2016-09

Modeling of engine parameters for condition-based maintenance of the MTU series 2000 diesel engine

Yue, Siew Peng

Monterey, California: Naval Postgraduate School

http://hdl.handle.net/10945/50512

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

MODELING OF ENGINE PARAMETERS FOR CONDITION-BASED MAINTENANCE OF THE MTU SERIES 2000 DIESEL ENGINE

by

Siew Peng Yue

September 2016

Thesis Advisor:	Robert A. Koyak
Co-Advisor:	Fotis A. Papoulias
Second Reader:	Mark M. Rhoades

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2016	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
MODELING OF ENGINE PARAMETERS FOR CONDITION-BASED MAINTENANCE OF THE MTU SERIES 2000 DIESEL ENGINE

5. FUNDING NUMBERS

6. AUTHOR(S) Siew Peng Yue

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

Condition-based maintenance (CBM) entails performing maintenance only when needed to save on resources and cost. Formulating a model that reflects the behavior of the marine diesel engine in its "normal" operating conditions would aid in making predictions of the behavior of a condition monitoring parameter. Modeling for CBM is a data-dependent process. Data acquisition, processing, and analysis are required for modeling the behavior of the "normal" operating conditions of the diesel engine. This thesis leverages on existing data collected through sensors on a diesel engine to describe these conditions using regression analysis. The proposed data selection criteria ensure that data used for modeling are suitable. To model the behavior of the engine, an autoregressive distributed lag (ARDL) time series model of engine speed and exhaust gas temperature is derived. The lag length for ARDL is determined by whitening of residuals using the autocorrelation function. Due to non-normality of the residuals, a nonparametric quantile regression approach is adopted, and the derived model allows us to predict the parameter (exhaust gas temperature) that we consider.

14. SUBJECT TERMS
condition-based maintenance, regression, autoregressive distributed lag, marine diesel engine, modeling, prediction, nonparametric

15. NUMBER OF PAGES
89

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500
Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

MODELING OF ENGINE PARAMETERS FOR CONDITION-BASED MAINTENANCE OF THE MTU SERIES 2000 DIESEL ENGINE

Siew Peng Yue
Civilian, Singapore Technologies Marine Limited
B.Eng., Nanyang Technological University, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 2016

Approved by: Robert A. Koyak
Thesis Advisor

Fotis A. Papoulias
Co-Advisor

Mark M. Rhoades
Second Reader

Ronald E. Giachetti
Chair, Department of Systems Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Condition-based maintenance (CBM) entails performing maintenance only when needed to save on resources and cost. Formulating a model that reflects the behavior of the marine diesel engine in its “normal” operating conditions would aid in making predictions of the behavior of a condition monitoring parameter. Modeling for CBM is a data-dependent process. Data acquisition, processing, and analysis are required for modeling the behavior of the “normal” operating conditions of the diesel engine. This thesis leverages on existing data collected through sensors on a diesel engine to describe these conditions using regression analysis. The proposed data selection criteria ensure that data used for modeling are suitable. To model the behavior of the engine, an autoregressive distributed lag (ARDL) time series model of engine speed and exhaust gas temperature is derived. The lag length for ARDL is determined by whitening of residuals using the autocorrelation function. Due to non-normality of the residuals, a nonparametric quantile regression approach is adopted, and the derived model allows us to predict the parameter (exhaust gas temperature) that we consider.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AI	artificial intelligence
ARDL	autoregressive distributed lag
CBM	condition-based maintenance
CUSUM	cumulative sum
ECU	engine control unit
EMU	engine monitoring unit
ES	expert systems
FSRM	facilities sustainment, restoration and modernization
FY	fiscal year
HSSC	high speed surface craft
IMO	International Maritime Organization
IPDSS	intelligent predictive decision support system
LM	linear model
OLS	ordinary least square
QQ	quantile-quantile
RNN	recurrent neural networks
RPM	revolutions per minute
RQ	quantile regression
SCMS	ship control and monitoring system
SPC	statistical process control
STM	Singapore Technologies Marine Limited

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

In a new ship construction project, the main engines constitute about 10 to 20 percent of the total acquisition cost (Banks et al. 2001). Maintaining the main engine in an operationally ready condition is crucial to the availability of the vessel for its dedicated mission. To increase the availability of the vessel, it is reasonable to devise a maintenance program that provides high availability of the main engine.

Corrective maintenance requires the performance of maintenance only when a failure occurs; this process consists of system diagnostics, fault finding, and parts replacement. Preventive maintenance, on the other hand, aims to increase availability by preventing failure through the detection and unveiling of potential failure modes (Tsang 1995).

Condition-based maintenance (CBM) is a preventive maintenance policy that recommends performing maintenance only when needed to save on resources and cost. Recommendations concerning the need for maintenance typically are derived from a statistical model that reflects the behavior of the diesel engine in its "normal" operating condition that allows making predictions about the behavior of a condition monitoring parameter.

This thesis aims to document the process for finding an appropriate statistical model that may be used to characterize the data collected from a high speed surface craft (HSSC), which may be used to forecast the condition of a marine diesel engine. Our approach includes determining the following:

- selection of useful data for model formulation and estimation;
- selection of a condition monitoring parameter that is indicative of system health status;
- approach for characterizing the data;
- method for deriving the model.

We describe the steps taken to analyze data collected from the main engine of the HSSC to formulate the model and analyzes the residuals of the predictions done through

the model on a new set of data. The derived model can predict the exhaust gas temperature for the engine with a set of known engine speeds. Future work could include derivation of nonparametric control charts proposed by Li, Tang and Ng (2010). Data parameters collected could also include engine running hours to determine mechanical wear and tear of the engine. This would give more insight into the engine behavior with respect to the engine running hours.

The methodology for characterising the data collected and deriving the model for prediction of the selected engine parameter is shown in Figure 1. There are six steps in the process. The first step is to select the data collected using the selection criteria and then to analyse the data to determine their properties. The next step is to categorize the data according to their properties. Based on the different groups, models can be derived for each group. The model is then used for prediction, and the results of the prediction are verified against a new set of data.

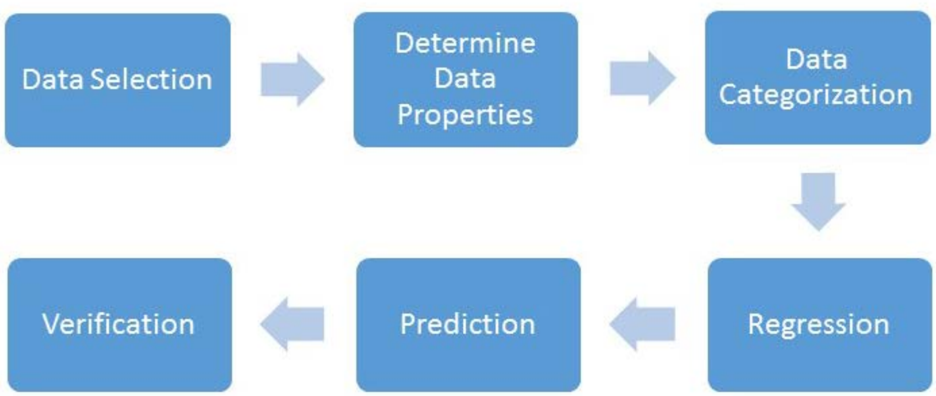

Figure 1. Methodology for Regression Analysis

Modeling for CBM is a data dependent process. Data acquisition, processing, and analysis are required for modeling the behavior of the "normal" operating conditions of the diesel engine. Leveraging on existing data collected from sensors located on the diesel engine on the port side of the HSSC for CBM, we set up selection criteria that define the "normal" operating conditions of the HSSC and filtered out data that are not useful for analysis.

Analysis of data begins with the determination of the correlation coefficient between the engine speed and other temperature parameters from the data collected. The correlation coefficient for combined B exhaust gas temperature has the highest correlation coefficient among the rest of the temperature parameters. Thus, we select the combined B exhaust gas temperature as the dependent variable and engine speed as the independent variable for the regression model.

From the data collected, we divide the data into two categories, turbocharger mode and cruising mode. Each category has two sub categories, modeling datasets and prediction datasets. Modeling datasets are for deriving the regression model, and the prediction datasets are for verifying the regression model.

We derive an autoregressive distributed lag (ARDL) time series model of engine speed and exhaust gas temperature to model the behavior of the engine under "normal" operating conditions. We determine the lag length of the ARDL model by whitening the residuals autocorrelation. We use the R language statistical programming language (R Core Team 2015) for data analysis. One can express an ARDL time series model (Greene 2000) as:

$$Y_t = \mu + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=0}^{r} \beta_j X_{t-j} + \varepsilon_t$$

with

$$\varepsilon_t \sim i.i.d\ N(0, \sigma_\varepsilon^2)$$

and, p and r are the lag length for the dependent variable, Y_t, and independent variable, X_t, respectively.

The results of linear regression, using ordinary least square methods, show non-normality nature of the residuals; hence, a nonparametric quantile regression approach is more suitable. With the derived models, we use the prediction dataset to predict the trend of exhaust gas temperature through the set of engine speed input. We use residual analysis through normality quantile-quantile (QQ) plot and robust autocorrelation function using Spearman's rank correlation, to verify the prediction results.

References:

Banks, Jeffrey, Jason Hines, Mitchell Lebold, Robert Campbell, Colin Begg, and Carl Byington. 2001. "Failure Modes and Predictive Diagnostics Considerations for Diesel Engines." *New Frontiers in Integrated Diagnostics and Prognostics. Proceedings of the 55th Meeting of the Society for Machinery Failure Prevention Technology*, 93–102.

Greene, William H. 2000. *Econometric Analysis*, 4th ed. Englewood Cliffs, NJ: Prentice Hall.

Li, Su-Yi, Loon-Ching Tang, and Szu-Hui Ng. 2010. "Nonparametric CUSUM and EWMA Control Charts for Detecting Mean Shifts." *Journal of Quality Technology* 42(2): 209–226.

R Core Team. 2015. "R: A language and environment for statistical computing." Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Tsang, Albert H. C. 1995. "Condition-Based Maintenance: Tools and Decision Making." *Journal of Quality in Maintenance Engineering* 1(3): 3–17.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to Professor Robert Koyak for his guidance, coaching, and patience while working on this thesis. His advice helped me overcome many challenges along the way for the successful completion of this thesis. His expertise on the topic of statistical analysis for preventive maintenance provided insights into this area of study and allowed me to build on the methodology for data analytics of ship data.

I would like to thank Professor Fotis Papoulias for his support and advice in the areas of systems engineering and mechanical systems. I would also like to thank Professor Mark Rhoades for taking time off from his busy schedule to give his valuable comments on this thesis. I would also like to thank Professor Barbara Berlitz for taking time to proofread the thesis despite her busy schedule.

I would like to thank my former superior, Mr. Ang Gim Kee, director (Special Projects), and Mr. Lim Kwang Yong, manager of automation (Product Development), for their recommendations and support given for me to attend the 18-month dual masters' degree program jointly organized by Temasek Defence Systems Institute of the National University of Singapore and Naval Postgraduate School.

I would like to thank my current superior, Mr. Tan Ching Eng, senior vice president of Engineering Design Center, and Mr. Sim Chee Chong, assistant director, Engineering Design Center (Automation), for their support and advice throughout this course. I also appreciate their provision of ship data for use in this thesis. I would like to thank my colleague, Ms. Lee Shu Min, senior engineer of automation (Product Development), for coordinating with me to acquire ship data for analysis in this thesis.

I would also like to thank my fellow course mates from Temasek Defense Systems Institute and the Singaporean community in Monterey for their support and help given in one way or another throughout the past year. The stay in Monterey would not have been as memorable without their company.

I would like to thank my parents, Yue Kam Weng and Wong Yuet Yong Joyce, as well as my siblings, Yue Siew Chin and Yue Heng Cheong, for their unwavering care, trust, and support over the years. Their encouragement has kept me motivated in achieving my goal.

I. INTRODUCTION

A ship and in particular a naval ship can be thought of as a system-of-systems. These systems include the propulsion system which produces a mechanical force that thrusts the ship forward or backward; the navigation system which gives locational and directional information of the ship; the electrical system which produces electrical power for equipment onboard the ship; and others. Each system onboard a ship has a unique role to play, and all systems are integrated. One of the crucial elements of the propulsion system is the main engine.

In a new ship construction project, the main engines constitute about 10 to 20 percent of the total acquisition cost (Banks et al. 2001). The function of the main engine is to provide propulsion and thrust to the vessel, giving it mobility to travel across the surface of the sea.

Maintaining the main engine in an operationally ready condition is crucial to the availability of the vessel for its dedicated mission. To increase the availability of the vessel, it is reasonable to devise a maintenance program that provides high availability of the main engine. One can broadly divide maintenance into two main types, namely, preventive maintenance and corrective maintenance (Tsang 1995).

Corrective maintenance requires the performance of maintenance only when a failure occurs; this process consists of system diagnostics, fault finding, and parts replacement. This type of run-to-failure maintenance is unscheduled, and easy to manage and implement; but it will incur a longer downtime since the part that is required to be replaced may not be readily available.

Preventive maintenance, on the other hand, aims to increase availability by preventing failure through the detection and unveiling of potential failure modes (Tsang 1995). One method of preventive maintenance of a system involves performing scheduled maintenance at a fixed time interval, before a failure occurs. In fixed interval maintenance, the condition of the part that the operator will replace is not tested. Therefore, the operator replaces the part regardless of whether it is worn out or still good for operation. While the replacement of a worn out part is timely and can prevent a failure from occurring, replacing a part that is still good increases waste and costs.

The other method of preventive maintenance is condition-based maintenance (CBM). CBM involves the continuous monitoring of a parameter that correlates to the occurrence of the failure of the equipment. The parameter should be measurable, and an operating range should be identified (Tsang 1995). Whenever the measured value of the parameter falls outside of the operating range, the system issues a warning to alert the operators for the relevant maintenance.

A. NEEDS ANALYSIS INFLUENCES ON MAINTENANCE PLANNING

The construction of a naval ship can cost up to billions of dollars (Peer 2012). It begins with the awarding of the contract to the shipbuilder, followed by the design of the ship hull and structure, the selection of systems to be installed on the ship, the detailed design of piping, wiring, and layout of systems, indicators, and sensors on the ship.

The shipbuilder, the program manager, and the end-users meet regularly to discuss their views on how the ship will operate and function. There are four primary stakeholders in the construction of a naval ship. They are the program manager, the end-user, the shipyard, and the system suppliers.

The program manager works closely with the end-users to define the requirements and operational concepts of the new naval vessel that is to be built. This individual manages the program, the budget and the shipyard to ensure that the shipbuilding project is within budget and meets the operational requirements and specifications. Together with the end-users, program managers are the decision makers for selecting the shipyards and their subcontractors, both in terms of technological compliance and financial aspects. Their effective need is to ensure that the cost estimated for the shipbuilding program is sufficient throughout the life cycle of the vessel, from cradle to grave.

The end-user for a naval vessel is the navy. The navy fleet must be operationally ready for the missions assigned to them. Increasing the operational readiness and availability of the fleet is an effective need for any navy. Another effective need is to keep the fleet’s operations and maintenance at a low cost. When there is an unlimited budget for maintenance, the availability of the fleet can be keep at a very high level. However, this is not possible in the real world. There is always a limited budget for operations and maintenance. Therefore, we can see that the need for keeping maintenance cost low and the navy fleet’s availability high is conflicting.

In the construction of a new ship project, the end-user is able to specify the use of equipment that meets a certain reliability to reduce maintenance costs throughout the lifetime of the ship. A preventive maintenance plan of fixed interval maintenance is the usual approach to keeping the fleet’s availability high. But, this approach increases cost and waste as previously discussed.

The shipyard is also a main contractor of any shipbuilding project. Shipyard personnel must ensure that the construction project is on schedule to meet the delivery milestones, and that seamless integration occurs between the many systems on board the ship. They also have to ensure that the equipment is positioned in the most efficient location to carry out its specified functions. The shipyard submits proposals to ship construction project tenders. The objective of the proposal is to ensure that the price is competitive while complying with all requirements, and still able to maximize the profits of the shipyard.

The shipyard must also take care of the structural maintenance of the ship hull over the lifetime of the vessel through a maintenance contract with the navy. The effective need of the shipyard is to maximize its profits through the ship construction and maintenance project they undertake.

The system suppliers’ effective need is to increase sales to meet their sales target. They provide pre-sales and after-sales support of their equipment to the navy. After-sales support includes servicing and repair of equipment, as well as the sales of spare parts. As they may also be the original equipment manufacturer, they are the subject matter expert for the system and may provide technical advice regarding maintenance procedures and the frequency of such maintenance.

A summary of the roles and effective needs of the primary stakeholders is shown in Table 1.

Table 1. Role and Effective Needs of Stakeholders

<table>
<tr><th>Stakeholder</th><th>Role</th><th>Effective Needs</th></tr>
<tr><td rowspan="3">Program manager</td><td>Define requirements and operational concepts</td><td rowspan="3">Ensure that the cost estimated is sufficient throughout the life cycle from cradle to grave</td></tr>
<tr><td>Manage program, the budget, and the shipyard</td></tr>
<tr><td>Select shipyard and their subcontractors</td></tr>
<tr><td rowspan="3">End-users</td><td rowspan="3">Define operational concept and requirement</td><td>Increase availability of fleet</td></tr>
<tr><td>Keep the operations and maintenance cost low</td></tr>
<tr><td>Be operational ready for missions</td></tr>
<tr><td rowspan="4">Shipyard</td><td>Ensure project is on schedule to meet the delivery milestones</td><td rowspan="4">Maximize profits in the ship construction and maintenance project</td></tr>
<tr><td>Ensure seamless integration between the many systems on board the ship</td></tr>
<tr><td>Ensure equipment is positioned in the most efficient location to carry out its functions</td></tr>
<tr><td>Provide structural maintenance of the ship hull over the lifetime of the vessel</td></tr>
<tr><td rowspan="2">System suppliers</td><td>Offer service and repair of equipment, and sale of spare parts</td><td rowspan="2">Increase sales to meet target</td></tr>
<tr><td>Provide technical advice on maintenance procedures and the frequency of such maintenance</td></tr>
</table>

From the stakeholder needs analysis, we find the end-users have two needs that conflict with each other. They need to keep the operations and maintenance cost low, while increasing the availability of the fleet. Adopting a preventive maintenance policy can keep the availability of the fleet high, but replacing parts at fixed intervals, especially parts that are not worn out, with new ones to prevent the occurrence of a failure will increase costs and waste.

Adopting the run-to-failure corrective maintenance policy, which requires performing maintenance only when a failure occurs, can reduce cost and waste. However, this approach greatly affects and, in fact, reduces the availability of the fleet due to the increased downtime required to troubleshoot problems and the procurement lead time required to purchase the replacement parts.

