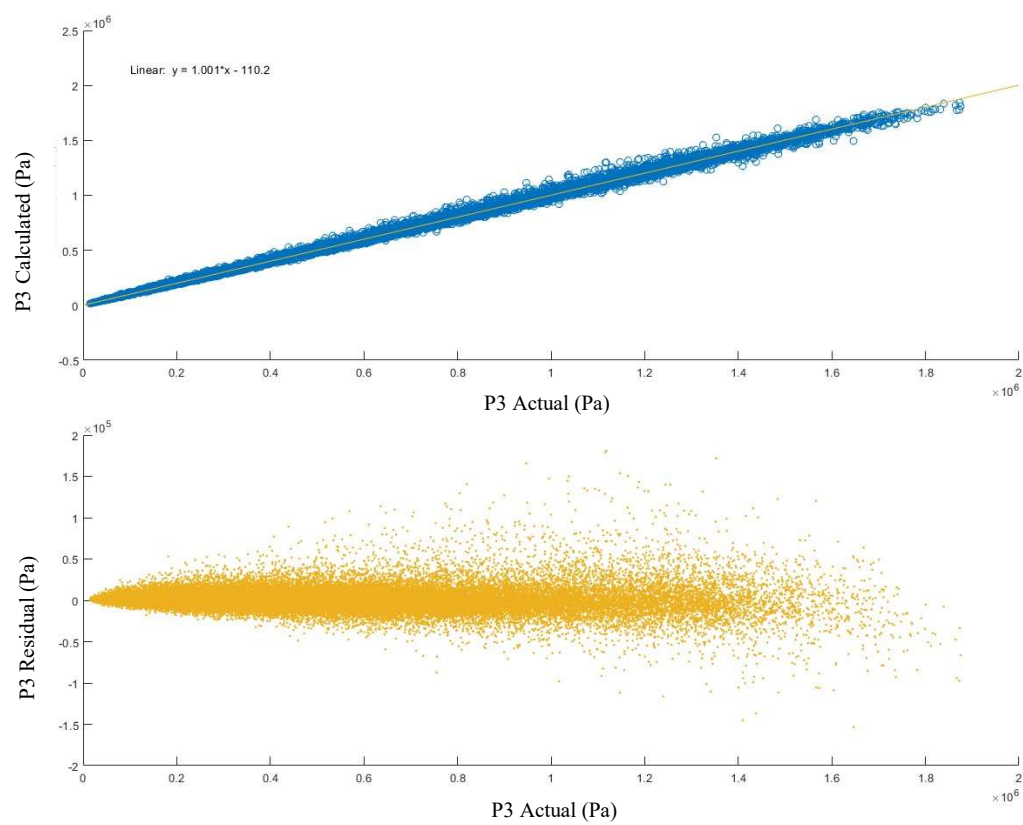


Figure 6.13: P3 Random Forest Model Predicted Versus Actual And Residual

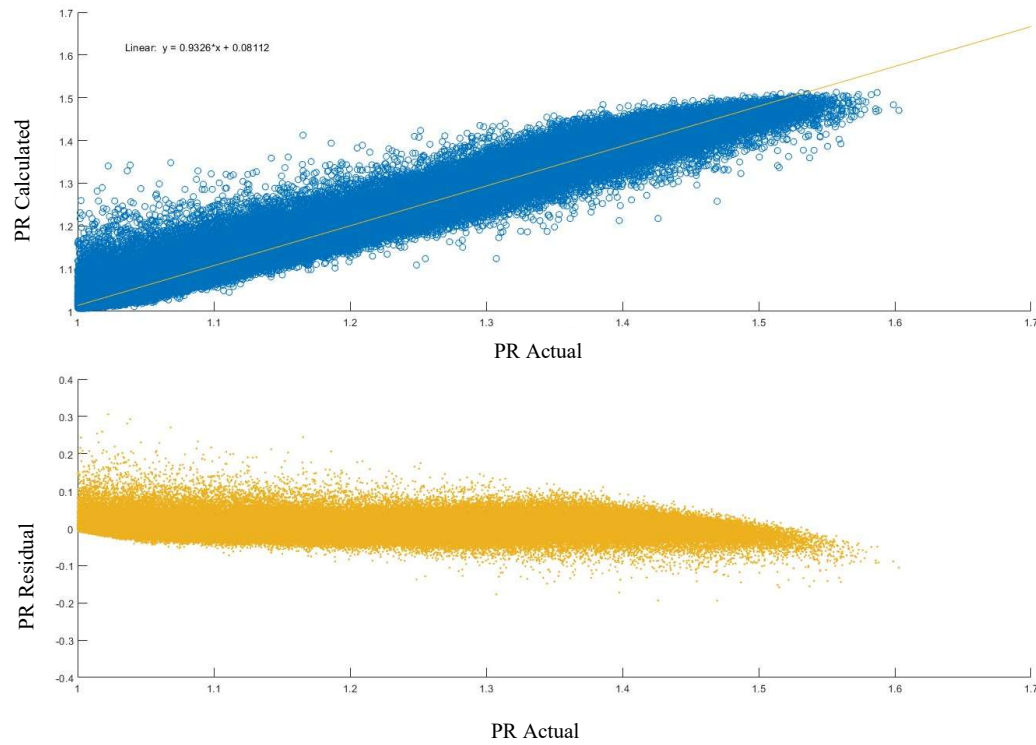


Figure 6.14: PR Random Forest Model Predicted Versus Actual And Residual

4- Neural Network

Table 6.4: Neural Network model scores

Model Score Neural Network	RMSE	MSE	MAE	R^2
T_3 (K)	4.229	17.884	3.075	0.9965
P_3 (KPa)	7.737	59.863	3.478	0.9989
PR	0.0295	0.000871	0.0209	0.9324

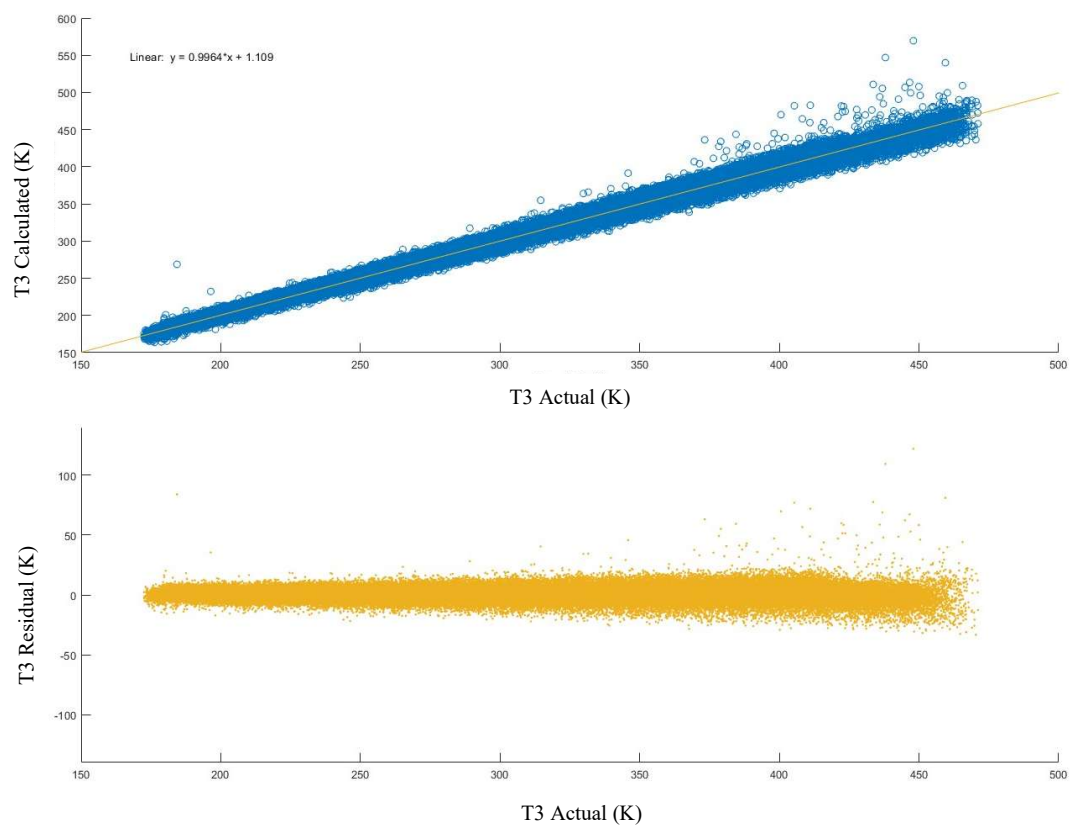


Figure 6.15: T3 Neural Network Model Predicted Versus Actual And Residual

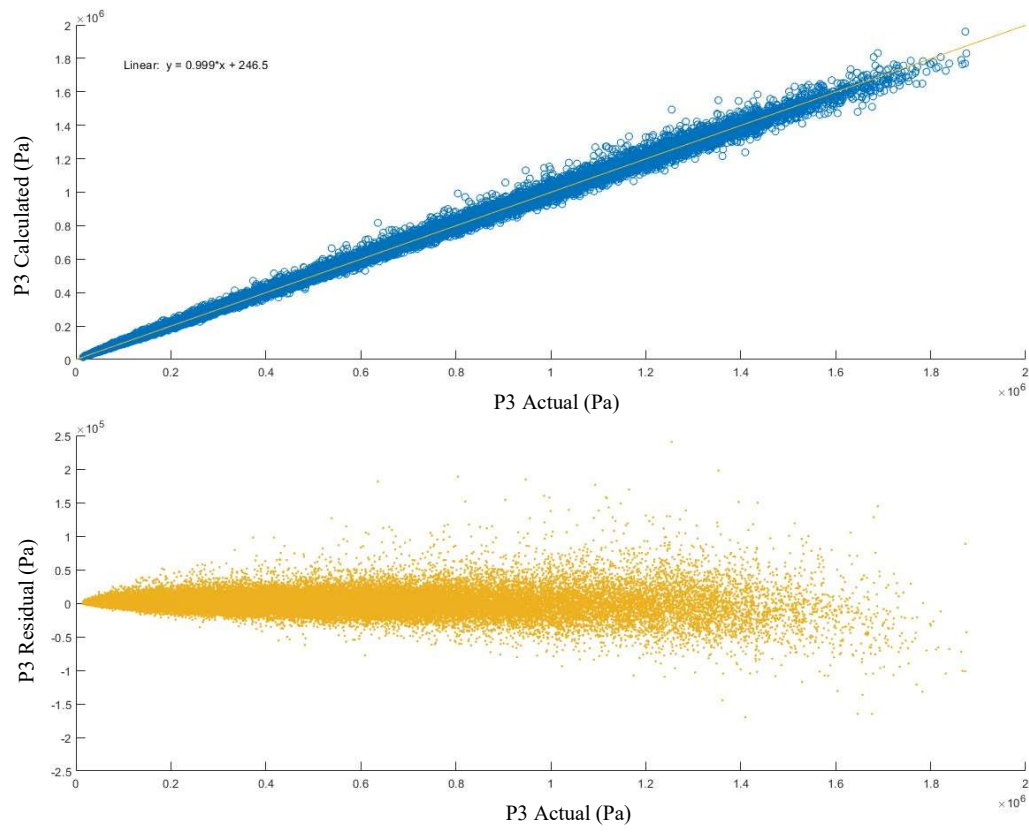


Figure 6.16: P3 Neural Network Model Predicted Versus Actual And Residual

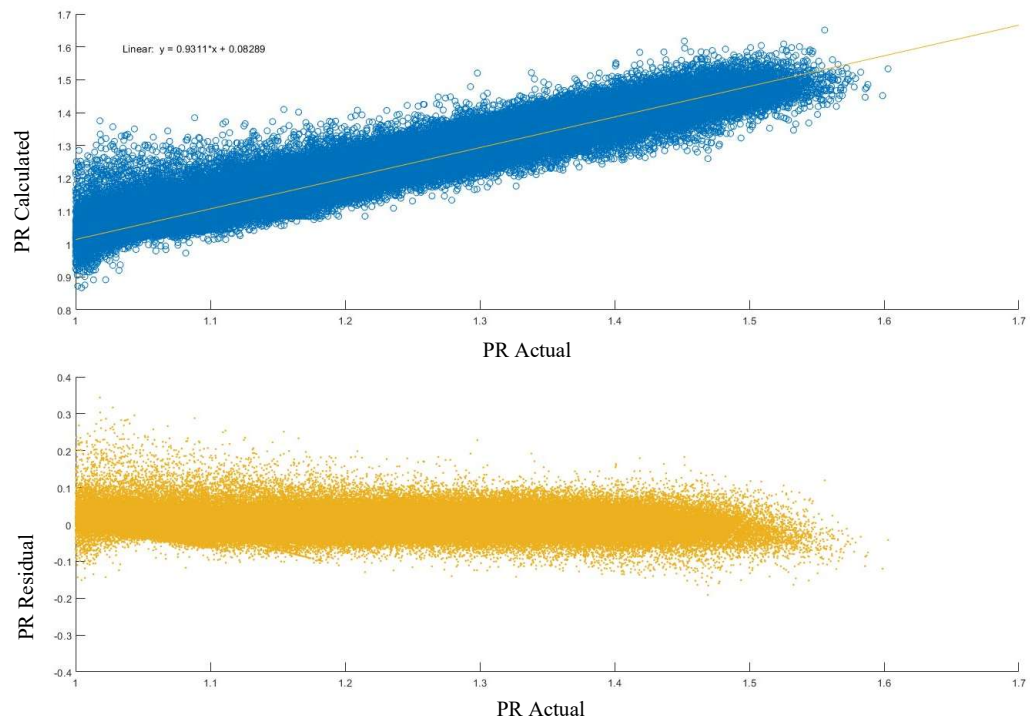


Figure 6.17: PR Neural Network Model Predicted Versus Actual And Residual

CHAPTER 7

Hybrid Theoretical and Machine Learning Modelling

As demonstrated in tables 7.1 to 7.3, Clearly predicting outcomes from real example is more effective than ideal thermodynamic calculations. Random forest model has 8.8 less RMSE than thermodynamic one in predicting outlet temperature by, 6.424 in outlet pressure and 0.0231 in pressure ratio. As presented thermodynamic calculation models have different approach than machine learning. While machine learning depends purely on real example, thermodynamic calculation depends on physics rules. Both models are missing something. Thermodynamic one missing realistic heat loss, friction, environmental condition etc. Machine learning is trained in specific conditions with a wide range of probabilities. The thermodynamic base needs detailed information about geometrical design which usually is not handy. Machine leaning can be run even with limited number of variables but could lead to unexplainable misleading conclusion. Also, Noises sensor, useability could capture by machine learning models which affect model accuracy. Theoretical model has design boundary which could filter out noises and instabilities.

