OFFSHORE CREW BOAT SAILING TIME PREDICTION MODELS

Miss Atiporn Chongterdtoonskul

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Industrial Engineering
Department of Industrial Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2019

แบบจำลองการทำนายเวลาเเล่นเรือกลางทะเล

น.ส.อธิพร จงเทิดทูนสกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมอุตสาหการ ภาควิชาวิศวกรรมอุตสาหการ
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Thesis Title	OFFSHORE CREW BOAT SAILING TIME PREDICTION MODELS
By	Miss Atiporn Chongterdtoonskul
Field of Study	Industrial Engineering
Thesis Advisor	Associate Professor WIPAWEE THARMMAPHORNPHILAS, Ph.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Engineering

.... Dean of the FACULTY OF ENGINEERING
(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

.... Chairman
(Associate Professor NARAGAIN PHUMCHUSRI, Ph.D.)

.... Thesis Advisor
(Associate Professor WIPAWEE THARMMAPHORNPHILAS, Ph.D.)

.... Examiner
(Nantachai Kantanantha, Ph.D.)

.... External Examiner
(Associate Professor Peerapon Siripongwutikorn)

อธิพร จงเทิดทูนสกุล : แบบจำลองการทำนายเวลาแล่นเรือกลางทะเล. (OFFSHORE CREW BOAT SAILING TIME PREDICTION MODELS) อ.ที่ปรึกษาหลัก : รศ. ดร.วิภาวี ธรรมาภรณ์พิลาศ

งานวิจัยนี้ศึกษาวิเคราะห์ความสัมพันธ์ระหว่างเวลาแล่นเรือกลางทะเลและปัจจัยหลักที่คาดว่าน่าจะมีผล เพื่อปรับปรุงแบบจำลองที่ใช้ทำนายเวลาแล่นเรือ เนื่องจากข้อมูลที่นำมาใช้คำนวณเวลาแล่นเรือในปัจจุบันไม่มีการนำข้อมูลมาวิเคราะห์จึงทำให้มีความคลาดเคลื่อนในการทำนายสูงสำหรับเส้นทางการเดินเรือช่วงเช้าซึ่งเป็นช่วงหลัก ของทั้งเรือ AและเรือB ปัจจัยที่นำมาศึกษาประกอบด้วย ปัจจัยภายใน ได้แก่ ระยะทาง ความเร็วเรือ และปัจจัยภายนอก คือ ความสูงของคลื่น ทิศทางคลื่น ความเร็วลม และทิศทางลม ข้อมูลที่ใช้วิเคราะห์เป็นของปี 2017-2019 โดยแบ่งข้อมูลดังนี้ ข้อมูลปี 2017-2018 ไว้สำหรับ train และ validate ส่วนปี 2019 ใช้สำหรับ test แบบจำลองที่ดีที่สุด โดยเปรียบเทียบระหว่างสมการถดถอยแบบต่างๆ (สมการถดถอยแบบMultiple linear, Stepwise, Partial least squares, Lasso, and Elastic net) โดยใช้กลุ่มข้อมูล 2 แบบคือ ชุดข้อมูลอันดับแรก และชุดข้อมูลอันดับแรกร่วมกับชุดข้อมูลปฏิสัมพันธ์เพื่อหาแบบจำลองที่ดีที่สุดสำหรับสำหรับเรือA เรือ B และแบบจำลองแบบรวมลำเรือ โดยพิจารณาประสิทธิภาพของการทำนายจากการเปรียบเทียบจากค่า MAPE ที่น้อยที่สุด เพื่อการทำนายเวลาเดินเรือสำหรับการทำนายระยะสั้น(แบบล่วงหน้าวันต่อวัน)และระยะยาว(แบบ1เดือนล่วงหน้า)

แบบจำลองอีลาสติกเน็ตสมการถดถอยมีค่า MAPE ต่ำที่สุดสำหรับเรือทุกลำทั้งการทำนายระยะสั้นและระยะยาว ชุดข้อมูลอันดับแรกเหมาะสำหรับการทำนายระยะสั้น ให้ค่า MAPE เพียง 5.5% สำหรับเรือAและ7.43%เรือB นอกจากนั้นปัจจัยสำคัญสำหรับแบบจำลองระยะสั้นคือ ระยะทาง ความเร็วเรือ ความสูงของคลื่น ทิศทางคลื่น และความเร็วลม สำหรับแบบจำลองการทำนายระยะยาว สมการของเรือAมีค่า MAPE ต่ำที่สุดที่ 8.51% และ 10.09% สำหรับเรือB โดยใช้ชุดข้อมูลอันดับแรกร่วมกับชุดข้อมูลปฏิสัมพันธ์ ตัวทำนายที่มีความสำคัญสำหรับการทำนายระยะยาวของเรือทุกลำคือ ระยะทาง ความเร็วเรือ และทิศทางคลื่น อีกทั้งยังมีปฏิสัมพันธ์ระหว่าง ความเร็วเรือและทิศทางลม หลังจากนำแบบจำลองที่ดีที่สุดมาประเมินอีกครั้งโดยใช้ชุดข้อมูลใหม่ของปี 2019 พบว่าแบบจำลองที่ใช้มีความเหมาะสมสำหรับการทำนายเวลาแล่นเรือกลางทะเลและจะช่วยเพิ่มประสิทธิภาพสำหรับระบบการแล่นเรือของบริษัท AA ได้

สาขาวิชา	วิศวกรรมอุตสาหการ	ลายมือชื่อนิสิต ....
ปีการศึกษา	2562	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

6170320021 : MAJOR INDUSTRIAL ENGINEERING
KEYWORD: Nautical, Crew boat, Prediction model, Sailing time, Regression, Lasso regression, Elastic net regression

Atiporn Chongterdtoonskul : OFFSHORE CREW BOAT SAILING TIME PREDICTION MODELS. Advisor: Assoc. Prof. WIPAWEE THARMMAPHORNPHILAS, Ph.D.

In this study, the relationship between travelling time of crew boats and the potential related factors was studied to improve prediction model for travelling time calculation. According to lack of data analysis, the existing calculation method had very high forecast error for crew boat's morning route which is the main route for boat A and boat B. The selected variables were total distance and boat speed as internal factors while using significant wave height, wave direction, wind speed, and wind direction as external factors. 2017-2019 data were used in this analysis by dividing into 2017-2018 as train and validate data. Moreover, 2019 data was applied as test data for model evaluation of the optimal prediction model. Those input variables were analyzed with various regression models (Multiple linear regression, Stepwise regression, Partial least squares regression, Lasso, and Elastic net regression) using two different datasets of 1st order and 1st order with Interaction terms to find the optimal model for boat A, boat B, and combined boat. The performance of the forecasting models was compared using MAPE. The models were developed to forecast sailing time for short-term and long-term period.

Elastic net regression model provided the lowest MAPE for all boats both short-term and long-term prediction. 1st order dataset is appropriate for short-term prediction, yielding MAPE of 5.55% for boat A and 7.43% for boat B. Moreover, the significant input variables for short-term model are total distance, boat speed, significant wave height, wave direction, and wind speed. On the other hand, boat A's equation obtained the lowest MAPE of 8.51% and 10.09% for boat B for long-term model with 1st order with interaction terms dataset. The predictors that have effect for all boats were total distance, boat speed, and wave direction with interaction between boat speed and wind direction. After model re-evaluation with new dataset (2019), it was found that the developed models are suitable for travelling time prediction and will enhance efficiency for overall crew boat operation for AA company.

Field of Study: Industrial Engineering		Student's Signature ....
Academic Year:	2019	Advisor's Signature ....

ACKNOWLEDGEMENTS

I would like to express my deeply thanks and appreciation to all people below for kind support and encouragement throughout my Master's degree.

Firstly, I would like to gratefully thanks to my beloved parents and my 2 little daughters for their love, will power and supports in everything.

Secondly, I would like to give sincerely thanks to my thesis advisor, Assoc. Prof. Wipawee Tharmmaphornphilas, Ph.D. for her kind advices & valuable guidance throughout my study and this research, and support. It's my honour to work with and learn the great lessons from her. Then, I would like to give appreciate thanks to Assoc. Prof. Peerapon Siripongwutikorn, Ph.D., Assoc. Prof. Naragain Phumchusri, Ph.D., and Prof. Nantachai Kantanantha, Ph.D. for all suggestions and knowledge for my thesis.

In addition, I would like to thank all staffs in the Industrial Engineering department as well as in Sasin Graduate Institute of Business Administration for their help during my study.

Lastly, I would like to thank all my friends for their love and support in everything.

Atiporn Chongterdtoonskul

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

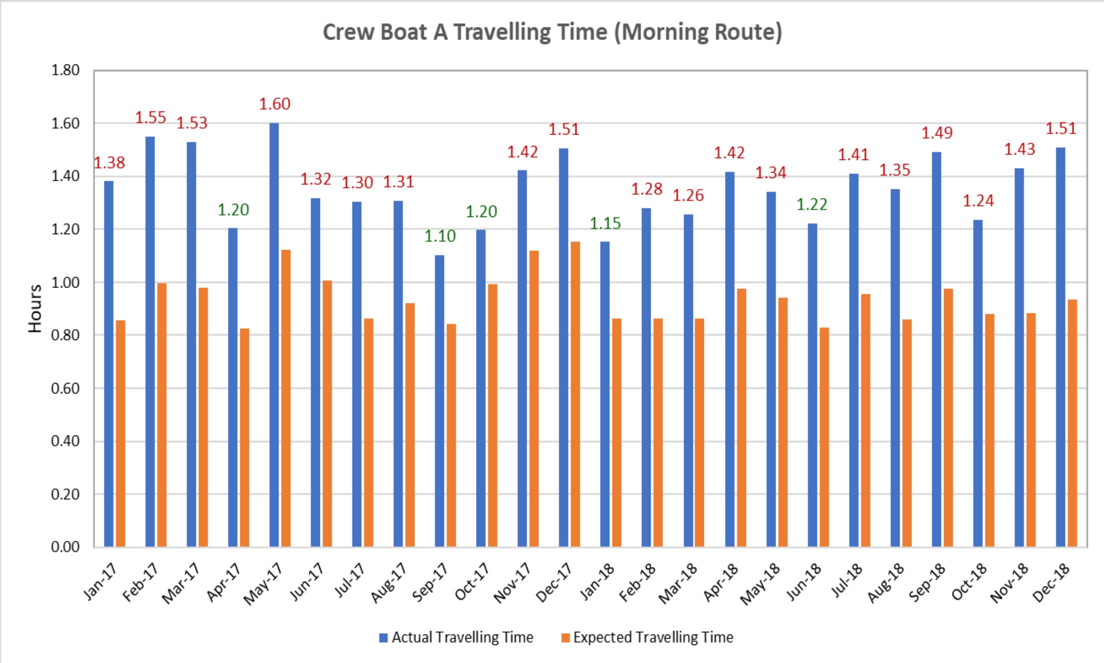
CHAPTER I
INTRODUCTION

1.1 Statement of Problem

AA Company is an oil & gas company that utilizes two crew boats, Boat A and Boat B, as the main transportation from living quarter to wellhead platforms. Objective is to perform routine activities and projects such as sending workforces to wellhead platforms for preventive and corrective maintenance, supporting drill rigs, and delivering cargos. Morning route which is generally between 6 A.M. to 10 A.M. is the main route. Current model that AA Company uses for travelling time forecasting is as per equation 1 below with assumption of constant boat speed of 16 kt (KPI) and constant wave height of 1.5 m. Thus, the travelling time is directly proportional to distance only with no seasonal effect. This equation is applied for both boat A and boat B.

$$Travelling\ Time = \frac{Distance\ (in\ NM)}{Boat\ Speed * 60 * Wave} \qquad (1)$$

However, AA Company found that crew boat's planning team missed target for annual key performance indicator for total travelling time of crew boats for two years in 2017 and 2018. The annual target is set at 1.22 hrs. Clear gaps between expected results and actual results for the total travelling time were observed after further investigated the results as shown in Figure 1. Current model had high MAPE of 39% for Boat A and 35% for Boat B in 2017-2018, indicating room for improvement. As per current practice that the company uses the same equation for both boats, short-term prediction models for each boat are required for daily operational adjustment i.e. boat speed, etc. and long-term prediction models for each boat are required for future planning 1 month in advance for crew boats to support these two different purposes. All of above information provides an incentive for improvement for better forecasting model taking into account other important operational parameters besides distance for short-term application and long-term use to increase accuracy in travelling time prediction for crew boats.



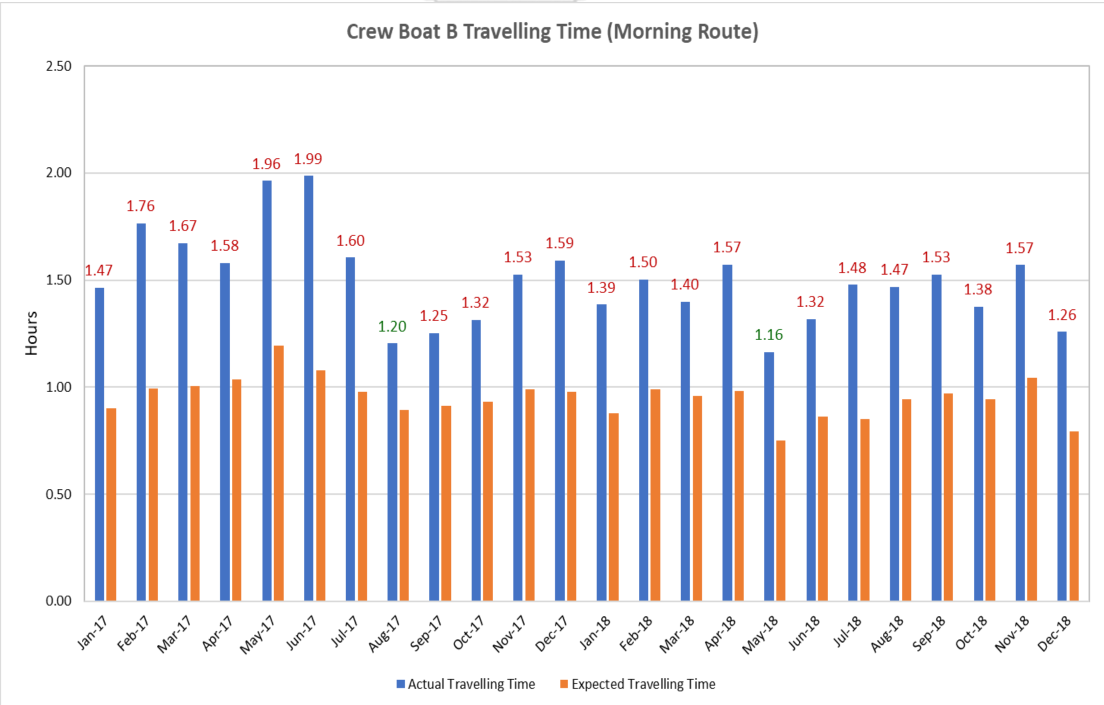


Figure 1 Actual VS Expected travelling time for crew boats A and B

1.2 Objectives

The main objective of this study is to assess the operational parameters affecting the crew boats' travelling time performance and to find the better prediction models for Boats A, B, and combined model for both boats.

1.3 Scope of Work

1.3.1. Factors

The main interested internal factors are distance and boat speed while the external factors are wave height, wave direction, wind speed, and wind direction.

1.3.2. Data

The datasets used for this study are 1st order and 1st order with interaction. The data are divided as 80% of 2017-2018 for train data and 20% of 2017-2018 for validate data in model built-up step. Furthermore, 2019 data are used as test data in model evaluation part.

1.3.3 Regression Models

More advanced regression models are applied to further investigate the improvement of forecasting performance in comparison to a baseline model. Other models that are applied apart from multiple regression model are Stepwise regression, PLS regression, Lasso regression, and Elastic net regression.

1.3.4 Forecast error

MAPE is used to compare forecasting error of the models to determine the optimal regression model.

1.3.5 Short-term & Long-term Models

There are two main purposes of forecasting, short-term and long-term. Short-term forecast is identified from daily historical data to predict for the next day. Objective for short-term model is for operational level adjustment of boat speed to achieve travelling time at 1.22 hrs as per goal. On the other hand, long-term model uses a quarter historical data to plan one month ahead for crew boats.

The most appropriate model for Boat A and Boat B which enhances the lowest MAPE comparing to current equation for sailing time forecasting is selected for future use of AA company.

1.4 Methodology Review

1.4.1 Short-term Prediction

- Data Collection
- Data Analysis and Clean-up
- Data Selection
  - Daily historical data for travelling time and distance, 4-hr average (6-10 A.M.) for boat speed, 4-hr average (before) for wave data, and 5 A.M. data for wind data
- Dividing Data into Train Set, Validation Set, and Test Set
- Running Multiple Regression, Stepwise Regression, PLS Regression, Lasso Regression, and Elastic Net Regression
- MAPE Evaluation of All Forecasting Models using Validation Set
- Optimal Forecasting Model Selection
- Optimal Model Evaluation using Test Set
- Plug Data in the Optimal Forecasting Model to obtain Boat Speed for Operational Level Adjustment

1.4.2 Long-term Prediction

- Data Collection
- Data Analysis and Clean-up
- Data Selection
  - Daily historical data for travelling time and distance, 4-hr average (6-10 A.M.) for boat speed, and average quarter basis for wave and wind data
- Dividing Data into Train Set, Validation Set, and Test Set

- Running Multiple Regression, Stepwise Regression, PLS Regression, Lasso Regression, and Elastic Net Regression
- MAPE Evaluation of All Forecasting Models using Validation Set
- Optimal Forecasting Model Selection
- Optimal Model Evaluation using Test Set
- Plug data in the Optimal Forecasting Model
- Calculate Crew Boat's Travelling Time for future operation planning

1.5 Results of the thesis

Prediction models for two purposes of applications as follows.