Considering that the navy needs to be operational ready and available, a preventive maintenance policy is definitely more suitable. The maintenance of the ship hull requires maintenance at fixed intervals to ensure the seaworthiness and safety of the ship, as regulated by International Maritime Organization (IMO). The mechanical parts of the equipment on-board the ship are also subject to wear and tear over time. The degree of wear and tear is dependent on the usage level of the equipment and the vessel.

The interval between replacing parts is advised by the system supplier, who may also be the original equipment manufacturer, whose effective need is also to increase the sales of their products and spare parts. Thus, for the navy, correctly identifying the parts that are required to be replaced is one way to reduce waste and keep maintenance costs low.

CBM is one type of preventive maintenance that can help the navy efficiently determine whether maintenance is required by monitoring certain predefined parameters that are correlated to the occurrence of a failure. Using this predictive method, naval personnel can perform maintenance just before the occurrence of a failure. Conducting a thorough business-case analysis prior to the implementation of CBM is also required to ensure that there is positive cost savings to the maintenance cost (Koyak 2013).

B. LIFE CYCLE COST AND MAINTENANCE COST

In the acquisition of a defense system, the life cycle cost of the system may sometimes be overlooked. The primary consideration in the acquisition project is usually the "short term" cost, which is the initial procurement and acquisition cost

(Blanchard 2014). Apart from initial procurement and acquisition cost, many other related costs may be hiding below the surface. This can be illustrated by the "iceberg" effect shown in Figure 1. Operations and maintenance costs contribute to a large percentage of the total cost of the system throughout the whole life span of the system. For some systems it is estimated to be about 70 to 75 percent of the total cost (Blanchard 2014).

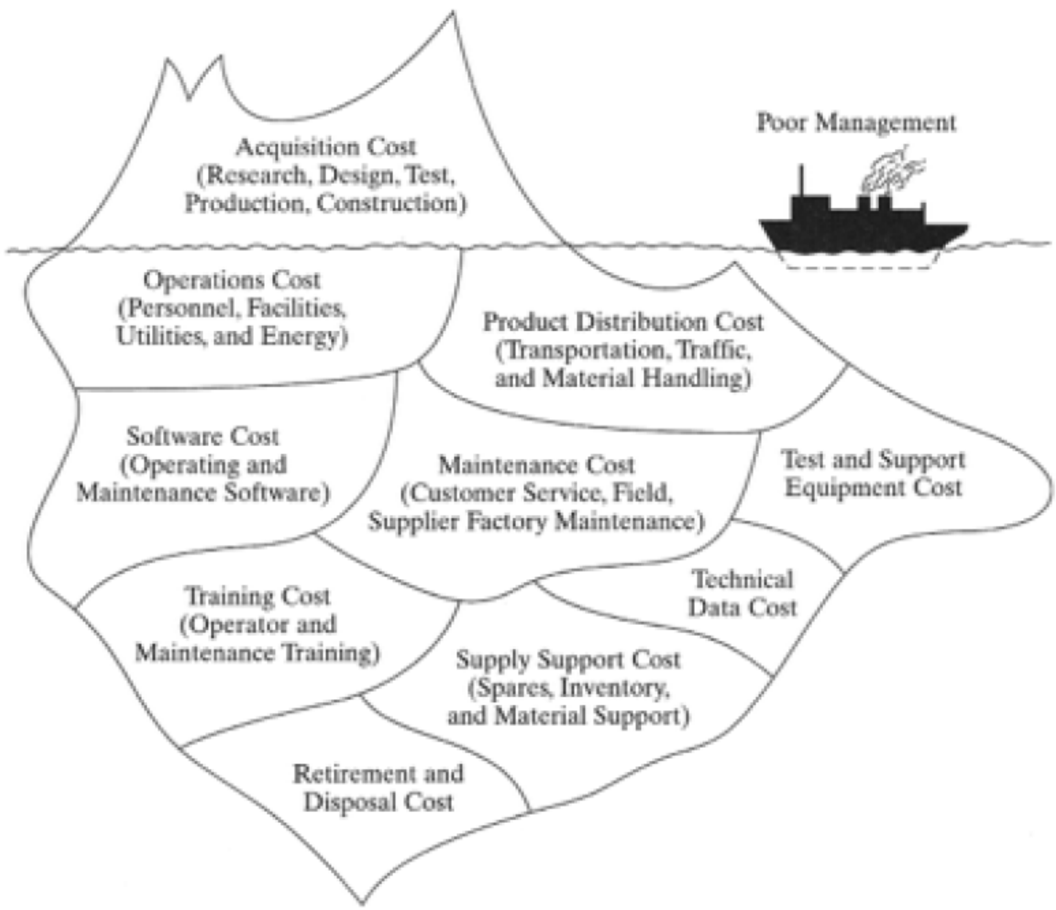

Figure 1. Total Life Cycle Cost Visibility. Source: Blanchard (2014).

The Department of Defense has published the operation and maintenance overview for the fiscal year (FY) 2017 (Office of the Under Secretary of Defense 2016). As mentioned in the report, the FY 2017 estimate for depot maintenance for the U.S. Navy is $6,262 million. Depot maintenance includes overhaul, repair and maintenance of ships for the U.S. Navy. Apart from depot maintenance, a separate amount is requested for day-to-day maintenance under the Facilities Sustainment, Restoration and Modernization (FSRM) program. The Navy is budgeting $1,642.7 million for FSRM for FY 2017. This makes a total of $7,904.7 million for all maintenance for the U.S. Navy.

C. HIGH SPEED SURFACE CRAFT

The vessel from which data is collected for this study is designed and built by Singapore Technologies Marine Limited (STM). This is a new-built project in STM. The vessel, which is designed and built with an aluminium hull, is a light weight, high speed surface craft (HSSC). The vessel is equipped with two MTU 16V 2000 M94 diesel engines, one on the port side and the other on the starboard side. These two engines are connected to two waterjets that thrust the vessel forward or astern. The HSSC is capable of maneuvering sharp turns at high speed while maintaining stability throughout the course.

The HSSC was launched in July 2014, and the commissioning of other platform systems was completed by late 2014. This was then followed by internal sea trials and official sea trials. The data used in this study were collected from this vessel over a period of 13 months, from January 2015 until February 2016.

D. MTU DIESEL ENGINE

As previously mentioned, the diesel engine that powers the HSSC is a MTU series 2000 general purpose four-stroke diesel engine with 16 cylinders. The engine's sensors are connected to the engine control unit (ECU) and engine monitoring unit (EMU). The ECU and EMU use the sensors for engine control and monitoring (MTU Friedrichshafen GmbH [MTU] 2012). The electronic engine control unit provides closed-loop, open-loop, and monitoring of various parameters of the engine through these sensors.

The placement of the sensors by the diesel engine manufacturer is shown in Figures 23 to 27 in the Appendix. These sensors measure exhaust gas temperature, oil temperature, fuel temperature, coolant temperature, oil pressure, coolant pressure, fuel pressure, engine speed, coolant level, and fuel level. One of the parameters that the ECU controls is to maintain the desired engine speed as the operator drives the vessel. As the engine speed changes, parameters correlated to the engine speed also change.

E. SHIP CONTROL AND MONITORING SYSTEM

The ship control and monitoring system (SCMS) is a centralized information system that connects to various platform systems on board the ship. The SCMS issues both audible and visual alarms to the operator when any of the predetermined anomaly conditions are triggered. The operator is also able to remotely control and view the status of the monitored equipment and platform systems through the SCMS. A database in the SCMS logs all events and alarms triggered during the operation period. The database also stores the measured readings of condition monitoring values.

The SCMS is connected to the MTU diesel engine local equipment panel through a serial RS422 interface. Values, status, alarms messages, and warning messages of the MTU diesel engine are transmitted to the SCMS for display to the operator and storage in the SCMS database.

F. PROBLEM STATEMENT

The aims of this thesis are to determine the appropriate statistical method to characterize the data collected from the HSSC and to derive a model for forecasting the behavior of a marine diesel engine. This includes determining the following:

- selection of useful data for model formulation and estimation;
- selection of a condition monitoring parameter that is indicative of system health status;
- approach for characterizing the data;
- method for deriving the model.

After model development, a decision-making process similar to that of a control chart can be used to set the range of normal operation and out-of-control thresholds.

G. ORGANIZATION OF THE THESIS

The thesis is organized in the following manner. Chapter I gives the background and the motivation for the study and analysis. Chapter II presents a literature review of the previous research on CBM and on marine engines. Chapter III describes the tools and methodology adopted to analyze the data collected. Chapter IV presents the results of the analysis and offers some discussion of the results. Chapter V summarizes the contributions of this work and concludes with recommendations for further studies.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

As a maintenance policy, CBM is useful for diagnostic and prognostic purposes. The objective of diagnostic CBM is to identify and find the cause of a failure. One of the common causes of failures is equipment ageing and deterioration. Prognostic CBM aims to estimate the remaining useful life of an entity, allowing time for maintenance to be carried prior to an actual failure.

There are different methods and techniques to perform CBM. Jardine, Lin and Banjevic (2006) present a review of diagnostics and prognostics CBM for the mechanical systems. Diagnostic CBM takes place after a failure has occurred, to find the cause of failure. By contrast, prognostic CBM tries to predict the occurrence of a future failure with currently available data. Methods and techniques related to CBM range from data acquisition to data analysis and decision making. These are the three main steps for CBM.

A. DATA ACQUISITION

Data acquisition is an important process in CBM. It consists of the collection of data from the equipment and then storing them for analysis of the health condition of the equipment. Two types of data are useful for performing analysis in CBM. Event data records what has happened to the equipment, while condition monitoring data records the actual measurements taken that reflect the condition of the equipment at a particular moment, over a certain time period.

Jardine, Lin and Banjevic (2006) mention the use of event data and condition monitoring data for CBM data analysis. They also emphasize the importance of event data towards CBM, since event data contributes to the performance evaluation of the condition monitoring data.

Data is essential to the subsequent steps of CBM. Yam et al. (2001) implemented CBM for a power generation plant, by monitoring the vibration of the planetary gear train of one of the motor-pumps within the power generation plant. Bank et al. developed the Diesel Enhanced Mechanical Diagnostic Test Bed to study the characteristics of diesel engines for healthy and faulty data (Bank et al. 2001). The test bed generates operational data of pressure, temperature, vibration, and displacement for various speeds and loads.

Jiang and Yan study the level of deterioration in marine diesel engines through the concentration of wear particles in the analysis of engine oil samples (Jiang and Yan 2008). Lee monitors the exhaust gas temperature of the diesel engine for a roll-on-roll-off-passenger commercial vessel (Lee 2013). Jardine, Lin and Banjevic note other monitoring parameters, such as acoustic, moisture, humidity, weather or environmental data as condition monitoring data through various types of sensors for CBM (Jardine, Lin and Banjevic 2006).

There are no strict rules on which parameter to use for condition monitoring. From several previous works on CBM of diesel engines, we find the monitoring parameters are not the same. It depends on the data available as well as the availability of resources or special equipment to perform the analysis.

B. DATA PROCESSING AND ANALYSIS

The next step in CBM is data processing and analysis. There are different methods and techniques to process and analyze data. It is dependent on the data collected, whether it is value type, waveform type, or multi-dimension type.

Signal processing is commonly used for waveform and multi-dimension type data, while for value type data, multivariate analysis or trend analysis techniques can be adopted (Jardine, Lin and Banjevic 2006).

Jiang and Yan derive composite scale models for the prediction of failure of the diesel engine based on the data collected from oil analysis (Jiang and Yan 2008). Lee, in his thesis, uses the moving average time series technique to predict future values based on the model derived (Lee 2013).

C. DECISION MAKING

In the work by Yam et al. (2001), the authors develop the intelligent predictive decision support system (IPDSS) which is trained with domain-specific knowledge through recurrent neural networks (RNN). IPDSS provides analytical decisions for fault diagnosis and trend prediction for the deterioration of the equipment. Condition monitoring of the vibration of the planetary gear train is used for prognostic analysis of the equipment. Yam et al. compared the trend of the faults against the predicted trends to verify the performance of their CBM.

Jiang and Yan (2008) construct a multivariate control chart, by setting a threshold for alarms and anomalies to aid in maintenance decision support for CBM. Lee (2013) uses a cumulative sum (CUSUM) control chart to determine whether the process under monitoring is still in control.

Among the common techniques for maintenance decision support, as presented by Jardine, Lin and Banjevic (2006), are artificial intelligence (AI) and statistical approaches. Statistical process control (SPC) is one of the statistical approaches use for CBM. SPC determines if a process is in control by comparing the conditioning monitoring parameter against a reference. The process is in control if the condition monitoring parameter is within a predefined set of control limits.

One of the AI approaches use for CBM is expert systems (ES). ES can provide decision support in a particular domain through an inference engine, based on a high quality domain expert knowledge base. The reasoning methods of the inference engine can be rule-based, case-based or model-based.

THIS PAGE INTENTIONALLY LEFT BLANK

III. TOOLS AND METHODOLOGY

A. R LANGUAGE

R (R Core Team 2015) is an open source statistical programming language, commonly used in statistical research. The advantage of using R is that the developer released R under the GNU general public license, which makes R available at no cost to the user. The software is available for download from the R project website. The license allows the user to obtain the R source code, which the user may modify and distribute to other users of R (Free Software Foundation 2007).

Although R is open source, it is in no way inferior to commercially released statistical software. This powerful programming language can perform data analysis, data manipulation, and statistical modeling; and the results can be plotted in a variety of graphical formats. The main consideration for selection of R language as the analysis tool for this project is the portability of the written code and the availability of the analysis software after the completion of this Master’s thesis, allowing further work related to this project.

B. REGRESSION ANALYSIS

Regression analysis is a technique for deriving a mathematical model that represents the relationship between a dependent variable with one or more independent variables (Ragsdale 2012). The changes of independent variables can influence the values of dependent variables when a relationship between them exists. The model derived through regression analysis should closely reflect the behavior of the dependent variable with respect to the independent variables. One can use this model to predict the response of the dependent variable based on known values of the independent variables.

1. Time Series Analysis

From basic mechanical operating principles, we know that when the engine speed increases, the temperature of the other engine parameters will react in response to the increase in engine speed. It is therefore intuitive to acknowledge that there is autocorrelation in the behavior between the engine speed and the other temperature parameters. Hence, one can use a time series regression approach to express the behavior of the engine in the presence of the variable speed.

2. Autoregressive Distributed Lag (ARDL) Time Series

The observations on a measurable variable acquired over a period of time form a time series (Ragsdale 2012). One type of time series is an autoregressive distributed lag (ARDL) time series. This ARDL time series is a commonly used statistical and econometrical analysis technique. An ARDL time series model can be used to describe the behavior of a dependent variable, Y_t, expressed in terms of a constant y-intercept, μ, a disturbance, ε_t, an independent variable, X_t, their past, X_{t-j}, and its own past, Y_{t-i}. Lags in the time series are values of past periods. The number of lags used in the model determines how many past periods can affect the dependent variable.

An ARDL time series model (Greene 2000) can be expressed as:

$$Y_t = \mu + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=0}^{r} \beta_j X_{t-j} + \varepsilon_t$$

with

$$\varepsilon_t \sim i.i.d\ \ N(0, \sigma_\varepsilon^2)$$

and, p and r are the lag length for the dependent variable, Y_t, and independent variable, X_t, respectively. One can rewrite the ARDL time series model (Greene 2000) as:

$$C(L)Y_t = \mu + B(L)X_t + \varepsilon_t$$

where

$$C(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \cdots - \gamma_p L^p$$

and

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_r L^r$$

Here, L refers to the lag operator.

3. Assumptions of Linear Regression

One of the assumptions of linear regression performed using ordinary least square (OLS) method is that the residuals are normally distributed. A nonparametric robust regression approach may be appropriate if the residuals have a heavy-tailed distribution. Instead of minimizing the sum of squared errors in OLS, the nonparametric robust regression approach minimizes the sum of absolute errors. Furthermore, a robust autocorrelation function that uses Spearman ranked correlation in the residual analysis can be used to measure the autocorrelation in the residuals in a nonparametric manner.

4. Independent and Dependent Variables for Regression

In this thesis we derive a regression model based on the data collected from an MTU series 2000 general purpose diesel engine on the port side of the HSSC. The data, sampled at an interval of one second, are measured by sensors placed on various parts of the engine by the engine manufacturer. The selection of a dependent variable with one or more independent variable is required for the regression analysis. The independent variable is the known and controlled variable. In this case, among the data collected, the obvious choice for the independent variable is the engine speed parameter. The engine speed is controlled by the helmsman as the driver of the vessel.

The dependent variable is the variable that is of interest for prediction. The selection of the dependent variable should be appropriate for the system under consideration. A previous study in the area of CBM for marine diesel engines is Lee (2013), who uses exhaust gas temperatures as the monitoring parameter for CBM. Lee anticipates that high exhaust gas temperatures could lead to possible failure of the main engine, as it may be an indication of problems in the combustion chamber. High exhaust gas temperature would also add to the level of thermal stress for the surrounding components and materials. From the data collected, the operating temperature for exhaust temperature can be as high as 700 degrees Celsius. This temperature range is higher than for other measured parameters. Thus, we take exhaust temperature as the dependent variable in our regression model.

C. METHODOLOGY

The methodology to characterize the data collected and derive the model for prediction of the selected engine parameter is shown in Figure 2. There are six steps in the process. The data collected are first selected according to the selection criteria, then analyzed to determine their properties. The next step is to categorize the data according to their properties. Based on the different groups, we can derive models for each group. We can then use the models for prediction, and we can verify the results of the prediction against a new set of data.

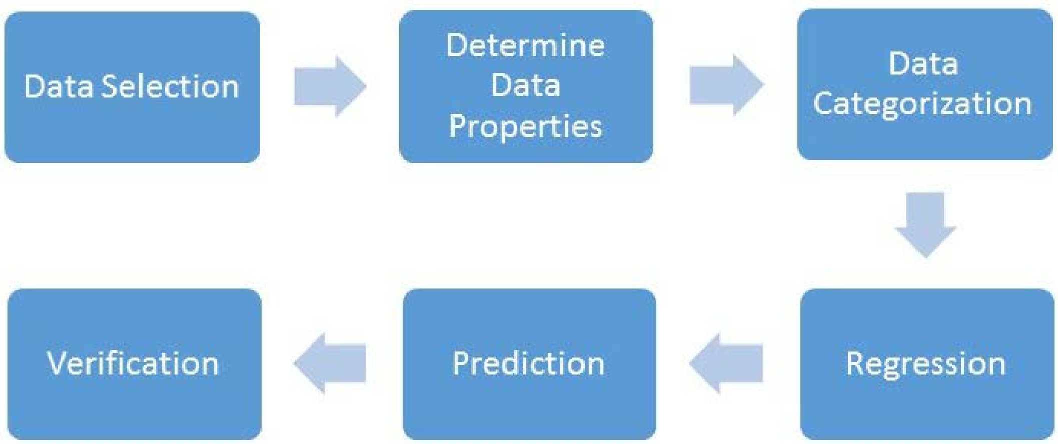

Figure 2. Methodology for Regression Analysis

1. Data Selection Criteria

It is not meaningful to analyze the data when the main engine is idling or when there is no load attached to the engine. To ensure that our analysis is meaningful, the data we use for the analysis should only be obtained when the HSSC is in operation and for a sufficient period to observe the operation of the vessel under usage conditions. Therefore, the following data selection criteria for the selection of useful data for analysis are derived:

- The data should be selected from the period when the clutch is engaged.
- The engine should be sufficiently warmed up.
- The engine speed should be above its idling speed.
- The clutched-in duration should be sufficiently long to display the relationship between variables.