Combining Thermodynamic base and Artificial Intelligence base model may have better prediction. It covers the negative sides of both. Wide probabilities can be narrowed by feeding basic design concept and theoretical models into machine learning models. It is a way of leading machine learning algorithm of halfway and letting it only compute the rest with more confident range. In other words, instead of letting software find a relation blindly, ideal physical relation can be provided. Machine learning models algorithm only objective then is to calibration and adopt ideal models into real example.

Table 7.1: P3 Theoretical Model And Random Forest Scores

Model Score P_3	RMSE (Pa)	MSE (Pa ²)	MAE (Pa)	R^2
Calculated Theoretical	12.851	165.145	8.26	0.9970
Random Forest	6.427	41.312	2.701	0.9992
difference	6.424	123.833	5.559	0.0022

Table 7.2: T3 Theoretical Model And Random Forest Scores

Model Score T_3	RMSE (K)	MSE (K ²)	MAE (K)	R^2
Calculated Theoretical	12.202	148.905	10.594	0.9711
Random Forest	3.397	11.541	2.362	0.9977
difference	8.805	137.364	8.232	0.0266

Table 7.3: PR Theoretical Model And Random Forest Scores

Model Score PR	RMSE	MSE	MAE	R^2
Calculated Theoretical	0.0509	0.0026	0.0520	0.7986
Random Forest	0.0278	0.000776	0.0195	0.9398
difference	0.0231	0.00182	0.0325	0.1412

7.1 Hybrid Modelling Methodology

As raw data pulled from sensors, Temperature, pressure flow rate is measured. When using those raw data directly as predictors, error will be high. To understand the differences between theoretical models and machine learning, linear regression is considered because it easily provides simple linear relationship between predictors and outcome. equation (A) is pulled from linear regression model that has been computed in last section. Equation (B) is derived from thermodynamic rules.

$$(A) \quad T_3 = 2.873 + 0.0861\beta_1 - 0.8995r_2 - 0.0428\beta_2 + 1.0653T_0 - 2.29e^{-6}p_0 \\ - 1.230m + 5.23e^{-4}N + 6.6735e^{-5}Power$$

$$(B) \quad T_3 = T_0 + \frac{1}{2C_p}u_0^2 - \frac{1}{2C_p}u_3^2 + \frac{1}{2C_p}U_2V_{\theta 2}$$

observation

- Number of input variables: First clear observation is that theoretical model is relying on less variable compared to linear regression. While linear regression has 8 predictors, the theoretical model has 4 variables.
- Orders of input variables: Secondly, variable in linear is in first order with no couple variables while theoretical model variable has some in first order, second order, coupled
- Dependent input variables: some input predictors in linear regression can derived or calculated from other one

So, the target then to improve modeling is to make both theoretical equation and linear regression similar. That can be done by reducing predictors, changing power order. Feature engineering, which creates new variables calculated from raw variables, is the key. In case of outlet temperature, the new predictors will be Fluid Inlet velocity, fluid outlet velocity,

blade velocity and fluid projected velocity on blade velocity which can be calculated as presented in previous section from raw data and geometrical details. Then make them in second order by raising fluid inlet and outlet to the power of 2 and couple blade velocity and projected fluid velocity. The final predictors list for machine learning models will be as following: $T_0, \frac{1}{2c_p}u_0^2, \frac{1}{2c_p}u_3^2, \frac{1}{2c_p}U_2V_{\theta 2}$

7.2 Result and discussion of hybrid models

Over fitted for outlet temperature linear regression model. It has almost 0 errors which is not realistic. These types of overfit models often fail to predict new sets of data. Other algorithms show reduced in errors. They can predict with average outlet temperature with less than 0.1 difference. The best regression is Neural Network.

1- Hybrid Linear regression:

Table 7.4: Hybrid Linear Regression Model Scores

Model Score Linear Regression	RMSE	MSE	MAE	R^2
T_3 (K)	7.18E-13	5.150e-25	6.224e-13	0.9999
P_3 (KPa)	0.195	0.0381	0.0603	0.9999
PR	0.000354	1.252e-7	0.000224	0.9999

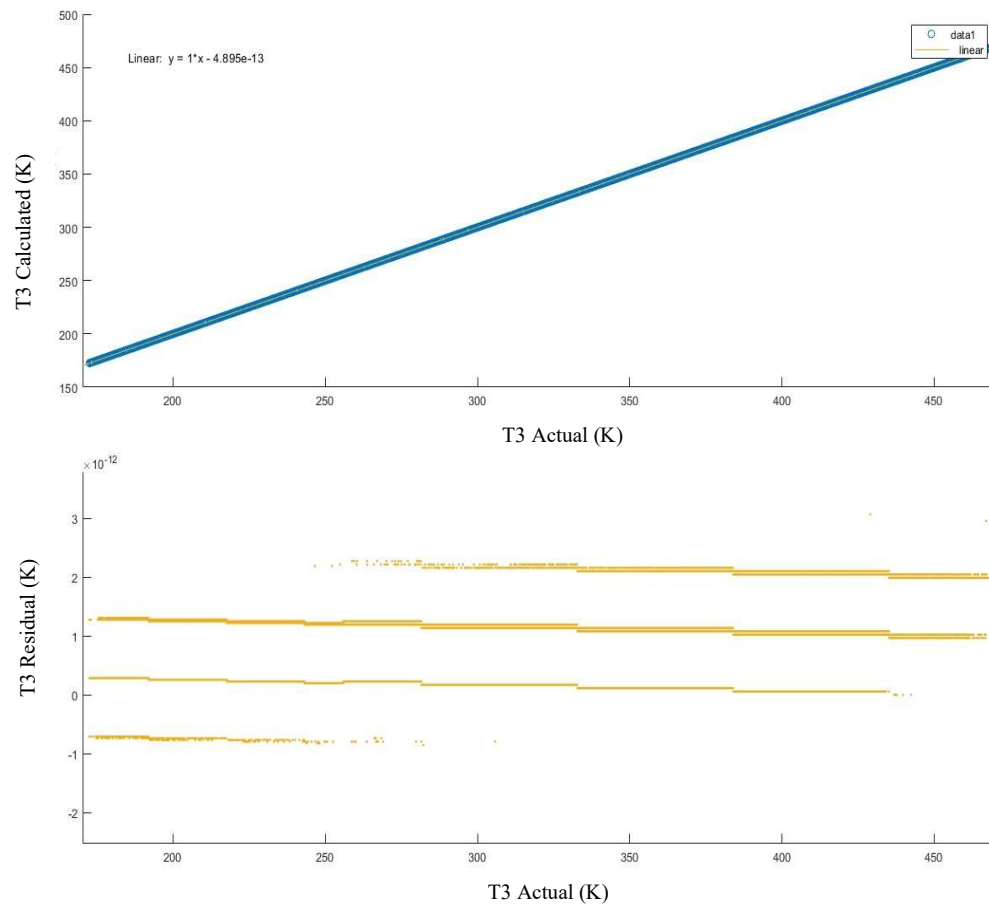


Figure 7.1: T3 Hybrid Linear Model Predicted Versus Actual And Residual

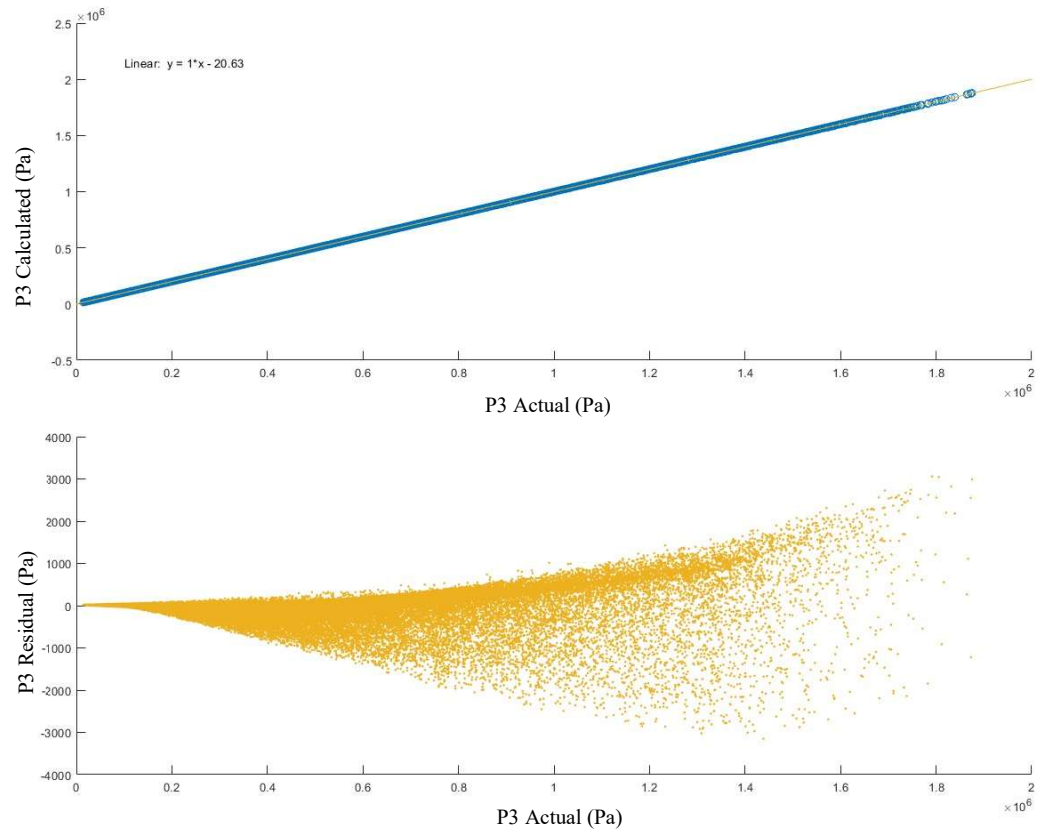


Figure 7.2 [38]: P3 Hybrid Linear Model Predicted Versus Actual And Residual

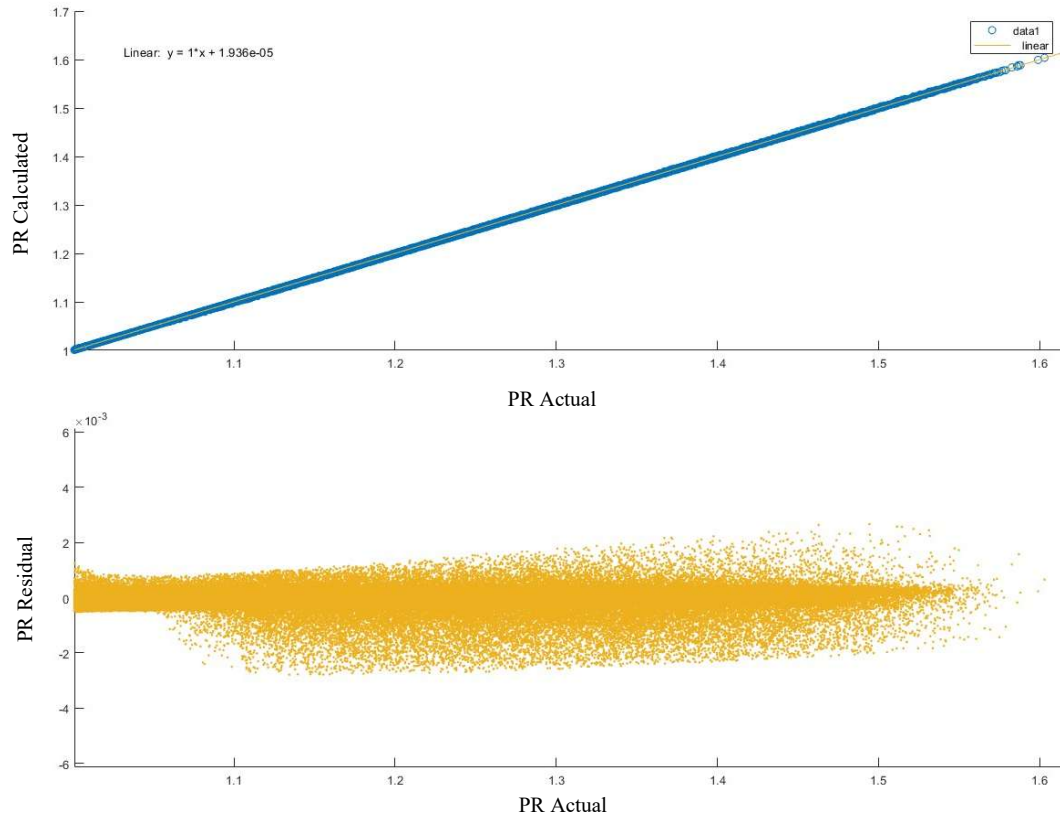


Figure 7.3 [38]: PR Hybrid Linear Model Predicted Versus Actual And Residual

2- Hybrid Decision Tree

Table 7.5: Hybrid Decision Tree Model Scores

Model Score Decision Tree	RMSE	MSE	MAE	R^2
T_3 (K)	1.499	2.249	1.205	0.9995
P_3 (KPa)	0.3533	0.1248	0.1599	0.9999
T_3 (K)	0.001122	1.258e-6	0.000889	0.9999