- Best model for short-term crew boat's travelling time prediction for boat A, boat B, and combined boat for operational level adjustment
- Best model for long-term crew boat's travelling time prediction for boat A, boat B, and combined boat for future planning

1.6 Benefits of the thesis

- Help planning team to achieve annual key performance indicator through proper forecasting model
- Gain benefit from different operational parameters' adjustment i.e. reduce boat speed in some period and save diesel fuel consumption
- Lower forecast error and gain more accuracy in sailing time prediction.

1.7 Thesis Timeline

Thesis Timeline													
YEAR	2019					2020							
CALENDAR MONTH	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
PROJECT MONTH	1	2	3	4	5	6	7	8	9	10	11	12	13
MILESTONES													
Literature Review													
1st Scope Model Selection													
Paper for Conference Proceeding													
Proposal Presentation													
Presentation at ICIEA 2020 Conference													
Final Defense Presentation													
Final Thesis Package Submission													
A. LITERATURE REVIEW													
B. RUNNING REGRESSION MODELS (1ST SCOPE)													
C. PAPER FOR ICIEA CONFERENCE PREPARATION													
D. THESIS PROPOSAL AND PRESENTATION PREPARATION													
E. RUNNING REGRESSION MODELS (2ND SCOPE)													
F. THESIS DEFENSE BOOK AND PRESENTATION PREPARATION													
G. FINAL DEFENSE BOOK UPDATE AND SUBMISSION													
PROJECT MONTH	1	2	3	4	5	6	7	8	9	10	11	12	13

CHAPTER II
THEORETICAL BACKGROUND AND LITERATURE REVIEWS

2.1 Sailing Time Forecasting for Crew Boat

2.1.1 General

Offshore vessels or crew boats are specially designed ships to daily transfer working crews or cargos to each offshore platform. This type of transportation is typically used in offshore industries. There were a number of studies about crew boats, especially in the areas of vehicle routing problem (VRP), scheduling, design, etc. However, very few researches on forecasting of travelling time which is one of the critical parameters that relate to the total operating cost were done.

Figure 2 Offshore operation and crew boat utilization

The approaches to determine crew boat sailing time might be as simple as using relationship between distance, speed, and time [1] up to more complex method for real time prediction. Vessel traffic management systems (VTMS) and vessel traffic monitoring information system (VTMIS) have been used for years to monitor travelling time and other parameters to adjust operation to increase efficiency and safety for operations at sea. Voyage optimization technology is employed to predict ship performance, including travelling time with consideration of energy efficiency in various sea states to assist captain in route selection [2]. Furthermore, machine learning and artificial neural networks were used in the study by Zissis et al. [3] to add predictive capacity to VTMIS. They implemented a publicly accessible, web-based system which can perform real-time learning and accurately predicting vessels future behavior. This system can provide critical information for operational adjustment, route planning, operational efficiency estimation, and anomaly detection. Moreover, a multi-method approach is used to validate monitoring system, combining with GPS, wind data logger, and cameras to measure route, speed, apparent wind velocity & direction, and position of crew members to determine equilibrium of the boat and estimate the loads applied to it [4].

2.1.2 Factors Affecting Sailing Time

There were several factors related to the sailing performance from previous researches, comprising of internal factors such as boat's specifications i.e. capacity, size, speed, fuel consumption, power, age, and machine [5] and external factors or environmental conditions i.e. wind speed, wave height, wave period, distance, tide, weather, number of fishing boat, etc. [6]. Many papers reported on that effect of wind speed and wave height were major environmental constraints [7]. Boat speed is inversely proportional to boat travelling time. The higher boat speed results in shorter travelling time. However, travelling time also depends on both internal and external attributes including boat's specification (maximum speed, cruising speed, propulsion system: main engines, main generators, propellers, etc.) and sea conditions (wind and wave) [8]. The lower resistance will obtain the larger velocity and ultimately enhance the sailing performance. Anderson [9] reported on the physics of sailing focusing on various components, including waves, eddies, turbulence in the water, and vortices produced in air by sailing to reduce resistance experienced from moving a sailing boat by optimizing hulls, keels, and sails. Pressure was found to be greater on the upwind side and lower on the downwind side as confirmed by measurements relative to the air pressure far from the sail. The boat's equilibrium speed is calculated from constant force of wind during sailing and resistance against the boat's motion through water.

2.1.2.1 Measurement Devices

2.1.2.1.1 Wave Sensor

This device measures both wave direction and magnitude then process and send data to record in the system. See photo in Figure 3.

Figure 3 Device for wave magnitude and direction measurement

2.1.2.1.2 Anemometer

These sensors are located at Northeast (Equipment 1 named as "E1") and Southwest (Equipment 2 named as "E2") corner of living quarter to measure wind speed and wind direction. See photos in Figure 4.

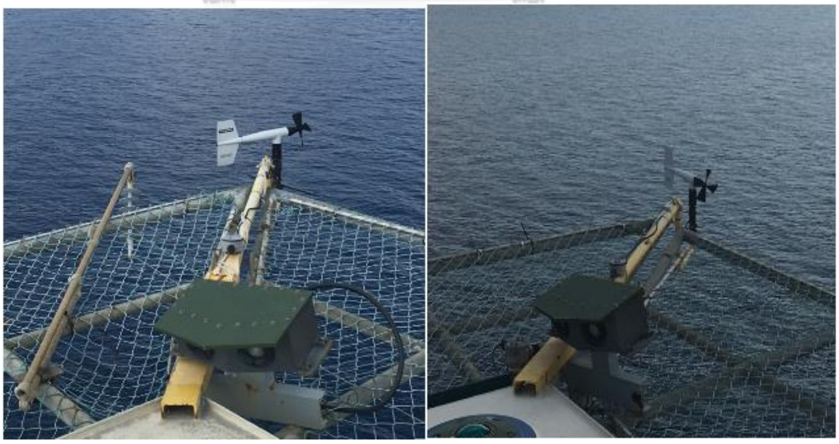


Figure 4 Anemometers located at NE (E1) and SW (E2) of living quarter

2.1.2.2 Crew Boats' Specification

2.1.2.2.1 Crew Boat A

DIMENSIONS	
Length Overall	38m
Breadth Moulded	7.6m
Depth Moulded	3.65m
Draft Loaded (max)	1.88m (approx.)

PERFORMANCE	
Maximum Speed	27 knots at 16 MT/ 24 hr
Cruising Speed	25 knots at 14 MT/ 24 hr

Figure 5 Crew Boat A's Specification [10]

2.1.2.2.2 Crew Boat B

DIMENSIONS	
Length overall	38 m
Breadth	7.6m
Depth	3.65m
Draft	1.88m
Gross tonnage	257
Net tonnage	77

SPEED	
Maximum/cruising	25/23 knots

Figure 6 Crew Boat B's Specification [10]

2.2 Quantitative Forecasting Methods

There are two general types for quantitative forecasting methods, which are time series and explanatory method (or casual model). Time series models identify historical patterns using time as reference to predict the future based on underlying pattems contained within those data. Forecasting models ranging from auto regressive moving average (ARMA), auto regressive integrated moving average (ARIMA), and hybrid models are progressively improved predictive performance.

On the other hand, the explanatory method, i.e. regression analysis, assumes that the variable being forecasted is related to other variables in the environment. Regression techniques have been applied in two dimensions as making conclusion from interpretation of outcomes of past studies and results estimation from analysis for further predicting other situations. Regression analysis helps to indicate significant relationship between dependent and independent variable as well as strength of impact of multiple independent variables on a dependent variable.

2.2.1 Quantitative Forecasting Model Application

Time series models are widely used for various applications, i.e. inflation forecasting [11], hydrological forecasting [12], and demand forecasting [13]. On the contrary, regression techniques have been applied on many fields such as value of travel time savings in transportation [14], epidemiology& biomedical [15], rack configuration of vehicle storage and retrieval system in warehouse [16], and many more using various advanced techniques such as heuristic regression [17], support vector regression [18], etc.

2.2.2 Data Preparation

2.2.2.1 Outliers Elimination

Observation which deviates so much from other observations must be removed since it diverges from overall pattern on a sample and can have a significant adverse impact on forecasts. One of the simplest methods for detecting outliers is the use of box plot as the graphical display for describing the distribution of the data. Time series plot can be used to show overall dominant outliers. There are several methods to drop the observed outliers in raw data depending on the dataset characteristics as below [19].

A) Remove the outlier from raw data which depict the significant difference. However, the removed data point must have no effect to the remaining data point.

B) Replace the outlier with the artificial value: their upper or lower bound or the arithmetic average of neighbors or their smoothed value.

2.2.2.2 Influential Points

The influential point is the outlier that has significant effect on analysis and negatively influence the regression as it does not represent other data points, not follow the trend, and mislead the result from prediction. Cook's distance in Regression analysis is the method to identify the influential point by determining unusual predicted value from measuring the distance between fitted value with and without ith observation [20]. The data points that should be eliminated are the one with Cook's distance greater than 4/n where n = total number of data points. Equation for Cook's distance is shown below [21].

$$D_i = \frac{\sum_{j=1}^{n}(y_j - y_{j(i)})^2}{(p+1)\sigma^2} \tag{2}$$

where D_i = cook's distance of observation i

y_j = estimated mean of y at observation j

p = number of regression coefficients

σ = estimated variance from the fit, based on all observations

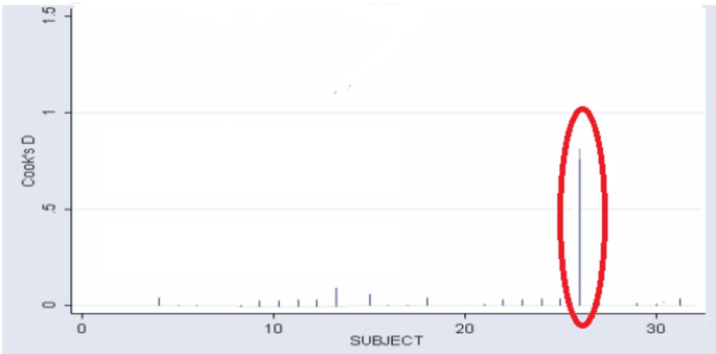


Figure 7 Example for potential outlier from Cook's distance

2.2.3 k-fold Cross-validation

Cross-validation is the re-sampling procedure to evaluate robustness of the data from using all data to test the model and helps to estimate model performance on the unseen data to ensure the less biased than other methods i.e. train/test split [22].

k refers to the number of groups that given data split into. Choosing value for k is important as they might result in a mis-representative skill of the model e.g. a score with high variance or high bias. It is usually set at 5 or 10. From experimentation, k = 10 is a value that generally results in model skill estimate with low bias with a modest variance. Thus, k = 10 is applied in our research [23].

Below schematic shows general procedure for k-fold cross-validation (k=10) [24].

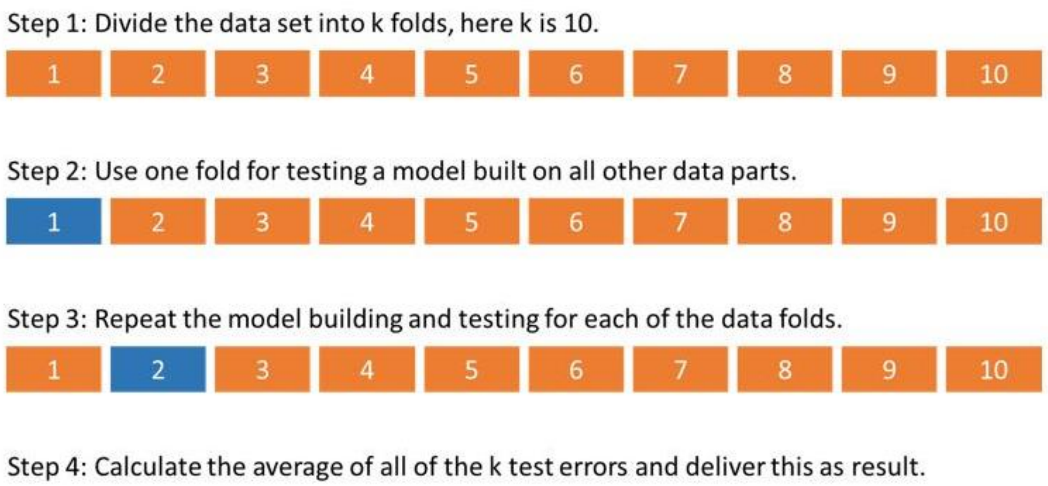


Figure 8 Procedure for 10-fold cross-validation

2.2.4 Regression Analysis Techniques

There are various kinds of regression techniques that can be applied for predictions depending on 3 scenarios which are type of dependent variables, number of independent variables, and shape of regression line. Table 1 summarized regression techniques with proper data set to be applied.

Table 1 Summary for Regression techniques and proper data characteristic

Regression Technique	**Proper Data set**
1. Linear regression	- Dependent variable is continuous - Independent variable can be continuous or discrete - Linear relationship
2. Logistic regression	- Dependent variable is binary (0/1, T/F, Yes/No) - Independent variables can be continuous or binary.
3. Polynomial regression	- Power of independent variable is more than 1 (curve fit to data points) - Non-linear relationship between dependent and independent variables
4. Stepwise regression	Multiple Independent variables → handle higher dimensionality of data set
5. Ridge regression	Multicollinearity (independent variables are highly correlated)
6. Lasso regression	Multicollinearity
7. Elastic Net regression	Multicollinearity
8. Quantile regression	Data with outliers, high skewedness, and heteroscedasticity
9. Principal Components regression (PCR)	Many independent variables or multicollinearity exist in data
10. Partial Least Squares (PLS) regression	Highly correlated independent variables or large number of independent variables
11. Support Vector regression	- Solve both linear and non-linear models - Use non-linear kernel functions i.e. polynomial for non-linear models.
12. Ordinal regression	- Dependent variable is ordinal in nature i.e. survey response (1 of 6 scale) - Predict ranked values - Change in level of dependent variable does not necessarily to be equivalent throughout range of the variable.
13. Poisson regression	Dependent variable has count data (assume distribution of count having variance equals to its mean)
14. Negative Binomial regression	Count data (not assume variance equal to its mean)
15. Quasi Poisson regression	Overdispersed count data (Variance is linear function of meanwhile negative binomial model is a quadratic function of mean)
16. Cox regression	Time-to-event data

From this table, characteristics of crew boats' related data are multiple independent variables with potential high correlation among variables. Thus, the proper regression techniques to be applied are Linear regression, Stepwise regression, Ridge regression, Lasso regression, Elastic net regression, and Partial least squares regression.

2.2.4.1 Linear Regression

It is the simplest and most widely known form of regression. The dependent variable is continuous while the independent variables can be either continuous or discrete with linear nature of regression line. It represents by equation as follows:

$$y = \beta_0 + \beta_1 x \quad (3)$$

where β_0 is y-intercept and β_1 is a slope. The coefficients are calculated using least square method for fitting a regression line [25]. See example for linear regression in Figure 9.

Linear regression is applied under assumptions as per below:

- Linear relationship between independent and dependent variables
- Sample observations must be independent
- According to very sensitive to outliers, need to ensure that no outliers present to avoid effect on the forecasted values.
- No multicollinearity, heteroscedasticity, and autocorrelation
- Normally distributed with mean 0 and constant variance for error terms

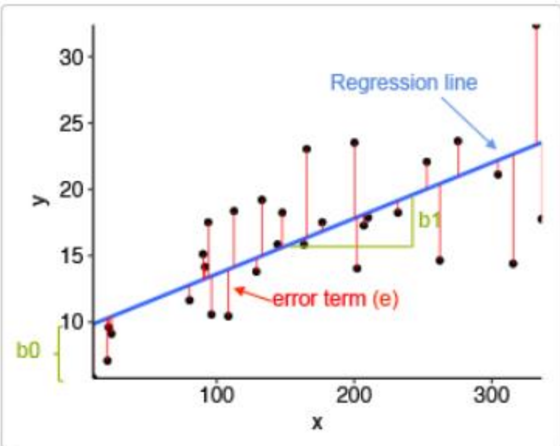


Figure 9 Relationship between x and y for linear regression

As mentioned earlier that regression analysis investigated the casual effect relationship between the variables, used for forecasting, and time series modelling. It allows comparison of the effects of variables measured on the different scales. These will help researchers to eliminate and identify the best set of parameters to be used in forecasting models.

According to relationship of travelling time to other independent variables (distance, boat speed, wind speed, etc.) from research study, multiple regression method was selected for the first part of this study as baseline model to find the best forecasting model.

2.2.4.2 Stepwise Regression

This technique normally uses for multiple independent variables by applying automatic process to select significant independent variables. R-squared, F-test, and t-test from statistic indicate significant variables that should be added and co-variates to be dropped. Objective of this technique is to maximize prediction power with minimum number of predictor variables. It is suitable for data set with higher dimensionality [26].

There are 3 main approaches which are forward selection, backward elimination, and bi-directional elimination. Figure 10 illustrated general process schematic of steps for stepwise regression.

- **Forward selection** starts with no variable in the model, test under fitting criteria before adding most statistically significant variable into the model and repeat until there is none that further improve the model.
- **Backward elimination** starts with all predictors in the model then test per fitting criteria to eliminate insignificant variables that reduce model fit and repeat this step until no further variables can be deleted.
- **Bidirectional elimination** combines both forward selection and backward elimination and test at each step for variables to be added or removed. Thus, a variable might be added in Step 3, dropped in Step 5, and added again in Step 8.