2. Data Properties

The collected data are analyzed to derive a predictive model for the dependent variable. This analysis quantifies the linear relationship between variables using the correlation coefficient. The measured strength of the relationship ranges from -1 to 1, with a positive value indicating an up slope direction and a negative value indicating a down slope direction. An absolute value that is closer to 1 indicates stronger relationship between the two variables.

There are two exhaust temperature parameters in the data collected, combined A exhaust temperature and combined B exhaust temperature. Combined A exhaust temperature measures exhaust gas temperature of cylinders located on the left side of the engine, and combined B exhaust temperature measures exhaust gas temperature of cylinders on the right side of the engine. For our analysis, the parameter with a higher correlation coefficient is chosen as the dependent variable for regression analysis.

3. Data Categorization

The data to be used for regression is dependent on the definition of “normal” operating characteristics of the vessel. Selecting the sets of data that exhibit normal operating characteristics is critical to the derivation of a model for predicting the behavior of the dependent variable during normal operation. We choose engine speed as the independent variable for regression. Correctly categorizing the datasets into groups having similar operating conditions creates better models for prediction. The data may be grouped into different categories. There are different categories of “normal” operation. One way of categorizing the datasets could be in terms of engine speed profile. The MTU series 2000 general purpose diesel engine can operate with sequential turbocharger once the engine speed exceed 2000 RPM. The first category uses datasets for engine operation without the turbocharger. These datasets show the engine speed not exceeding 2000 RPM on average and sustained speeds over the analyzed duration. The other category uses datasets for the engine operating with turbocharger, with average and sustained speeds over the analyzed duration of over 2000 RPM.

The duration of the dataset also affects the effectiveness of the model. A longer operating duration can capture more details throughout the operation. Therefore, the datasets are categorized into modeling sets and prediction sets. The datasets used for regression modeling are longer than 30 minutes, while the datasets used for prediction are not required to be longer than 30 minutes.

4. Regression Model

To derive the ARDL time series model, we must estimate the following unknowns:

- coefficients of lagged dependent variable;
- coefficients of current and lagged independent variable;
- intercept value;
- lag length of the model.

For each speed profile category, a separate regression model is estimated. Once we have derived all the regression models for all the datasets within the same category, we derive the overall regression model for that particular category by taking the average of all the corresponding coefficients. Using this method, we can derive the intercept value and coefficients of current and lagged dependent and independent variables.

The overall regression model can also be expressed mathematically as:

$$Y_{t_{final}} = \bar{\mu} + \sum_{i=1}^{p} \bar{\gamma}_i Y_{t-i} + \sum_{j=0}^{r} \bar{\beta}_j X_{t-j} + \varepsilon_t$$

with

$$\bar{\mu} = \frac{1}{k} \sum_{s=1}^{k} \mu_s$$

and

$$\bar{\gamma}_i = \frac{1}{k} \sum_{s=1}^{k} \gamma_{i_s}$$

and

$$\bar{\beta}_j = \frac{1}{k} \sum_{s=1}^{k} \beta_{j_s}$$

where k is the total number of datasets in the category.

The lag length can be estimated by examining the autocorrelation of the residuals. After estimating the regression model on each dataset, we inspect the autocorrelation function of the residuals, adding lags to the regression model until the residuals approach the behaviour of white noise (Parker 2012).

5. Prediction

Some datasets are set aside for the purpose of prediction to test an estimated model. The independent variable of these datasets set aside for prediction is injected into the derived model to predict the values for the dependent variable. We discuss this further in Chapter IV, Section D3.

6. Verification

The simplest way to verify the prediction is by plotting the predicted values against the actual values for comparison. Next, we plot the residuals of the results of the prediction to determine deviation of predicted results from the actual values. Then, there are residual analysis techniques that we can adopt to verify the results of the prediction. One of these residual analysis techniques is performing autocorrelation function on the residuals. This is to see if there is any autocorrelation present in the residuals. Another technique is normal quantile-quantile (QQ) plots of the residual, to test for the normality of the residuals.

IV. DATA ANALYSIS AND RESULTS

A. COLLECTED DATA

The data used in this analysis is collected from the main engine on the port side of the HSSC, over a period of 13 months from January 2015 to February 2016. Sensors on the MTU diesel engine measure these values, which are transferred to and stored in the SCMS database. There are two types of data collected. They are events data and condition monitoring data.

Events data logs the status of the equipment. These data keep track of what happened during the day, for example, what time the main engine is powered up, what time the clutch is engaged or disengaged, whether the clutch is engaged for ahead or astern.

The other type of data collected automatically is condition monitoring data. These are the data that can be used for analysis. The condition monitoring data, sampled at a time interval of one second, is a measure of analog values of parameters such as temperature, pressure, voltages, and speed.

Events data related to usage can add meaning to condition monitoring data. It can aid in understanding for which period of time the condition monitoring data is useful for analysis. The selection of useful data for analysis is usually performed prior to data processing and analysis.

1. Dataset

The events data collected from the HSSC for the period from January 2015 to February 2016 can be used to determine whether a particular period in time is for analysis and derivation of the prediction model. Based on the selection criteria defined in Chapter III, Section C1, we should only use the data collected when the clutch is engaged. Table 2 lists the dates and duration when the HSSC had its port side engine and clutch engaged.

Table 2. Data and Duration of Engaged Clutch of HSSC

Dataset	Date	Duration
1	9 Apr 2015	02:31:36
2	18 Jun 2015	02:24:10
3	18 Jun 2015	02:15:07
4	24 Jun 2015	00:47:23
5	25 Jun 2015	01:27:34
6	25 Jun 2015	00:20:44
7	26 Jun 2015	02:00:35
8	26 Jun 2015	00:21:17
9	26 Jun 2015	00:28:44
10	26 Jun 2015	01:08:39
11	30 Jun 2015	01:07:34
12	30 Jun 2015	01:11:22
13	30 Jun 2015	02:16:22
14	2 Jul 2015	01:58:50
15	2 Jul 2015	01:31:24
16	3 Jul 2015	03:01:29
17	6 Jul 2015	01:36:06
18	6 Jul 2015	01:57:22
19	8 Jul 2015	01:38:08
20	8 Jul 2015	01:04:02
21	8 Jul 2015	02:55:12
22	15 Jul 2015	00:36:51
23	26 Aug 2015	00:16:08
24	26 Aug 2015	01:00:47
25	27 Aug 2015	13:28:28
26	20 Nov 2015	00:24:49
27	23 Nov 2015	00:29:27
28	25 Nov 2015	00:31:06
29	26 Nov 2015	00:19:12
30	7 Dec 2015	02:27:29
31	19 Jan 2016	00:31:22
32	20 Jan 2016	02:59:04
33	10 Feb 2016	00:30:18
34	25 Feb 2016	00:18:29

Table 2 shows that the HSSC does not have a fixed operating time and period every day. There are days when the clutch is engaged and disengaged multiple times, and there are days that there is no operation at all. The duration of operation also varies, ranging from the longest at 13 hours to the shortest at 16 minutes. There are periods when the HSSC clutch is engaged for 10 minutes or less. Those data are not listed in Table 2.

For Dataset 25, the HSSC had its clutch engaged for 13 hours. This extended period of clutch engagement is an unusual pattern of operation when compared with the rest of the datasets. This may suggest that the HSSC is involved in a long operational activity or sea-trial on that day. As the HSSC is in its testing and evaluation phase during the data collection period, it is not surprising that the HSSC is scheduled for sea-trial to test equipment performance at open sea for an extended period of time in a day. We should consider this an independent event, not a usual operation pattern. Since Dataset 25 meets the selection criteria, this dataset is not discarded.

2. Data Variables

The data collected from the MTU diesel engine are analog values of measured variables. The variables, with their respective units, and the range of the values are summarized in Table 3. The engine speed is measured and recorded in the range from 0 to 3000 RPM. The idling speed for the engine is at 600 RPM. The engine goes into its idling speed once the engine is powered up and no load is applied to the engine. The temperature of variables in fluid form, such as coolant, lube oil, and fuel, are measured and recorded in the range from -20 degrees Celsius to 120 degrees Celsius. Charge air temperature is measured and recorded in the range from 0 degrees Celsius to 120 degrees Celsius. For exhaust temperature, the measured and recorded temperature range is not documented in the manual (MTU 2007). But based on the data collected, it could reach up to 600 or 700 degrees Celsius. Coolant pressure and gear lube oil pressure, in bars, are not documented as well. Lube oil pressure is between 0 to 10 bars, and gear control oil pressure is between 0 to 30 bars. Power supply voltage is measured and recorded between 0 and 35 volts.

Table 3. Diesel Engine Variables Collected. Adapted from MTU (2007).

Variables	Units	Minimum Value	Maximum Value
Engine speed	RPM	0	3000
Coolant temperature	Degree Celsius	-20	120
Coolant pressure	Bar	-	-
Lube oil temperature	Degree Celsius	-20	120
Lube oil pressure	Bar	0	10
Exhaust temperature	Degree Celsius	-	-
Fuel temperature	Degree Celsius	-20	120
Charge air temperature	Degree Celsius	0	120
Gear control oil pressure	Bar	0	30
Gear lube oil pressure	Bar	-	-
Power supply voltage	Volt	0	35

The sensors on the engine measure these variables and then electrically transmit them to the engine local control panel and to the SCMS through a serial RS422 interface at the sampling interval of one second, for display, storage and offline analysis.

3. Data Plots

The data collected are dependent on the operation of the HSSC. Intuitively, the temperature and pressure parameters are dependent on the engine speed. As the load or demand on the engine increases, the engine speed increases. This in turn leads to the change in temperature and pressure parameters. On the other hand, we do not expect the power supply voltage to change with respect to the engine speed. Yet, only through further analysis can we determine how each individual variable is related to the engine speed. Engine speed from sample dataset, Dataset 19, is plotted to observe the behavior of the engine and its relationship with the other parameters.

a. Engine Speed

Dataset 19 is collected over a period of 1.5 hours from about 09:30 hr until 11:05 hr, as shown in Figure 3. During that period of time, we observe that at around 09:30 hr, the engine is picking up speed, and it reaches about 2200 RPM for a period of 10 minutes, then fluctuating between 2500 RPM and 1300 RPM, and then settling down at about 1800 RPM at around 10:30 hr. At about 10:50 hr, the engine speed once again increases to about 2500 RPM, before returning back to its idling speed at around 11:00 hr. The engine speed is controlled by the helmsmen as he drives the ship, moving the throttle forward to increase the engine speed, and moving the throttle backward to reduce speed. The engine speed may also affect other parameters.

b. Temperature

The temperature parameters of Dataset 19 are plotted against time, as shown in Figure 4. We observe that lube oil temperature, coolant temperature, combined A exhaust gas temperature, and combined B exhaust gas temperature exhibit a similar trend with each other. From the plots, we also observe that the operating temperature of combined A exhaust gas temperature and combined B exhaust gas temperature can reach up to 700 degrees Celsius. Charge air temperature reflects an inverse trend with that of combined A exhaust gas temperature and combined B exhaust gas temperature. Fuel temperature seems to have a longer delayed reaction than the other temperature parameters.

c. Pressure

The plots of pressure against time are shown in Figure 5. From the plots, we observe that gear control oil has a jump in pressure once the clutch is engaged. The trend of gear control oil pressure, gear lube oil pressure, and coolant pressure exhibit a similar trend closely resembling that of engine speed. Lube oil pressure does not show a clear similarity to the other pressure plots. Another point to note is that gear control oil pressure operates at a range much higher than the other pressure variables. Gear control oil pressure operates above 20 bars, while the other pressure parameters operate at less than 10 bars.

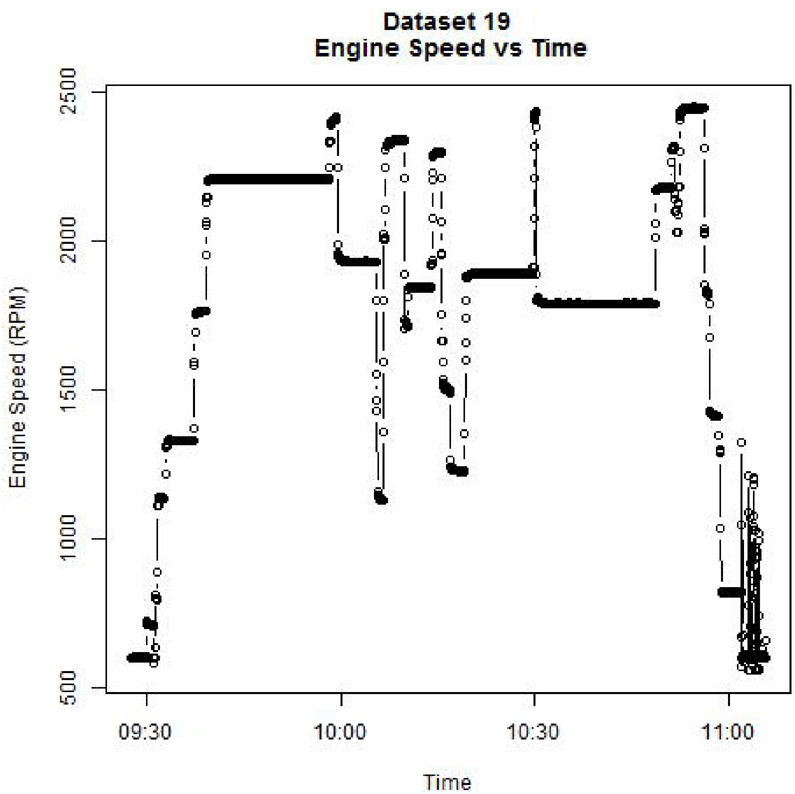

Figure 3. Engine Speed Plotted against Time for Dataset 19

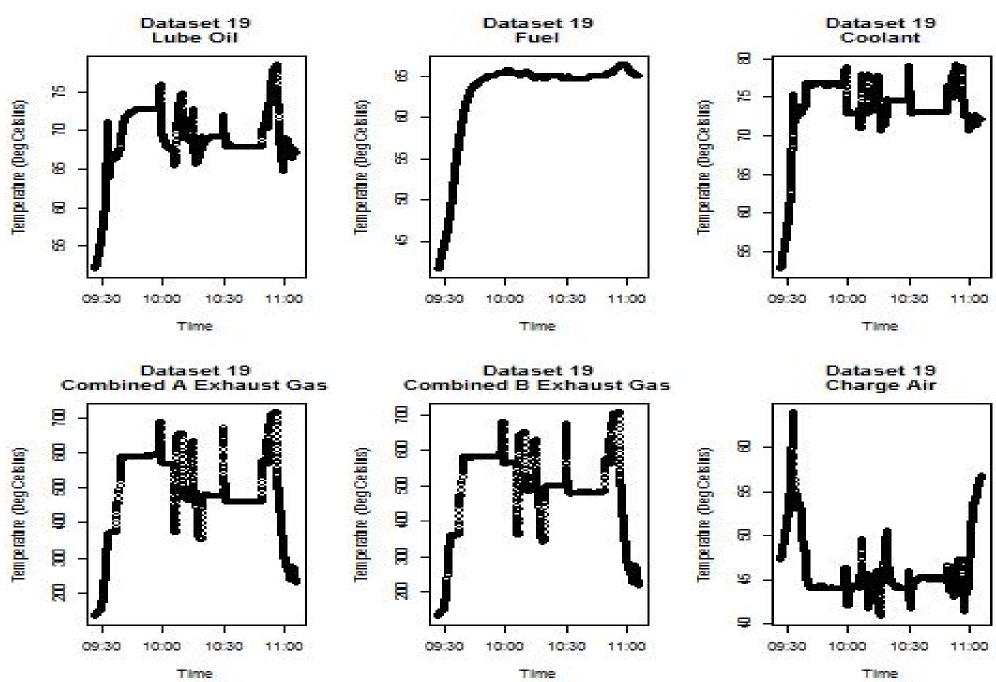

Figure 4. Temperature Parameters Plotted against Time for Dataset 19

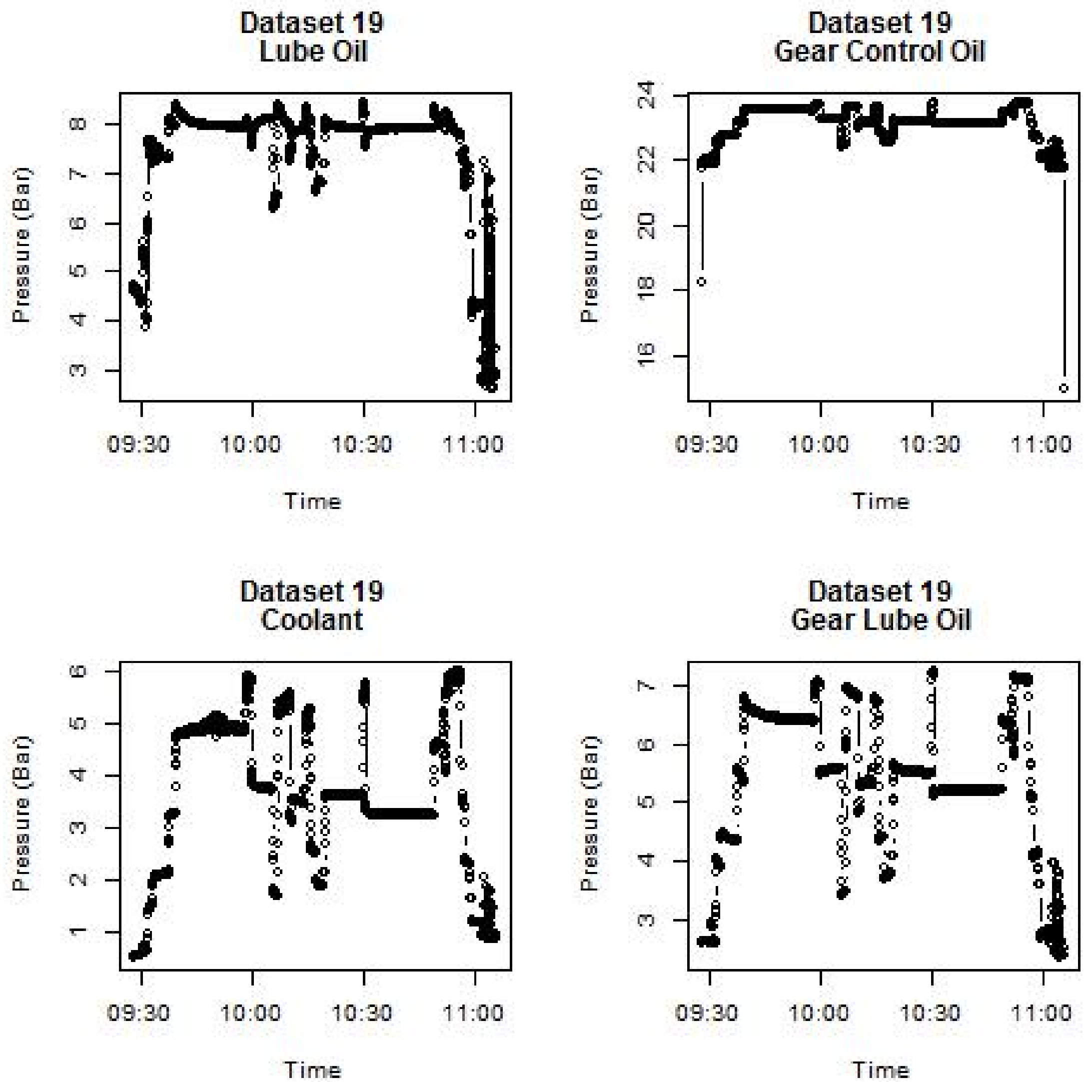

Figure 5. Pressure Parameters Plotted against Time for Dataset 19

d. Voltage

The power supply voltage plot is shown in Figure 6. The power supply voltage fluctuates between 25 V and 27 V. There is no obvious similarity in trends between the power supply voltage and the engine speed. The fluctuation in voltage could be due to noise or other factors.