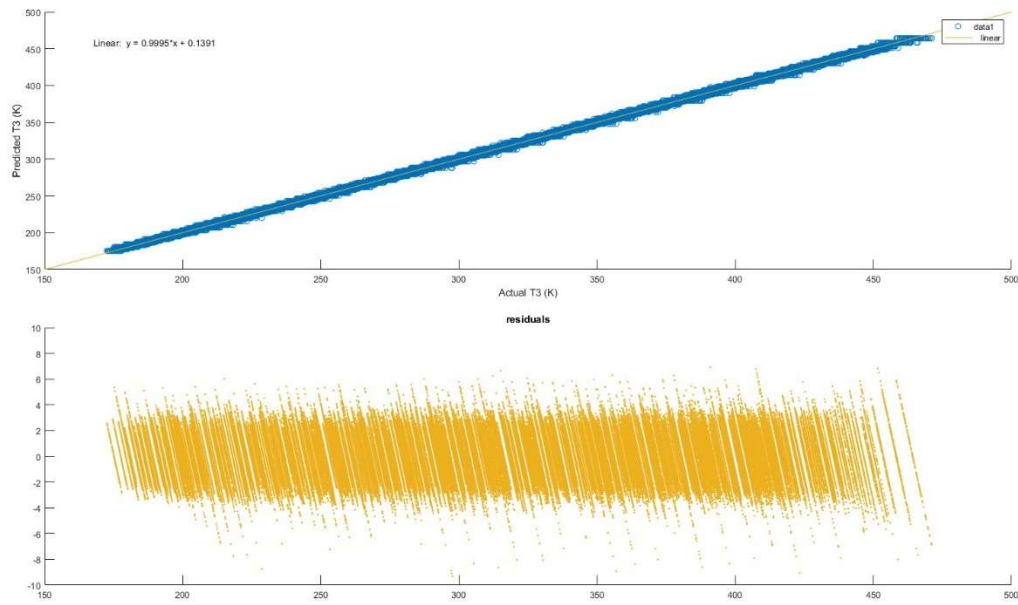


Figure 7.4: T3 Hybrid Decision Tree Model Predicted Versus Actual And Residual

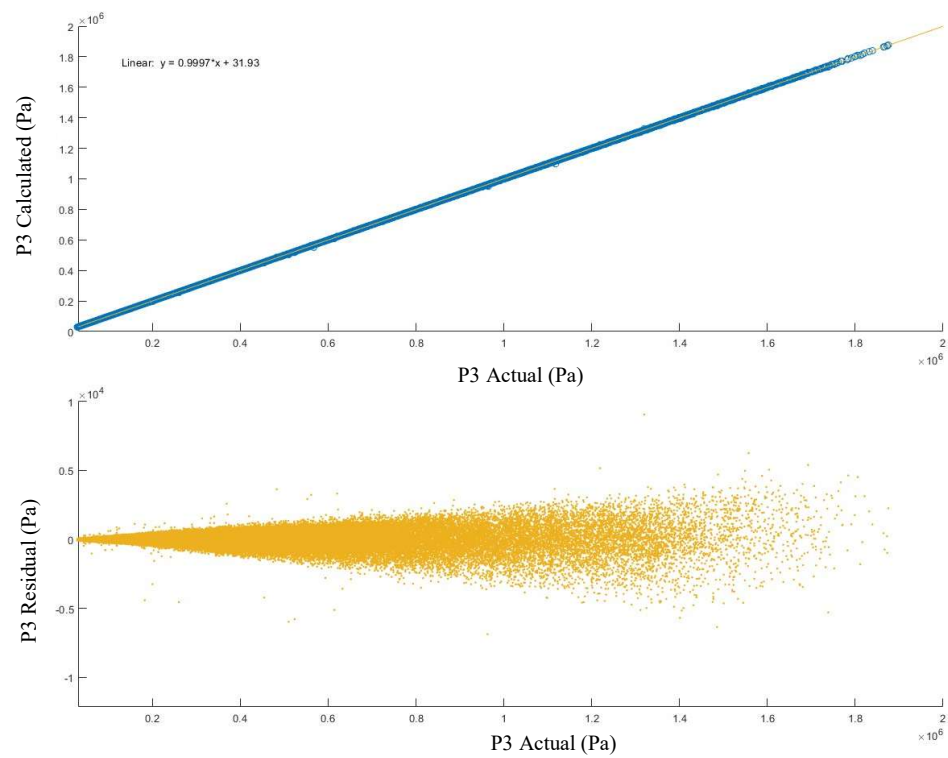


Figure 7.5: P3 Hybrid Decision Tree Model Predicted Versus Actual And Residual

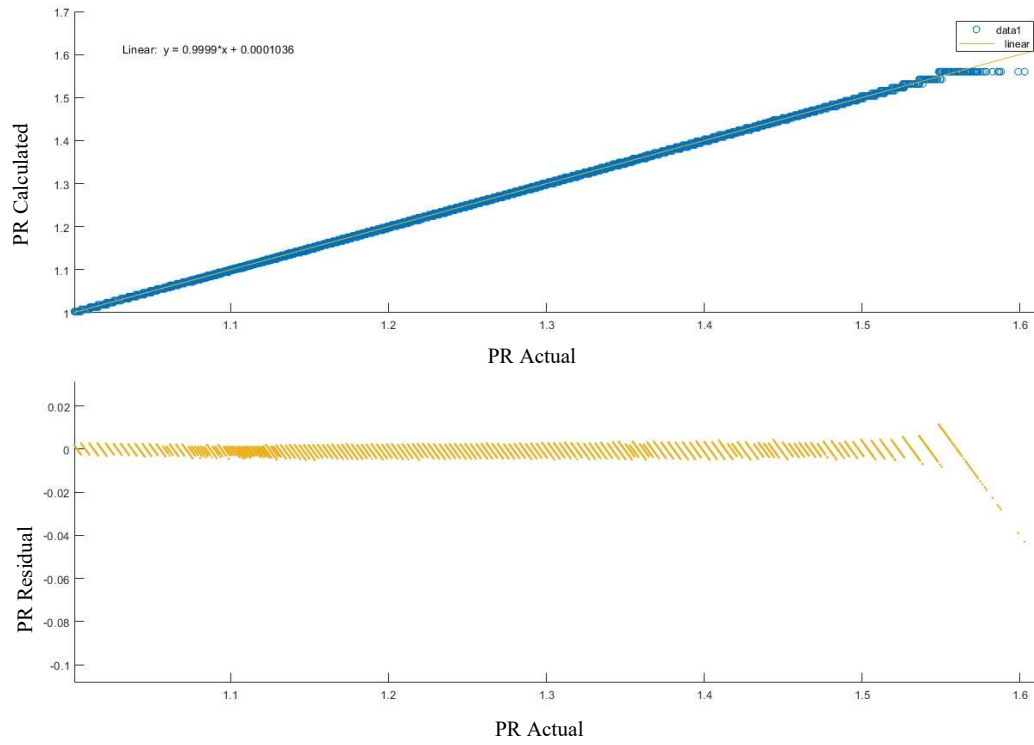


Figure 7.6: PR Hybrid Decision Tree Model Predicted Versus Actual And Residual

3- Hybrid Random Forest

Table 7.6: Hybrid Random Forest Model Scores

Model Score Random Forest	RMSE	MSE	MAE	R^2
T_3 (K)	0.676	0.457	0.496	0.9999
P_3 (KPa)	0.2841	0.0807	0.1112	0.9999
PR	0.000733	5.373e-7	0.000546	0.9999

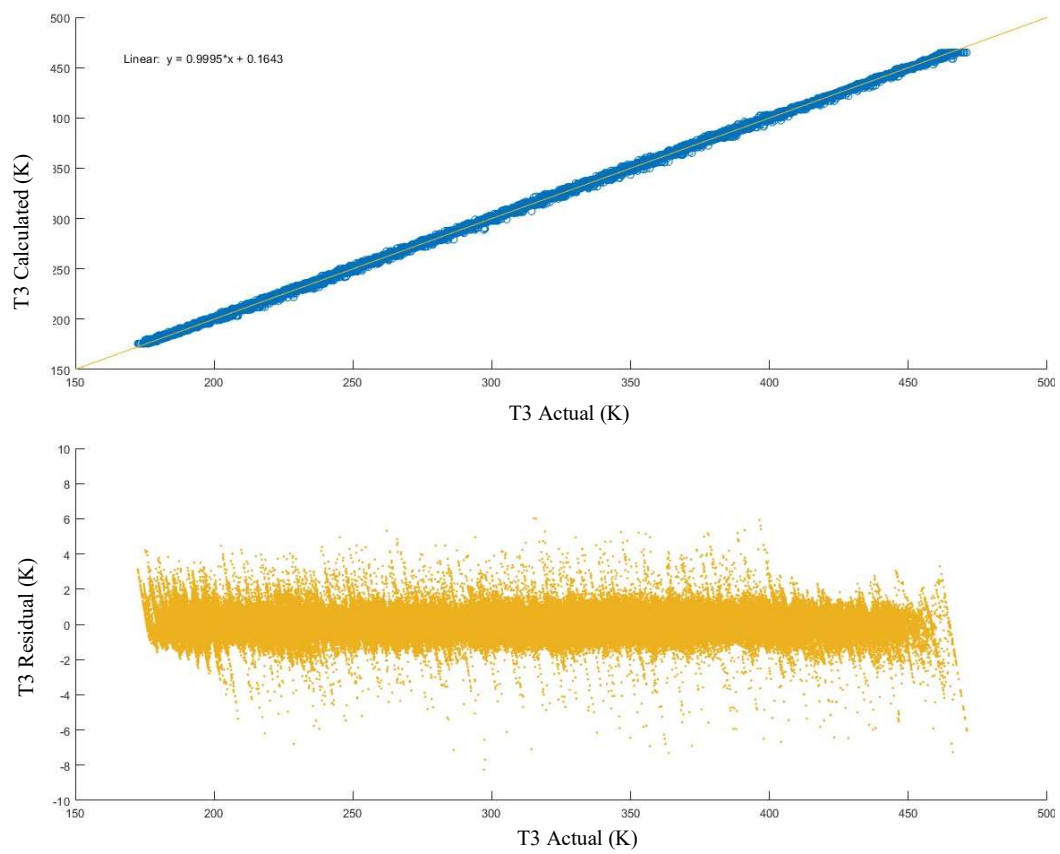


Figure 7.7: T3 Hybrid Random Forest Model Predicted Versus Actual And Residual

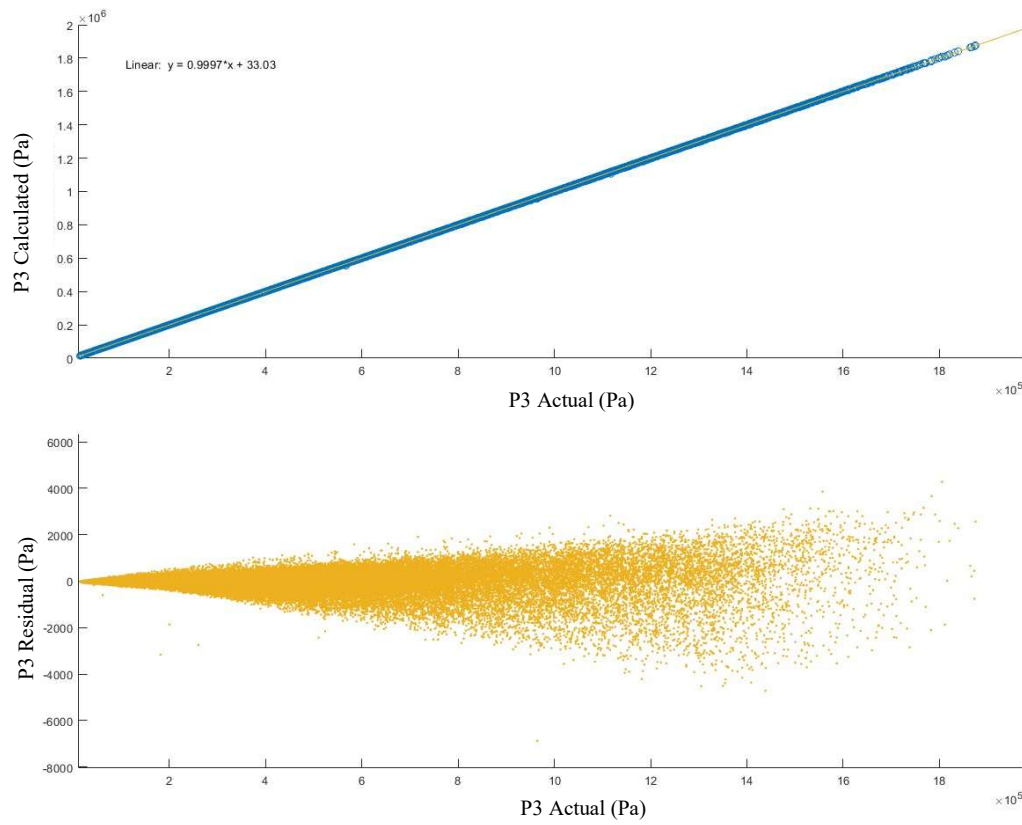


Figure 7.8 P3 Hybrid Random Forest Model Predicted Versus Actual And Residual

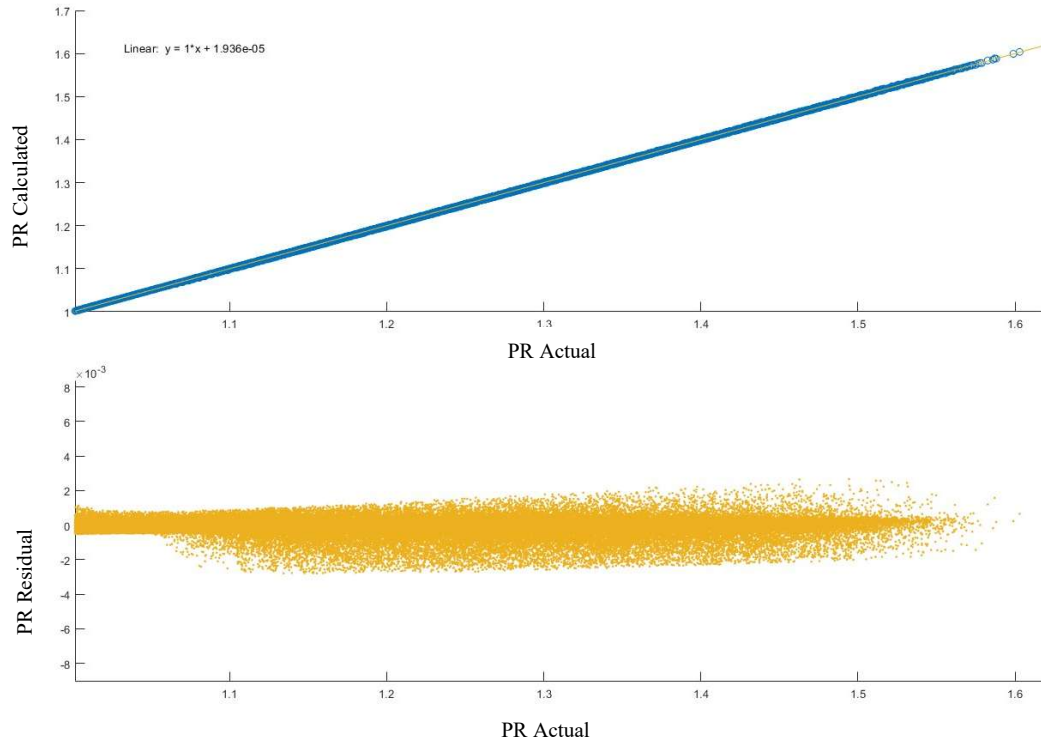


Figure 7.9: PR Hybrid Random Forest Model Predicted Versus Actual And Residual

4- Hybrid Neural Network

Table 7.7: Hybrid Neural Network Model Scores

Model Score Neural Network	RMSE	MSE	MAE	R^2
T_3 (K)	0.00259	6.756e-6	0.00051	0.99999
P_3 (KPa)	0.3526	0.1243	0.1266	0.99993
PR	0.001009	1.018e-6	0.000653	0.999921