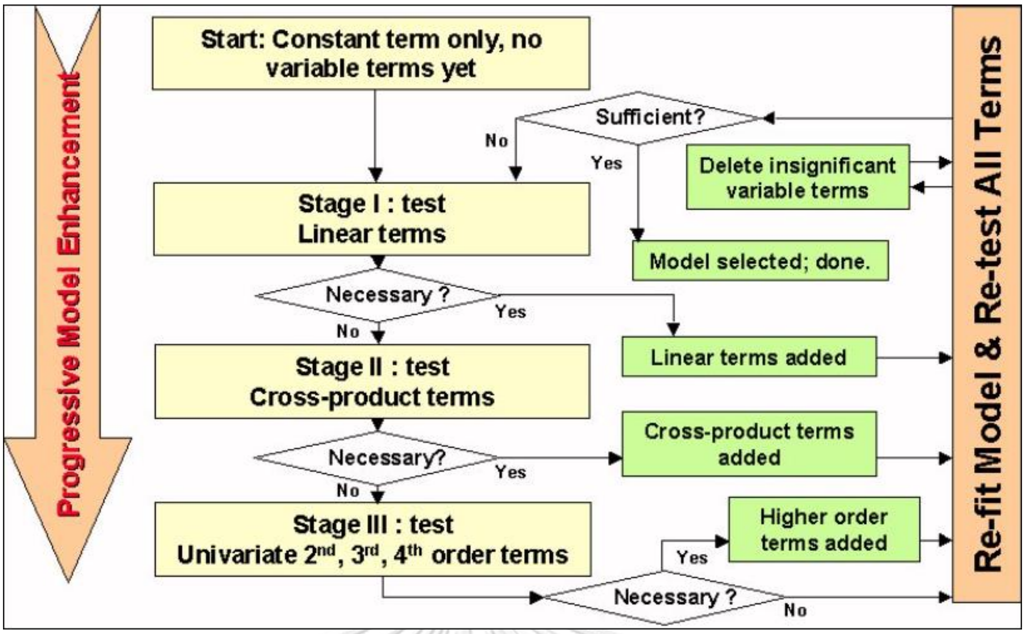


Figure 10 Simplified diagram for steps for stepwise regression

However, there was criticism on stepwise approach i.e. stepwise estimates are not invariant to inconsequential linear transformations, a local optimization obtained by including variables one-by-one is not necessarily to be a global optimization, and sometimes select explanatory variables that do not directly affecting the dependent variable but are systematically related to variables that affect the dependent variable. Some of explanatory variables that have effect on dependent variable might be ignored (not statistically significant) while nuisance variables might be coincidentally significant, resulting in good fitting for data in-sample but do poorly for out-of-sample. Smith [27] applied Monte Carlo simulations to proof that a stepwise approach might select nuisance variables rather than true variables. Moreover, model's accuracy for the out-of-sample was found to be worse than the in-sample and less effective for the larger the number of potential explanatory variables.

According to above arguments on the stepwise procedure, other regression techniques are applied in comparison to this technique to determine the optimal model.

2.2.4.3 Ridge Regression

This technique is used to analyze multiple regression data that suffer from multicollinearity which independent variables are highly correlated. Although least squares are unbiased, variances are large making observed value far from true value when multicollinearity occurs. Ridge regression aims to reduce the standard errors. General equation for ridge regression model is as per below [28].

$$y = \beta_0 + \beta_1 x + e \tag{4}$$

For multiple independent variables,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + e \tag{5}$$

where y is the dependent variable, x is the independent variables, β_0 & β_1 are the regression coefficients to be estimated, and e represents the errors

Error term (e) is the value needed to correct for prediction error between observed value and predicted value. If x is centered and scaled matrix, the cross product matrix (x'x) is nearly singular when x columns are highly correlated. Ridge regression proceeds by adding a ridge parameter, λ, to the cross product matrix, forming a new matrix (x'x + λI). The new formula to find the coefficients is as follows.

$$\beta = (x'x + \lambda I)^{-1} x'y \tag{6}$$

where	y	=	dependent variable
	x	=	independent variable
	β	=	coefficient of independent variable
	λ	=	tuning parameter or regularization penalty
	I	=	identity matrix

The objective function is to minimize the sum squares of errors as per below.

$$\sum_{j=1}^{n}(y_i - \sum_j x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}(\beta_j)^2 \tag{7}$$

where y = dependent variable

x = independent variable

β = coefficient of independent variable

λ = tuning parameter

Ridge regression is applied under concepts as per below:

- Same assumptions as least squared regression (regression model is linear in the coefficients & error term, all independent variables are uncorrelated with the error term, observations of the error term are uncorrelated with each other, and the error term has constant variance), except not assume normality
- Shrink the value of coefficients but not reach to zero which indicates no selection feature
- Regularization method (Process of adding information to solve an ill-posed problem or to prevent overfitting) and use L2 regularization

L2 regularization element or $\lambda\sum_{j=1}^{p}(\beta_j)^2$ term in equation 7 is adding squared magnitude of coeffiecient as penalty term to the loss function. Loss function is calculated from sum of squared difference between the actual value and the predicted value. This regularization term is added to keep the weights small, making the model to be simpler and avoid overfitting. λ, the penalty term or regularization parameter, determines how much to penalize the weights. λ is a positive quantity less than one (usually less than 0.3) [29]. In addition, Ridge regression performs better when all the input features influence the output and all with weights are of roughly equal size [30].

Note:

- As $\lambda \rightarrow 0$, β_{ridge} becomes close to β_{OLS}
- As $\lambda \rightarrow \infty$, $\beta_{ridge} \rightarrow 0$
- As λ becomes larger, the variance decreases, and the bias increases.

Generally, there are two ways to find the optimal value for λ. One is the traditional approach based on information criterion, i.e. AIC or BIC, is the smallest. The other is machine learning-like approach to perform cross-validation and select λ that minimizes sum of squared residuals or forecasting error. Figure 11 shows the effect of tuning parameter on forecasting error. As tuning parameter increases, MSE decreases to a particular point then exponentially increases as the tuning parameter further increases.

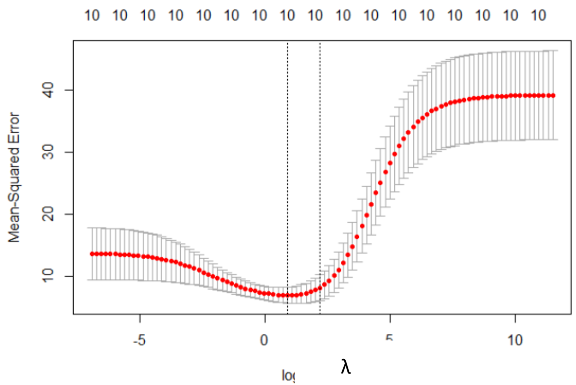


Figure 11 Effect of tuning parameter variation to the model error

2.2.4.4 Lasso Regression (Least Absolute and Selection Operator)

Similar to Ridge regression but apply L1 regularization in objective function, this Lasso technique reduces the variability and improves the accuracy of linear regression models. In L1 regularization, the absolute value in penalty function in $\lambda \sum_{j=1}^{p} |\beta_j|$ term in equation 8 is used for Lasso regression instead of squares as do in Ridge regression, causing some of parameter estimates to be exactly zero from equivalently constraining the sum of the absolute values of the estimates. This means that the larger penalty being applied, the further the estimates getting shrunk towards absolute zero, leading to variable selection out of given n variables. This technique picks only one of predictor from the group of highly correlated predictors and shrinks others to zero. In summary, the variables that coefficient converge to zero are considered to be insignificant and can be removed for model build-up. Feature selection's characteristic of Lasso helps to filter only variables that have contribution to the output, minimize the number of independent variables so it reduces overfitting, and can handle high-dimensional data [31].

Lasso regression is applied under concepts as per below:

- Same assumptions as least squared regression, except not assume normality
- Shrink coefficients to zero (exactly zero) which helps in feature selection
- Regularization method and use L1 regularization

Solutions of this regression are in quadratic form. The objective function is to minimize:

$$\sum_{j=1}^{n}(y_j - \textstyle\sum_{j=1}^{n} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \quad (8)$$

where y_j = dependent variable

x_{ij} = independent variable

β_j = coefficient of independent variable

λ = tuning parameter

As goal is to minimize sum of squares, some of βs shrink to exactly zero and finally obtain this regression model, resulting in easy for interpretation. Unchange for intercept term if includes in the method [32].

λ, a tuning parameter, indicates strength of L1 penalty or amount of shrinkage.

- λ= 0, not eliminate any parameter → Equal to linear regression
- As λ increases, more coefficients are set to zero & eliminated → λ =∞, all coefficients are eliminated.
- λ increases, bias increases
- λ decreases, variance increases

Figure 12 illustrates that coefficients of all variables converge to zero as λ increases and ultimately become zero due to high penalty strength added [33]. The optimal tuning parameter is the one that provides the model with the lowest forecasting error. Moreover, the error varies as changing tuning parameter. It tends to slightly decrease to one point then will substantially increase as the tuning parameter increases as shown in Figure 13 on different MSE on each fold from cross-validation when the tuning parameter increases [34].

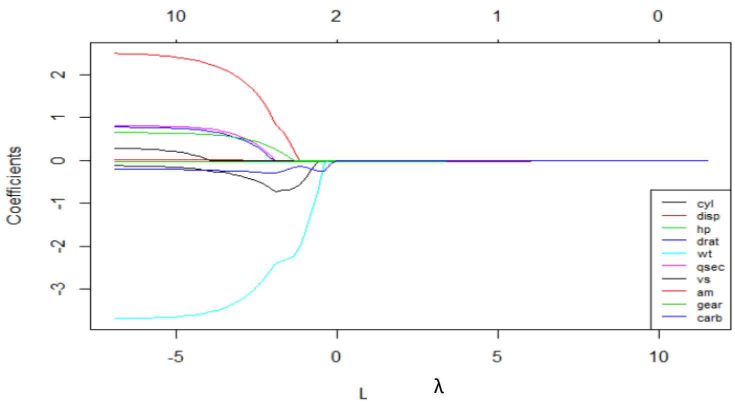


Figure 12 Effect of tuning parameter to the variable coefficient

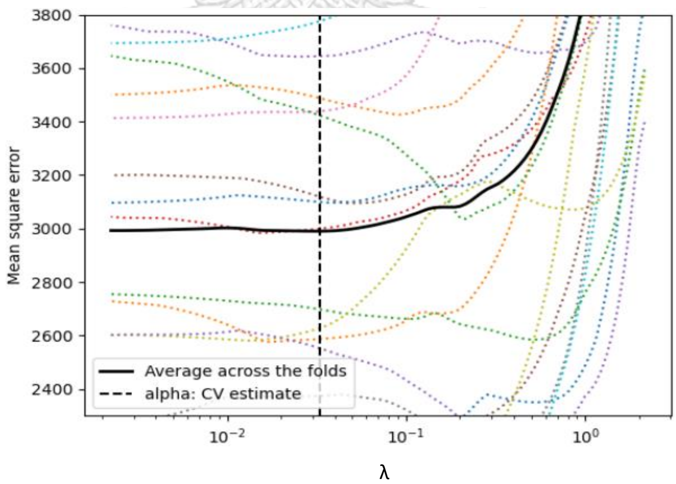


Figure 13 Effect of tuning parameter to the forecasting error for k-fold cross-validation

The optimal tuning parameter is determined using k-fold cross-validation technique to find the model input. See more detail for k-fold cross-validation in Section 2.2.3.

Lasso regression is better than Ridge regression in the way that it can perform both in-built variable selection and parameter shrinkage while Ridge regression ends up with getting all variables but with shrunk parameters. Thus, only Lasso model is decided to use for further model study.

2.2.4.5 Elastic Net Regression

This regression is hybrid of Ridge and Lasso techniques, combining both L1 and L2 regularization. It is generally applied for highly correlated independent variables. The main difference from Lasso is that Lasso picks one of these randomly while Elastic net picks both. Elastic net having practical advantage from trading-off between Ridge and Lasso regressions as inheriting some of Ridge's stability under rotation [32].

The objective function is to minimize:

$$\sum_{i=1}^{n}(y_i - \textstyle\sum_{i=1}^{n} x_{ij}\beta_j)^2 \sum_{j=1}^{p}\lambda_1(\beta_j)^2 + \lambda_2\sum_{j=1}^{q}|\beta_j| \qquad (9)$$

where y_i = dependent variable

x_{ij} = independent variable

β_j = coefficient of independent variable

λ_1 = tuning parameter of L2 regularization (Ridge)

λ_2 = tuning parameter of L1 regularization (Lasso)

Elastic net regression is applied under assumptions as per below:

- Not assume normality
- No limitation on number of selected variables

This regression encourages group effect in case of higher correlated variables [35].

2.2.4.6 Partial Least Squares Regression (PLS)

PLS is applied to describe the relationship between a set or predictors and response when predictors are highly collinear or have more predictors than observations. This technique reduces the predictors to a smaller set of uncorrelated components and performs least squares regression on these components instead of on the original data. The predictors can be measured with error according to not assume that they are fixed, unlike multiple linear regression, making this model more robust to measurement uncertainty.

PLS uses nonlinear iterative partial least squares (NIPALS) algorithm that uses technique similar to principal components analysis to reduce number of predictors to extract a set of components that describes maximum correlation between the predictors and response variables. Normally, cross-validation is used to identify the smaller set of components to enhance the best prediction ability [36]. Equation for coefficients of PLS is shown below [37].

$$\beta_{ij} = \frac{\operatorname{cov}(x_j, y)}{\sqrt{\sum_{j=1}^{p} \operatorname{cov}^2(x_j, y)}} = \frac{\langle x_j, y\rangle}{\sqrt{\sum_{j=1}^{p} \langle x_j, y\rangle^2}} \tag{10}$$

where y = dependent variables

x_j = explanatory variables

β_{ij} = coefficient

Note: PLS regression leads to a linear model the same as the OLS and PCR do.

2.2.5 Model Selection

There are two main topics for model selection, model fitting and forecast error.

2.2.5.1 Model Fitting

Objective is to determine how well the forecasting model generalizes to similar data in which it was trained.

2.2.5.1.1 R-squared

R-squared (R^2) represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model [38].

$$R^2 = 1 - \frac{\Sigma e_i^{\,2}}{\Sigma(y_i - y^\wedge)^2} \tag{11}$$

where y_i = actual value

$y^\wedge$ = predicted value

e_i = error = $y_i - y^\wedge_i$

2.2.5.1.2 Adjusted R-squared

It is a modified version of R-squared that is adjusted for the number of predictors in the model, indicating how well the model fits with the data. It increases only if the new added parameter improves the model so the bias in prediction from adding more parameters in R-squared is resolved. Theoretically, Adjusted r-squared was calculated from following equation [39]:

$$R^2_{adj} = 1 - \left(\frac{\Sigma(yi - \hat{y}i)^2}{\Sigma(yi - \hat{y})^2}\right)\left(\frac{n-1}{n-p-1}\right) \tag{12}$$

where y_i = i^{th} observed response value

$\hat{y}_i$ = i^{th} fitted response

$\hat{y}$ = Mean response

n = Number of observations

p = Number of terms in the model

Adjusted R-sq is always between 0 and 100%.

- 0% means the model explains none of the variability of the response data around its mean.
- 100% means the model explains all the variability of the response data around its mean.

2.2.5.1.3 Predicted R-squared

The R-sq(pred.) represents the prediction ability of the model with the new dataset from regression analysis when the observations are systematically removed in dataset. The higher value of the R-sq(perd.) indicates that the model has high ability to predict accurately. The equation is shown below [39].

$$R^2_{pred} = 1 - \left(\frac{\Sigma\left(\frac{ei}{1-hi}\right)^2}{\Sigma(yi - \hat{y})^2}\right) \tag{13}$$

where	y_i	=	i[th] observed response value
	$\hat{y}$	=	Mean response
	n	=	Number of observations
	e_i	=	i[th] residual
	h_i	=	i[th] diagonal element of $x(x'x)^{-1}x'$
	x	=	Predictor matrix

2.2.5.2 Forecasting Error

According to statistics, a forecast error is the difference between actual and the predicted value of phenomenon of interest. This statistical measure also employs to assess the error of forecasting model. In a linear equation, prediction errors can be decomposed into two sub-components. First is due to the bias and second is due to the variance.

2.2.5.2.1 Mean absolute error (MAE)

MAE is an average of all absolute error [40].

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |A_t - F_t| \tag{14}$$

where	A_t	=	actual values
	F_t	=	corresponding forecasted or predicted values

2.2.5.2.2 Mean squared error (MSE)

MSE indicates how close a regression line to a set of data points. The squaring is required to remove negative signs and emphasize more weight to larger differences. Values closer to zero are the better [40].

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (A_t - F_t)^2 \tag{15}$$

2.2.5.2.3 Mean absolute percentage error (MAPE)

MAPE which is a measure of prediction accuracy of a forecasting statistical method is robust to the effect of outliers from using absolute value. Moreover, MAPE alleviates problem with MAE and MSE that the values depend on the magnitude of item being forecasted [41]. Thus, MAPE is selected to assess performance of models in each scenario in this study [42].