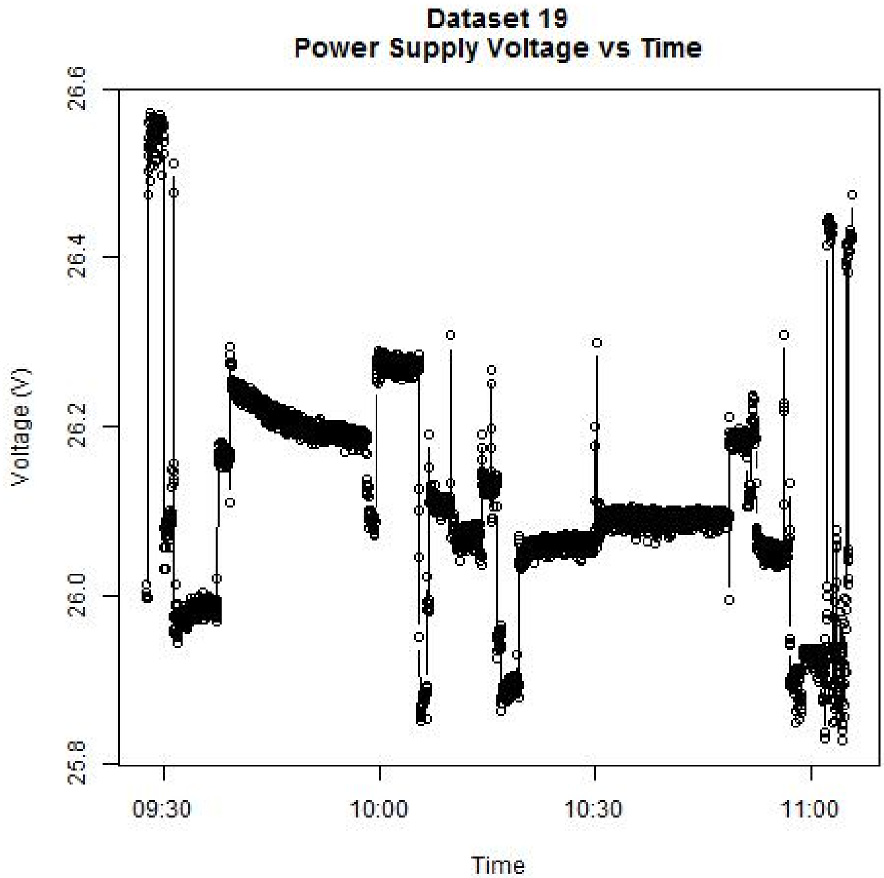

Figure 6. Power Supply Voltage Plotted against Time for Dataset 19

B. DATA SELECTION

The selection criteria for data to be used for analysis are discussed in Chapter III, Section C1. The selection of the duration for Dataset 1 to Dataset 34 is based on the events data that logged the exact time the clutch is engaged. Each dataset records a minimum operating duration of at least 15 minutes.

Coolant is the fluid that flows through the engine to cool the running engine. The heat exchange between the running engine and the coolant causes the temperature of the coolant to increase. Therefore, the coolant temperature is a good indicator to show that the engine is warmed up. To ensure that the engine is sufficiently warmed up, a coolant temperature of 70 degrees Celsius is used to mark the beginning of the selected data within each dataset. Table 4 shows the number of observations in each dataset after the engine is warmed up.

Table 4. Number of Observations of Selected Data for Analysis

Dataset	Number of Observations	Dataset	Number of Observations
1	9097	18	7042
2	8084	19	5542
3	8108	20	3843
4	2844	21	10,513
5	4916	22	1457
6	1245	23	968
7	6788	24	3647
8	1278	25	48,243
9	1725	26	792
10	4119	27	849
11	3623	28	1866
12	4283	29	743
13	8183	30	8298
14	6756	31	1882
15	5485	32	10,563
16	10,495	33	1239
17	5475	34	816

C. CORRELATION COEFFICIENT

Table 5 shows the correlation coefficient between the engine speed and the temperature parameters. The datasets highlighted in red have very low correlation coefficients. These datasets shows that the HSSC is operating at the engine idling speed, and do not meet the selection criteria. Therefore, these datasets are discarded. Among the temperature parameters shown in Table 5, combined B exhaust gas temperature has the highest correlation coefficient for most of the datasets. Selecting combined B exhaust gas temperature for the dependent variable is supported by the results of the correlation coefficient.

Table 5. Correlation Coefficient of Engine Speed and Temperature Variables

Dataset	Lube Oil Temp	Fuel Temp	Coolant Temp	Exhaust Gas A	Exhaust Gas B	Charge Air Temp
1	0.7016	0.4016	0.9058	0.9777	**0.9830**	-0.8735
2	0.1079	-0.3156	0.3814	0.7267	**0.7375**	-0.4583
3	0.2070	0.1634	0.6892	0.8060	**0.8320**	-0.4118
4	0.0552	0.2629	0.7259	0.8017	**0.8203**	-0.3846
5	0.7484	0.0356	0.7166	0.8296	**0.8534**	-0.2431
6	0.7038	0.2938	0.7467	0.9002	**0.9156**	-0.4611
7	*0.1299*	*0.1668*	*0.1102*	*0.3734*	*0.3923*	*-0.0552*
8	*0.1591*	*-0.0475*	*0.0187*	*0.1406*	*0.1619*	*0.0800*
9	-0.3981	0.7070	0.6799	0.9144	**0.9234**	-0.7510
10	0.5753	0.1031	0.7813	0.8901	**0.9167**	-0.6468
11	0.8460	0.1682	0.9014	0.9573	**0.9616**	-0.7162
12	0.8675	0.5597	0.9254	0.9458	**0.9533**	-0.3343
13	0.8477	0.7049	0.9200	0.9615	**0.9672**	-0.7113
14	0.7806	0.0296	0.8416	0.9254	**0.9325**	-0.5630
15	0.7716	0.6156	0.9004	0.9688	**0.9740**	-0.7308
16	0.7639	0.2209	0.8790	0.9584	**0.9662**	-0.7035
17	0.7789	0.3755	0.7493	0.8670	**0.8826**	-0.2803
18	0.6656	0.3508	0.7844	0.9330	**0.9554**	-0.7376
19	0.7024	0.1548	0.7869	0.9211	**0.9445**	-0.5676
20	0.6893	0.6454	0.8304	0.9423	**0.9500**	-0.7190
21	0.5505	0.4663	0.8618	0.9670	**0.9687**	-0.8690
22	-0.1663	-0.4421	0.4072	0.8289	**0.8477**	-0.5016
23	*-0.0253*	*0.0410*	*-0.0396*	*0.0256*	*0.0247*	*-0.0274*
24	0.1819	0.0785	0.8008	0.8871	**0.8966**	-0.5934
25	0.6372	0.1267	0.8929	0.9645	**0.9701**	-0.8919
26	0.4243	-0.1786	0.8548	0.9202	**0.9290**	-0.4060
27	0.3906	-0.5012	0.7062	0.9434	**0.9515**	-0.3173
28	0.4128	0.7634	0.8187	**0.9662**	0.9642	-0.8142
29	0.3533	-0.5127	0.8204	0.9319	**0.9439**	-0.3894
30	0.8198	-0.0538	0.8925	0.9447	**0.9528**	-0.6727
31	0.1100	-0.0341	0.6525	0.8433	**0.8550**	-0.5659
32	0.7199	0.2431	0.7953	0.9111	**0.9202**	-0.6337
33	-0.1652	-0.8665	0.6785	0.9044	**0.9319**	-0.6818
34	0.0588	-0.5947	0.7329	**0.9710**	0.9674	-0.2685

Rows with very low correlation coefficient are highlighted in red. Highest correlation coefficient for each dataset is highlighted in bold.

D. ENGINE SPEED PROFILE

The engine speed is an uncontrolled, independent variable. As discussed in Section A3, there is correlation between the behavior of the engine speed parameter and the temperature and pressure parameters; the temperature and pressure parameters are dependent on the engine speed.

The engine speed is dependent on the position of the throttle at the bridge of the vessel. The helmsman increases the throttle level to increase the speed of the engine so that the vessel can travel at a higher speed and decreases the throttle level to reduce the speed of the engine, which in turn reduces the speed of the vessel. The speed profile of the engine is critical for the analysis of the behavior of the engine.

Categorizing the engine speed into different speed profiles aids in the data analysis and processing. The engine speed data can be categorized by their speed and duration. At an engine speed greater than 2000 RPM, the engine operates with turbocharger. So the first category is for datasets with sustained engine speed greater than 2000 RPM. Another category contains datasets of sustained engine speeds lower than 2000 RPM. In terms of duration, one category reflects when the clutch is engaged for more than 30 minutes, and another is for when the clutch is engaged for less than 30 minutes. A summary of the speed profile is shown in Table 6.

Table 6. Engine Speed Profile

Dataset	Sustained Speed > 2000 RPM	Sustained Speed < 2000 RPM	No sustained speed	Clutch Engaged > 30 minutes	Clutch Engaged < 30 minutes
1	X	-	-	X	-
2	-	-	X	-	-
3	-	-	X	-	-
4	-	X	-	X	-
5	-	X	-	X	-
6	X	-	-	-	X
7	-	-	X	-	-
8	-	-	X	-	-
9	-	X	-	-	X
10	-	X	-	X	-
11	X	-	-	X	-
12	X	-	-	X	-
13	X	-	-	X	-
14	X	-	-	X	-
15	X	-	-	X	-
16	X	-	-	X	-
17	X	-	-	X	-
18	X	-	-	X	-
19	X	-	-	X	-
20	X	-	-	X	-
21	X	-	-	X	-
22	-	X	-	-	X
23	-	-	X	-	-
24	-	X	-	X	-
25	X	-	-	X	-
26	X	-	-	-	X
27	-	X	-	-	X
28	X	-	-	-	X
29	X	-	-	-	X
30	X	-	-	X	-
31	-	X	-	-	X
32	X	-	-	X	-
33	-	X	-	-	X
34	-	X	-	-	X

1. Discarded Datasets

Five datasets are discarded. These five datasets show spikes at various engine speeds, but they do not exhibit any sustained engine speed above the engine idling speed. These engine speed graphs do not reflect any load on the engine. As discussed in Chapter III, Section C1 on data selection criteria, the engine speed should be above its idling speed; these five datasets do not meet these criteria. Therefore, these five datasets could not be used for the analysis.

2. Datasets Used for Modeling

The datasets used for deriving the regression model are those that have the clutch engaged for a duration of more than 30 minutes. We separate these datasets further into datasets for turbocharger mode operation and cruising mode operation based on whether their operating engine speed is a sustained engine speed greater than 2000 RPM. The model for the turbocharger mode of operation uses the datasets with sustained engine speeds greater than 2000 RPM for derivation. The model for cruising mode of operation uses those datasets with sustained engine speeds lower than 2000 RPM for derivation. We use a total of 15 datasets for deriving the regression model for turbocharger mode, and four datasets for deriving the regression model for cruising mode, as shown in Table 7.

Table 7. Regression Model Datasets for Turbocharger and Cruising Mode

Mode	Datasets
Turbocharger	Dataset 1, Dataset 11, Dataset 12, Dataset 13, Dataset 14, Dataset 15, Dataset 16, Dataset 17, Dataset 18, Dataset 19, Dataset 20, Dataset 21, Dataset 25, Dataset 30, Dataset 32.
Cruising	Dataset 4, Dataset 5, Dataset 10, Dataset 24.

3. Datasets Used for Prediction

The datasets that we use for predictions have not been used for modeling. These datasets are those that have the clutch engaged durations of less than 30 minutes. From Table 8, there are four datasets for predictions in turbocharger mode and six datasets for predictions in cruising mode.

Table 8. Prediction Datasets for Turbocharger and Cruising Mode

Mode	Datasets
Turbocharger	Dataset 6, Dataset 26, Dataset 28, Dataset 29.
Cruising	Dataset 9, Dataset 22, Dataset 27, Dataset 31, Dataset 33, Dataset 34.

E. LAGGED TIME SERIES

The number of lags in the lagged time series can be estimated by taking the autocorrelation of the residuals after linear regression. Lags are added to the model until the autocorrelation in the residuals is removed and the results of the autocorrelation of the residuals becomes like white noise.

The number of lags to add starts from one. This is then followed by adding another lag after each residual autocorrelation, and increases up to 15 lags. This is repeated for all datasets.

The results of autocorrelation and normal QQ plots from Dataset 19 are presented here to illustrate the effects of lag length on autocorrelation and QQ plots. The autocorrelation results for 1, 5, 10, and 15 lags are shown in Figures 7, 9, 11, and 13 respectively. And the QQ plots for 1, 5, 10, and 15 lags are shown in Figure 8, 10, 12, and 14, respectively. The normal QQ plots show roughly a straight line when the distribution of the data is normal.

1. ARDL with One Lag

By introducing one lag to the ARDL time series model, we obtain the autocorrelation of the residuals of the regression, which is shown in Figure 7. With only one lag, the residuals are autocorrelated across most time periods. The QQ plot of the residuals for normality shows an upside down ‘S’ with heavy tail, indicating that the residuals are not normally distributed, and there are outliers with both extreme positive and negative values as shown in Figure 8.