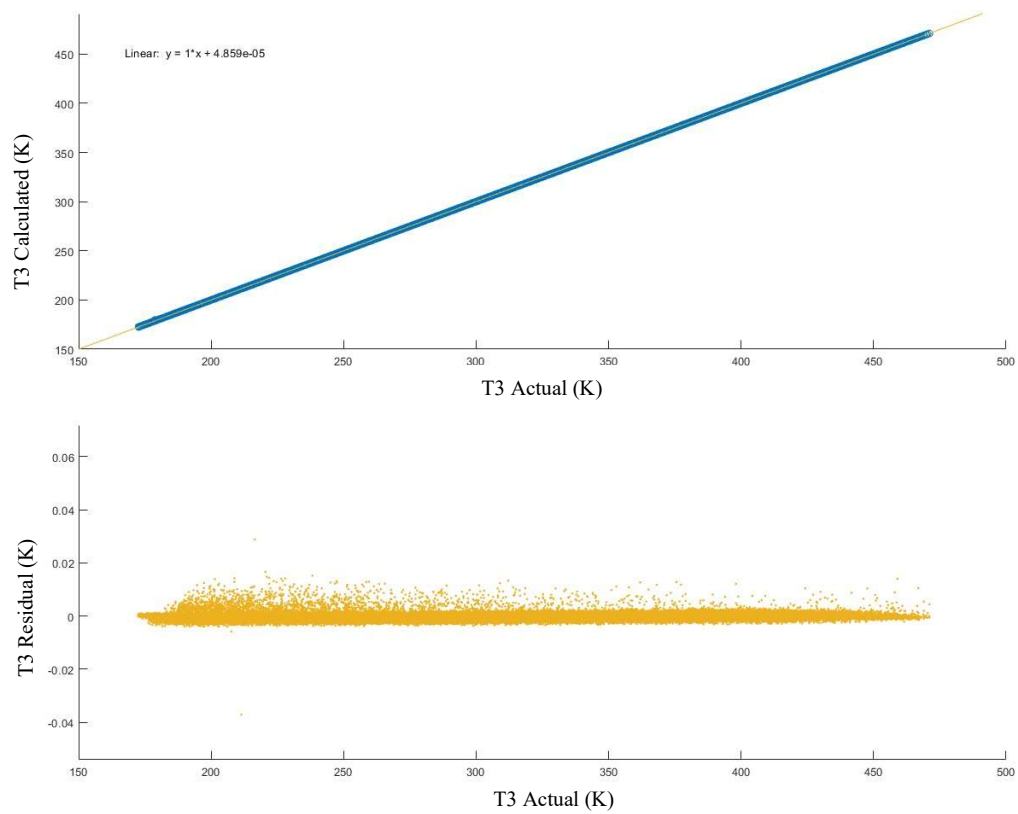


Figure 7.10: T3 Hybrid Neural Network Model Predicted Versus Actual And Residual

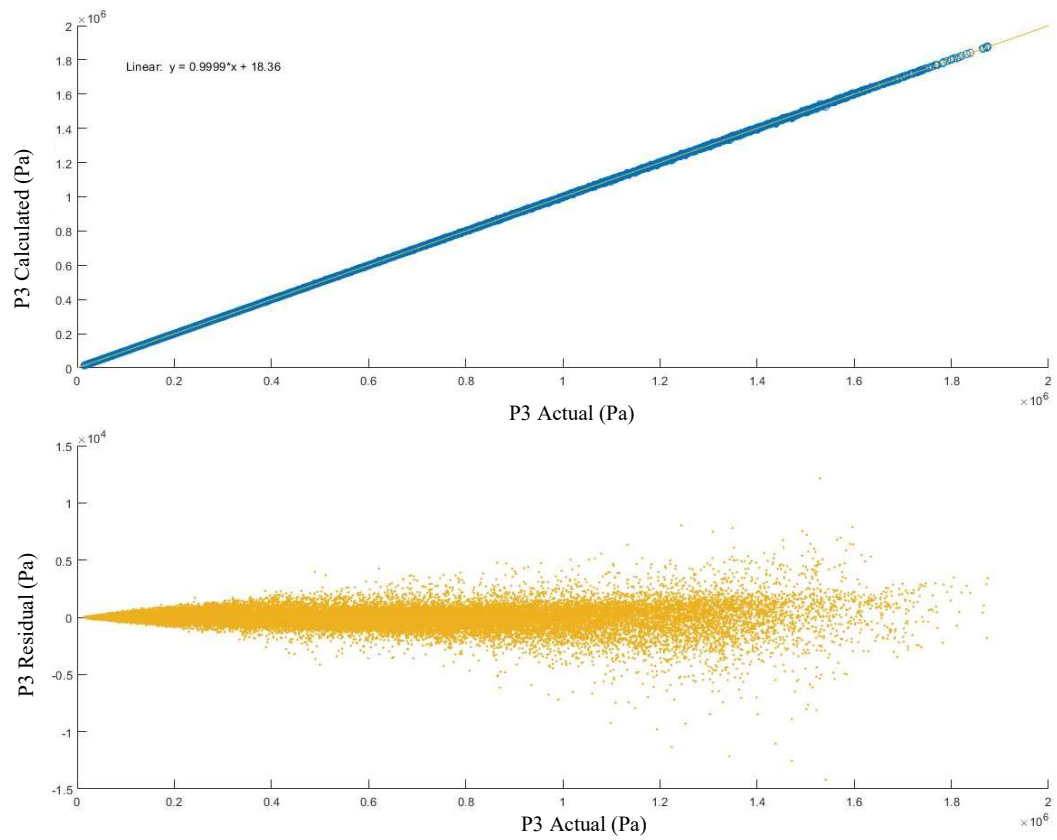


Figure 7.11 P3 Hybrid Neural Network Model Predicted Versus Actual And Residual

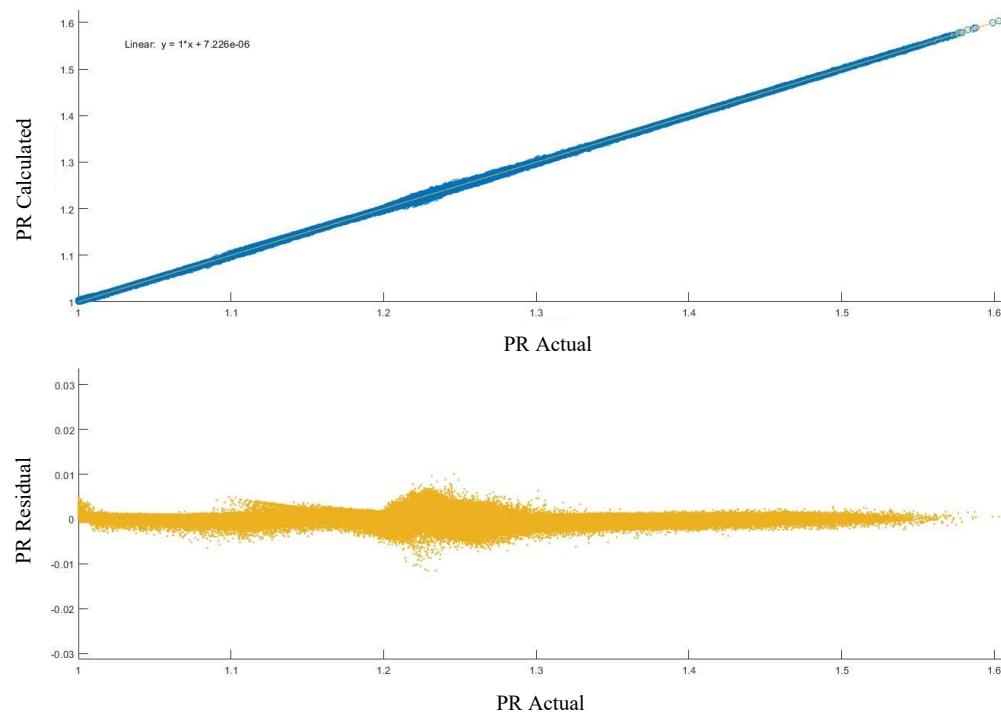


Figure 7.12: PR Hybrid Neural Network Model Predicted Versus Actual And Residual

7.3 The Proposed Hybrid Modelling Compared To Latest Advance AI Modelling

In condition monitoring field, AI techniques have been recognized and proven their effectiveness within past years [49]. Conventional techniques includes mainly Support vector machines (SVM), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), Bayesian Networks (BN) and Artificial Neural Network (ANN). On the other hand, there are new generation of artificial intelligence algorithms which are leverage cutting-edge algorithms uses deep learning, transfer learning, and combined models that merge various AI approaches to tackle complex and dynamic compressor health monitoring challenges such as: Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Graph Neural Networks (GNNs), Reinforcement Learning (RL) and Multi-Scale Attention Pyramid Vision Transformer (MSAPVT). Different advance methods can be found for performance modelling using artificial intelligence which might handle more complexity, signal noises and missing data. However, they might need more computational time and cost. Also, they might need experts to model and monitor. Simplicity and cost effectiveness are essential in modelling to be practically feasible in a wide range of applications. Simple rule of thumb, building and optimizing models should not exceed operational and maintenance cost. Otherwise, performance modeling won't be the preferred strategy. The proposed method is a simple to utilize technique since it is using common conventional algorithms with simple ideal gas models and can predict performance within low range error. To assess the proposed method, five recent published papers are considered to compare error score. While this comparison method might not be accurate since they have different applications and data, it give general idea at which range

of error other researchers reach. Table 7.8 summarized each paper including application, methodologies and error score.

Table 7.8: Most Recent Advance Performance Modelling Techniques with MAE score

Citation	Author and Title	Application and Modeling Methodology	MAE
[50]	Xiaomo Jiang et al. A thermodynamics-informed deep learning approach for lightweight modeling of gas turbine performance	Model: Thermodynamic-base model and Deep Operator Networks (DeepONet). Data: two various real-world heavy-duty gas turbine (HDGT) machines.	1.4
[51]	Zepeng Wang et al. Online adaptive transfer learning for aeroengine on-board gas path modeling under domain shifts	Model: online adaptive transfer learning (OATL) method component-level model (CLM), a stacked LSTM network Data: simulated and actual flight data	0.25
[52]	Heng Wei et al. Research on modeling strategy of centrifugal air compressor in vehicle PEMFC's air supply subsystem based on machine learning	Model: the support vector regression (SVR) technique. particle swarm optimization (PSO) and dung beetle optimization (DBO) Data: Experimental study on the performance of a centrifugal air compressor in a 100-kW PEM fuel cell system.	0.71
[53]	Nagoor Basha Shaik et al. XGBoost based enhanced predictive model for handling missing input parameters: A case study on gas turbine	Model: advanced XGBoost :ensemble learning approach Data: A case study on gas turbine	0.71
[54]	Yulin Ma et al. Enhancing accuracy of one-dimensional characteristic predictions for axial compressors using deep learning	Model: Embedded deep learning model into the axial flow compressor 1D calculation program (HARIKA), Data: generated from Computational Fluid Dynamics (CFD) simulations base on NACA65 2D National Advisory Committee for Aeronautics 65-series (NACA65) blade profiles.	0.0048

Each of these papers have a test in different loads and scenario. For the sake of comparison, load and scenario with the lowest absolute error were considered. From summarized papers in table 7.8, it can be said that mean absolute error fall in range 0.0048 to 1.4. The proposed hybrid model can predict as low as 0.00051 which is lower than all of reviewed papers. However, this thesis test the proposed model only in simulated single stage centrifugal compressor with limited input range. The proposed models are not necessary work with this low error in other higher range and other applications. Since this can simplify the problem and can be utilized easier, it is recommended to have it as a first trial model to start with in any rotating performance modelling. If the error is not acceptable, then other advance deep learning can be considered to improve prediction.

CHAPTER 8

Degradation Condition Monitoring

Degradation detection in most plants is a straightforward comparison of sensor performance and the performance displayed on the manufacturer's maps. As previously stated, the maps are created by conducting trials on a rig under particular operating settings when the compressor is in perfect condition. They are then dimensionalized and are thought to be reflective of a wider range of operating conditions. But even in perfect condition, they are expected to have a discrepancy of up to 4% between the freshly installed compressor's on-site performance and the manufacturer maps' performance [55]. Additionally, following every major overhaul, the manufacturer's performance maps will gradually become less representative.

Hybrid thermodynamic and AI models could provide a solution to this problem, as this thesis proposes. By adapting to the compressor while it is operating on-site, the models would be able to precisely replicate the performance. The predicted error is around 1%. The models accurately depict every operational circumstance that the compressor ran into during the on-site installation. However, directly comparing the model and data is insufficient for the detection of degradation. This will be discussed more in the following section.

8.1 Degradation Detection workflow

Residual deviations which are the result of comparing the predicted and baseline performance are normally distributed around 0 in the absence of any fault. Also, it is assumed that a systematic departure from 0 in the repeated comparison of measurements

and baseline over time is a sign of slowly deteriorating gas-path components. Condition monitoring can be done mainly by three steps as following:

8.1.1. Outlier removing

Before calculating deviations from models, it is important to eliminate outlier which are number of points that do not follow 99% distribution of data. Even with the minimum of outlier it could affect model predictions. The suggested way to remove outlier is using multivariate statistical analysis on collected measurements [56]. The key concept is to examine the variance of measured variables and establish a threshold. Below threshold measurement can be accepted as being in steady-state conditions, while above measurement be rejected as being in non-steady-state conditions. The threshold is decided based on a range of historical data where we are confident that no instabilities nor sensors malfunction exist. Examine existing of outlier can be done using moving-window approach. A fixed number of data points of the process variable, n_s , are used to define a moving window of data. The window's contents are updated recursively. New data is added and outdated data is removed at each stage. Every S number of data points are moved by the window. The window containing data $(t-n_s, +1, t)$ is represented by the variable τ . For the intended use, the data-set's sampling rate, n_s , and S are parameters that must be adjusted. The following figure demonstrates how moving window works.