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \frac{|At - Ft|}{At} \ \%, \tag{16}$$

where A_t = actual values

F_t = corresponding forecasted or predicted values (y)

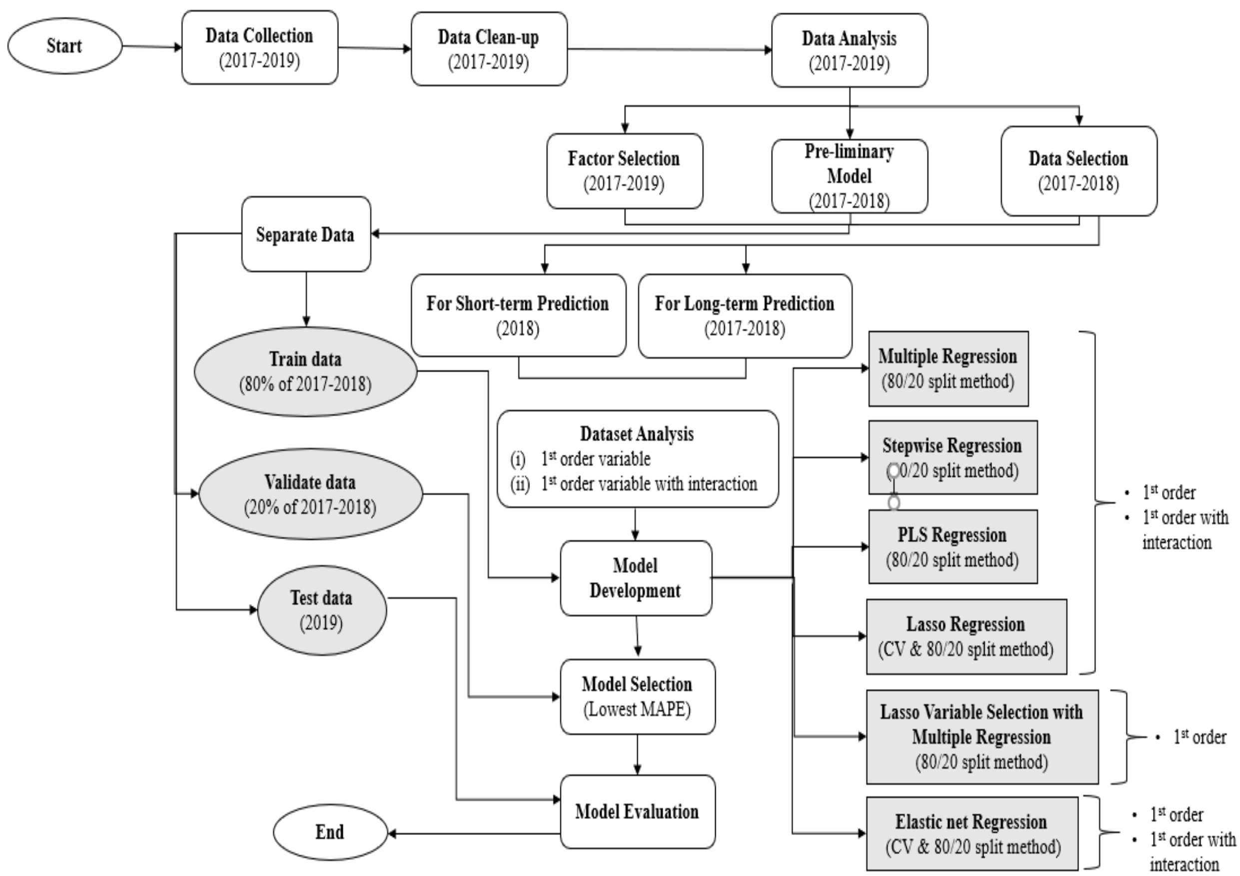
n = no. of data [40]

CHAPTER III
METHODOLOGY

At the beginning, crew boat's operation, internal and external factors which potentially had effect on crew boat's sailing time were studied in detail for more understanding of possible relationship and critical parameters to develop predictive models to match with current operation. After removing outliers, cleaning data, and selecting data, several regression models (Multiple regression, Stepwise regression, PLS regression, Lasso regression, and Elastic net regression) of collected data were run. The obtained forecasting models were evaluated by MAPE to compare performance to yield the optimal forecasting model for AA company.

Experimental Procedures

The overall steps are illustrated in schematic below.



3.1 Data Collection

Hourly operational data of two main crew boats, Boat A and Boat B, used to transfer offshore personnel and cargos were collected known as ship reports. The data included in the ship reports are date, time, estimated time of departure from & arrival to each platform, distance, boat speed, wind direction, and total main engine diesel consumption. The possible impact of collected parameters on travelling time were studied through literature review and consulting with subject-matter expert (SME) at central engineering department. Regarding to literature review, distance, boat speed, wind speed, and wave height were used for model built-up. In addition, SME recommended that wave direction and wind direction potentially had effect on sailing time so they were added for prediction model. Thus, the additional parameters apart from ship reports which were wave height (both significant (Hs)[1] and maximum wave height (Hmax)[2]), wave direction, and wind speed (from both North-East and South-West sensors) need to be collected to forecast the total travelling time.

In summary, the internal factors for this analysis were distance, boat speed while the external factors were significant wave height, maximum wave height, wave direction, wind speed, and wind direction. The data of selected variables were collected from 2017 to 2019. See list for input data and expected effect to crew boat's travelling time in Table 2.

3.2 Data Clean-up

The collected hourly data (24-hr) was focused only time interval between 6:00 A.M. and 10 A.M which is the normal operating hours of morning route's crew boats to use for further analysis using 2017-2019 data. The total travelling time was calculated from the actual boat sailing time without time for dropping passengers and loading cargos. Total distance was obtained from travelling distance from starting point

---

[1] Average of 1/3 highest wave heights, expressed in metres.

[2] Most likely or actual maximum individual wave height over an 18 minute period, expressed in meters.

to ending platform. After getting raw data, the data preparation was done as per following steps.

Table 2 Input Parameters considered in Travelling Time Prediction

Category	**Parameters**	**Input Data**	**Effect on Travelling time**
Internal Factors	Distance (mile)	Total Distance	Longer distance, longer travelling time
	Speed (knot)	Boat Speed	Higher boat speed, shorter travelling time
External Factors	Wave height (m.) and direction (˚)	Significant Wave Height	Higher wave height will obstruct sailing and travelling time will increase.
		Maximum Wave Height	
		Wave Direction	If wave direction is in the same direction to the boat route, it will support sailing thus travelling time will be shorten.
	Wind speed (knot) and direction (˚)	Wind Speed (E1)	Boat can go faster (lower travelling time) with greater the relative wind, resulting on more force on the sails and greater dragging the boat forward. Note: Boat accelerates until the drag from the water balances the forward component of the force from the sails.
		Wind Speed (E2)	
		Wind Direction (E1)	The apparent wind angles behind the sail results in lower lift and increase drag as the predominant component of propulsion. On the contrary, a sail can propel a boat to a higher speed with a true wind velocity over the surface on points of sail when the entry point of the sail is aligned with the apparent wind, resulting in shorter travelling time.
		Wind Direction (E2)	

3.2.1 Outliers, Missing data, and Influential points

The abnormal data points considered as outliers were removed since input and output are highly sensitive with each other. Only small change in input might have significant effect on output. Hence, outliers were identified and cleaned before input in the model. Feasible and operating ranges were checked against all data and outliers (under or over range) were removed. For example, Boat A has specification of maximum boat speed of 27 knots. Boat speed data that was greater than 27 knots was removed.

The outliers of input data mainly come from either operating conditions or error of measurement devices.

A crew boat was sometimes stopped during a journey and the obtained boat speed during that period might cause deviation of average boat speed that is input data for model formation. Thus, those data which fallen into operating condition's category were eliminated as shown in Table 3.

On the contrary, error from measurement devices that did not show data was considered as outliers. According to Table 4, the number is obviously different from others. If they are included in the model, it will certainly cause deviation.

Table 3 The example of missing data during crew boat sailing from boat speed

Date	Travelling Time	Total Distance	Boat speed, 6 A.M.	Boat speed, 7 A.M.	Boat speed, 8 A.M.	Boat speed, 9 A.M.	Average Boat speed	Significant Wave Height
6/6/2018	1.550	24.1	36.60	0.00	0.00	15.10	17.23	0.60
10/23/2018	1.300	21.7	30.10	0.00	0.00	7.30	12.47	0.20
11/2/2018	1.650	23.6	17.30	0.00	0.00	6.20	7.83	0.88
7/17/2019	0.88	16.00	10.90	0.00	0.00	10.50	7.13	1.20

Table 4 The example of outliers and missing data from error of measurement devices

Date	Travelling Time	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed	Wind Direction
1/2/2017	1.650	24	24.70	1.08	100.8	#N/A	#N/A
2/20/2017	1.417	20.9	18.43	0.70	156.8	#N/A	#N/A
11/1/2017	0.367	4.5	20.93	1.23	168.8	#N/A	#N/A
4/23/2018	1.817	28.3	25.87	0.20	138.5	#N/A	#N/A
6/4/2018	0.650	10.4	18.03	0.30	197.8	#N/A	#N/A
10/20/2018	1.700	29.4	21.40	0.33	205.8	#N/A	#N/A
6/7/2018	1.467	23.4	25.50	15.0	189.4	3.8	225.3
6/8/2018	1.317	20.5	17.15	#N/A	#N/A	1.8	226.1
6/13/2018	1.200	21.5	19.80	#N/A	#N/A	4	229.7
6/14/2018	1.200	20.5	15.23	#N/A	#N/A	1.2	231.8
6/15/2018	0.983	15.2	14.35	#N/A	#N/A	2.2	298.3
6/16/2018	0.417	6.8	12.60	#N/A	#N/A	3.7	257.2

As identified outliers from box plot, there were some outliers as marked by x presented in the raw data for both boats as shown in Figure 14 and Figure 16. Then, the outlier in the hourly data either from internal or external factor was removed from the entire row for that exact hour of all factors. There was no outlier for the cleaned data as illustrated in Figure 15 and Figure 17.