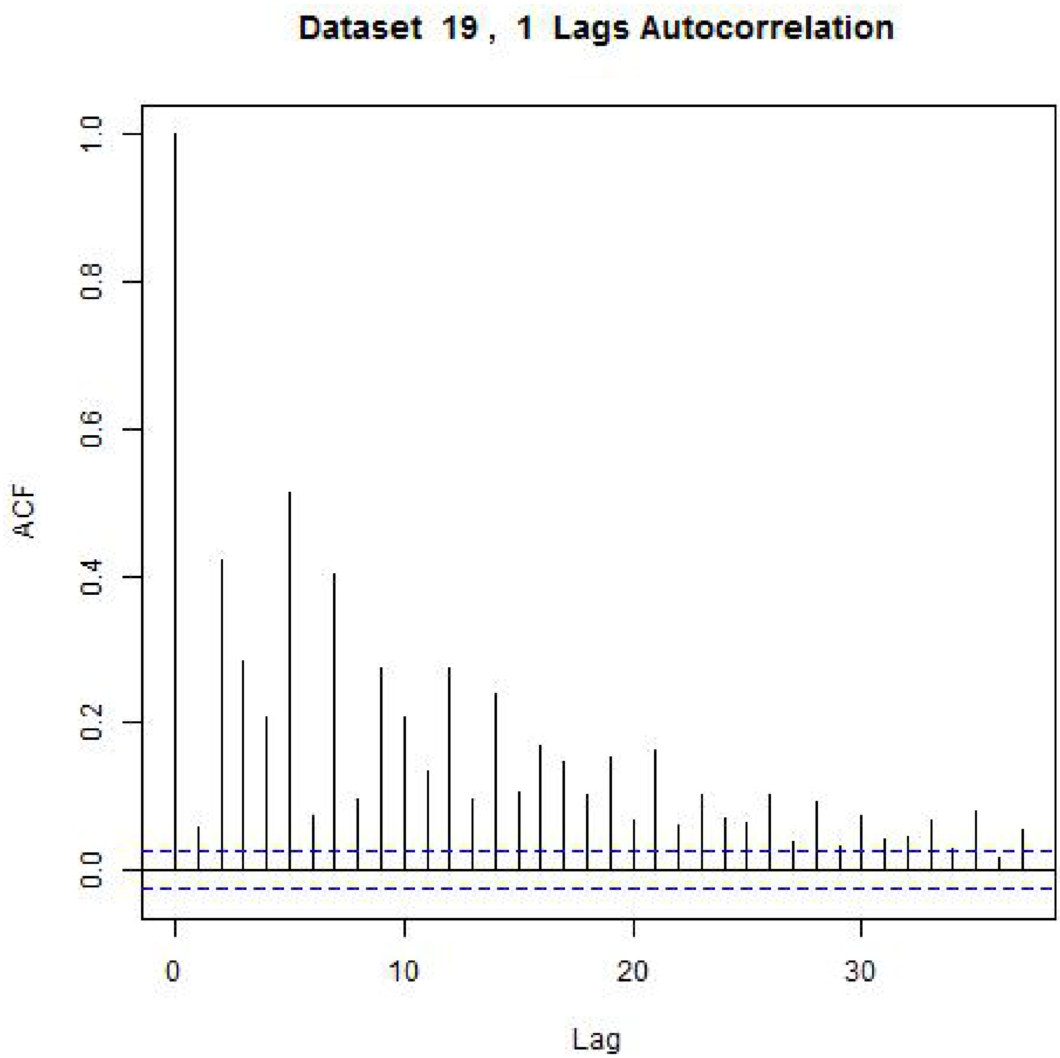

Figure 7. Autocorrelation Function for Time Series with One Lag

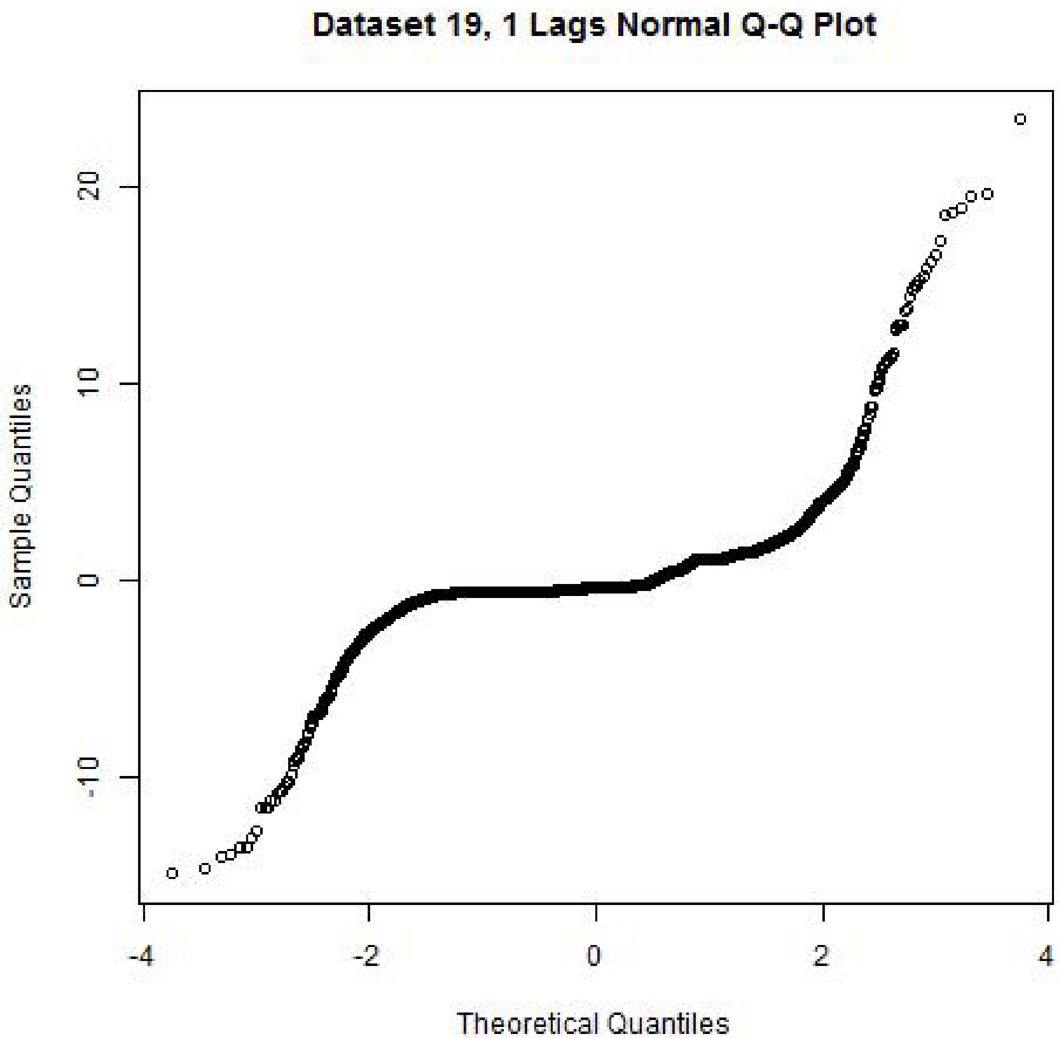

Figure 8. Normal QQ Plot for Time Series with One Lag

2. ARDL with Five Lags

The autocorrelation of the residuals of the regression with five lags is shown in Figure 9. There is a significant reduction in the autocorrelation, which is statistically significant, between time periods. The QQ plot of the residuals for normality still exhibit heavy-tailed characteristics, as shown in Figure 10. Therefore, the residuals are also not normally distributed.

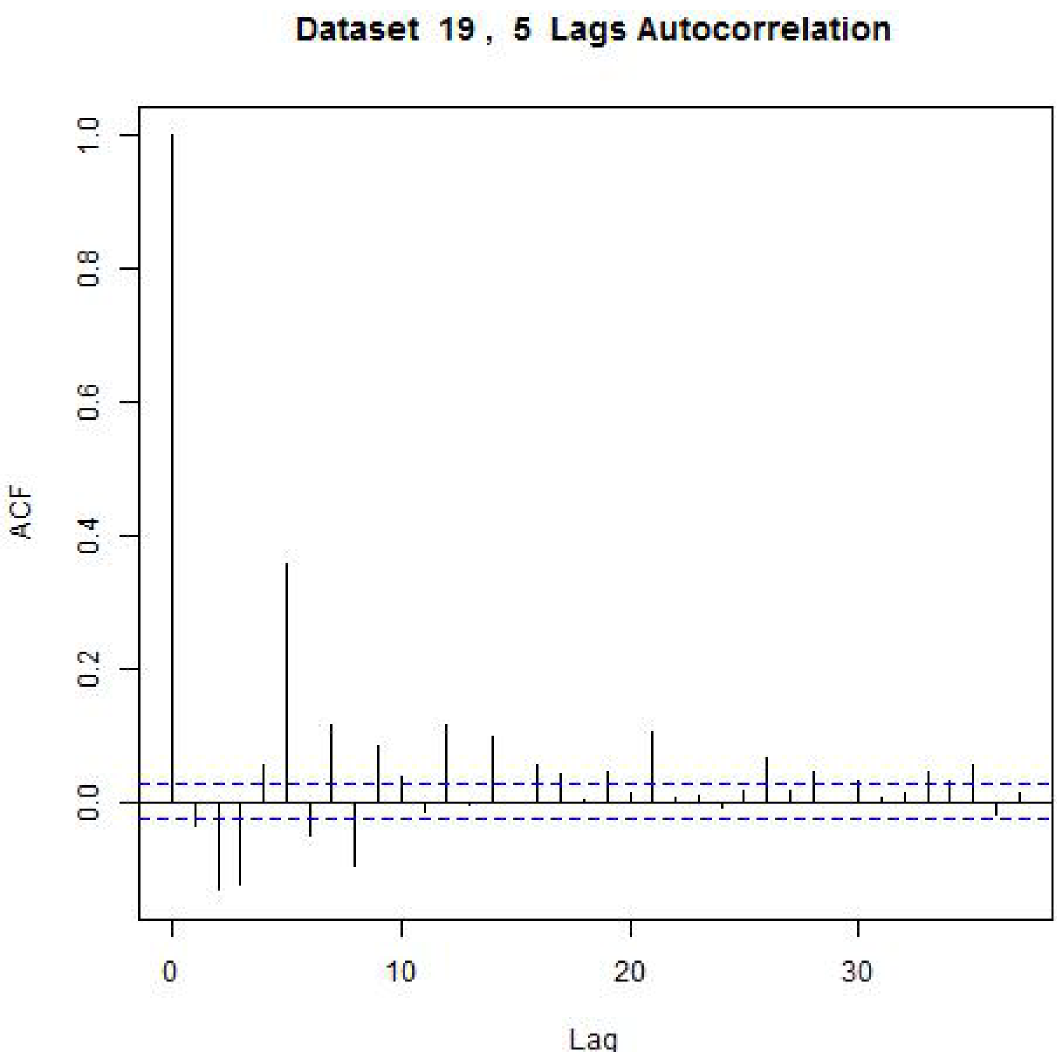

Figure 9. Autocorrelation Function for Time Series with Five Lags

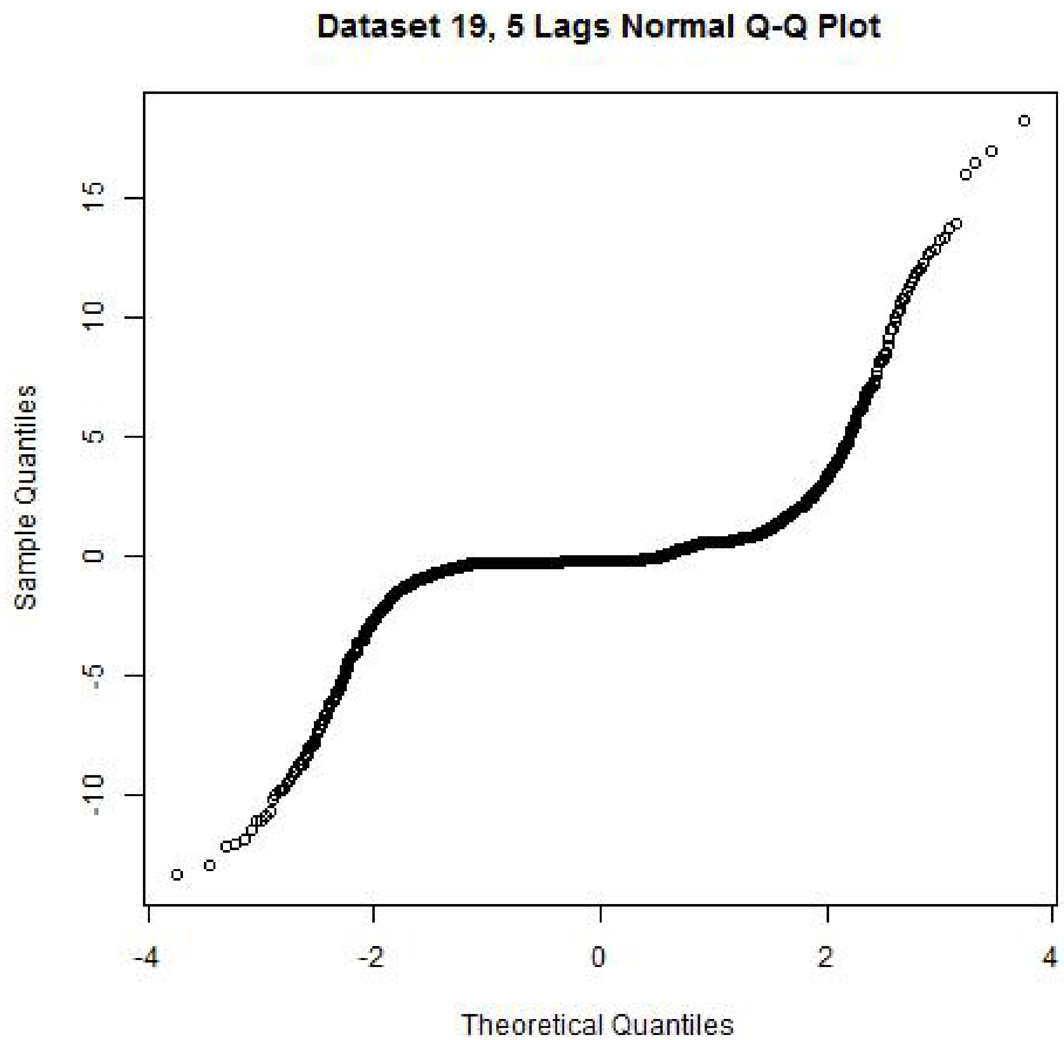

Figure 10. Normal QQ Plot for Time Series with Five Lags

3. ARDL with 10 Lags

When there are 10 lags in the ARDL time series, the autocorrelation of the residuals of the regression is as shown in Figure 11. There is a further significant reduction in the autocorrelation, with only three time periods that are statistically significant. The QQ plot of the residuals shown in Figure 12 exhibits heavy-tailed characteristics, with an upside down ‘S’ shape, not tracking a straight line at all. This shows that the residuals are not normally distributed.

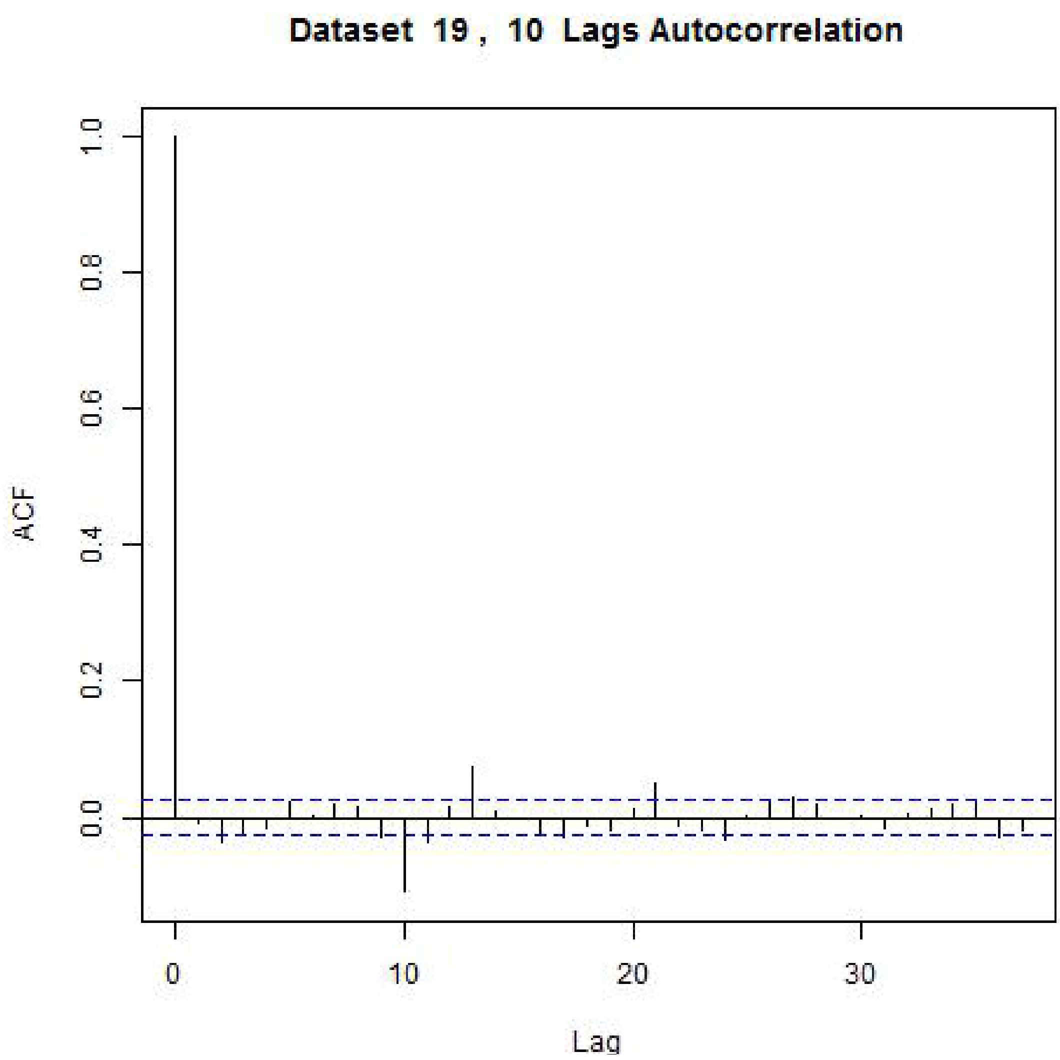

Figure 11. Autocorrelation Function for Time Series with 10 Lags

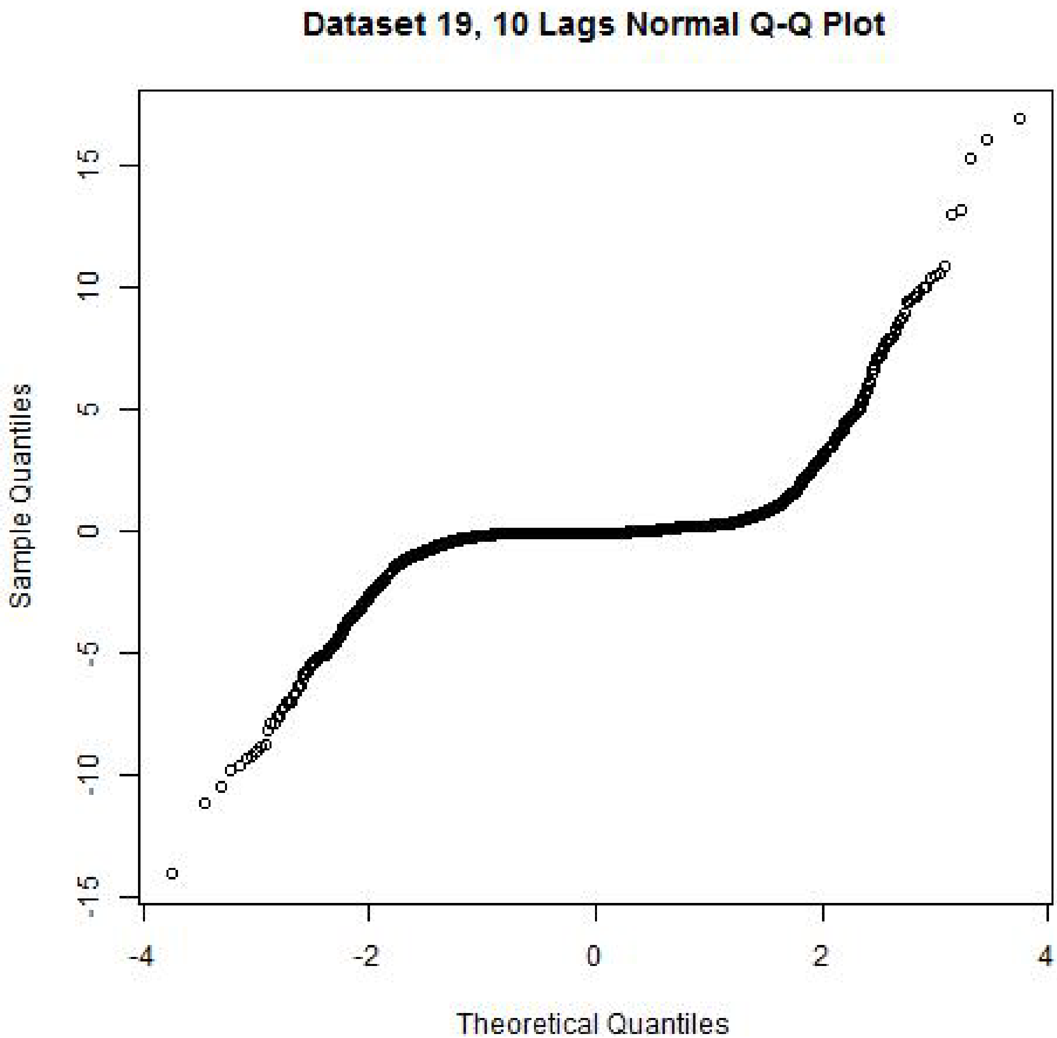

Figure 12. Normal QQ Plot for Time Series with 10 Lags

4. ARDL with 15 Lags

When there are 15 lags in the ARDL time series, the autocorrelation of the residuals of the regression is as shown in Figure 13. There is no significant autocorrelation shown between time periods. The QQ plot of the residuals for normality still exhibits heavy-tailed characteristics, as shown in Figure 14.