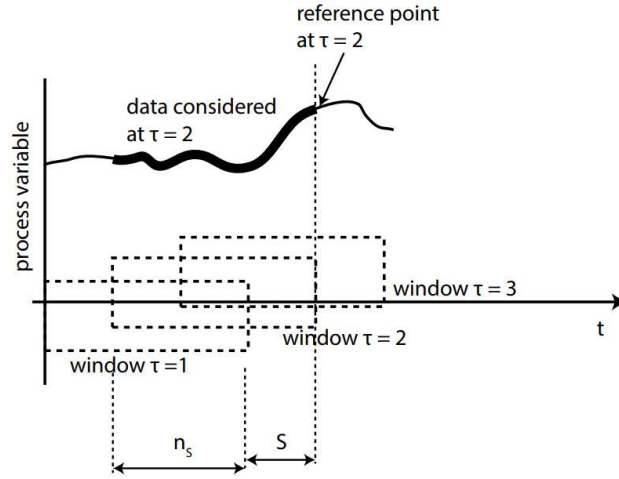


Figure 8.1: Moving-Window For Outlier Remover

8.1.2. Performance parameters calculation

After applying pre-processing moving window and make sure there is no outlier, the next step is to find actual performance parameters: pressure ratio PR_{actual} and efficiency η_{actual} from measured output parameters, T_3 and P_3 . Then predicted performance parameters $PR_{predicted}$ and efficiency $\eta_{predicted}$ are simulated from inlet parameters and dimensional details using hybrid models as presented before. Since the models trained and calibrated in healthy data, they represent performance without degradation. UD Subscript given for those from now to end of this report as UnDegraded. Comparisons are made between the undegraded and actual performance parameters with normalized deviations.

8.2 Degradation detection Result and discussion

Simulated degradation data is considered to test degradation prediction. Outlet pressure and temperature could be monitored directly but since they are influenced by process and environmental change, it is not easy to see degrading trends. Figures 8.2 and 8.3 show both

outlet pressure and temperature predicted undegraded and actual degraded. Performance parameters trend can have clearer vision to degrading. Simulated data has 18 months trends. PR_{actual} and η_{actual} are calculated from actual output parameters. Then with same data trends, input parameter used to compute undegraded outlet parameters and performance parameters PR_{UD} and η_{UD} using Neural Network hybrid model. Both calculation trends are plotted below in one chart for comparison. At early stage, undegraded parameters are close to actual one as it just starts up from maintenance. But, within time degradation rate increases exponentially which makes deviation between actual and undegraded is large which reaches to 4% deviation percentage. Deviation percentage is calculated as following:

$$d = \frac{y - y_i}{y} * 100 \quad \text{Eq 8-1}$$

There are 4 stops for maintenance as can be seen in figures (8.2 to 8.7). Every time there is a stop there is a step change in both pressure ratio and efficiency. Predictive models are adjusted to follow the new base line value. Due to limited amount of data for normal operation, it is not possible to retrain the model to the new base line.

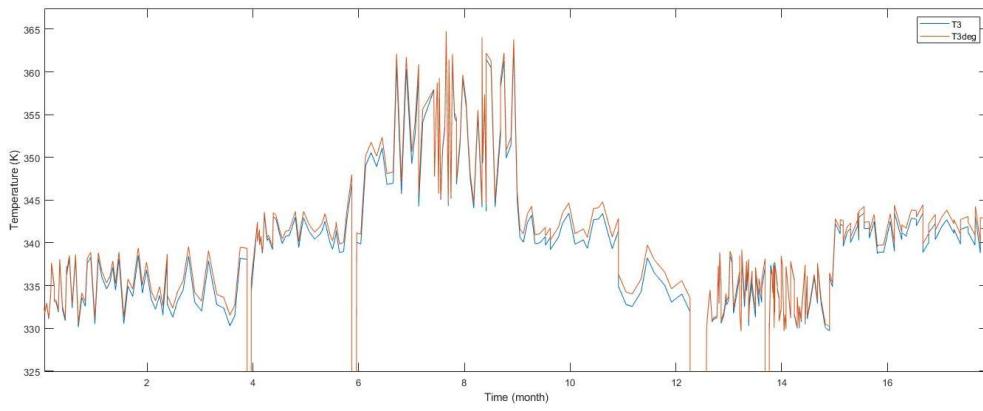


Figure 8.2: Outlet Temperature T3 Undegraded Predicted And Degraded Actual

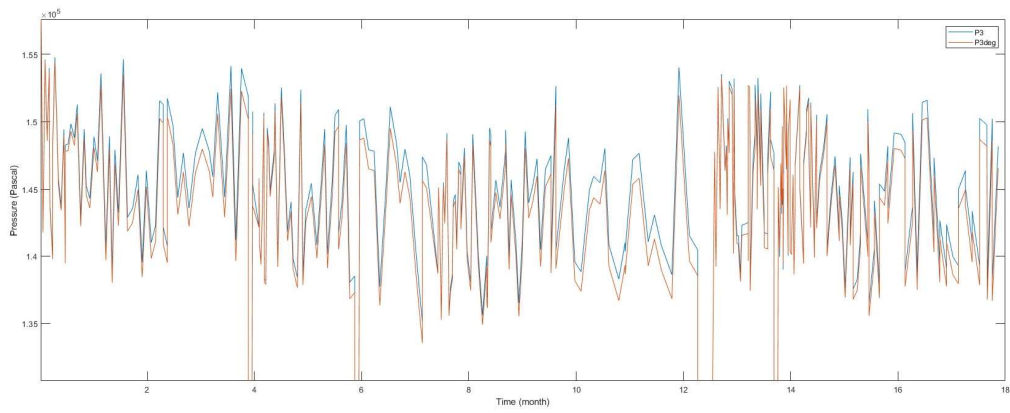


Figure 8.3: Outlet Pressure P3 Undegraded Predicted And Degraded Actual

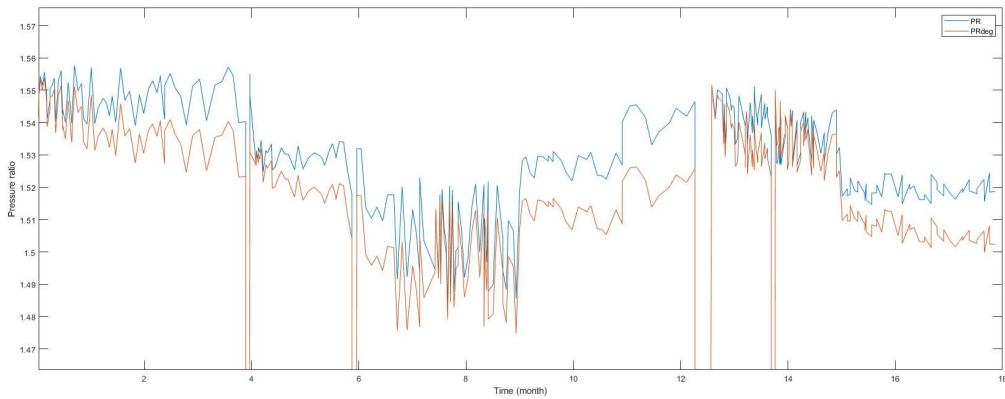


Figure 8.4: Pressure Ratio PR Undegraded Predicted And Degraded Actual

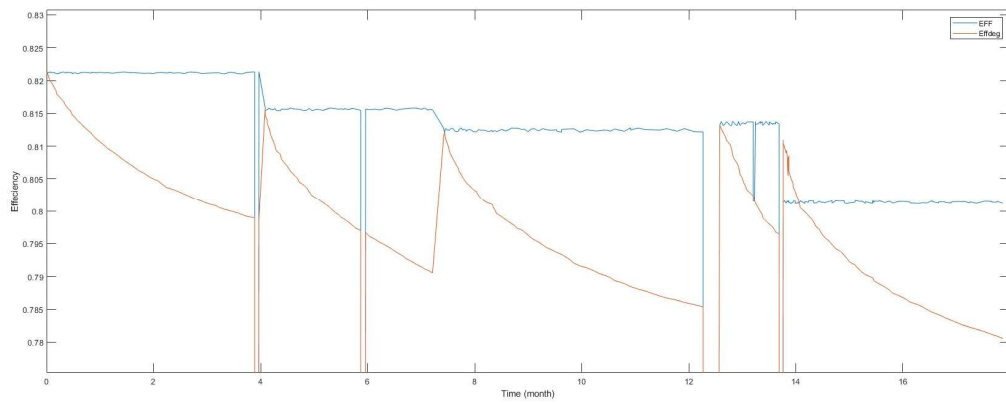


Figure 8.5: Efficiency η Undegraded Predicted And Degraded Actual

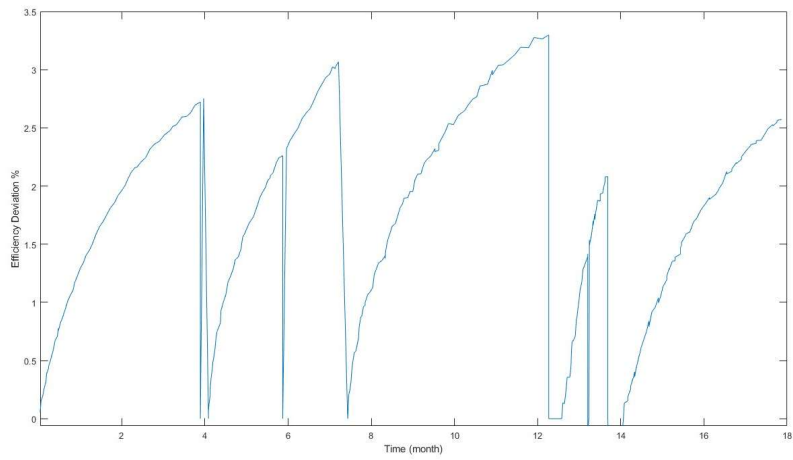


Figure 8.6: Efficiency η Degradation Deviation Percentage Indicator

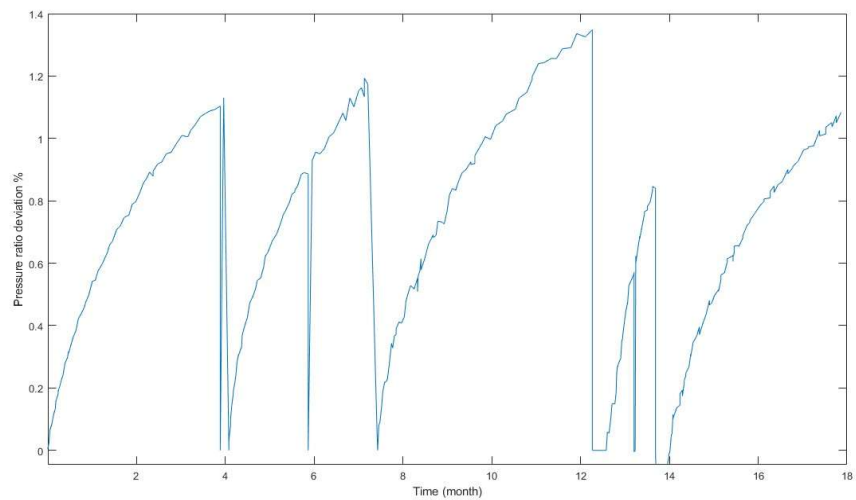


Figure 8.7: Pressure Ratio PR Degradation Deviation Percentage Indicator

CHAPTER 9

Conclusion and Perspectives

9.1 Conclusion

Relying on ideal thermodynamic calculation or experimental performance map includes high range of error. Also, building machine learning models alone as black box without any prior knowledge of thermodynamic relation still leads to unacceptable errors. To reach a better model both of the approaches are important and need to be applied in one algorithm to minimize error.

The suggested hybrid model has been proved their effectiveness in simulated degradation trend test because of following:

- Even with limited information about compressor design and limited historical data, hybrid models can be built with acceptable errors to estimate degradation.
- The proposed approaches can be used to any turbomachinery arrangement, even if we were to discuss centrifugal compressors in this thesis. As it link to basic general physics law.
- Another advantage of this approach is that it is simple to do analysis and does not need experts. Unlike many approaches which may have better accuracy, they include complex calculation and many variable.
- Also, combination of both physics-base and data-base models make diagnoses of changes are directly explainable and directly link to design variable. For example, reduction in efficiency due to increased inlet flow, temperature or pressure.

While this proposal seems a great deal, it has some challenges and limitation that could mislead to fault alerts which can be listed as the following:

- The models are sensitive to training historical data. The model rely on stable steady state condition to be as base line to trigger degradation level. If models trained on instable or transient operation changes, they may miss hit alert at the right time and lead to unplanned shutdown. Vice versa, if models train on stable steady state, they may generate false alert when there are many transient events.
- Another challenge is the quality and noisiness of collected data. Sensors that on machine are crucial factor to mirror real condition of machine. Noisy instrument, disconnecting signal and wrong reading could result in inaccurate estimation.

To reduce the impact of these problem, the thesis suggests using statistical parameter with theoretical models to test captured data whether they are within reasonable range or not before using them in models.

Due to these limitation and challenges, it is not recommended to fully depend on the proposed approach without considering other maintenance strategies. It can be used as complementary to support other findings, optimize maintenance frequencies and operation procedures. it is a great tool to quantify the severity found problems comparing them with the base line.

9.2 Perspectives

Artificial intelligence (AI) has proven its effectiveness to solve complex problems in different fields. Cost and evolving computational processors and availabilities of data make decision to utilize (AI) is more cost effective than following classical maintenance methods. However, there are still some gaps that force us to follow the classical way to do maintenance. More researches are needed to cover these gaps. Here are some recommended topics for future expansion of this research:

To make degradation monitoring more effective and accurate, more studies are needed to test having multi group of models for each operational mode, pattern recognition and instrumental problem. Categorical models could be used to select based on operational input which models need to be used. For instance, low roll models are used for mode 1 when valve is opened from 10 to 35 %. Having only one group of models for all modes of operation could affect the fitness of model. Models would take an average between high and low roll which increase residual between actual and predicted trend.

In addition, more advance method is needed other than only relying on statistical parameters to identify instrument or signal problems. As models are trained in normal acceptable operational range, another group of models may be built on abnormal faulty instrument. These models could be used to flag instruments' malfunction. Since these instruments become more critical in taking decision to evaluate internal condition of machine, continuous monitoring is needed to filter-out unwanted data and let instrument technicians take an action in right time.

Faults recognition and isolation are still a gray area that need more study. Models need to be developed to capture data trends pattern and tell if we have corrosion, erosion or overload. Also, they can highlight where could be the most effective part of machine. To do these type of study, real historical data are needed prior to each failure type. Categorical models can be built by training each different pattern. life span can be estimated as well based on data To know how long operational time until hitting unwanted efficiency.