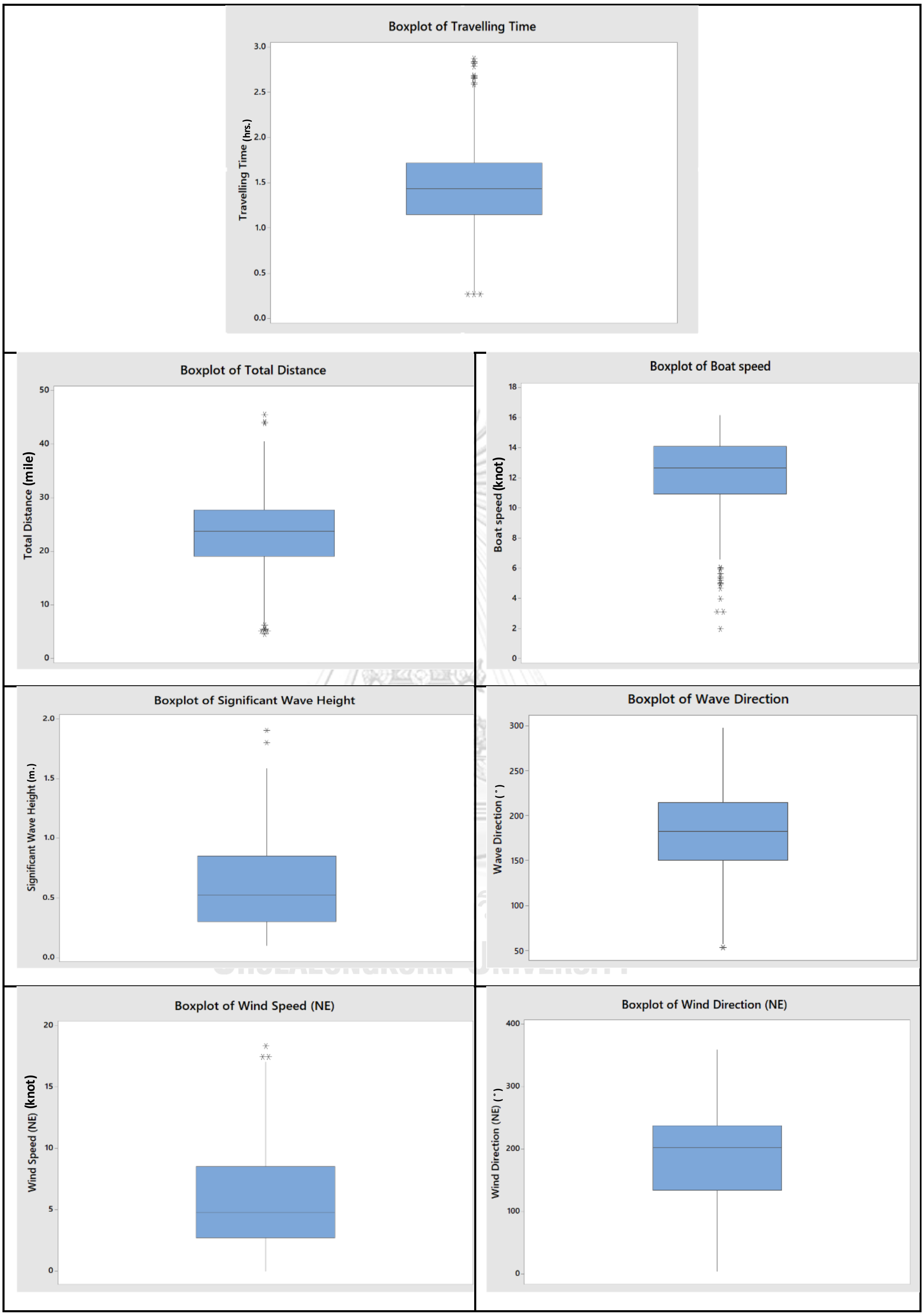


Figure 14 Box plot for raw dataset of Boat A

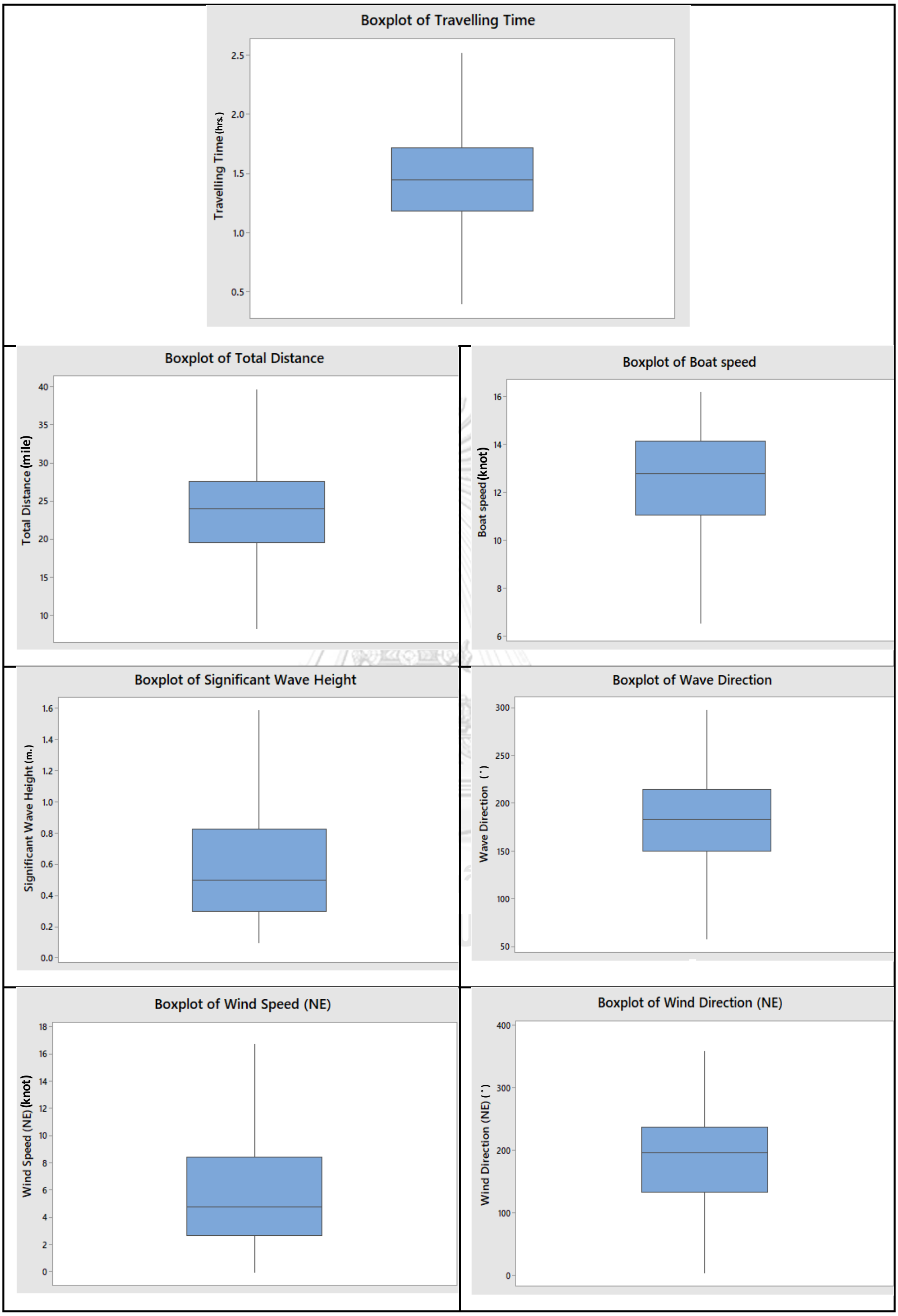


Figure 15 Box plot for dataset of Boat A after eliminating outliers

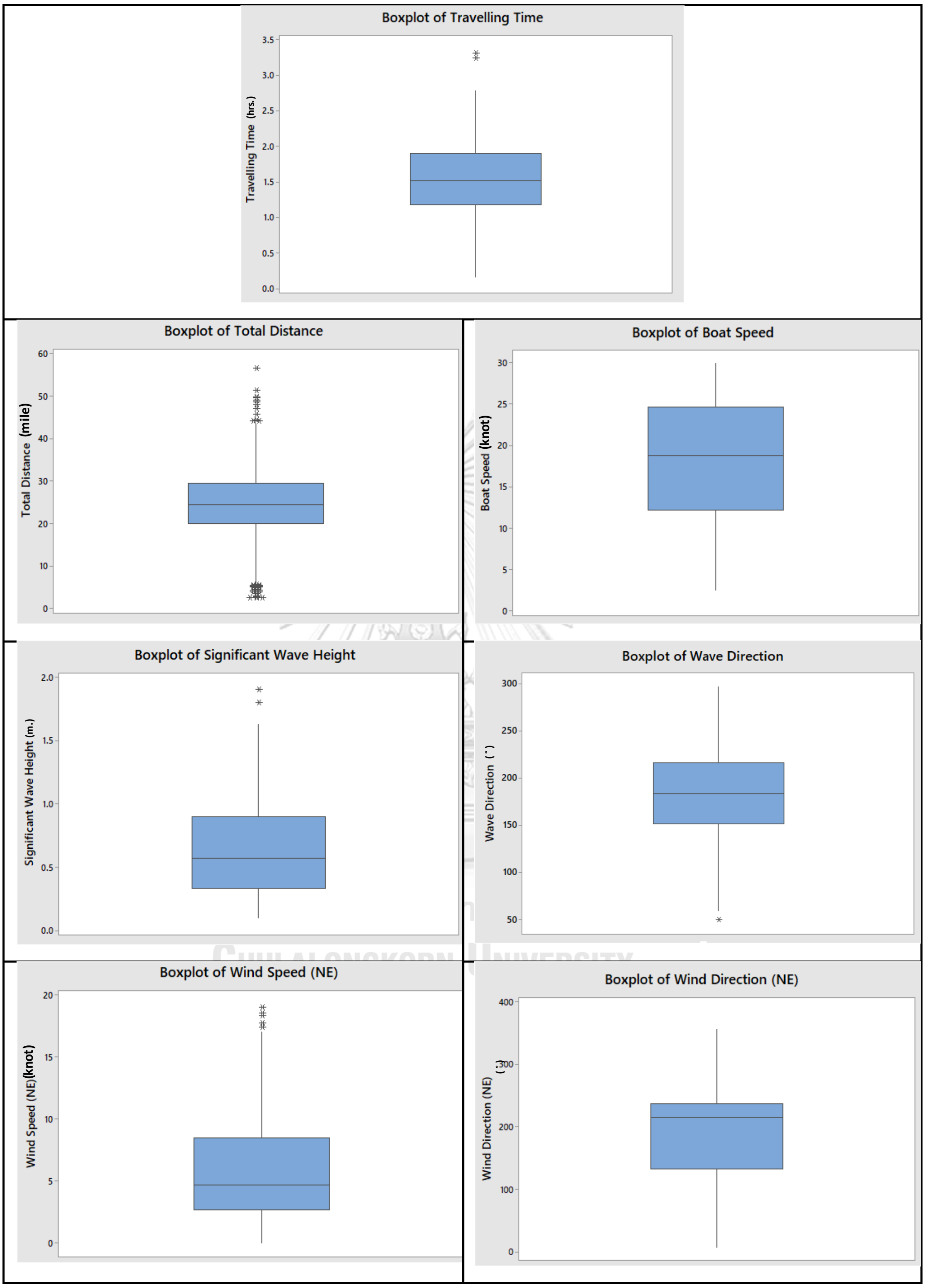


Figure 16 Box plot for raw dataset of Boat B

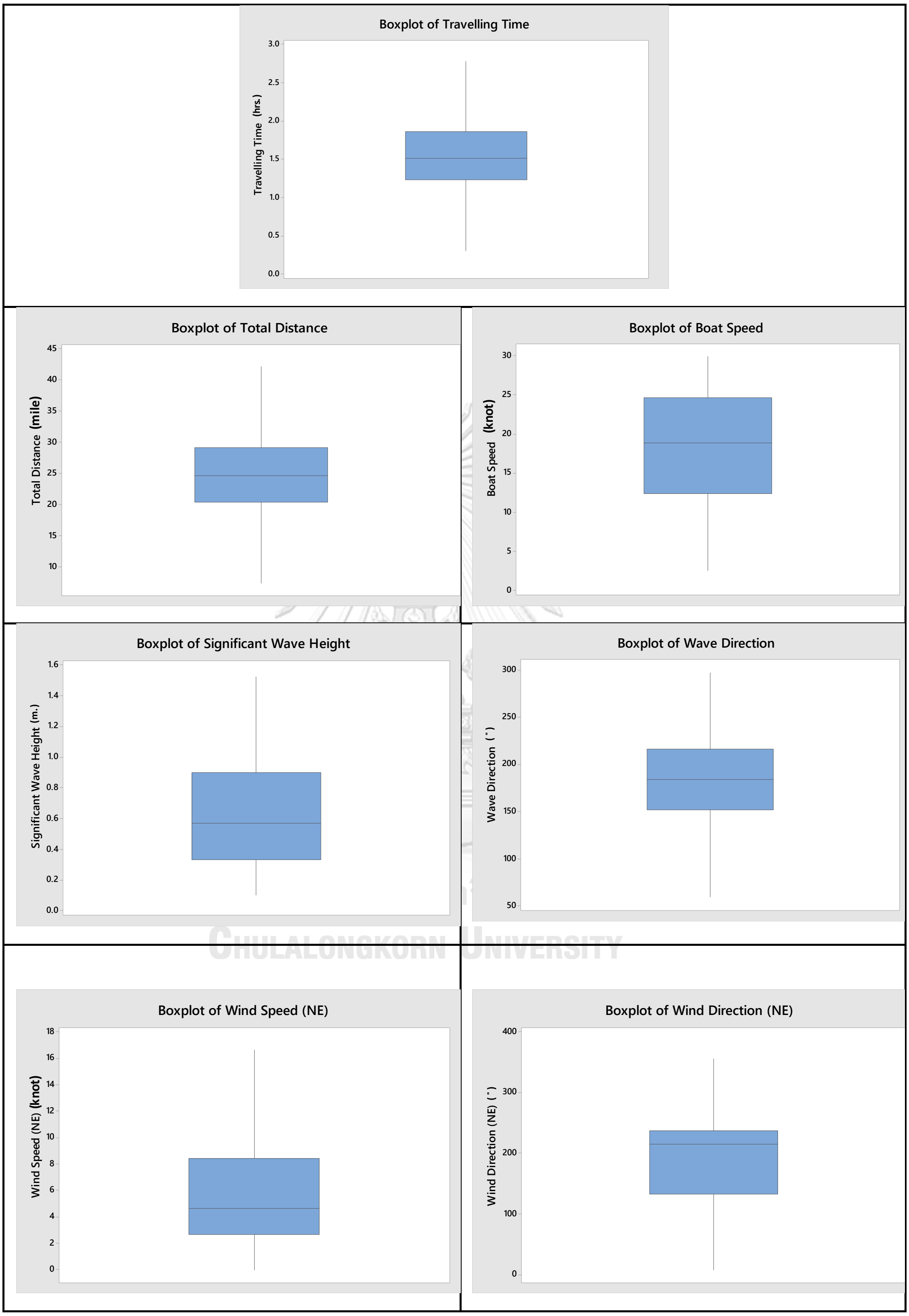


Figure 17 Box plot for dataset of Boat B after eliminating outliers

Furthermore, the influential points were eliminated from the dataset to avoid the misleading prediction using cook's distance analysis. The removed period is considered based on the influential point of travelling time (endogenous variable), the analysis is shown in Figure 18. Cook's values higher than 0.0051 (4/783 observations) for Boat A, higher than 0.0065 (4/618 observations) for Boat B, and higher than 0.0029 (4/1401 observations) for Combined boat were eliminated from the dataset. The final plots were shown in Figure 19 where no influential point presented in the dataset.

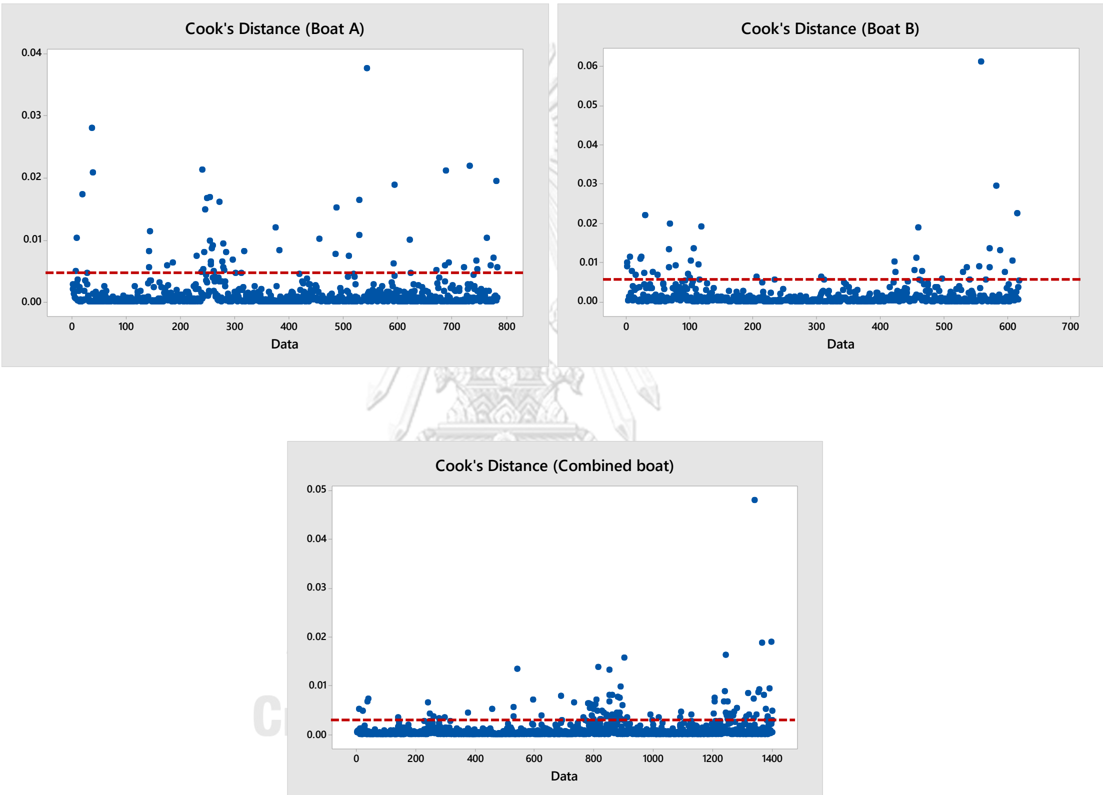


Figure 18 Cook's distance plot from raw dataset of Boat A, B, and Combined boat

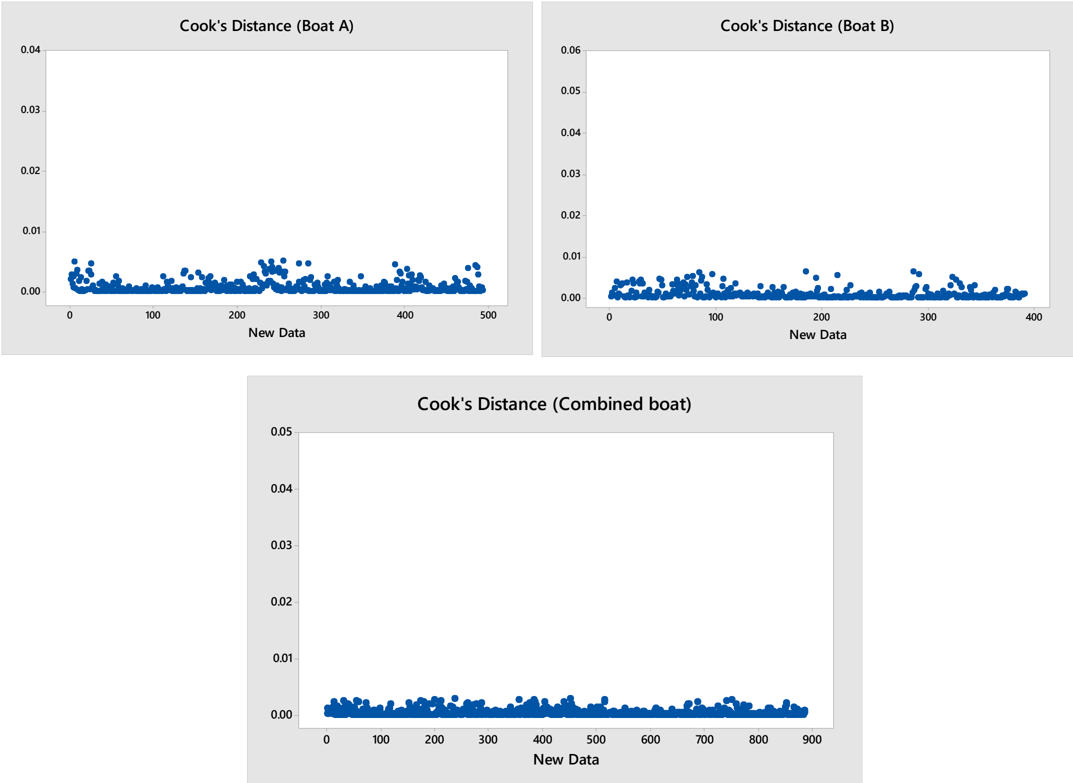


Figure 19 Cook's distance plot from dataset of Boats A, B, and Combined boat after eliminating influential points

3.3 Data Analysis

3.3.1 Factor Selection

The cleaned dataset was analyzed for variables selection and the best time period for the data of external factors for short-term and long-term prediction models was determined in this section.

Firstly, final parameters to be included in the model were determined using correlation coefficient to assess the strength of the relationship between pair variables. Pearson's correlation coefficient which is higher than 0.7 indicates high degree of correlation and suggests that one of those variables should be removed from the model [43]. Thus, this criteria of greater than 0.7 was applied in this study using 2017-2019 data. Some parameters i.e. maximum wave height, wind speed (E2) and wind direction (E2) were eliminated from the model due to highly correlated with other parameters which were significant wave height, wind speed (E1), wind direction (E1), respectively as shown in Table 5. Maximum wave height was removed from the model built-up because significant wave height was the better representative of major wave height data than the peak one shown by maximum wave height. Moreover, E1's data, both wind speed and wind direction, were selected rather than E2's data according to higher reliability of measurement device as rechecked from maintenance data. The final parameters for the model were listed in Table 6.

Table 5 Pearson Correlation among input parameters

	Significant Wave	Maximum Wave Height	Wave Direction (E1)	Wind Speed (E1)	Wind Direction (E1)	Wind Speed (E2)
Maximum Wave Height	**0.995**					
Wave Direction	-0.017	-0.014				
Wind Speed (E1)	0.214	0.207	-0.100			
Wind Direction (E1)	-0.075	-0.069	0.107	-0.422		
Wind Speed (E2)	0.109	0.110	0.072	**0.689***		
Wind Direction (E2)	-0.003	0.002	0.159	-0.510		
Wind Speed (E2)					0.314	
Wind Direction (E2)					**0.740**	0.475

Note: * = Remove due to close to 0.7 and significant higher value than others.

Then, all selected variables were plotted to see change overtime. Time series plots in Figure 20 (boat A) and Figure 21 (boat B) for daily data depicted that there was no trend, cycle or seasonal for these variables, except for wind direction. Moreover, 2019 data were used as test set to estimate an expected error from the best predicting model when further applied to unknown data and evaluate the optimal model.

According to different scales from various measurement units, data were standardized to convert all variables into the same scales prior plugging in models.

Table 6 Final Input Parameters for in Travelling Time Prediction

Category	Parameters	Input Data
Internal Factors	Distance	Total Distance
	Speed	Boat Speed
External Factors	Wave	Significant Wave Height
		Wave Direction
	Wind	Wind Speed (E1)
		Wind Direction (E1)

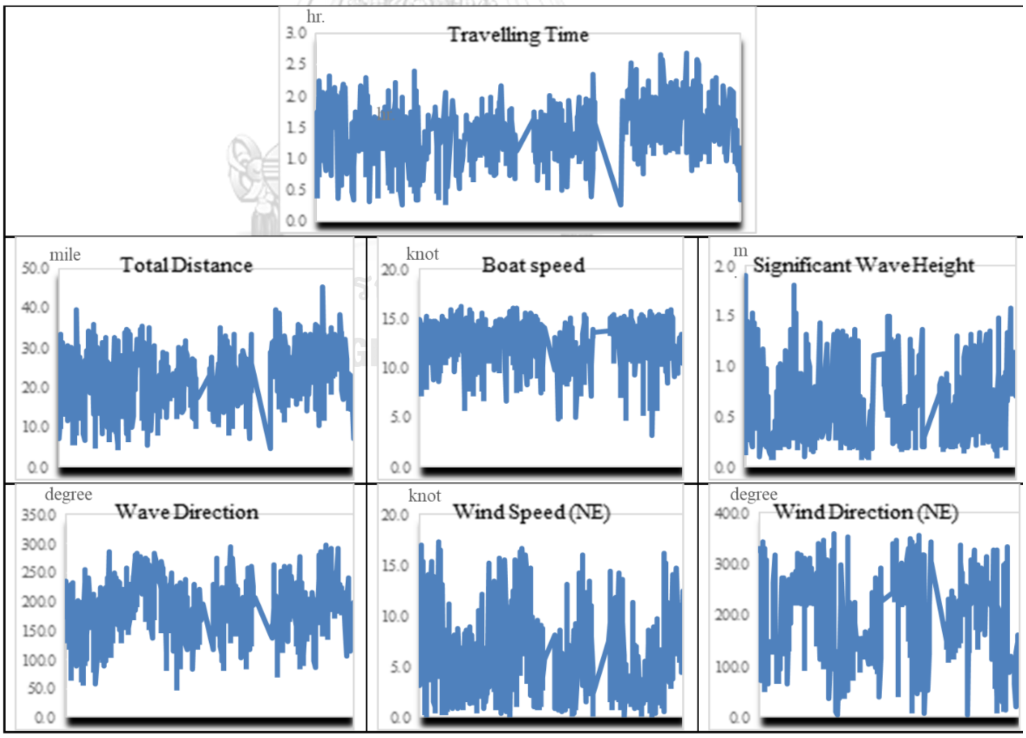


Figure 20 Time series plot of all variables (Boat A)

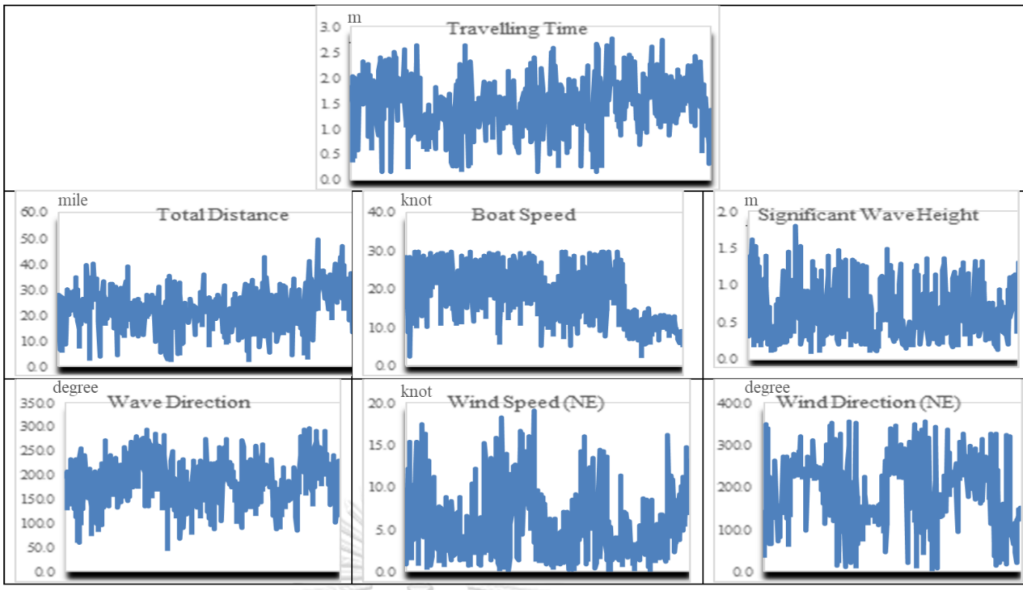


Figure 21 Time series plot of all variables (Boat B)

3.3.2 Preliminary Model

Initially, preliminary multiple regression models were built-up using 1st order dataset of 2017-2018 data for boats A and B to confirm whether the exact time period that crew boats sailed (6-10 A.M.) was the best representative dataset for further processing. Input data for these models were daily travelling time, daily total distance, 4-hr average (6-10 A.M.) for boat speed, significant wave height, wave direction, wind speed, and wind direction. MAPEs obtained for boat A was 8.69% while boat B was 10.73%, showing low MAPE results for both boats. According to this trial with multiple regression, it indicated that the exact time interval as the crew boats sailed was the best in representing the actual condition. Hence, this time interval was used for the baseline data for forecasting to build-up the models.