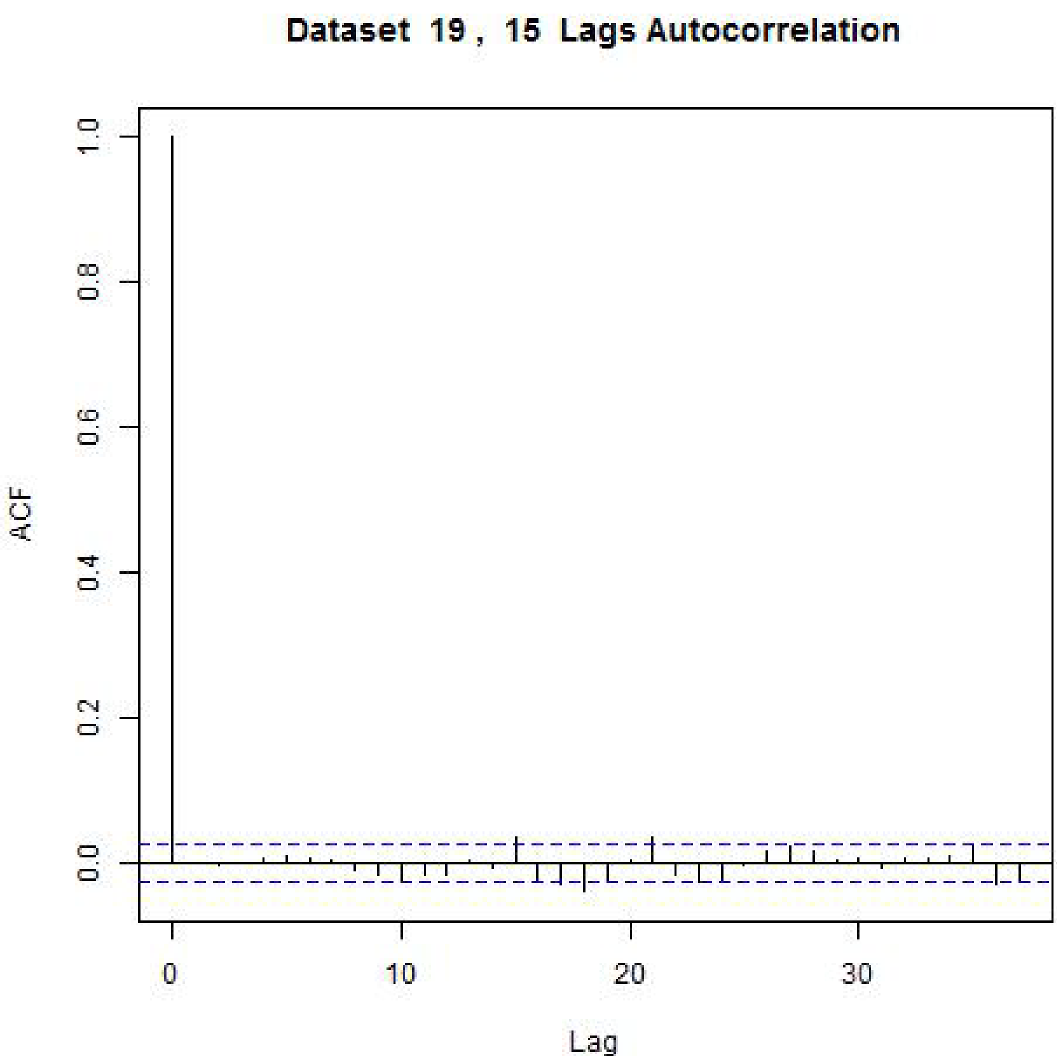

Figure 13. Autocorrelation Function for Time Series with 15 Lags

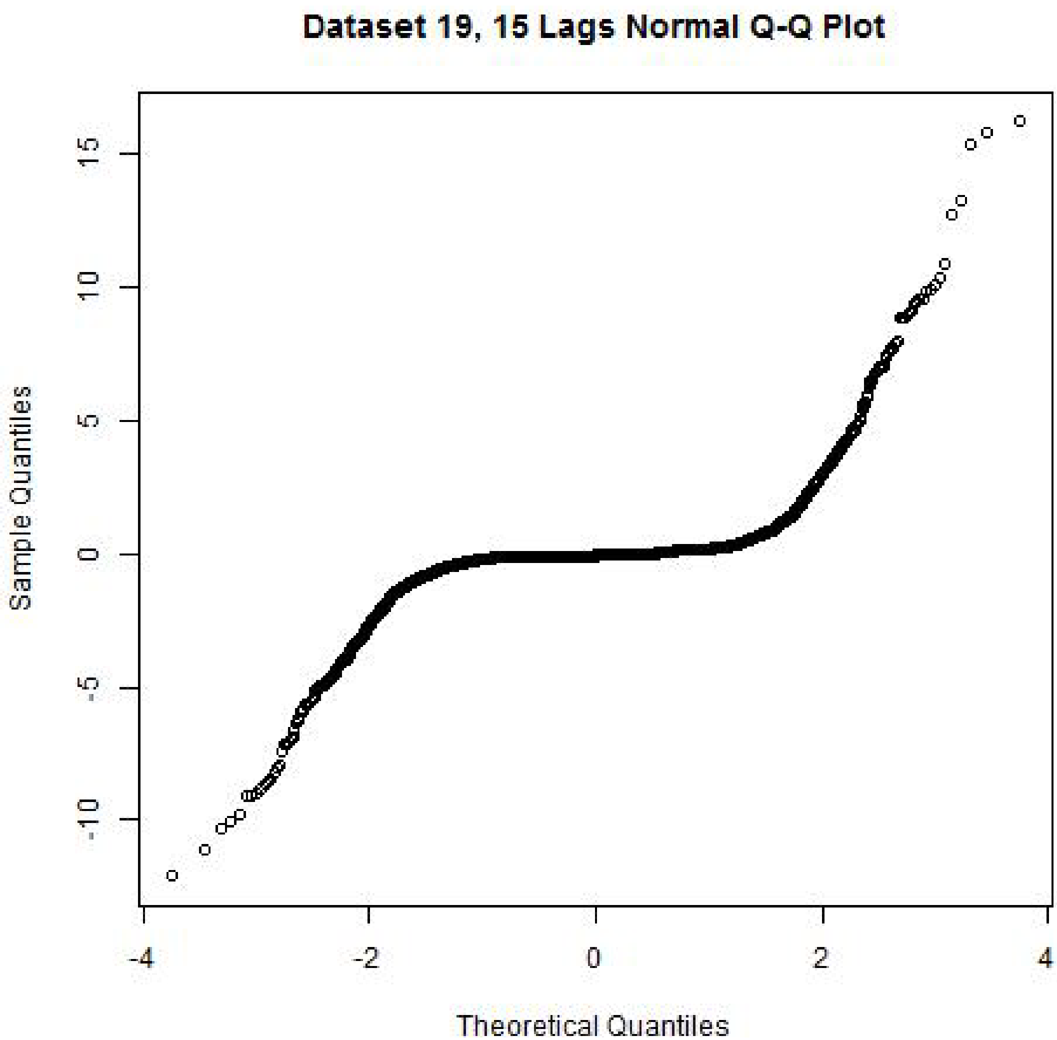

Figure 14. Normal QQ Plot for Time Series with 15 Lags

5. Lag Length Results Summary

From the results of the autocorrelation and QQ plots, with 15 lags, we observe that the autocorrelation of residuals is reduced to an approximate white noise condition. The residuals are not normally distributed and exhibit heavy-tailed characteristics, with outliers at both positive and negative ends. Therefore, the ARDL time series model will use 15 lags for the modeling and prediction of the datasets. Due to the non-normality of the residuals, a nonparametric regression analysis may be more robust and appropriate for the analysis of the collected data.

F. ARDL MODEL

The ARDL time series model is set up with 15 lags. We divide the datasets into two speed categories, one for turbocharger mode and another for cruising mode. For turbocharger mode, we use 15 datasets in deriving the model and four datasets for deriving the model for cruising mode.

LM is an R language command that fits a linear model for a set of data according to the model specification provided using OLS, and RQ is the R language command that fits a quantile regression model (Koenker 2016). RQ is a nonparametric approach and allows the user to specify the quantile to be estimated, p, which is between 0 and 1. We use the RQ command with p = 0.5, which is equivalent to minimizing the sum of absolute values of the regression errors. The independent variable, X_{t-j}, used for the model is the engine speed, and the dependent variable, Y_{t-i} , is for combined B exhaust gas temperature.

1. Regression Model for Turbocharger Mode

The regression model for turbocharger mode is derived from 15 datasets during operation and with sustained engine speeds above 2000 RPM. Each dataset is processed with linear regression using 15 lags with respect to both engine speed and exhaust temperature. The coefficients derived by the regression for each dataset are then averaged to derive the overall regression model for the turbocharger mode.

The derived model for turbocharger mode using the LM command is shown in Table 9, and the derived model for turbocharger mode using the RQ command is shown in Table 10. The derived model lists the intercept and coefficients of the dependent and independent variables with 15 lags.

Table 9. LM Regression Model for Turbocharger Mode

Terms	Coefficients	Terms	Coefficients
μ	0.224	X_t	0.003
Y_{t-1}	0.783	X_{t-1}	-0.002
Y_{t-2}	0.319	X_{t-2}	0.003
Y_{t-3}	0.090	X_{t-3}	0.002
Y_{t-4}	-0.031	X_{t-4}	0
Y_{t-5}	0.195	X_{t-5}	0.006
Y_{t-6}	-0.275	X_{t-6}	-0.007
Y_{t-7}	0.004	X_{t-7}	0.002
Y_{t-8}	-0.115	X_{t-8}	-0.003
Y_{t-9}	0.004	X_{t-9}	-0.001
Y_{t-10}	-0.067	X_{t-10}	-0.001
Y_{t-11}	0.038	X_{t-11}	0
Y_{t-12}	0.056	X_{t-12}	-0.001
Y_{t-13}	0.001	X_{t-13}	-0.001
Y_{t-14}	0.016	X_{t-14}	0.002
Y_{t-15}	-0.023	X_{t-15}	-0.001

Table 10. RQ Regression Model for Turbocharger Mode

Terms	Coefficients	Terms	Coefficients
μ	-0.013	X_t	0.001
Y_{t-1}	0.795	X_{t-1}	0
Y_{t-2}	0.335	X_{t-2}	0.002
Y_{t-3}	0.011	X_{t-3}	0.001
Y_{t-4}	-0.036	X_{t-4}	0.001
Y_{t-5}	0.238	X_{t-5}	0.003
Y_{t-6}	-0.346	X_{t-6}	-0.004
Y_{t-7}	0.067	X_{t-7}	0.002
Y_{t-8}	-0.073	X_{t-8}	-0.002
Y_{t-9}	0.004	X_{t-9}	0
Y_{t-10}	-0.029	X_{t-10}	-0.001
Y_{t-11}	0.021	X_{t-11}	0
Y_{t-12}	0.031	X_{t-12}	0
Y_{t-13}	-0.017	X_{t-13}	0
Y_{t-14}	0.016	X_{t-14}	0.001
Y_{t-15}	-0.018	X_{t-15}	-0.001

2. Regression Model for Cruising Mode

The regression model for cruising mode is derived from four datasets during operation and with sustained engine speeds below 2000 RPM. Each dataset goes through linear regression with 15 lags. The coefficients derived by the regression for each dataset are then averaged to derive the overall regression model for the cruising mode. The derived model for cruising mode using the LM command is shown in Table 11, and the derived model for cruising mode using the RQ command is shown in Table 12.

Table 11. LM Regression Model for Cruising Mode

Terms	Coefficients	Terms	Coefficients
μ	0.141	X_t	0.001
Y_{t-1}	0.727	X_{t-1}	0
Y_{t-2}	0.378	X_{t-2}	0.002
Y_{t-3}	0.062	X_{t-3}	0.002
Y_{t-4}	0.014	X_{t-4}	0.003
Y_{t-5}	0.230	X_{t-5}	0.006
Y_{t-6}	-0.280	X_{t-6}	-0.009
Y_{t-7}	0.016	X_{t-7}	0.002
Y_{t-8}	-0.177	X_{t-8}	-0.004
Y_{t-9}	-0.058	X_{t-9}	0
Y_{t-10}	-0.016	X_{t-10}	-0.001
Y_{t-11}	0.032	X_{t-11}	-0.003
Y_{t-12}	0.075	X_{t-12}	0.001
Y_{t-13}	-0.003	X_{t-13}	0
Y_{t-14}	0.016	X_{t-14}	0.002
Y_{t-15}	-0.021	X_{t-15}	0

Table 12. RQ Regression Model for Cruising Mode

Terms	Coefficients	Terms	Coefficients
μ	0.038	X_t	0.001
Y_{t-1}	0.745	X_{t-1}	0
Y_{t-2}	0.408	X_{t-2}	0.001
Y_{t-3}	0.015	X_{t-3}	0.001
Y_{t-4}	-0.084	X_{t-4}	0.001
Y_{t-5}	0.305	X_{t-5}	0.004
Y_{t-6}	-0.363	X_{t-6}	-0.005
Y_{t-7}	0.065	X_{t-7}	0.002
Y_{t-8}	-0.095	X_{t-8}	-0.002
Y_{t-9}	-0.019	X_{t-9}	0
Y_{t-10}	-0.006	X_{t-10}	-0.002
Y_{t-11}	0.004	X_{t-11}	0
Y_{t-12}	0.046	X_{t-12}	0
Y_{t-13}	-0.022	X_{t-13}	0
Y_{t-14}	0.027	X_{t-14}	0
Y_{t-15}	-0.027	X_{t-15}	-0.001

G. PREDICTION RESULTS

The prediction results using the derived models for both speed modes and both regression approaches are presented in this section. Verification of the prediction results through comparison of the actual dataset and residual analysis is also presented.

1. Turbocharger Mode

The turbocharger mode uses Datasets 6, 26, 28, and 29 as new datasets to apply to the derived model, to verify the prediction abilities of the models. The prediction result for Dataset 28 is presented here for illustration of the prediction capability of the models.

a. Prediction

Figure 15 shows the prediction results for Dataset 28. The graph at the top is the actual data of combined B exhaust gas temperature. The graph in the middle is the prediction result using the parametric regression and prediction approach. The graph at the bottom is the prediction result using the robust nonparametric regression and prediction approach. The three graphs seem to show a similar general trend.

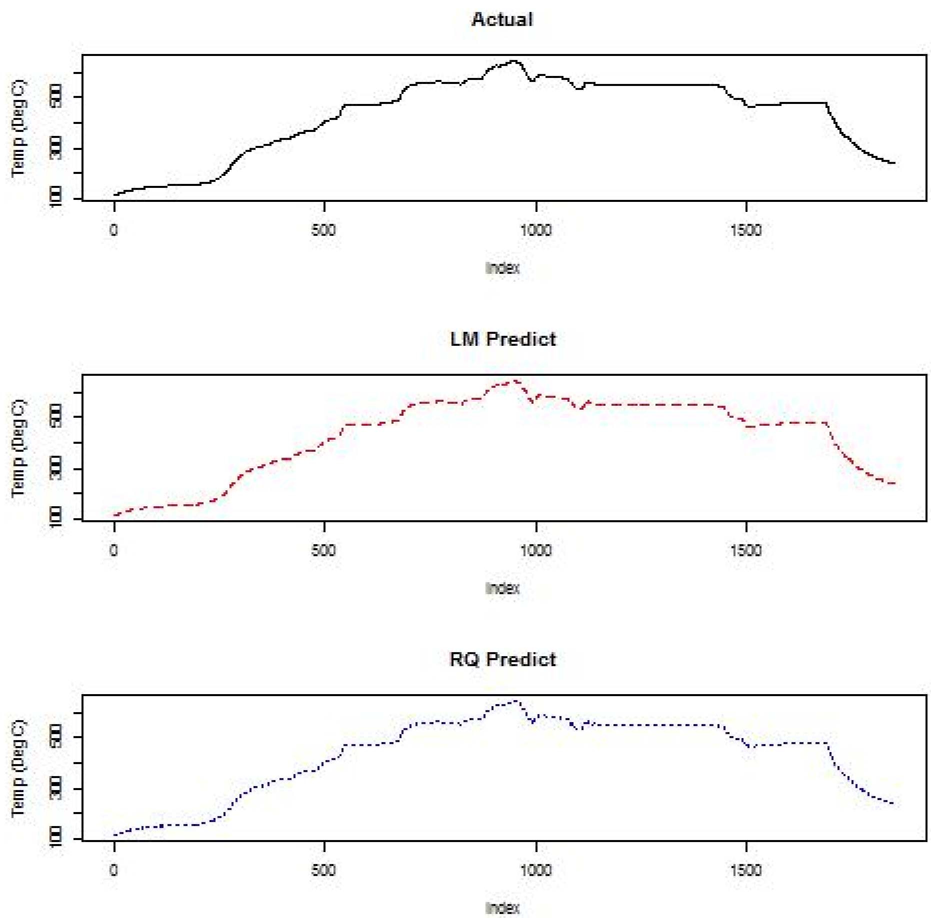

Figure 15. Prediction of Exhaust Gas Temperature in Turbocharger Mode Using Dataset 28 (Actual versus LM Regression Model versus RQ Regression Model)

b. Residuals

From the prediction results in Figure 15, we do not observe much deviation between the parametric and nonparametric prediction from the actual dataset. A further step for analyzing the prediction results is to observe the residuals. The residuals represent the difference between the actual plot and the prediction results. The graph at the top in Figure 16 shows the residuals of parametric prediction using the LM regression model, and the graph at the bottom in Figure 16 shows the residuals of nonparametric prediction using the RQ regression model. The deviation from the actual dataset plot is less than 6 degrees Celsius for both predictions. Both parametric and nonparametric models and predictions seem to perform well based on the prediction results.