REFERENCES

- [1] M. Tahan, E. Tsoutsanis, M. Muhammad and Z. a. A. Karim, "Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review," *Applied Energy*, p. 122–144, 2017.
- [2] A. D. Fentaye, S. I. Gilani and A. T. Baheta, "Gas turbine gas path diagnostics: A review," *MATEC Web of Conferences*, no. 74, 2016.
- [3] S. Nithin, K. Hemanth, V. Shamanth, R. S. Mahale, P. Sharath and A. Patil, "Importance of condition monitoring in mechanical domain," *Mateials Today Proceedings*, vol. 54(2), pp. 234-239, 2022.
- [4] S. Shouyi, Y. Zhufeng, L. Lei, Z. Mengchuang and Y. Weizhu, "Preliminary design of centrifugal compressor using multidisciplinary optimization method," *Mechanics & Industry*, vol. 20, no. 628, 2019.
- [5] RotMach, "CENTRIFUGAL COMPRESSORS UPGRADES & MODERNIZATION," [Online]. Available: <https://www.rotmach.com/services-centrifugal-compressors-and-blowers/>. [Accessed 06 05 2025].
- [6] W. Forsthoffer, "Compressor Best Practices," in *Forsthoffer's Best Practice Handbook for Rotating Machinery*, Elsevier, 2011, pp. 93-228.

- [7] W. B. F. (President), "23 - Turbo-compressor performance condition monitoring," in *Forsthoffer's Rotating Equipment Handbooks*, Elsevier Ltd. , 2005, pp. 279-287.
- [8] R. Kurz and K. Brun, "Fouling Mechanism in Axial Compressors," *Journal of Engineering for Gas Turbines and Power* , vol. 134 (3), 2011.
- [9] J. Sun, S. Chen, Y. Qu and J. Li, "Review on stress corrosion and corrosion fatigue failure of centrifugal compressor impeller," *Chinese Journal of Mechanical Engineering volume*, vol. 28, pp. 217-225, 2015.
- [10] J. Alqallaf, N. Ali, J. A. Teixeira and A. Addali, " Solid Particle Erosion Behaviour and Protective Coatings for Gas Turbine Compressor Blades—A Review," *processes*, vol. 8, no. 984, 2020.
- [11] J. Li and Y. Ying, "Gas turbine gas path diagnosis under transient operating conditions: A steady state performance model based local optimization approach," *Applied Thermal Engineering*, vol. 170, no. 115025, 2020.
- [12] S. Kim, J. H. Im, M. H. Kim, J. Kim and Y. I. Kim, "Diagnostics using a physics-based engine model in aero gas turbine engine verification tests," *Aerospace Science and Technology*, vol. 133, no. 108102, 2023.
- [13] F. Chu, F. Wang, X. Wang and S. Zhang, "Performance modeling of centrifugal compressor using kernel partial least squares," *Applied Thermal Engineering*, vol. 44, p. 90–99, 2012.

- [14] X. Li, Y. Ying, Y. Wang and J. Li, "A component map adaptation method for compressor modeling and diagnosis," *Advances in Mechanical Engineering*, vol. 10(3), no. 168781401876716, 2018.
- [15] D. Zhou, D. Huang, H. Zhang and J. Yang, "Periodic analysis on gas path fault diagnosis of gas turbines," *ISA Transactions*, vol. 129, p. 429–441, 2022.
- [16] E. Tsoutsanis, N. Meskin, M. Benammar and K. Khorasani, "A component map tuning method for performance prediction and diagnostics of gas turbine compressors," *Applied Energy*, vol. 135, p. 572–585, 2014.
- [17] P. Jin, F. Lü, J. Huang, X. Kong and M. Fan, "Life cycle gas path performance monitoring with control loop parameters uncertainty for aeroengine," *Aerospace Science and Technology*, vol. 115, no. 106775, 2021.
- [18] F. Lü, J. Huang and Y. & Lv, "Gas path health monitoring for a turbofan engine based on a nonlinear filtering approach," *Energies*, vol. 6(1), p. 492–513, 2013.
- [19] F. Lü, T. Gao, J. Huang and X. Qiu, "A novel distributed extended Kalman filter for aircraft engine gas-path health estimation with sensor fusion uncertainty," *Aerospace Science and Technology*, vol. 84, pp. 90-106, 2019.
- [20] L. Lin, J. L. H. Guo, Y. L. and C. , "Sample adaptive aero-engine gas-path performance prognostic model modeling method," *Knowledge Based Systems*, vol. 224, no. 107072, 2021.

- [21] M. Talaat, M. H. Gobran and M. Wasfi, "A hybrid model of an artificial neural network with thermodynamic model for system diagnosis of electrical power plant gas turbine," *Engineering Applications of Artificial Intelligence*, vol. 68, p. 222–235, 2018.
- [22] X. Zhou, X. Fu, M. Zhao and S. Zhong, "Regression model for civil aero-engine gas path parameter deviation based on deep domain-adaptation with Res-BP neural network.," *Chinese Journal of Aeronautics*, vol. 34(1), p. 79–90, 2021.
- [23] K. Wang, Y. Guo, W. Zhao, Q. Zhou and P. Guo, "Gas path fault detection and isolation for aero-engine based on LSTM-DAE approach under multiple-model architecture," *Measurement*, vol. 202, no. 111875, 2022.
- [24] Z. Wang, Y. Wang, X. Wang, B. Zhao and Y. Zhao, "Engine gas path component fault diagnosis based on a sparse deep stacking network," *Heliyon*, vol. 9(8), no. e19252, 2023.
- [25] I. Odokwo, "Computerization of Gas Turbine Performance Using Gas Path Analysis and Artificial Intelligence," *American Journal of Engineering Research (AJER)*, vol. 8(12), p. 44–52, 2019.
- [26] J. Liu, "Gas path fault diagnosis of aircraft engine using HELM and transfer learning," *Engineering Applications of Artificial Intelligence*, vol. 114, no. 105149, 2022.

- [27] S. Tang, H. Tang and M. Chen, "Transfer-learning based gas path analysis method for gas turbines," *Applied Thermal Engineering*, vol. 155, p. 1–13, 2019.
- [28] T. Breikin, Г. Куликов, V. Arkov and P. Fleming, "DYNAMIC MODELLING FOR CONDITION MONITORING OF GAS TURBINES: GENETIC ALGORITHMS APPROACH," *IFAC Proceedings Volumes*, vol. 38(1), p. 739–742, 2005.
- [29] Y. Hu, Z. Sun, L. Cao, Y. Zhang and P. Pan, "Optimization configuration of gas path sensors using a hybrid method based on tabu search artificial bee colony and improved genetic algorithm in turbofan engine," *Aerospace Science and Technology*, vol. 112, no. 106642, 2021.
- [30] H. Bao, J. Fu, X. Sun, C. Sun, J. Kuang, W. X. and J. Liu, "Performance prediction of the centrifugal air compressor for fuel cells considering degradation characteristics based on the hierarchical evolutionary model," *Thermal Science and Engineering Progress*, vol. 46, no. 102212, 2023.
- [31] J. Zhao, Y. Li and S. Sampath, "A hierarchical structure built on physical and data-based information for intelligent aero-engine gas path diagnostics," *Applied Energy*, vol. 332, no. 120520, 2023.
- [32] M. A. Zaidan, R. Relan, A. R. Mills and R. F. Harrison, "Prognostics of gas turbine engine: An integrated approach," *Expert Systems With Applications*, vol. 42(22), p. 72–83, 2015.

- [33] Z. Wang, X. Fu, R. Zhang, Z. X. X. Bai and W. Jiang, "Knowledge and data jointly driven aeroengine gas path performance assessment method," *Chinese Journal of Aeronautics*, vol. 37(5), pp. 533-557, 2024.
- [34] M. Stajuda, D. G. Cava and G. Liśkiewicz, "Aerodynamic instabilities detection via empirical mode decomposition in centrifugal compressors," *Measurement*, vol. 199, no. 111496, 2022.
- [35] D. Huang, D. Zhou, X. Wei, H. Wang and X. Zhao, "Gas path deterioration observation based on stochastic dynamics for reliability assessment of aeroengines," *Reliability Engineering & System Safety*, vol. 238, no. 109458, 2023.
- [36] D. Huang, D. Zhou, X. Jia, H. Jiang and Q. Fang, "Gas path deterioration assessment for turbofan engines based on stochastic dynamics responses in the thermodynamic cycle," *Mechanical Systems and Signal Processing*, vol. 178, no. 109314, 2022.
- [37] M. Kiaee and A. M. Tousi, "Vector-based deterioration index for gas turbine gas-path prognostics modeling framework," *Energy*, vol. 216, no. 119198, 2021.
- [38] S. Tang, H. Tang and M. Chen, "Multi-state data-driven gas path analysis method," *Energy Procedia*, vol. 158, p. 1565–1572, 2019.
- [39] D. S.L., "Basic Thermodynamics, Fluid Mechanics: Definitions of Efficiency," in *Fluid Mechanics and Thermodynamics of Turbomachinery*, Woburn, Butterworth-Heinemann, 1966, pp. 23-30.

- [40] R. I. Lewis, in *Turbomachinery Performance Analysis*, Elsevier Science & Technology Books, 1996, p. 7.
- [41] S. A. Korpela, "Principles Of Thermodynamics And Fluid Flow," in *Principles of Turbomachinery*, Hoboken, New Jersey, John Wiley & Sons, 2012, pp. 17-54.
- [42] C. Picard, J. Schiffmann and F. Ahmed, "DATED: Guidelines for Creating Synthetic Datasets for Engineering Design Applications," in *ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2023.
- [43] M. Zagorowska, F. S. Spüntrup, A. Ditlefsen, L. Imsland, E. Lunde and N. F. Thornhill, "Adaptive detection and prediction of performance degradation in off-shore turbomachinery," *Applied Energy*, vol. 268, no. 114934, 2020.
- [44] W. P. Philip and P. Fletcher, "Properties and Charts for Dry Air, Combustion Products and other Working Fluids," in *Gas Turbine Performance*, Oxford, UK, Blackwell Science Ltd, 2004, pp. 102-108.
- [45] Ahmed., H. M. B. and W. H. , "Chapter 8 introduction to Fans and Compressors," in *PUMPING MACHINERY THEORY AND PRACTICE*, John Wiley & Sons, 2015.
- [46] S. Ramlakhan, R. Saatchi, L. Sabir, D. Ventour, R. Hughes, O. Shobayo and Y. Singh, "Building artificial intelligence and machine learning models : a primer for emergency physicians," *Sheffield Hallam University Research Archive*, 2022.

- [47] M. Piri, "Review of Regression Algorithms," in *Congress on Engineering, Technology and Innovation*, 2023.
- [48] D. Kyuchukova, G. Hristov, P. Kyuchukov, G. Georgiev, П. Захариев and R. Daskalov, "Machine Learning Algorithms for Regression Analysis and Predictions of Numerical Data," pp. 1-6, 2021.
- [49] A. Kumar, "Research overview and prospect in condition monitoring of compressors," *Elsevier Expert Systems With Applications*, no. 277, 2025.
- [50] Y. L. M. W. X. C. Z. W. Xiaomo Jiang, "A thermodynamics-informed deep learning approach for lightweight modeling of gas turbine performance," *Engineering Applications of Artificial Intelligence: ELSEVIER*, no. 143, 2025.
- [51] J. X. X. W. K. Y. Y. Z. Zepeng Wang, "Online adaptive transfer learning for aeroengine on-board gas path modeling under domain shifts," *Measurement: ELSEVIER*, no. 253, 2025.
- [52] C. D. F. K. ., X. L. ., J. Z. Heng Wei, "Research on modeling strategy of centrifugal air compressor in vehicle PEMFC's air supply subsystem based on machine learning," *Fuel: ELSEVIER*, no. 381, 2025.
- [53] K. J. ., K. B. Nagoor Basha Shaik, "XGBoost based enhanced predictive model for handling missing input parameters: A case study on gas turbine," *Case Studies in Chemical and Environmental Engineering: ELSEVIER*, no. 10, 2024.

- [54] Z. D. Q. X. Yulin Ma, "Enhancing accuracy of one-dimensional characteristic predictions for axial compressors using deep learning," *Engineering Applications of Artificial Intelligence*, no. 149, 2025.
- [55] L. A. Urban and A. Volponi, "Mathematical Methods of relative Engine Performance Diagnostics," 1992.
- [56] M. Ciccotti, A. E. F., P. Xenos and N. F. Thornhill, "Preprocessing of raw data for developing steady-state data-driven models for optimizing compressor stations," in *2014 UKACC International Conference on Control (CONTROL)*, Loughborough, UK, 2014.