3.3.3 Data Selection for Short-term Prediction

Short-term prediction is inclusion of recent past data to forecast for the next day (24 hours). The main objective of this prediction is for daily operational adjustment on the proper boat speed to meet the required sailing time of 1.22 hrs per annual goal. The hourly 2018 data of all variables were considered in this section.

The input data selection can be divided into two groups, group A and group B. Group A was selected based on the exact operating time of daily crew boats which was for travelling time, distance, and 4-hr average (6-10 A.M.) for boat speed. These factors can be directly plugged in the model for future application.

On the other hand, group B was for environmental data (significant wave height, wave direction, wind speed, and wind direction). Real time data of these factors cannot be directly used in the model in the real application since there is no environmental data at that time (6-10 A.M.) when crew boats sail for prediction. Hence, the 24-hr hourly data before crew boats sail and 4-hr average (before at 2-6 A.M.) were used to compare with the actual values (4-hr average data between 6-10 A.M.) which previously found to be the best in showing the exact condition to calculate MAPE as shown in section 3.3.2. The hourly time period that provided the lowest MAPE was selected to represent the actual environmental data when the boats sailed in the prediction models. MAPE is calculated from average absolute percent error for each hourly period (10 A.M. the day before until 5 A.M.) minus actual values divided by actual values. It was found that 4-hr average time before sailing (2-6 A.M.) had the lowest MAPE for significant wave height and wave direction from Table 7. On the other hand, data at 5 A.M. shown to have lowest MAPE for wind speed and wind direction. Thus, data for 4-hr average (before) and 5 A.M. time were used for wave and wind data, respectively for further forecasting.

3.3.4 Data Selection for Long-term Prediction

Long-term prediction is inclusion of the past data to provide information about expected future condition. The main objective of this prediction is for future planning for total travelling time of crew boats that AA company aims to forecast 1 month in advance. In this analysis, the daily data of 2017 and 2018 were used. For data group A, the same basis as short-term to use the exact operating time of daily crew boats for travelling time & distance and 4-hr average (6-10 A.M.) for boat speed was applied for long-term prediction.

Table 7 MAPE comparison of environmental data for short-term prediction

Time	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
10 A.M.	35.2%	25.0%	95.9%	35.3%
11 A.M.	36.9%	55.2%	69.9%	25.9%
12 P.M.	36.0%	34.4%	70.2%	100.6%
13 P.M.	36.8%	34.1%	112.4%	36.1%
14 P.M.	36.8%	43.5%	87.0%	39.1%
15 P.M.	36.5%	54.3%	84.8%	54.2%
16 P.M.	34.7%	51.1%	80.3%	34.6%
17 P.M.	35.2%	62.2%	96.4%	35.1%
18 P.M.	34.7%	60.2%	91.3%	74.4%
19 P.M.	38.5%	38.0%	123.4%	139.7%
20 P.M.	35.7%	38.8%	96.2%	107.1%
21 P.M.	34.0%	40.5%	90.1%	135.6%
22 P.M.	37.4%	42.1%	88.8%	129.3%
23 P.M.	37.3%	40.2%	64.8%	20.5%
0 A.M.	35.1%	51.9%	63.8%	24.5%
1 A.M.	32.1%	98.1%	69.6%	21.4%
2 A.M.	26.2%	34.9%	58.5%	15.4%
3 A.M.	21.1%	52.6%	52.2%	14.8%
4 A.M.	16.1%	21.4%	44.6%	22.1%
5 A.M.	13.7%	18.5%	33.1%	14.2%
4-hr (before)	8.8%	10.3%	38.9%	22.2%

However, daily data cannot be used for environmental data as there is no information available for the sailing period and it is not practical to use for monthly pre-planning. Thus, the best time interval was determined by trialing from weekly, bi-weekly, monthly, bi-monthly, quarter, and half year based on the lowest MAPE. Firstly, the best time period obtained from short-term prediction (4-hr average (before) for significant wave height & wave direction and 5 A.M. time for wind speed & wind direction) was used for each daily environmental data (group B). Then, the data were further processed into each trial period as mentioned earlier and compared with the same intervals in 2017 and 2018 to calculate MAPE. For weekly interval calculation as an example, average daily data for 7 days i.e. 1st week of Jan'17 VS 1st week of Jan'18 were compared forecast error of each environmental data using the same MAPE method as short-term to select the proper time interval. The same interval for each year is the suitable data because there is no seasonal trend depicted from time series plot. The time interval that provided the lowest MAPE would be used for environmental data to forecast travelling time for long-term. From the results in Table 8, Quarter data depicted the lowest MAPE for overall parameters so they were selected to plug in the models.

Table 8 MAPE comparison of environmental data for long-term prediction

Data	**Significant Wave Height**	**Wave Direction**	**Wind Speed (E1)**	**Wind Direction (E1)**
Weekly	53.6%	20.0%	44.4%	16.7%
Bi-Weekly	29.7%	15.7%	30.9%	20.7%
Monthly	21.9%	12.6%	26.3%	17.8%
Bi-Monthly	13.3%	11.3%	19.1%	12.4%
Quarter	12.8%	11.4%	10.2%	11.6%
Half year	13.1%	11.9%	10.5%	12.2%

Examples of standardized data for the final input parameters for short-term and long-term prediction shown in Table 9 and Table 10.

Table 9 Example for final input variables of boat A for short-term prediction

Date	**Travelling Time**	**Total Distance**	**Boat speed**	**Significant Wave Height**	**Wave Direction**	**Wind Speed (E1)**	**Wind Direction (E1)**
2/1/2017	1.100	-0.471	-0.061	0.057	-1.053	0.148	-0.761
2/2/2017	1.483	0.245	0.354	-0.519	-2.240	-0.339	2.146
2/3/2017	1.067	-0.545	0.082	0.417	1.067	1.073	-0.887
2/5/2017	0.800	-1.276	1.223	-0.231	-0.829	-1.483	-0.432
2/8/2017	1.500	0.290	0.333	-0.303	-0.001	0.197	1.829
2/9/2017	1.233	-0.575	0.071	-0.735	1.139	-1.117	0.605
2/10/2017	1.217	-0.426	-2.255	-1.022	-0.293	0.002	-1.621
2/11/2017	2.067	0.722	-0.995	-0.159	-0.548	0.830	-0.485
2/14/2017	1.150	-0.933	-1.046	2.503	-0.241	0.002	-1.725
2/17/2017	1.950	0.439	-0.513	2.503	-1.813	1.584	-0.729

Table 10 Example for final input variables of boat A for long-term prediction

Date	Travelling Time	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
2/1/2017	1.100	-0.471	-0.061	1.196	-1.277	1.346	-0.921
2/2/2017	1.483	0.245	0.354	1.196	-1.277	1.346	-0.921
2/3/2017	1.067	-0.545	0.082	1.196	-1.277	1.346	-0.921
2/5/2017	0.800	-1.276	1.223	1.196	-1.277	1.346	-0.921
2/8/2017	1.500	0.290	0.333	1.196	-1.277	1.346	-0.921
5/2/2017	1.317	0.215	0.907	-1.051	-0.550	-0.732	0.760
5/3/2017	1.217	0.171	-0.061	-1.051	-0.550	-0.732	0.760
5/8/2017	1.583	0.737	0.993	-1.051	-0.550	-0.732	0.760
5/9/2017	1.500	0.439	0.825	-1.051	-0.550	-0.732	0.760
5/10/2017	1.583	0.558	-0.278	-1.051	-0.550	-0.732	0.760

3.4 Model Development

Same methods were applied for both short-term and long-term prediction. According to potential influence on model input, interaction term was also included and excluded from the model to ensure to include the most appropriate input data in section 3.4.2.

3.4.1 Dataset Analysis

The interaction terms between each variable were plugged in the model in comparison to one without interaction (only 1st order) to obtain the most appropriate dataset for the optimal model. 2nd order term was not included in the model as there was no curvature characteristic from matrix plot as shown in Figures 43-45, Appendix 1. Hence, the variable dataset was as follows:

(i) 1st order variable

(ii) 1st order variable with interaction terms.

K-fold cross-validation technique was applied during 1st step of model built-up for Lasso regression and Elastic net regression to find the optimal tuning parameter. This method equally divided number of data into 10 folds, took out one 1 group at a time as validate set, and other remaining was used as train set. Ultimately, all folds were switchingly used as train and validate sets for the model built-up. Detail for cross-validation technique was explained in Chapter 2, Section 2.2.3. Moreover, 80/20 split method which divided 80% of 2017-2018 data as train data (Jan'17-July'18) and the remaining 20% as validate data (Aug'18-Dec'18) was used to develop prediction models for Multiple regression, Stepwise regression, PLS regression, Lasso regression, and Elastic net regression. The lowest MAPE from input tuning parameter and dataset represented the best result that will be used to compare between each selected model to find the optimal prediction model. In addition, 2019 data were used as test set to evaluate the optimal prediction model.

3.4.2 Model Selection

To obtain the best prediction models, 6 models were compared using dataset from section 3.4.1.

3.4.2.1 Multiple Regression Model

(i) Run Multiple regression model with 2 types of input:
1st order and 1st order with interaction terms

(ii) Forecast crew boat sailing time

(iii) Calculate MAPE from comparison between the predicted values and actual validate data

3.4.2.2 Stepwise Regression Model

(i) Run Stepwise regression model with 2 types of input:
1st order and 1st order with interaction terms by setting stepwise in input with input alpha of 0.05 for variables elimination

(ii) Forecast crew boat sailing time

(iii) Calculate MAPE from comparison between the predicted values and actual validate data

3.4.2.3 Partial Least Squares Regression Model (PLS)

(i) Run PLS regression model with 2 types of input:
1st order and 1st order with interaction

(ii) Forecast crew boat sailing time

(iii) Calculate MAPE from comparison between the predicted values and actual validate data

3.4.2.4 Lasso Model

First, optimal tuning parameter was determined then build Lasso model with the optimal tuning parameter.

Determine optimal tuning parameter

(i) Build Lasso model by vary 20 tuning parameters from 0-1 using cross-validation technique (k = 10)

(ii) Find MAPE of each tuning parameter input

(iii) The tuning parameter provides lowest MAPE is considered optimal

Develop Lasso model with best tuning parameter input

(i) Input train data using 80/20 split method for 2017-2018 data to build Lasso model using the optimal tuning parameter input from the previous step

(ii) Forecast crew boat sailing time

(iii) Calculate MAPE from comparison between the predicted values and actual validate data

3.4.2.5 Lasso Variable Selection with Multiple Regression Model

(i) Input all train dataset and build Lasso model by varying tuning parameter from 0-1

(ii) Determine and observe coefficient trend of each variable

(iii) Eliminate variables which coefficient converge to zero, the remaining variables considered significant variables

(iv) Run Multiple regression model with selected significant variables

(v) Forecast crew boat sailing time

(vi) Calculate MAPE from comparison between the predicted value and actual validate data

3.4.2.6 Elastic Net Regression Model

(i) Build-up Elastic net model with all variables input using cross-validation technique (k = 10), fix L1 ratio at 0.5 equally weighted to L2 at 0.5, and vary λ_1 from 0-1 to find the best tuning parameters, λ_1 and λ_2

(ii) Input train data using 80/20 split method for 2017-2018 data to build Elastic net model using the optimal tuning parameters obtained from step (i)

(iii) Forecast crew boat sailing time

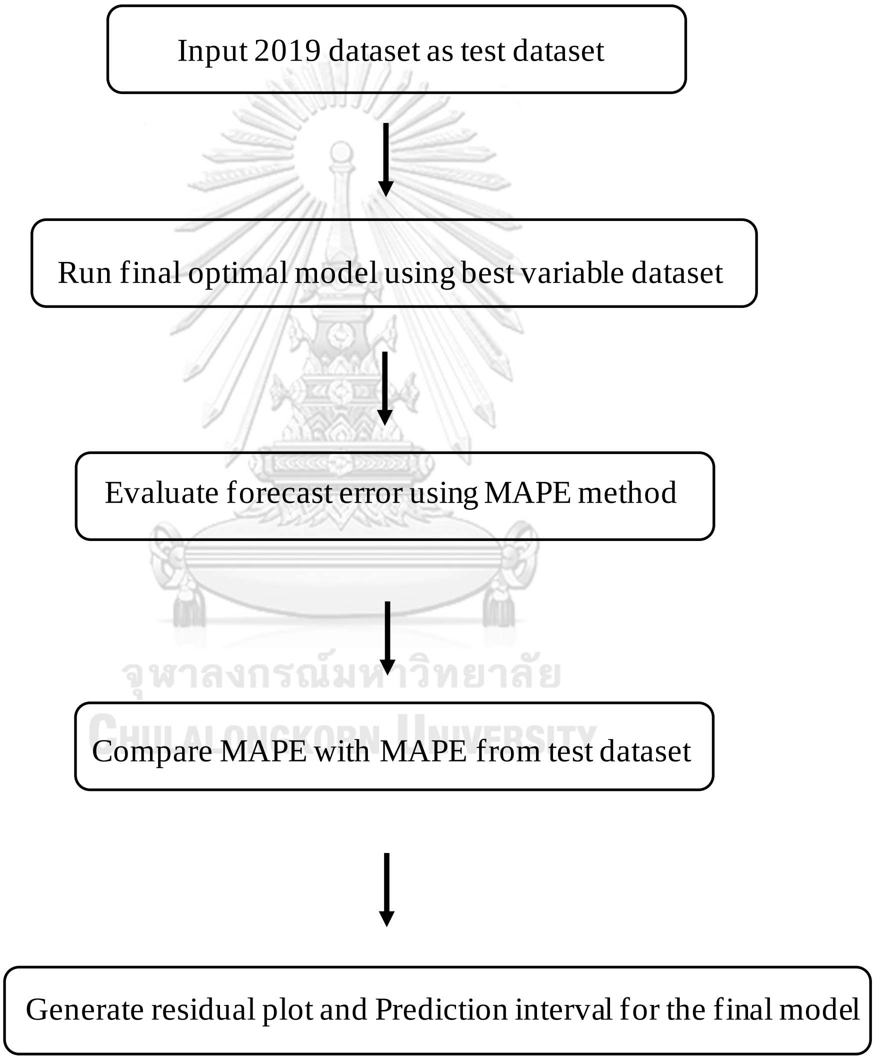
(iv) Calculate MAPE from comparison between the predicted values and actual validate data

(v) Calculate MAPE from comparison between the predicted values and actual test data

After that, the residual plots were generated for the final optimal model which provided the lowest MAPE among all models for both short-term and long-term prediction.

3.5 Model Evaluation

After obtaining the final optimal model for crew boat sailing time prediction, the model was evaluated with 2019 test dataset. Objective of re-evaluation with new dataset is to cross-check that prediction model was not overfitted and simulate real-time daily input data for prediction. Flowchart below summarizes the steps for model evaluation's process.



The optimal tuning parameter and the selected prediction model obtained from section 3.4.2 were used in this step by substituting all input variables for the new dataset in. For short-term prediction, the used data were daily data for travelling time &

distance, 4-hr average (6-10 A.M.) for boat speed, 4-hr average (before) for wave data and 5 A.M. time for wind data, On the other hand, for long-term prediction, daily data for travelling time & distance, 4-hr average (6-10 A.M.) for boat speed, and quarter data for wave and wind data. Then, the predicted values for travelling time were determined and compared with the actual values in 2019 to calculate MAPE for the model evaluation. The selected prediction model was run based at 2 different timeframe, short-term and long-term. The same aforementioned methodology was applied for both time period.