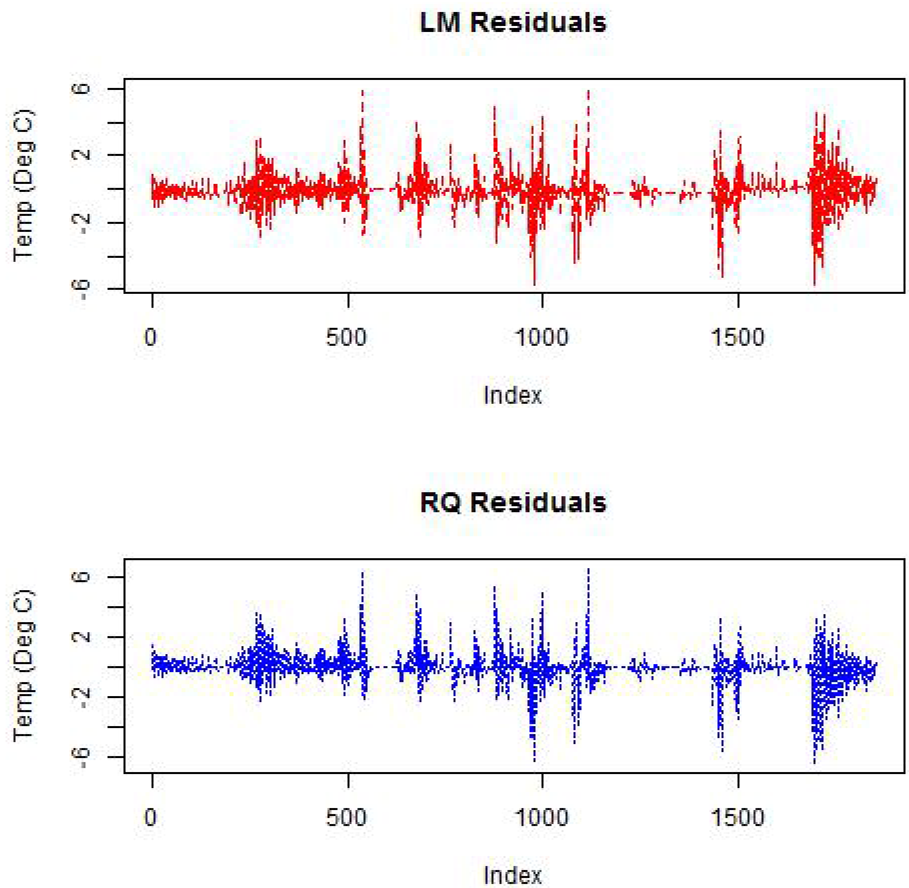

Figure 16. Prediction Residuals for Turbocharger Mode Using Dataset 28 (LM Regression Model and RQ Regression Model)

c. Autocorrelation Function

The autocorrelation function of the predicted data using the derived LM model shows generally low autocorrelation between time periods. There are only eight time periods that show an autocorrelation that is statistically significant, as shown in Figure 17. These autocorrelation have a value of less than 0.2.

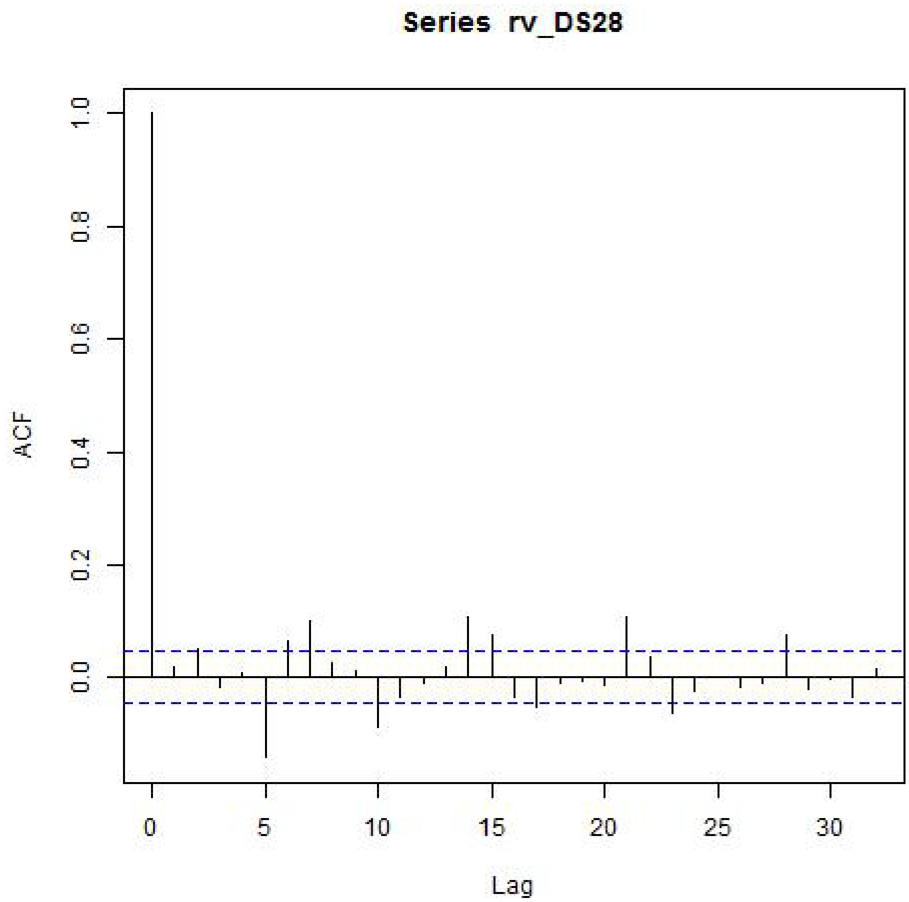

Figure 17. Autocorrelation Function Plot for Turbocharger Mode Prediction Results (LM Regression Model)

The autocorrelation function of the predicted data using the derived RQ model also shows generally low autocorrelation between time periods. There are also eight time periods that show autocorrelations that are statistically significant, as shown in Figure 18. These autocorrelations have a value of less than 0.2, but they seem to be slightly higher than those of the autocorrelations for the LM model.

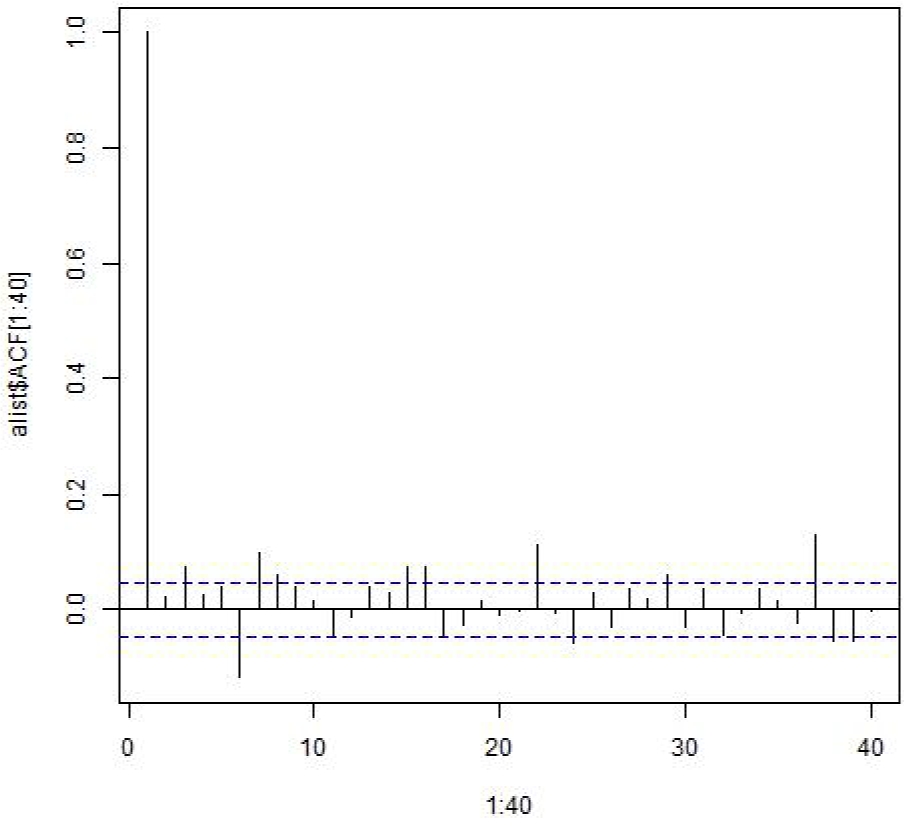

Figure 18. Autocorrelation Function Plot for Turbocharger Mode Prediction Results (RQ Regression Model)

d. Measures of Variability

The prediction results can be quantified by determining the R-squared, standard deviation, and interquartile range. R-squared measures the fitness of the model to the prediction, standard deviation and interquartile range measures the variability of the residuals. All computed R-squared values are very high, at above 0.999. This shows that engine speed is obviously needed to predict the dependent variable. Table 13 shows the variability of the residuals.

Table 13. Prediction Results Measures of Variability for Turbocharger Mode

Dataset	LM Regression Model		RQ Regression Model	
	Standard Deviation	Interquartile Range	Standard Deviation	Interquartile Range
6	2.944	0.645	3.157	0.528
26	1.384	0.633	1.451	0.491
28	0.991	0.537	1.037	0.396
29	1.316	0.707	1.367	0.543

2. Cruising Mode

The cruising mode uses Datasets 9, 22, 27, 31, 33, and 34 as new datasets to apply to the derived model, to verify the prediction abilities of the models. The prediction result for Dataset 33 is presented here for illustration of the prediction capability of the models.

a. Prediction

Figure 19 shows the prediction results for Dataset 33. The graph at the top is the actual data of combined B exhaust gas temperature. The graph in the middle is the prediction result using the parametric regression and prediction approach. The graph at the bottom is the prediction result using the robust nonparametric regression and prediction approach. The three graphs seem to show a similar general trend.

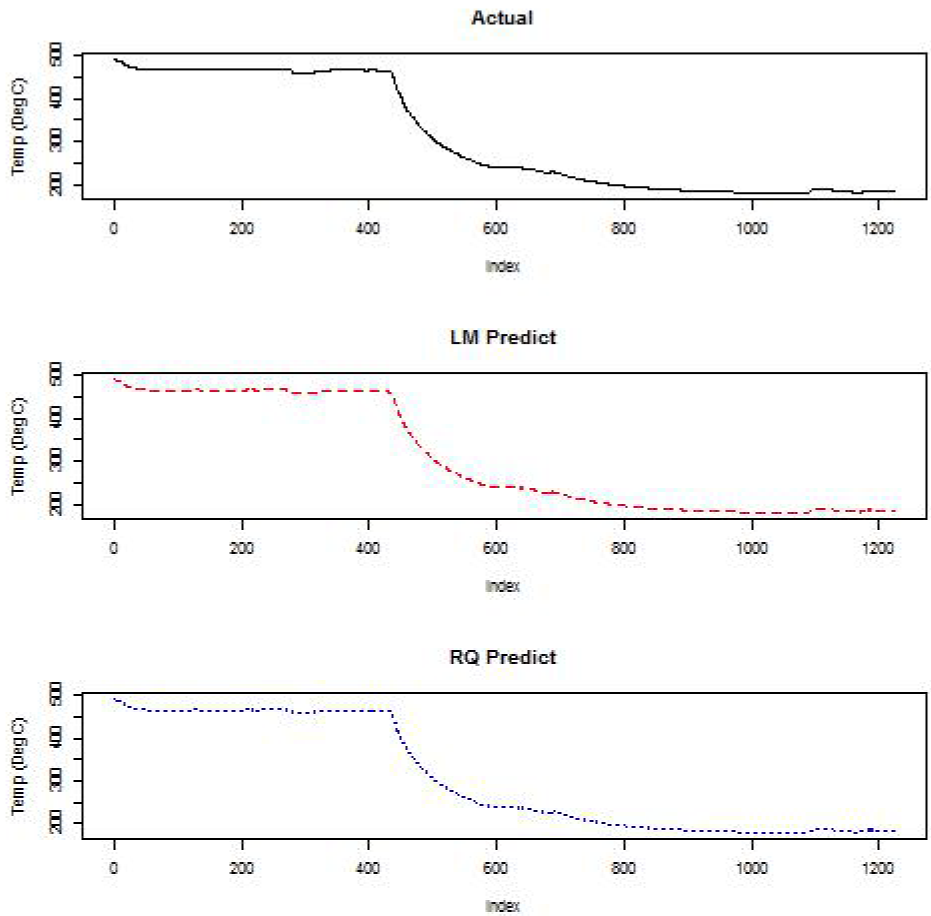

Figure 19. Prediction of Exhaust Gas Temperature in Cruising Mode Using Dataset 33 (Actual versus LM Regression Model versus RQ Regression Model)

b. Residuals

From the prediction results in Figure 19, we do not observe much deviation between the parametric and nonparametric prediction from the actual dataset. A further step in analyzing the prediction results is to observe the residuals. The residuals represent the difference between the actual plot and the prediction results. The graph at the top in Figure 20 shows the residuals of parametric prediction using the LM regression model, and the graph at the bottom in Figure 20 shows the residuals of nonparametric prediction using the RQ regression model. The deviation from the actual dataset plot is less than 5 degrees Celsius for both predictions. Both parametric and nonparametric models and predictions perform well based on the prediction results.

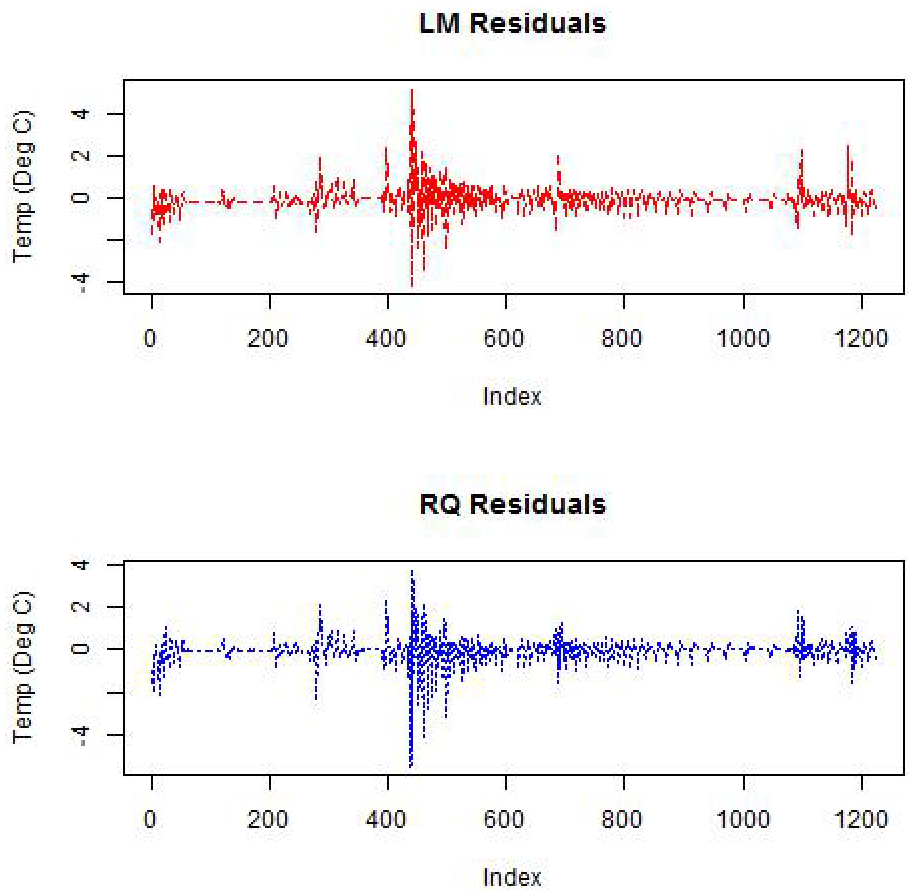

Figure 20. Prediction Residuals for Cruising Mode Using Dataset 33 (LM Regression Model and RQ Regression Model)

c. Autocorrelation Function

The autocorrelation function of the predicted data using the derived LM model shows moderate autocorrelation between time periods. The autocorrelation function value for the second lag has a value of about 0.3. More lags that are statistically significant are distributed over a longer time period, as shown in Figure 21.

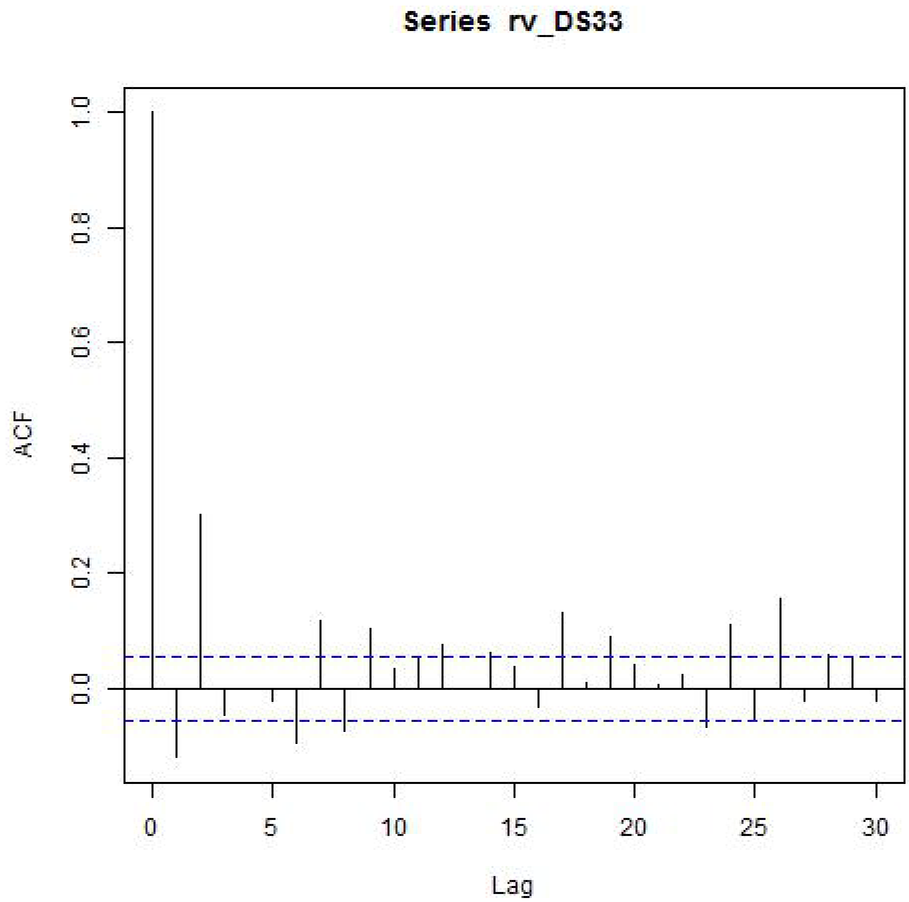

Figure 21. Autocorrelation Function Plot for Cruising Mode Prediction Results (LM Regression Model)

The autocorrelation function of the predicted data using the derived RQ model shows moderate autocorrelation between time periods. The autocorrelation function value for the first, second, and fifth lags has a value exceeding the absolute value of 0.2. The lags that are statistically significant are mainly the first few lags, as shown in Figure 22.