APPENDIX

MATLAB CODE:

Variable Dimension details

- r0: inlet pipe or nozzle
- r1=r1h-r1s: Inducer inlet radius
- r1h: impeller hub radius
- r1s: impeller shroud radius
- beta1: mean-line impeller inlet blade angle,
- r2: impeller outlet radius
- beta2: mean-line impeller outlet blade angle,
- b2: impeller outlet blade height
- r3: diffuser outlet radius
- b3: diffuser channel width
- Zb: number of blade
- PR: pressure ratio
- eff: compressor isentropic efficiency
- N: rotation speed RPM

```
datedrealAir = movevars(datedrealAir, "beta2", "Before", "r3");  
datedrealAir = movevars(datedrealAir, "beta1", "Before", "r2");  
datedrealAir = movevars(datedrealAir, "b2", "Before", "r3");  
datedrealAir = movevars(datedrealAir, "m", "Before", "in_m_in0");  
datedrealAir = movevars(datedrealAir, "Power", "Before", "in_m_in0");  
datedrealAir = movevars(datedrealAir, "N", "Before", "Power");  
datedrealAir = movevars(datedrealAir, "eff", "Before", "in_m_in0");  
datedrealAir = movevars(datedrealAir, "PR", "Before", "N");  
datedrealAir.r1=datedrealAir.r1s-datedrealAir.r1h  
datedrealAir = movevars(datedrealAir, "r1", "Before", "beta1");  
datedrealAir.Properties.VariableNames(16) = "m_N";  
datedrealAir.Properties.VariableNames(24) = "m";  
datedrealAir = movevars(datedrealAir, "m", "Before", "m_N");  
datedrealAir = movevars(datedrealAir, "m_N", "Before", "comp_valid");
```

Air gas Constant properties

Given:

1. Specific Gas constant $R=287 \text{ J/kg.K}$
2. Isobaric Specific heat $C_p= 1007 \text{ J/(kg.K)}$
3. specific heat ratio $\Gamma= 1.4$

$$\text{unit play } \frac{1}{C_p} * V = \left(\frac{\text{Kg} \cdot K}{\text{Kg} * \frac{m^2}{s^2}} \right) * \left(\frac{m^2}{s^2} \right) = K$$

R=287

C_p=1007

Gamma=1.4

Variable Operational condition

T0: inlet temperature (K)

P0: inlet pressure (kPa)

m: mass flow rate (kg/s)

P3: outlet Pressure (kPa)

T3: outlet temperature (K)

PR: Pressure ratio

```
datedrealAir.P3=datedrealAir.PR.*datedrealAir.P0
datedrealAir.T3=((((datedrealAir.Power)./(datedrealAir.m)))+(C_p.*datedrealAir.T0))./(C_p)
```

Theoretical Performance Models of Centrifugal Compressors:

P3 Calc: outlet Pressure (kPa)

T3 Calc: outlet temperature (K)

PR Calc: Pressure ratio

Zone 1: (0-1) Inlet Nozzle (Duct):

Z0=1

```
datedrealAir.mu0=(Z0.*datedrealAir.P0)./(R.*datedrealAir.T0)
datedrealAir.u0=(datedrealAir.m)
./(pi*datedrealAir.mu0.*(datedrealAir.r0.^2))
datedrealAir.u1=(datedrealAir.r0./datedrealAir.r1).*datedrealAir.u0
datedrealAir.calc_T1=(datedrealAir.T0)+(1/(2*C_p))*(datedrealAir.u0.^2)-(1/(2*C_p))*(datedrealAir.u1.^2)
```

Zone 2: (1-2) Impeller (Rotor)

```
datedrealAir.Ub1=datedrealAir.N.*datedrealAir.r1s
datedrealAir.Vr1=sqrt((datedrealAir.Ub1.^2)+(datedrealAir.u1.^2))
datedrealAir.Ub2=datedrealAir.N.*datedrealAir.r2
```

```

datedrealAir.Vteta=(datedrealAir.Power)./(datedrealAir.m.*datedrealAir.Ub2
)
datedrealAir.u2=sqrt(((datedrealAir.Ub2-
datedrealAir.Vteta)./(cotd(datedrealAir.beta2)).^2)+(datedrealAir.Vteta.^2
))
datedrealAir.Vr2=(datedrealAir.Ub2-
datedrealAir.Vteta)./(cosd(datedrealAir.beta2))
datedrealAir.calc_T2=(datedrealAir.calc_T1)+(1/(2*C_p))*(datedrealAir.u1.^
2)-
(1/(2*C_p))*(datedrealAir.u2.^2)+(1/(2*C_p))*(datedrealAir.Ub2.*datedrealA
ir.Vteta)

```

Zone 3: (2-3) Diffuser (Duct)

```

datedrealAir.u3=datedrealAir.u2.*((datedrealAir.b2)./(datedrealAir.r3))
datedrealAir.calc_T3=(datedrealAir.calc_T2)+(1/(2*C_p))*(datedrealAir.u2
.^2)-(1/(2*C_p))*(datedrealAir.u3 .^2)

```

$$\frac{P_{02}}{P_{01}} = \left(1 + \frac{\eta_c S_f U_2 V_{\theta 2}}{C_p T_{01}} \right)^{\frac{\gamma}{\gamma-1}}$$

```

datedrealAir.calc_PR=(1+((datedrealAir.Ub2.*datedrealAir.Vteta)./(C_p.*dat
edrealAir.T0))).^(Gamma/(Gamma-1))
datedrealAir.Calc_P3=datedrealAir.calc_PR.*datedrealAir.P0
datedrealAir.Calc_DPR=datedrealAir.calc_PR-datedrealAir.calc_PR

```

Featuring Variables Base on Theoretical

$$T_3 = T_0 + \frac{1}{2C_p} u_0^2 - \frac{1}{2C_p} u_3^2 + + \frac{1}{2C_p} U_2 V_{\theta 2}$$

Variable list

$$T_0$$

$$(datedrealAir.calc_T1)+(1/(2*C_p))*(datedrealAir.u1.^2)+(1/(2*C_p))*(datedrealAir.Ub2.^2)-(1/(2*C_p))*(datedrealAir.u2.^2)+(1/(2*C_p))*(datedrealAir.Ub2.*datedrealAir.Vteta)$$

$$\frac{1}{2C_p} u_0^2$$

$$datedrealAir.u0s=(1/(2*C_p))*(datedrealAir.u0.^2)$$

$$\frac{1}{2C_p} u_3^2$$

$$datedrealAir.u3s=(1/(2*C_p))*(datedrealAir.u3.^2)$$

$$\frac{1}{2C_p} U_2 V_{\theta 2}$$

$$datedrealAir.Ub2Vteta=(1/(2*C_p))*(datedrealAir.Ub2.*datedrealAir.Vteta)$$

P_3

```
datedrealAir.PRo = (datedrealAir.PR).^((Gamma-1)/Gamma)
datedrealAir.calc_T3t=(datedrealAir.T0)+(datedrealAir.u0s)-
(datedrealAir.u3s)+(datedrealAir.Ub2Vteta)
```

Compare Calculated with actual

```
datedrealAir.calc_DPR=datedrealAir.calc_PR-datedrealAir.PR
datedrealAir.calc_DT3=datedrealAir.calc_T3-datedrealAir.T3
datedrealAir.calc_DP3=datedrealAir.Calc_P3-datedrealAir.P3
```

method 2

$$\left(\frac{P_{02}}{P_{01}}\right)^{\frac{\gamma-1}{\gamma}} = 1 + \frac{\eta_c S_f U_2 V_{\theta 2}}{C_p T_{01}}$$

```
datedrealAir.UVteta=(datedrealAir.Ub2.*datedrealAir.Vteta)./(C_p.*datedrealAir.T0)
datedrealAir = datedrealAir(datedrealAir.calc_DP3 >= -6667.0799 |
ismissing(datedrealAir.calc_DP3),:);
datedrealAir = datedrealAir(datedrealAir.calc_DPR >= -0.0042 |
ismissing(datedrealAir.calc_DPR),:)
```

Remove outlier

```
datedrealAir = datedrealAir(datedrealAir.calc_DP3 >= -6667.0799 |
ismissing(datedrealAir.calc_DP3),:);
datedrealAir = datedrealAir(datedrealAir.calc_DPR >= -0.0042 |
ismissing(datedrealAir.calc_DPR),:)
```

Part 4: Performance Machine Learning Supervised Models of Centrifugal Compressors:

P3 Linear: outlet Pressure (kPa) Linear Tree regression prediction

T3 Linear: outlet temperature (K) Linear Tree regression prediction

PR Linear: Pressure ratio linear Linear Tree prediction

P3 Tree: outlet Pressure (kPa) Decision Tree regression prediction

T3 Tree: outlet temperature (K) Decision Tree regression prediction

PR Tree: Pressure ratio Decision regression prediction

P3 NN: outlet Pressure (kPa) Neural Network regression prediction

T3 NN: outlet temperature (K) Neural Network regression prediction

PR NN: Pressure ratio Neural Network regression prediction

P3 LinearH: outlet Pressure (kPa) Hybrid Linear Tree regression prediction

T3 LinearH: outlet temperature (K) Hybrid Linear Tree regression prediction

PR LinearH: Pressure ratio linear Hybrid Linear Tree prediction

P3 TreeH: outlet Pressure (kPa) Hybrid Decision Tree regression prediction

T3 TreeH: outlet temperature (K) Hybrid Decision Tree regression prediction

PR TreeH: Pressure ratio Hybrid Decision regression prediction

P3 NNH: outlet Pressure (kPa) Hybrid Neural Network regression prediction

T3 NNH: outlet temperature (K) Hybrid Neural Network regression prediction

PR NNH: Pressure ratio Neural Hybrid Network regression prediction

Split data into two groups:

1. 50% Data for training and verification

1. 50% Data for testing

Seperate data two three groups:

```
rng(1)
datedrealAirPart=cvpartition(height(datedrealAir), "HoldOut",0.5)
datedrealAirtestIdx=test(datedrealAirPart)
datedrealAirtest=datedrealAir(datedrealAirtestIdx,:);
datedrealAirTrainIdx=training(datedrealAirPart)
datedrealAirTrain=datedrealAir(datedrealAirTrainIdx,:);
```

Hybrid Model AND direct: Thermodynamic-Based and Artificial Intelligence

```
datedrealAirresult.T3linear=T3linear.predictFcn(datedrealAir)
datedrealAirresult.T3tree=T3Tree.predictFcn(datedrealAir)
datedrealAirresult.T3NN=T3NN.predictFcn(datedrealAir)
datedrealAirresult.T3Forest=T3Forest.predictFcn(datedrealAir)
datedrealAirresult.T3linearH=T3linearH.predictFcn(datedrealAir)
datedrealAirresult.T3treeH=T3TreeH.predictFcn(datedrealAir)
datedrealAirresult.T3NNH=T3NNH.predictFcn(datedrealAir)
datedrealAirresult.T3ForestH=T3ForestH.predictFcn(datedrealAir)
datedrealAirresult.PRlinear=PRlinear.predictFcn(datedrealAir)
datedrealAirresult.PRtree=PRTree.predictFcn(datedrealAir)
datedrealAirresult.PRNN=PRNN.predictFcn(datedrealAir)
datedrealAirresult.PRForest=PRForest.predictFcn(datedrealAir)
datedrealAirresult.PRlinearH=PRlinearH.predictFcn(datedrealAir)
datedrealAirresult.PRtreeH=PRTreeH.predictFcn(datedrealAir)
datedrealAirresult.PRNNH=PRNNH.predictFcn(datedrealAir)
datedrealAirresult.PRForestH=PRforestH.predictFcn(datedrealAir)
```

```

datedrealAirresult.P3linear=(PRlinear.predictFcn(datedrealAir)).*(datedrealAirresult.P0)
datedrealAirresult.P3tree=PRTree.predictFcn(datedrealAir).*(datedrealAirresult.P0)
datedrealAirresult.P3NN=PRNN.predictFcn(datedrealAir).*(datedrealAirresult.P0)
datedrealAirresult.P3Forest=PRForest.predictFcn(datedrealAir).*(datedrealAirresult.P0)
datedrealAirresult.P3linearH=PRlinearH.predictFcn(datedrealAir).*(datedrealAirresult.P0)
datedrealAirresult.P3treeH=PRTreeH.predictFcn(datedrealAir).*(datedrealAirresult.P0)
datedrealAirresult.P3NNH=PRNNH.predictFcn(datedrealAir).*(datedrealAirresult.P0)
datedrealAirresult.P3ForestH=PRForestH.predictFcn(datedrealAir).*(datedrealAirresult.P0)

```

Plot result

normal modeling

```

datedrealAirCopy
datedrealAirCopy = datedrealAirCopy(datedrealAirCopy.P0 >= 100000 &
datedrealAirCopy.P0 <= 110000 | ismissing(datedrealAirCopy.P0),:);
datedrealAirCopy = datedrealAirCopy(datedrealAirCopy.T0 >= 286 &
datedrealAirCopy.T0 <= 318 | ismissing(datedrealAirCopy.T0),:);
datedrealAirCopy = datedrealAirCopy(datedrealAirCopy.eff >= 0.8 &
datedrealAirCopy.eff <= 0.9 | ismissing(datedrealAirCopy.eff),:);
datedrealAirCopy = datedrealAirCopy(datedrealAirCopy.PR >= 1.5 &
datedrealAirCopy.PR <= 1.5333 | ismissing(datedrealAirCopy.PR),:);
writetable(datedrealAirCopy, 'datedrealAirCopy.CSV', 'Delimiter', ',', 'QuotedStrings', 'all')
type 'datedrealAirCopy.csv'

```

Work in fixed Geometry Case Study

Zone 1: (0-1) Inlet Nozzle (Duct):

```

Z0=1
datedrealAirCopy.mu0=(Z0.*datedrealAirCopy.P0)./(R.*datedrealAirCopy.T0)
datedrealAirCopy.u0=(datedrealAirCopy.m)
./(pi*datedrealAirCopy.mu0.*(datedrealAirCopy.r0.^2))
datedrealAirCopy.u1=(datedrealAirCopy.r0./datedrealAirCopy.r1).*datedrealAirCopy.u0
datedrealAirCopy.calc_T1=(datedrealAirCopy.T0)+(1/(2*C_p))*(datedrealAirCopy.u0.^2)-(1/(2*C_p))*(datedrealAirCopy.u1.^2)

```

Zone 2: (1-2) Impeller (Rotor)

```

datedrealAirCopy.Ub1=datedrealAirCopy.N.*datedrealAirCopy.r1s
datedrealAirCopy.Vr1=sqrt((datedrealAirCopy.Ub1.^2)+(datedrealAirCopy.u1.^2))

```

```

datedrealAirCopy.Ub2=datedrealAirCopy.N.*datedrealAirCopy.r2
datedrealAirCopy.Vteta=(datedrealAirCopy.Power)./(datedrealAirCopy.m.*date
drealAirCopy.Ub2)
datedrealAirCopy.u2=sqrt(((datedrealAirCopy.Ub2-
datedrealAirCopy.Vteta)./(cotd(datedrealAirCopy.beta2)).^2)+(datedrealAirC
opy.Vteta.^2))
datedrealAirCopy.Vr2=(datedrealAirCopy.Ub2-
datedrealAirCopy.Vteta)./(cosd(datedrealAirCopy.beta2))
datedrealAirCopy.calc_T2=(datedrealAirCopy.calc_T1)+(1/(2*C_p))*(datedreal
AirCopy.u1.^2)-
(1/(2*C_p))*(datedrealAirCopy.u2.^2)+(1/(2*C_p))*(datedrealAirCopy.Ub2.*da
tedrealAirCopy.Vteta)

```

Zone 3: (2-3) Diffuser (Duct)

```

datedrealAirCopy.u3=datedrealAirCopy.u2.*((datedrealAirCopy.b2)./(datedrea
lAirCopy.r3))
datedrealAirCopy.calc_T3=(datedrealAirCopy.calc_T2)+(1/(2*C_p))*(datedreal
AirCopy.u2 .^2)-(1/(2*C_p))*(datedrealAirCopy.u3 .^2)

```

$$\frac{P_{02}}{P_{01}} = \left(1 + \frac{\eta_c S_f U_2 V_{\theta 2}}{C_p T_{01}} \right)^{\frac{\gamma}{\gamma-1}}$$

```

datedrealAirCopy.calc_PR=(1+((datedrealAirCopy.eff.*datedrealAirCopy.Ub2.*
datedrealAirCopy.Vteta)./(C_p.*datedrealAirCopy.T0))).^(Gamma/(Gamma-1))
datedrealAirCopy.Calc_P3=datedrealAirCopy.calc_PR.*datedrealAirCopy.P0