CHAPTER IV
RESULT AND DISCUSSION

4.1 Model Selection Result

1st order and 1st order with interaction terms datasets were applied using various regression models for further model development by comparing MAPE to find the best regression model for short-term and long-term prediction.

4.1.1 Multiple Regression Model

Multiple regression model was used for model built-up with 6 variables (1st order) and 21 variables (1st order with Interaction) fitted with 80% of 2017-2018 data. There were some models that found to have high VIFs (higher than 10 is considered as problem zone), indicating the existence of multicollinearity. Hence, the variables that have high VIF (>10) in the model must be removed then the model was re-run until VIFs of all variables were less than 10. For boat A with 1st order with interaction terms per example below, significant wave height was found to have the highest VIF at 114 so this term was removed from the model and the model was re-run. However, VIF was still high so other variables that were wind speed, wave direction, total distance, wind direction, speed*wave direction, speed*wind speed, speed*sig wave, speed*wind direction, distance*speed, and sig wave*wave direction were removed, respectively and re-ran the model every time that each variable was removed until VIFs of all variables were less than 10. Finally, the final model for multiple regression was obtained as shown below. Same approaches were applied for boat B and combined boat. See the results of final models for multiple regression in Appendix 2.1.1.

Multiple regression with all variable input analysis: Boat A, 1st order with Interaction (Before removing high VIF terms)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	21	59.8409	2.84956	170.32	0.000
Total Distance	1	1.1722	1.17225	70.06	0.000
Boat speed	1	0.0167	0.01670	1.00	0.318
Significant Wave Height	1	0.1359	0.13588	8.12	0.005
Wave Direction	1	0.0001	0.00006	0.00	0.952
Wind Speed (E1)	1	0.0101	0.01007	0.60	0.438
Wind Direction (E1)	1	0.0001	0.00015	0.01	0.925
Distance*Speed	1	0.0935	0.09354	5.59	0.019
Distance*Sig Wave	1	0.0096	0.00963	0.58	0.449
Distance*Wave Direction	1	0.1332	0.13322	7.96	0.005
Distance*Wind Speed	1	0.0139	0.01392	0.83	0.362
Distance*Wind direction	1	0.0026	0.00262	0.16	0.693
Speed*Sig Wave	1	0.3791	0.37913	22.66	0.000
Speed*Wave direction	1	0.0022	0.00223	0.13	0.715
Speed*Wind speed	1	0.0030	0.00297	0.18	0.674
Speed*Wind direction	1	0.0012	0.00124	0.07	0.786
Sig Wave*Wave direction	1	0.0053	0.00531	0.32	0.574
Sig Wave*Wind speed	1	0.0948	0.09483	5.67	0.018
Sig Wave*Wind direction	1	0.0023	0.00227	0.14	0.713
Wave direction*Wind speed	1	0.0001	0.00012	0.01	0.932
Wave direction*Wind direction	1	0.0014	0.00142	0.08	0.771

Wind speed*Wind direction	1	0.0008	0.00077	0.05	0.830
Error	356	5.9563	0.01673		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.129349	90.95%	90.41%	89.71%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.331	0.00665	193.81	0.000	
Total Distance	0.08075	0.0647	8.37	0.000	94.29
Boat speed	-0.0245	0.0502	-1.00	0.318	56.69
Significant Wave Height	0.583	0.0710	2.85	0.005	113.72
Wave Direction	0.00009	0.0696	0.06	0.952	109.13
Wind Speed (E1)	-0.0145	0.0767	-0.78	0.438	132.55
Wind Direction (E1)	-0.000092	0.0668	-0.09	0.925	100.66
Distance*Speed	-0.001149	0.0486	-2.36	0.019	53.27
Distance*Sig Wave	0.00279	0.0310	0.76	0.449	21.61
Distance*Wave Direction	-0.000066	0.0380	-2.82	0.005	32.61
Distance*Wind Speed	0.000297	0.0308	0.91	0.362	21.44
Distance*Wind direction	0.000006	0.0307	0.40	0.693	21.17
Speed*Sig Wave	-0.0552	0.0454	-4.76	0.000	46.43
Speed*Wave direction	0.000031	0.0598	0.37	0.715	80.71
Speed*Wind speed	0.00046	0.0526	0.42	0.674	62.40
Speed*Wind direction	-0.000017	0.0594	-0.27	0.786	79.38

Sig Wave*Wave direction	0.000277	0.0348	0.56	0.574	27.27
Sig Wave*Wind speed	0.01395	0.0280	2.38	0.018	17.68
Sig Wave*Wind direction	0.000128	0.0253	0.37	0.713	14.46
Wave direction*Wind speed	-0.000004	0.0374	-0.09	0.932	31.56
Wave direction*Wind direction	0.000001	0.0376	0.29	0.771	31.90
Wind speed*Wind direction	-0.000008	0.0252	-0.21	0.830	14.27

Regression Equation

Travelling Time = -0.331 + 0.08075 Total Distance - 0.0245 Boat speed
+ 0.583 Significant Wave Height + 0.00009 Wave Direction
- 0.0145 Wind Speed (E1) - 0.000092 Wind Direction (E1)
- 0.001149 Distance*Speed + 0.00279 Distance*Sig Wave
- 0.000066 Distance*Wave Direction + 0.000297 Distance*Wind Speed
+ 0.000006 Distance*Wind direction - 0.0552 Speed*Sig Wave
+ 0.000031 Speed*Wave direction + 0.00046 Speed*Wind speed
- 0.000017 Speed*Wind direction
+ 0.000277 Sig Wave*Wave direction
+ 0.01395 Sig Wave*Wind speed + 0.000128 Sig Wave*Wind direction
- 0.000004 Wave direction*Wind speed
+ 0.000001 Wave direction*Wind direction
- 0.000008 Wind speed*Wind direction

Multiple regression with all variable input analysis: Boat A, 1^{st} order with Interaction (After removing high VIF terms)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	10	55.8087	5.58087	205.06	0.000
Boat speed	1	0.7464	0.74636	27.42	0.000
Distance*Sig Wave	1	0.0209	0.02090	0.77	0.381
Distance*Wave Direction	1	2.4581	2.45811	90.32	0.000
Distance*Wind Speed	1	1.5802	1.58022	58.06	0.000
Distance*Wind direction	1	5.9043	5.90432	216.94	0.000

Sig Wave*Wind speed	1	0.1444	0.14444	5.31	0.022
Sig Wave*Wind direction	1	0.0000	0.00000	0.00	1.000
Wave direction*Wind speed	1	0.5555	0.55553	20.41	0.000
Wave direction*Wind direction	1	2.9969	2.99692	110.11	0.000
Wind speed*Wind direction	1	0.0749	0.07495	2.75	0.098
Error	367	9.9884	0.02722		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.164974	84.82%	84.41%	83.67%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.9465	0.00849	151.96	0.000	
Boat speed	-0.2429	0.00948	-5.24	0.000	1.24
Distance*Sig Wave	0.0272	0.0261	0.88	0.381	9.41
Distance*Wave Direction	0.00110	0.0189	9.50	0.000	4.96
Distance*Wind Speed	0.001959	0.0243	7.62	0.000	8.21
Distance*Wind direction	0.000122	0.0162	14.73	0.000	3.65
Sig Wave*Wind speed	0.01096	0.0227	2.30	0.022	7.16
Sig Wave*Wind direction	0.00002	0.0230	0.00	1.000	7.30
Wave direction*Wind speed	-0.000158	0.0271	-4.52	0.000	10.15
Wave direction*Wind direction	-0.00011	0.0177	-10.49	0.000	4.35
Wind speed*Wind direction	-0.00061	0.0244	-1.66	0.098	8.21

Regression Equation

Travelling Time = 0.9465 - 0.2429 Boat speed + 0.0272 Distance*Sig Wave
+ 0.00110 Distance*Wave Direction + 0.001959 Distance*Wind Speed
+ 0.000122 Distance*Wind direction + 0.01096 Sig Wave*Wind speed
+ 0.00002 Sig Wave*Wind direction
- 0.000158 Wave direction*Wind speed
- 0.00011 Wave direction*Wind direction
- 0.00061 Wind speed*Wind direction

For short-term prediction, VIFs of 1^{st} order data were low while they were high for 1^{st} order with interaction terms, indicating high correlation as shown in Appendix 2.1.1. The results of multiple regression were shown Table 11 comparing between 1^{st} order and 1^{st} order with interaction terms. Boat A had the lowest MAPE of 7.90% from 1^{st} order data, boat B of 10.13% from 1^{st} order data, and combined boat of 8.96% from 1^{st} order data as well.

On the other hand, for long-term prediction, VIFs were high for both 1^{st} order and 1^{st} order with interaction data. The lowest MAPEs obtained from 1^{st} order data for boat A was at 10.50%, boat B of 13.10%, and combined boat of 11.86%.

Moreover, R-sq(pred.) for all boats is higher than 87%, indicating high ability for prediction of models. The results indicated the trend that prediction ability decreases with longer time of prediction changing from short-term to long-term period.

Table 11 Multiple regression model results

Model	**Boat**	**Dataset**	**R-sq (pred.)**	**MAPE**
Short-term	**A**	1^{st} order	88.69%	7.90%
		1^{st} order with Interaction terms	83.67%	8.43%
	B	1^{st} order	87.20%	10.13%
		1^{st} order with Interaction terms	84.98%	11.48%
	Combined	1^{st} order	87.19%	8.96%
		1^{st} order with Interaction terms	84.37%	9.76%

Long-term	A	1st order	88.92%	10.50%
		1st order with Interaction terms	88.79%	10.69%
	B	1st order	82.97%	13.10%
		1st order with Interaction terms	82.51%	13.97%
	Combined	1st order	84.60%	11.86%
		1st order with Interaction terms	84.10%	12.12%

The equations for the best models which had the lower MAPE between 1st order and 1st order with interaction dataset for short-term and long-term prediction were summarized below.

<u>Best Multiple Regression Model</u>

<u>Short-term Prediction</u>

Boat A

Travelling Time (1st order) = 0.3777 + 0.4933 Total Distance - 0.2639 Boat speed + 0.3679 Significant Wave Height - 0.0412 Wave Direction + 0.0188 Wind Speed (E1) + 0.0013 Wind Direction (E1)

Boat B

Travelling Time (1st order) = -0.1106 + 0.5646 Total Distance - 0.2109 Boat Speed + 0.2977 Significant Wave Height – 0.0056 Wave Direction + 0.0148 Wind Speed (E1) + 0.0001 Wind Direction (E1)

Combined boat

Travelling Time (1st order) = 0.0827 + 0.5075 Total Distance - 0.3361 Boat speed + 0.2006 Significant Wave Height - 0.0192 Wave Direction + 0.0218 Wind Speed (E1) + 0.0016 Wind Direction (E1) - 0.1794 Boat Code

<u>Long-term Prediction</u>

Boat A

Travelling Time (1st order) = 1.0001 + 0.5923 Total Distance - 0.4044 Boat speed + 0.2185 Significant Wave Height - 0.003231 Wave Direction - 0.000181 Wind Direction (E1)

Boat B

Travelling Time (1st order) = 0.288 + 0.6145 Total Distance - 0.0746 Boat Speed + 0.169 Significant Wave Height - 0.01199 Wave Direction + 0.00646 Wind Speed (E1)

Combined boat

Travelling Time (1st order) = 0.6748 + 0.59600 Total Distance - 0.1098 Boat speed + 0.2591 Significant Wave Height - 0.002335 Wave Direction - 0.000446 Wind Direction (E1) - 0.1904 Boat Code

From the multiple regression equations, total distance, boat speed, and significant wave height had the effect on travelling time for all boats.

After that, Stepwise regression, Partial Least squares regression (PLS), Lasso regression and Elastic net regression were applied in the next step to compare performance with multiple regression.

4.1.2 Stepwise Regression Models

For short-term prediction, boat A had the lowest MAPE of 7.45% from 1st order with interaction data, boat B of 9.76% from 1st order data, and combined boat of 8.76% from 1st order data as shown in Table 12.

On the other hand, for long-term prediction, the lowest MAPE obtained for boat A was at 10.51% from 1st order data, boat B of 12.31% from 1st order data, and combined boat of 11.21% from 1st order with interaction terms.

In addition, Stepwise regression models provided lower MAPEs than multiple regression models for most of the boats. The results are in Appendix 2.2.1.

Table 12 Stepwise regression model results

Model	Boat	Dataset	R-sq (pred.)	MAPE
Short-term	**A**	1st order	88.76%	7.88%
		1st order with Interaction terms	90.30%	7.45%
	B	1st order	87.30%	9.76%
		1st order with Interaction terms	87.84%	11.46%
	Combined	1st order	87.22%	8.76%
		1st order with Interaction terms	87.66%	9.16%
Long-term	**A**	1st order	88.97%	10.51%
		1st order with Interaction terms	89.75%	10.53%
	B	1st order	82.99%	12.31%
		1st order with Interaction terms	83.15%	12.56%
	Combined	1st order	84.60%	11.49%
		1st order with Interaction terms	84.82%	11.21%

<u>Best Stepwise Regression Model</u>

<u>Short-term Prediction</u>

Boat A

Travelling Time	=	0.614 + 0.5506 Total Distance - 0.0331 Boat speed
(1^{st} order with		+ 0.2905 Significant Wave Height
Interaction)		+ 0.000831 Wave Direction
		- 0.00347 Wind Speed (E1)
		- 0.001455 Total Distance*Boat speed
		- 0.000072 Total Distance*Wave Direction
		- 0.04999 Boat speed*Significant Wave Height
		+ 0.01165 Significant Wave Height*Wind Speed (E1)

Boat B

Travelling Time	=	0.1195 + 0.5036 Total Distance - 0.0097 Boat speed
(1^{st} order)		+ 0.3073 Significant Wave Height

Combined boat

Travelling Time	=	0.0877 + 0.4079 Total Distance - 0.00471 Boat speed
(1^{st} order)		+ 0.2321 Significant Wave Height
		- 0.000365 Wave Direction
		+ 0.00526 Wind Speed (E1) - 0.1296 Boat Code

<u>Long-term Prediction</u>

Boat A

Travelling Time	=	0.4714 + 0.4927 Total Distance - 0.04040 Boat speed
(1^{st} order)		+ 0.2297 Significant Wave Height - 0.003300 Wave Direction

Boat B

Travelling Time	=	0.1937 + 0.6396 Total Distance -0.00714 Boat Speed
(1^{st} order)		

Combined boat

Travelling Time	=	0.549 + 0.4985 Total Distance - 0.0283 Boat speed
(1^{st} order with		+ 0.1645 Significant Wave Height
Interaction)		- 0.000366 Wind Direction (E1)
		- 0.00297 Wave Direction
		- 0.000093 Total Distance*Wave Direction
		- 0.0255 Boat speed*Significant Wave Height
		+ 0.000174 Boat speed*Wave Direction - 0.1912 Boat Code

4.1.3 Partial Least Squares Regression Model (PLS)

PLS model was built-up to compare forecasting performance with multiple regression model (ordinary least square method). The results for R-sq(pred.) and MAPE were shown in Table 13.

For short-term prediction, boat A had the lowest MAPE of 6.75% from 1^{st} order data, boat B of 9.01% from 1^{st} order data, and combined boat of 8.26% from 1^{st} order data. On the other hand, for long-term prediction, the lowest MAPE obtained for boat A was at 9.43% from 1^{st} order with interaction data, boat B of 11.86% from 1^{st} order data, and combined boat of 9.96% from 1^{st} order with interaction terms.

According to performance, PLS method provided lower MAPE than multiple regression and stepwise regression models for all datasets with slightly difference in R-sq(pred.). However, PLS regression model included all input predictors in the final model for travelling time prediction so the equations are lengthy for the optimal 1^{st} order with interaction models.