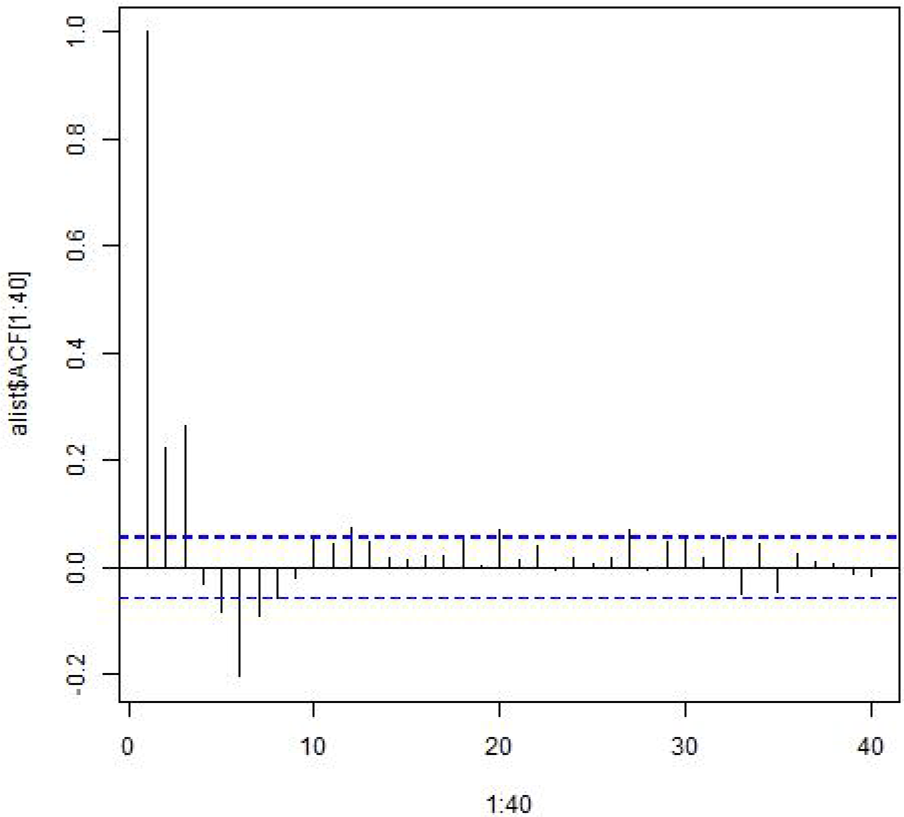

Figure 22. Autocorrelation Function Plot for Cruising Mode Prediction Results (RQ Regression Model)

d. Measures of Variability

The prediction results can be quantified by determining the R-squared, standard deviation and interquartile range. R-squared measures the fitness of the model to the prediction, standard deviation and interquartile range measure the variability of the residuals. All computed R-squared values are very high, at above 0.999. This shows that engine speed is obviously needed to predict the dependent variable. Table 14 shows the variability of the residuals.

Table 14. Prediction Results Measures of Variability for Cruising Mode

Dataset	LM Regression Model		RQ Regression Model	
	Standard Deviation	Interquartile Range	Standard Deviation	Interquartile Range
9	0.508	0.320	0.496	0.191
22	1.654	0.601	1.757	0.511
27	0.862	0.277	0.877	0.141
31	2.139	0.371	2.128	0.250
33	0.563	0.306	0.520	0.190
34	0.462	0.082	0.436	0.060

3. Summary of Prediction Results

The prediction results of both LM and RQ regression models for both modes of operation show good fit based on the R-squared values and the comparison of actual plots against predicted plots. There is randomness in the residuals from the residual plots, with no specific patterns shown. The autocorrelation plots show no autocorrelation for the turbocharger mode, but there is some autocorrelation for the cruising mode. It could be due to the data that is used for prediction; as shown in Figure 19, the exhaust gas temperature is reducing over time, which indicates the HSSC is slowing down and coming to a stop.

Comparing the prediction results for variability, we find the LM regression models produce residuals with lower standard deviation than RQ regression models. This is expected since LM regression uses the OLS method to minimize errors in models. The interquartile range is smaller for RQ regression models, which is more desirable.

LM is more risky due to the heavy tailed QQ plots of the data. There is non-normality with outliers at both the positive and negative end. From the data used in this study, we find RQ regression is a more robust method to model and predict the exhaust gas temperature.

V. CONCLUSION AND FUTURE WORK

Condition-based maintenance (CBM) is a preventive maintenance method that predicts the onset of a failure, allowing personnel to intervene with necessary maintenance actions to prevent the failure. Data analysis is a critical activity that contributes to the development of a regression model for making predictions in CBM.

In this thesis, we employ existing data captured by sensors located on the diesel engine of a high speed surface craft for regression modeling in CBM. We define the data selection criteria to ensure that the data used for the analysis are suitable and meaningful. We also categorize the datasets into two operating speed profiles of the HSSC: one for turbocharger mode and the other for cruising mode.

Regression is the finding of an appropriate mathematical model that can best fit the data. We use an autoregressive distributed lag time series model here due to the autocorrelative nature of the collected data. We select engine speed as the independent variable, and exhaust gas temperature as the dependent variable for the regression model. The condition monitoring parameter for CBM is exhaust gas temperature. Exhaust gas temperature has an operating temperature much higher than other operating temperatures. Possible failures in engines could lead to deviations in exhaust gas temperatures, causing higher than usual exhaust gas temperatures. This change in temperature may cause cascading effects, adding thermal stress to surrounding components.

Non-normality in regression results leads to a nonparametric approach using quantile regression and verification by Spearman's ranked robust autocorrelation function and normality QQ plots. With 15 lags in the ARDL time series model, regression models are derived for each of the speed category. We then use the predictions for new datasets to verify the models.

Future work could include derivation of nonparametric control charts proposed by Li, Tang and Ng (2010). Data parameters collected could include engine running hours to determine the mechanical wear and tear of the engine. This would give more insight into the engine’s behavior with respect to the engine running hours.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX. MTU SERIES 2000 DIESEL ENGINE DIAGRAMS

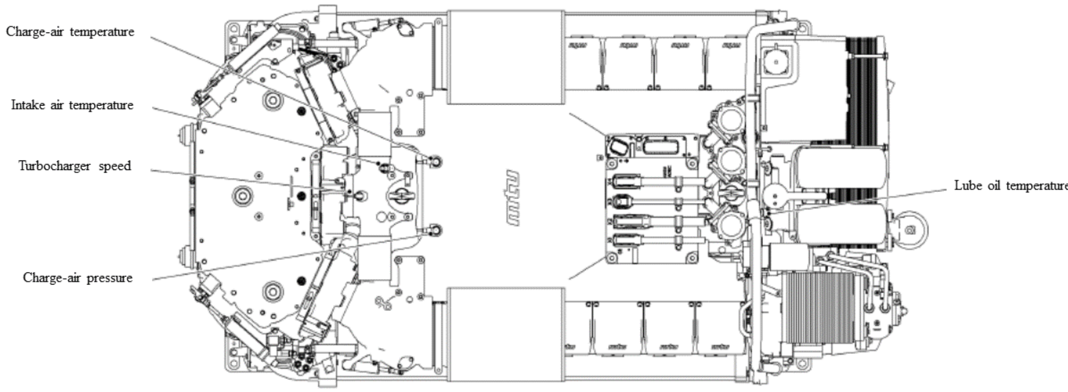

Figure 23. Sensors and Actuators (Engine Plan View). Adapted from MTU (2012).

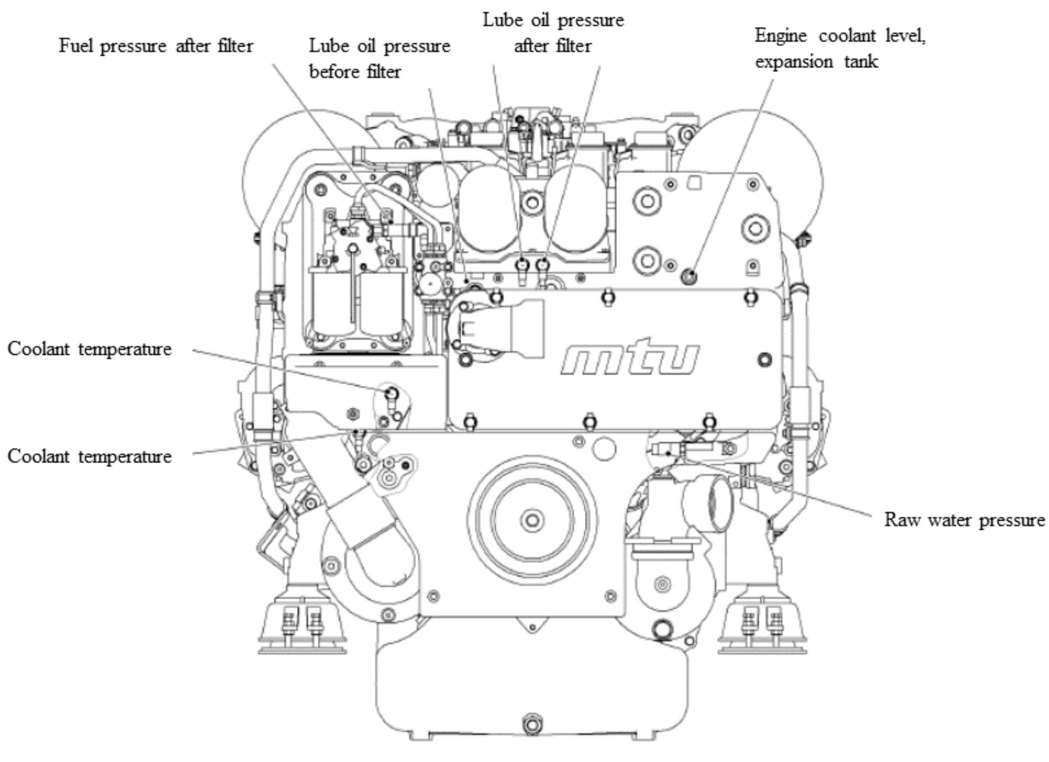

Figure 24. Sensors and Actuators (Engine Free End View). Adapted from MTU (2012).

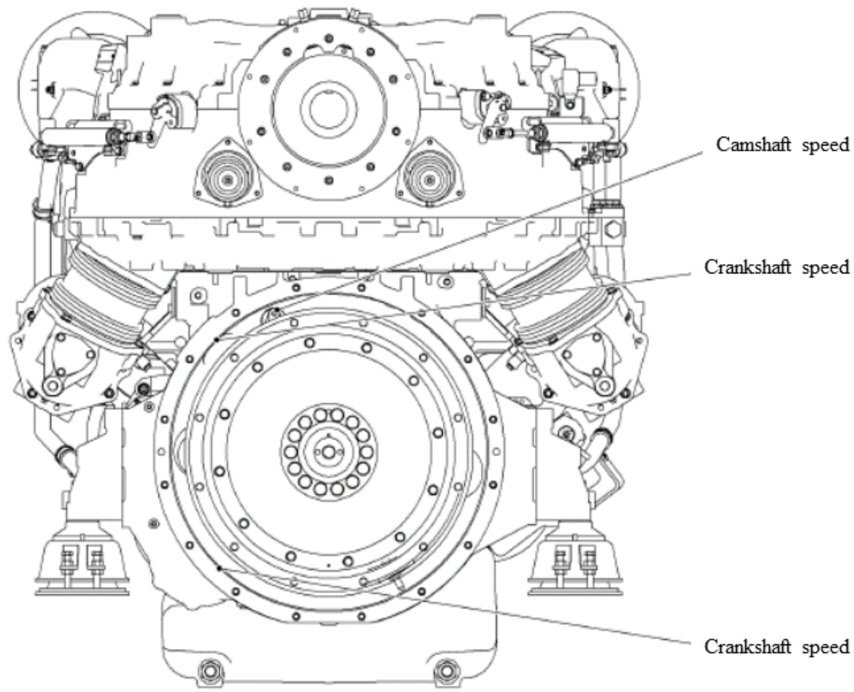

Figure 25. Sensors and Actuators (Engine Driving End View). Adapted from MTU (2012).

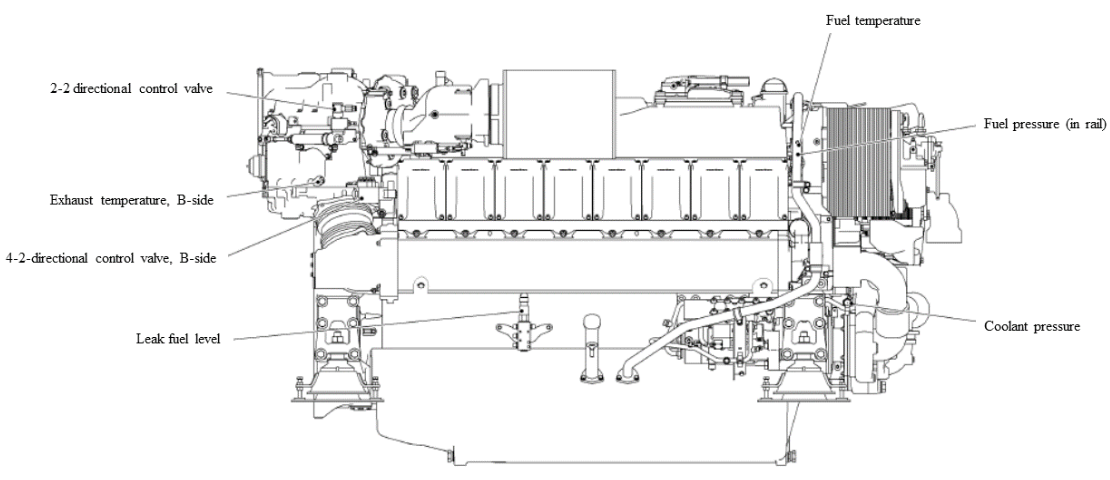

Figure 26. Sensors and Actuators (Engine, Right Side). Adapted from MTU (2012).

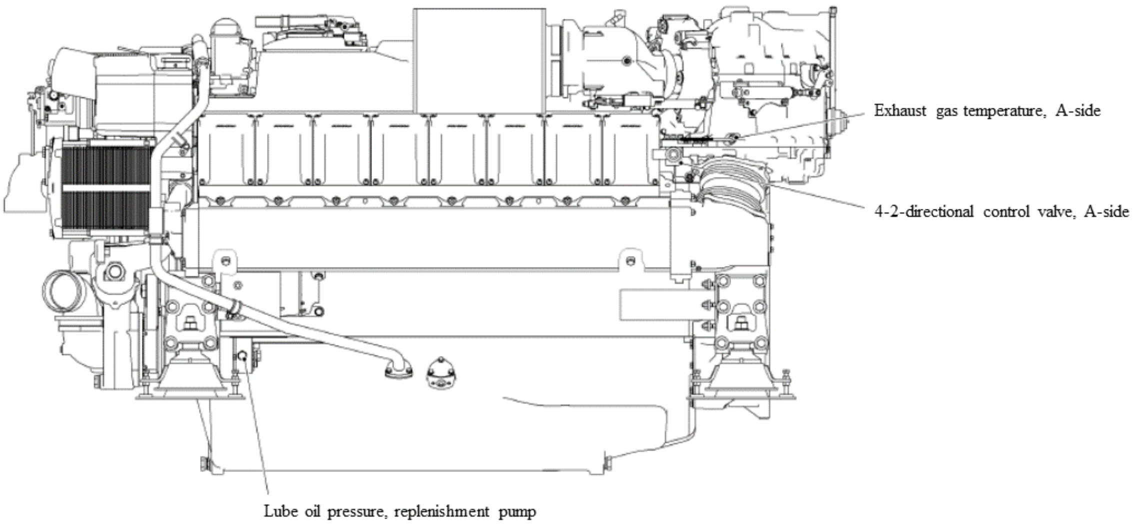

Figure 27. Sensors and Actuators (Engine, Left Side). Adapted from MTU (2012).

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Banks, Jeffrey, Jason Hines, Mitchell Lebold, Robert Campbell, Colin Begg, and Carl Byington. 2001. "Failure Modes and Predictive Diagnostics Considerations for Diesel Engines." *New Frontiers in Integrated Diagnostics and Prognostics. Proceedings of the 55th Meeting of the Society for Machinery Failure Prevention Technology*, 93–102.

Blanchard, Benjamin S. 2014. "Life-cycle Costing: An Effective Tool for Total Asset Management." *Journal of Reliability, Maintainability and Supportability in Systems Engineering* Winter 2014: 6–12.

Free Software Foundation. 2007. "GNU General Public License." https://www.gnu.org/licenses/gpl-3.0.en.html.

Greene, William H. 2000. *Econometric Analysis*, 4th ed. Englewood Cliffs, NJ: Prentice Hall.

Jardine, Andrew K. S., Daming Lin, Dragan Banjevic. 2006. "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance." *Mechanical systems and Signal Processing* 20(7): 1483–1510.

Jiang, Renyan, Xinping Yan. 2008. "Condition Monitoring of Diesel Engines." In *Complex System Maintenance Handbook*, edited by Khairy A. H. Kobbacy and D. N. Prabhakar Murthy, 533–557. London: Springer.

Koenker, Roger. 2016. "quantreg: Quantile Regression." *R package version 5.24.* https://CRAN.R-project.org/package=quantreg.

Koyak, Robert. 2013. "An Assessment of the Potential Benefits of Prognostics for AEGIS SPY-1 Radar Systems." *Unpublished paper prepared for AEGIS Program Executive Office, Integrated Warfare Systems.*

Lee, Guan Hock. 2013. "A Study on Predictive Analytics Application to Ship Machinery Maintenance." Master's thesis, Naval Postgraduate School.

Li, Su-Yi, Loon-Ching Tang, and Szu-Hui Ng. 2010. "Nonparametric CUSUM and EWMA Control Charts for Detecting Mean Shifts." *Journal of Quality Technology* 42(2): 209–226.

MTU Friedrichshafen GmbH. 2007. *Technical Documentation BlueVision Interface for non-MTU Monitoring System.* Friedrichschafen, Germany: MTU.

———. 2012. *Operating Instructions Diesel Engine 12 V 2000 M94 16V 2000 M94.* Friedrichschafen, Germany: MTU.

Office of the Under Secretary of Defense. 2016. "Operation and Maintenance Overview Fiscal Year 2017 Budget Estimates." http://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2017/fy2017_OM_Overview.pdf.

Parker, Jeffrey. 2012. "Learning Time-Series Econometrics." Reed College. http://www.reed.edu/economics/parker/s14/312/tschapters/S13_Ch_3.pdf.

Peer, David. 2012. "Estimating the Cost of Naval Ships." *Canadian Naval Review* 8(2): 4–8.

R Core Team. 2015. "R: A language and environment for statistical computing." Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ragsdale, Cliff T. 2012. *Spreadsheet Modeling & Decision Analysis*, 6th ed. Mason, OH: South-Western Cengage Learning.

Tsang, Albert H. C. 1995. "Condition-Based Maintenance: Tools and Decision Making." *Journal of Quality in Maintenance Engineering* 1(3): 3–17.

Yam, R. C. M., P. W. Tse, L. Li, and P. Tu. 2001. "Intelligent Predictive Decision Support System for Condition-Based Maintenance." *International Journal of Advanced Manufacturing Technology* 17(5): 383–391.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California