```

featuring variables base on theoretical

$$T_3 = T_0 + \frac{1}{2C_p} u_0^2 - \frac{1}{2C_p} u_3^2 + + \frac{1}{2C_p} U_2 V_{\theta 2}$$

Variable list

$$T_0$$

```

(datedrealAir.calc_T1)+(1/(2*C_p))*(datedrealAir.u1.^2)+(1/(2*C_p))*(datedrealAir.Ub2.^2)-
(1/(2*C_p))*(datedrealAir.u2.^2)+(1/(2*C_p))*(datedrealAir.Ub2.*datedrealAir.Vteta)

```

$$\frac{1}{2C_p} u_0^2$$

```

datedrealAirCopy.u0s=(1/(2*C_p))*(datedrealAirCopy.u0.^2)

```

$$\frac{1}{2C_p} u_3^2$$

```

datedrealAirCopy.u3s=(1/(2*C_p))*(datedrealAirCopy.u3.^2)

```

$$\frac{1}{2C_p} U_2 V_{\theta 2}$$

```

datedrealAirCopy.Ub2Vteta=(1/(2*C_p))*(datedrealAirCopy.Ub2.*datedrealAirC
opy.Vteta)

```

P_3

```
datedrealAirCopy.PRo = (datedrealAirCopy.PR).^((Gamma-1)/Gamma)

datedrealAirCopy.PRlinearH=PRlinearH.predictFcn(datedrealAirCopy)
datedrealAirCopy.T3linearH=T3linearH.predictFcn(datedrealAirCopy)
datedrealAirCopy.P3linearH=PRlinearH.predictFcn(datedrealAirCopy).*(datedr
ealAirCopy.P0)

writetable(datedrealAirCopy12,'datedrealAirCopy12.CSV','Delimiter',' ','Qu
oteStrings','all')
type 'datedrealAirCopy12.csv'
writetable(datedrealAir,'datedrealAir.CSV','Delimiter',' ','QuoteStrings',
'all')
type 'datedrealAir.csv'

writetable(datedrealAirresult,'datedrealAirresult.CSV','Delimiter',' ','Qu
oteStrings','all')
type 'datedrealAirresult.csv'

datedrealAirresultScore.T3mean=mean(datedrealAirresult.T3)
datedrealAirresultScore.calc_T3RMSE=sqrt(mean((datedrealAirresult.T3 -
datedrealAirresult.calc_T3).^2))
datedrealAirresultScore.calc_T3MSE=mean((datedrealAirresult.T3 -
datedrealAirresult.calc_T3).^2)
datedrealAirresultScore.calc_T3MAE=mean(abs(datedrealAirresult.T3 -
datedrealAirresult.calc_T3))
datedrealAirresultScore.calc_T3Rsquare=(sum((datedrealAirresult.T3 -
datedrealAirresultScore.T3mean).^2)-sum((datedrealAirresult.T3 -
datedrealAirresult.calc_T3).^2))./sum((datedrealAirresult.T3 -
datedrealAirresultScore.T3mean).^2)
datedrealAirresult.calc_PR=datedrealAirresult.calc_PR+(rand(1)*((0.05-
0.009)+0.009))
datedrealAirresult.Calc_P3=datedrealAirresult.calc_PR.*datedrealAirresult.
P0

datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)

datedrealAirresultScore.calc_P3RMSE=sqrt(mean((datedrealAirresult.P3./1000
- datedrealAirresult.Calc_P3./1000).^2))
datedrealAirresultScore.calc_P3MSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.Calc_P3./1000).^2)
datedrealAirresultScore.calc_P3MAE=mean(abs(datedrealAirresult.P3./1000 -
datedrealAirresult.Calc_P3./1000))
datedrealAirresultScore.calc_P3Rsquare=(sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)-sum((datedrealAirresult.P3./1000
- datedrealAirresult.Calc_P3./1000).^2))./sum((datedrealAirresult.P3./1000
- datedrealAirresultScore.P3mean./1000).^2)
datedrealAirresultScore.PRmean=mean(datedrealAirresult.PR)
datedrealAirresultScore.calc_PRRMSE=sqrt(mean((datedrealAirresult.PR -
datedrealAirresult.calc_PR).^2))
```

```

datedrealAirresultScore.calc_PRME=mean((datedrealAirresult.PR -
datedrealAirresult.calc_PR).^2)
datedrealAirresultScore.calc_PRMAE=mean(abs(datedrealAirresult.PR -
datedrealAirresult.calc_PR))
datedrealAirresultScore.calc_PRRsquare=(sum((datedrealAirresult.PR -
datedrealAirresultScore.PRmean).^2)-sum((datedrealAirresult.PR -
datedrealAirresult.calc_PR).^2))./sum((datedrealAirresult.PR -
datedrealAirresultScore.PRmean).^2)
datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3linearRMSE=sqrt(mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3linear./1000).^2))
datedrealAirresultScore.P3linearMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3linear./1000).^2)
datedrealAirresultScore.P3linearMAE=mean(abs(datedrealAirresult.P3./1000 -
datedrealAirresult.P3linear./1000))
datedrealAirresultScore.P3linearRsquare=(sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)-
sum((datedrealAirresult.P3./1000 -
datedrealAirresult.P3linear./1000).^2))./sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)
datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3treeRMSE=sqrt(mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3tree./1000).^2))
datedrealAirresultScore.P3treeMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3tree./1000).^2)
datedrealAirresultScore.P3treeMAE=mean(abs(datedrealAirresult.P3./1000 -
datedrealAirresult.P3tree./1000))
datedrealAirresultScore.P3treeRsquare=(sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)-sum((datedrealAirresult.P3./1000 -
datedrealAirresult.P3tree./1000).^2))./sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)
datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3NNRMSE=sqrt(mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3NN./1000).^2))
datedrealAirresultScore.P3NNMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3NN./1000).^2)
datedrealAirresultScore.P3NNMAE=mean(abs(datedrealAirresult.P3./1000 -
datedrealAirresult.P3NN./1000))
datedrealAirresultScore.P3NNRsquare=(sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)-sum((datedrealAirresult.P3./1000 -
datedrealAirresult.P3NN./1000).^2))./sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)
datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3linearHRMSE=sqrt(mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3linearH./1000).^2))
datedrealAirresultScore.P3linearHMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3linearH./1000).^2)

```

```

datedrealAirresultScore.PP3linearHMAE=mean(abs(datedrealAirresult.P3./1000
- datedrealAirresult.P3linearH./1000))
datedrealAirresultScore.P3linearHRSquare=(sum((datedrealAirresult.P3./1000
- datedrealAirresultScore.P3mean./1000).^2)-
sum((datedrealAirresult.P3./1000 -
datedrealAirresult.P3linearH./1000).^2))./sum((datedrealAirresult.P3./1000
- datedrealAirresultScore.P3mean./1000).^2)

datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3NNHRMSE=sqrt(mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3NNH./1000).^2))
datedrealAirresultScore.P3NNHMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3NNH./1000).^2)
datedrealAirresultScore.P3NNHMAE=mean(abs(datedrealAirresult.P3./1000 -
datedrealAirresult.P3NNH./1000))
datedrealAirresultScore.P3NNHRSquare=(sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)-sum((datedrealAirresult.P3./1000
- datedrealAirresult.P3NNH./1000).^2))./sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)
datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3treeHRMSE=sqrt(mean((datedrealAirresult.P3./1000
- datedrealAirresult.P3treeH./1000).^2))
datedrealAirresultScore.P3treeHMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3treeH./1000).^2)
datedrealAirresultScore.P3treeHMAE=mean(abs(datedrealAirresult.P3./1000 -
datedrealAirresult.P3treeH./1000))
datedrealAirresultScore.P3treeHRSquare=(sum((datedrealAirresult.P3./1000 -
datedrealAirresultScore.P3mean./1000).^2)-sum((datedrealAirresult.P3./1000
- datedrealAirresult.P3treeH./1000).^2))./sum((datedrealAirresult.P3./1000
- datedrealAirresultScore.P3mean./1000).^2)
datedrealAirresultScore.P3mean=mean(datedrealAirresult.P3)
datedrealAirresultScore.P3ForestHRMSE=sqrt(mean((datedrealAirresult.P3./10
00 - datedrealAirresult.P3ForestH./1000).^2))
datedrealAirresultScore.P3ForestHMSE=mean((datedrealAirresult.P3./1000 -
datedrealAirresult.P3ForestH./1000).^2)
datedrealAirresultScore.P3ForestHMAE=mean(abs(datedrealAirresult.P3./1000
- datedrealAirresult.P3ForestH./1000))
datedrealAirresultScore.P3ForestHRSquare=(sum((datedrealAirresult.P3./1000
- datedrealAirresultScore.P3mean./1000).^2)-
sum((datedrealAirresult.P3./1000 -
datedrealAirresult.P3ForestH./1000).^2))./sum((datedrealAirresult.P3./1000
- datedrealAirresultScore.P3mean./1000).^2)

```

Data Sample

First 9 Geometrical details

r0	r1h	r1s	r1	beta1	r2	beta2	b2	r3	b3
0.19	0.10	0.18	0.08	-42.43	0.24	-69.19	0.02	0.45	0.03
0.22	0.07	0.19	0.12	-44.14	0.25	-69.94	0.03	0.85	0.02
0.20	0.10	0.15	0.05	-53.10	0.24	-69.96	0.01	0.28	0.02
0.22	0.06	0.18	0.11	-45.56	0.25	-69.59	0.02	0.53	0.02
0.22	0.07	0.19	0.12	-44.14	0.25	-69.94	0.03	0.85	0.02
0.22	0.03	0.17	0.15	-48.02	0.23	-69.78	0.03	0.89	0.03
0.16	0.07	0.15	0.08	-50.09	0.19	-69.44	0.03	0.37	0.03
0.19	0.06	0.18	0.12	-31.93	0.23	-69.88	0.02	0.45	0.03
0.18	0.02	0.11	0.09	-38.89	0.18	-69.87	0.01	0.47	0.02

First 9 Input rows

N	Power	T0	P0	m
813.74	36162.95	204.63	100.22	7.94
843.56	54855.86	256.07	189.36	15.88
768.40	37847.28	174.27	105.93	3.72
924.66	10968.96	287.59	22.14	1.27
856.24	26223.27	267.96	47.55	3.67
807.90	31074.25	204.35	106.60	8.66
1064.57	18726.31	225.01	42.48	2.12
907.00	688123.90	249.23	779.59	63.25
1541.91	143090.24	394.52	352.14	8.42

First 9 rows output Actual and different predictive models

T3 ACTUAL	T3 calc	T3 linear	T3 tree	T3 NN	T3 Forest	T3 linearH	T3 NNH	T3 ForestH	T3 treeH
209.15	207.78	212.80	214.76	208.72	215.64	209.15	209.15	209.68	209.97
259.50	258.67	258.77	270.21	254.21	276.94	259.50	259.50	261.03	259.73
184.37	179.41	184.84	189.33	187.12	186.57	184.37	184.37	185.26	186.50
296.19	292.37	307.69	298.95	300.30	301.59	296.19	296.19	296.86	295.55
275.05	272.33	284.88	282.47	279.21	291.33	275.05	275.05	274.74	274.05
207.91	206.65	210.81	213.17	204.71	214.29	207.91	207.91	209.68	209.97
233.77	230.19	240.17	235.41	236.27	233.85	233.77	233.77	233.83	234.57
260.03	255.92	235.19	271.96	264.44	274.37	260.03	260.03	259.93	259.63
411.40	403.33	421.84	411.98	414.29	412.63	411.40	411.40	411.13	410.15

PR ACTUAL	PR calc	PR linear	PR tree	PR NN	PR Forest	PR linearH	PR NNH	PR ForestH	PR treeH
1.05	1.05	1.12	1.18	1.03	1.10	1.05	1.05	1.05	1.05
1.02	1.02	1.03	1.22	1.03	1.16	1.02	1.02	1.02	1.02
1.16	1.16	1.17	1.15	1.14	1.18	1.16	1.16	1.16	1.16
1.04	1.04	1.18	1.06	1.05	1.05	1.04	1.04	1.04	1.04
1.06	1.06	1.16	1.15	1.05	1.15	1.06	1.06	1.06	1.06
1.01	1.01	1.11	1.18	1.02	1.08	1.01	1.01	1.00	1.00
1.07	1.07	1.18	1.10	1.10	1.08	1.07	1.07	1.07	1.07
1.11	1.11	0.87	1.28	1.08	1.26	1.11	1.11	1.11	1.11
1.03	1.03	1.17	1.13	1.11	1.11	1.03	1.03	1.03	1.03

P3 ACTUAL	Calc P3	P3 linear	P3 tree	P3 NN	P3 Forest	P3 linearH	P3 NNH	P3 treeH	P3 ForestH
105.18	105.16	112.69	118.50	103.66	110.46	105.21	105.16	105.33	105.28
192.45	192.44	194.39	230.92	194.61	219.85	192.49	192.31	192.82	192.81
123.27	123.15	124.22	121.43	120.66	125.36	123.26	123.28	123.22	123.19
23.05	23.05	26.12	23.46	23.28	23.30	23.05	23.05	23.07	23.06
50.28	50.27	55.31	54.62	49.76	54.54	50.28	50.28	50.32	50.33
107.14	107.14	117.87	126.04	109.15	115.05	107.17	107.20	106.88	106.97
45.43	45.43	49.99	46.69	46.52	45.92	45.44	45.46	45.51	45.48
868.47	867.62	680.16	1001.45	845.68	980.70	868.47	866.44	868.19	868.17
363.97	363.95	411.84	398.89	390.10	390.22	363.91	363.81	363.72	363.71

Vitae

Name :Murtadha Alabdrab Alredha

Nationality :Saudi

Date of Birth :1/16/1989

Email :mhahaa@hotmail.com

Address :6431 2A AlNur Dist. Sayhat 32452 – 3462 KSA

Academic Background :Received Bachelor of Science in Mechanical Engineering from King Fahd University of Petroleum and Minerals 2013, Dhahran, KSA. Received Maintenance Management Certificate from The University of Kansas Professional & Continuing Education, Lawrence – Kansas US July 2015.

Professional Background : have 17 years of experience in Reliability and condition Monitoring field worked in Air separation unit and petrochemical refinery. Certified maintenance & reliability professional by (CMRP) by society of maintenance and reliability professional. Certified vibration Analyst CAT II by vibration institute.