Table 13 Partial Least Squares Regression Result

Model	Boat	Dataset	R-sq (pred.)	MAPE
Short-term	**A**	1^{st} order	88.69%	6.75%
		1^{st} order + Interaction	89.49%	6.90%
	B	1^{st} order	87.20%	9.01%
		1^{st} order + Interaction	87.40%	9.53%
	Combined	1^{st} order	87.19%	8.26%
		1^{st} order + Interaction	87.59%	8.44%
Long-term	**A**	1^{st} order	88.87%	9.74%
		1^{st} order + Interaction	88.60%	9.43%
	B	1^{st} order	83.04%	11.86%
		1^{st} order + Interaction	82.95%	12.77%
	Combined	1^{st} order	84.61%	10.21%
		1^{st} order + Interaction	84.78%	9.96%

<u>Best Partial Least Squares Regression Model</u>

<u>Short-term Prediction</u>

Boat A

Travelling Time (1^{st} order) = 1.28942 + 0.39332 Total Distance - 0.05232 Boat speed + 0.05843 Significant Wave Height - 0.03757 Wave Direction + 0.02284 Wind Speed (E1) + 0.00388 Wind Direction (E1)

Boat B

Travelling Time (1^{st} order) = 1.34243 + 0.42458 Total Distance - 0.01088 Boat speed + 0.09769 Significant Wave Height - 0.00556 Wave Direction + 0.01481 Wind Speed (E1) + 0.00012 Wind Direction (E1)

Combined Boat

Travelling Time (1^{st} order)	=	1.37990 + 0.40749 Total Distance - 0.02608 Boat speed + 0.08053 Significant Wave Height - 0.01720 Wave Direction + 0.02183 Wind Speed (E1) + 0.00153 Wind Direction (E1) - 0.12935 Boat Code

Long-term Prediction

Boat A

Travelling Time (1^{st} order with Interaction)	=	1.28942 + 0.39746 Total Distance - 0.08381 Boat speed + 0.01234 Significant Wave Height - 0.03762 Wave Direction + 0.00378 Wind Speed (E1) + 0.00296 Wind Direction (E1) + 0.00277 Distance*Speed + 0.00042 Distance*Sig Wave - 0.01355 Distance*Wave Direction + 0.00663Distance*Wind Speed - 0.00118 Distance*Wind direction + 0.00656Speed*Sig Wave - 0.01378 Speed*Wave direction + 0.01110 Speed*Wind speed - 0.00349 Speed*Wind direction - 0.01308 Sig Wave*Wave direction + 0.00212 Sig Wave*Wind speed + 0.00412 Sig Wave*Wind direction - 0.01503 Wave direction*Wind speed - 0.00815 Wave direction*Wind direction + 0.00587 Wind speed*Wind direction

Boat B

Travelling Time (1^{st} order)	=	1.34243 + 0.42946 Total Distance - 0.04471 Boat speed + 0.04161 Significant Wave Height - 0.03251 Wave Direction - 0.10512 Wind Speed (E1) - 0.12104 Wind Direction (E1)

Combined Boat

Travelling Time (1^{st} order with Interaction)	=	1.39740 + 0.26035 Total Distance - 0.01421 Boat speed + 0.04167 Significant Wave Height - 0.04767 Wave Direction - 0.05841 Wind Speed (E1) - 0.00901 Wind Direction (E1) +

0.04275 Distance*Speed + 0.00882 Distance*Sig Wave - 0.11833 Distance*Wave Direction + 0.09960 Distance*Wind Speed + 0.17899 Distance*Wind direction - 0.07941 Speed*Sig Wave + 0.17299 Speed*Wave direction - 0.05527 Speed*Wind speed - 0.14345 Speed*Wind direction + 0.06636 Sig Wave*Wave direction - 0.00249 Sig Wave*Wind speed - 0.00725 Sig Wave*Wind direction - 0.02705 Wave direction*Wind speed - 0.09788 Wave direction*Wind direction - 0.02440 Wind speed*Wind direction - 0.16499 Boat Code

4.1.4 Lasso Model

The optimal tuning parameter for both datasets was determined from Lasso model with cross-validation technique by varying tuning parameter, λ, from 0 to 1 with the same inputs to see the effect of the parameter. MAPE results from different dataset were summarized in Table 26 for short-term prediction and Table 27 for long-term prediction in Appendix 3. As illustrated in Figures 22-24, MAPEs continuously decreased to certain point then turned to increase as the tuning parameter increased. The optimal tuning parameter results shown in Table 14 varied for each boat.

For short-term prediction, boat A with the optimal tuning parameter of 0.12 had the lowest MAPE of 5.64% from 1st order with interaction data, boat B with λ of 0.28 provided the lowest MAPE at 7.54% from 1st order data, and combined boat of 6.64% with λ of 0.10 from 1st order with interaction terms.

On the other hand, for long-term prediction, the lowest MAPE obtained for boat A was at 8.55% with λ of 0.20 from 1st order with interaction data, boat B of 10.23% with λ of 0.16 from 1st order with interaction data, and combined boat of 9.85% with λ of 0.12 from 1st order data.

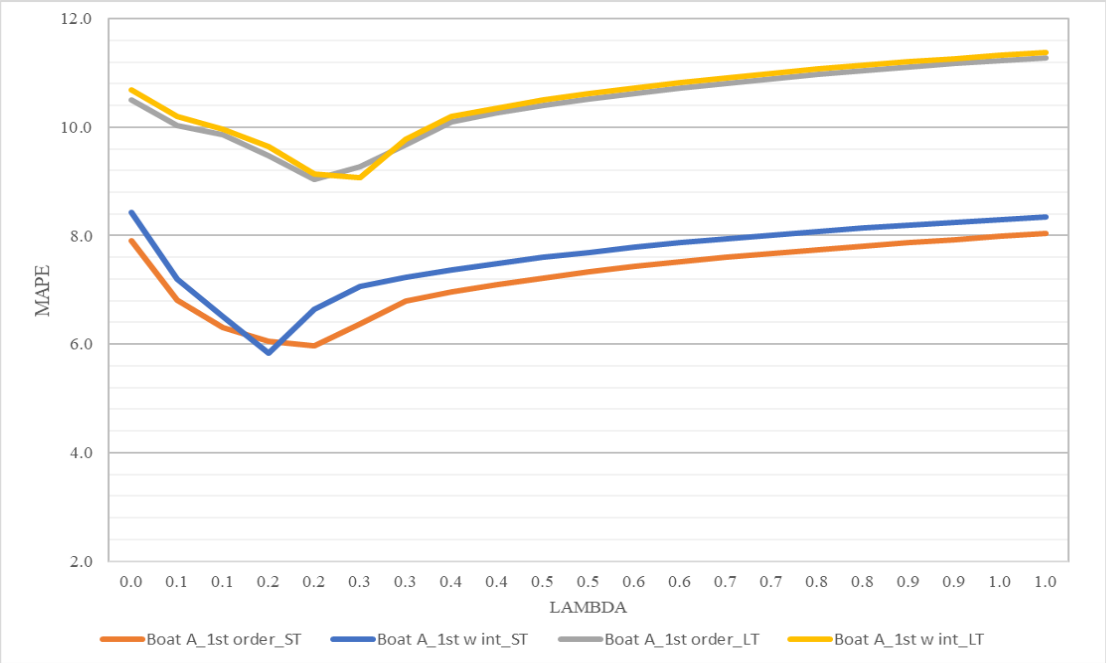


Figure 22 Lasso MAPE with tuning parameter variation for boat A

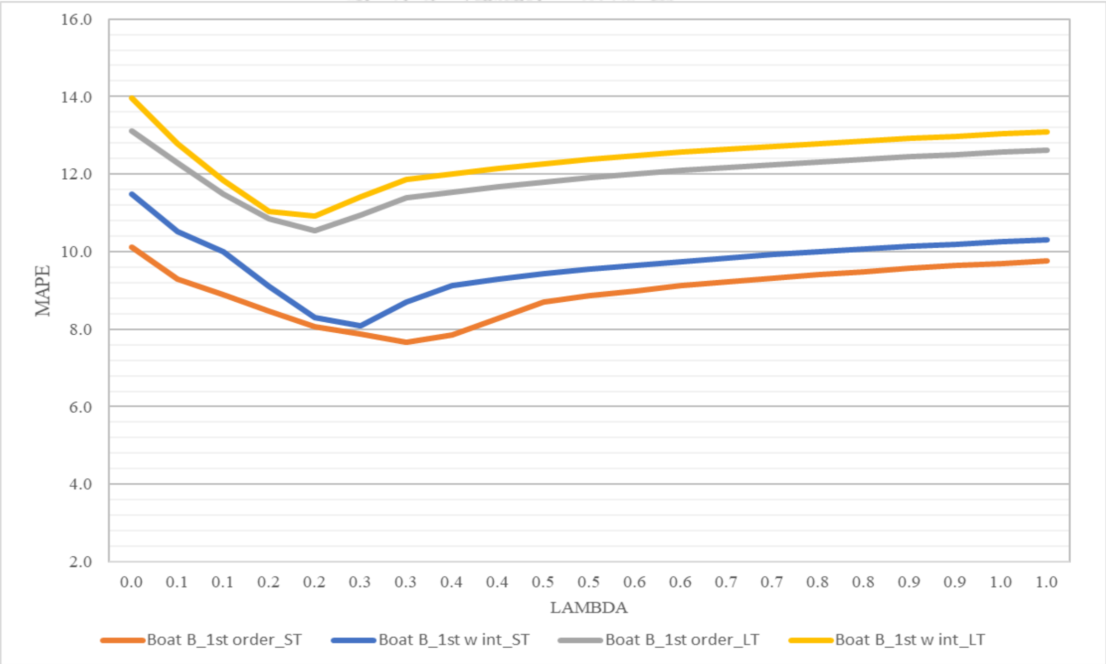


Figure 23 Lasso MAPE with tuning parameter variation for boat B

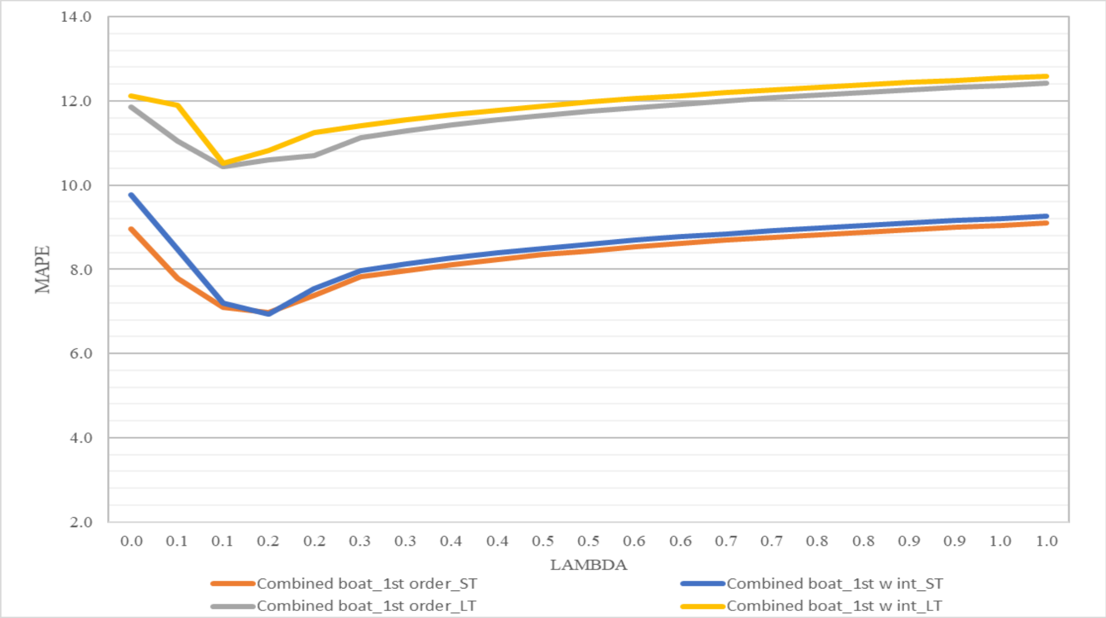


Figure 24 Lasso MAPE with tuning parameter variation for combined boat

It was found that MAPEs from Lasso model are lower than all previous models, showing significant improvement in forecasting for both short-term and long-term periods.

The regression equations of Lasso models were shown below. Main predictors that included in all equations are total distance and boat speed with significant wave height contained in most equations. Interaction between boat speed & significant wave height, boat speed & wind direction were observed when compared models for each boat for both prediction periods.

Table 14 Summary Result from Optimal Dataset and Tuning Parameter Analysis

Model	**Boat**	**Dataset**	**R-sq (pred.)**	**Lasso minimum MAPE**	**Best tuning parameter, λ**
Short-term	**A**	1st order	89.11%	5.76%	0.19
		1st order + Interaction	90.69%	5.64%	0.12
	B	1st order	87.57%	7.54%	0.28
		1st order + Interaction	88.55%	7.69%	0.25
	Combined	1st order	87.43%	6.66%	0.14
		1st order + Interaction	88.04%	6.64%	0.10
Long-term	**A**	1st order	89.29%	8.62%	0.24
		1st order + Interaction	89.33%	8.55%	0.20
	B	1st order	82.95%	10.3%	0.20
		1st order + Interaction	83.19%	10.23%	0.16
	Combined	1st order	84.87%	9.85%	0.12
		1st order + Interaction	85.38%	9.88%	0.10

<u>Best Lasso Model</u>

<u>Short-term Prediction</u>

Boat A

Travelling Time (1st order with Interaction) = 1.2894 + 0.4306 Total Distance -0.0234 Boat speed + 0.1111 Significant Wave Height + 0.0218 Wind Speed (E1) -0.02155 Distance*Speed

+0.0418 Distance*Sig Wave -0.0612 Distance*Wave Direction +0.0184 Distance*Wind Speed
-0.0142 Speed*Sig Wave +0.0128 Sig Wave*Wave direction +0.0661 Sig Wave*Wind speed
+0.011 Sig Wave*Wind direction -0.0109 Wind speed*Wind direction

Boat B

Travelling Time (1st order)	=	1.3424 + 0.4207 Total Distance -0.1469 Boat speed + 0.0965 Significant Wave Height -0.0009 Wave Direction +0.0107 Wind Speed (E1)

Combined boat

Travelling Time (1st order with Interaction)	=	1.3726 + 0.3885 Total Distance -0.0289 Boat speed -0.0161 Wave Direction +0.0233 Distance*Sig Wave +0.015 Distance*Wind Speed +0.003 Distance*Wind direction -0.0255 Speed*Sig Wave -0.0102 Speed*Wind direction +0.0117 Sig Wave*Wave direction +0.0712 Sig Wave*Wind speed +0.0448 Sig Wave*Wind direction -0.0345 Wind speed*Wind direction -0.1146 Boat code

<u>Long-term Prediction</u>

Boat A

Travelling Time (1st order with Interaction)	=	1.2894 + 0.3946 Total Distance -0.0791 Boat speed + 0.0195 Significant Wave Height -0.0695 Wave Direction -0.0023 Distance*Wave Direction +0.0005 Speed*Wind speed -0.0094 Speed*Wind direction

Boat B

Travelling Time (1st order with Interaction)	=	1.3424 + 0.3956 Total Distance -0.0018 Boat speed +0.0181 Distance*Sig Wave -0.02 Speed*Wave Direction +0.0112 Speed*Wind direction

Combined boat

Travelling Time (1st order)	=	1.4043 + 0.4068 Total Distance -0.0566 Boat speed + 0.0253 Significant Wave Height -0.0458 Wave Direction - 0.0141 Wind Direction (E1) -0.1791 Boat code

4.1.5 Lasso Variable Selection with Multiple Regression Model

This method for Lasso model was implemented to identify significant variables. 1st order dataset was selected to build-up Lasso model. The same range of tuning parameter from 0-1 which is the occupied range of the optimal tuning parameter was selected for this purpose. The results of Lasso model coefficient summarized in Appendix 2, Tables 28-30 for short-term prediction and Tables 31-33 for long-term prediction.

From Figures 25-27 for short-term and Figures 28-30 for long-term, the coefficients depicted the same trend to slightly decreased with higher tuning parameter until some of them converged to zero in the particular point. Those variables that the value converged to zero considered to be insignificant according to addition of high penalty from L1 regularization to those variables. Thus, they were removed from the model as they did not add additional information to the model output. In addition, the same characteristic was observed for both short-term and long-term prediction. The final variables selected for each boat and prediction period were shown in Table 15. Total distance, boat speed, and significant wave height were important variables for short-term model while only total distance and boat speed considered to include for long-term prediction.

Then, the selected variables from Table 15 were input parameters to run Multiple regression model. From Table 16, the results shown that MAPEs for Lasso variable selection are higher than original Lasso model for all cases. For short-term prediction, boat A's MAPE increased from 5.76% to 8.41%, boat B from 7.54% to 9.83%, and combined boat from 6.66% to 8.75%. For long-term prediction, MAPE for boat A was at 10.41%, boat B at 12.31%, and combined boat increased to 11.26%. According to no improvement gained from this method, trial was done for only 1st order dataset and sought for other regression method that potentially decreased MAPE instead.

<u>Short-term Prediction</u>

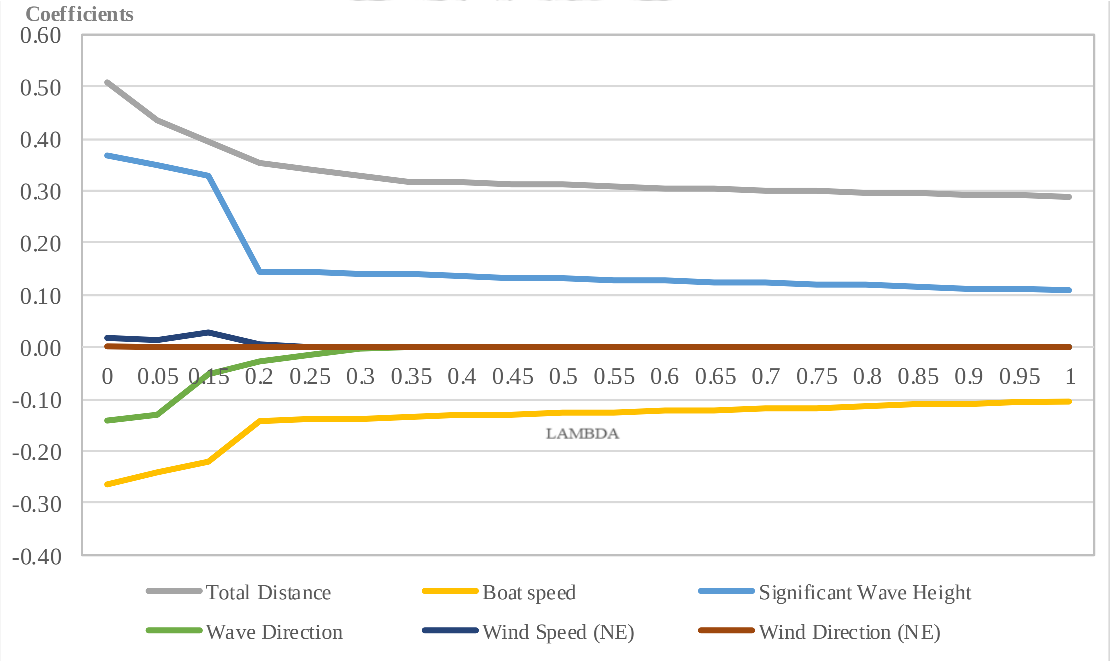


Figure 25 Lasso Significant Variable Selection for Boat A

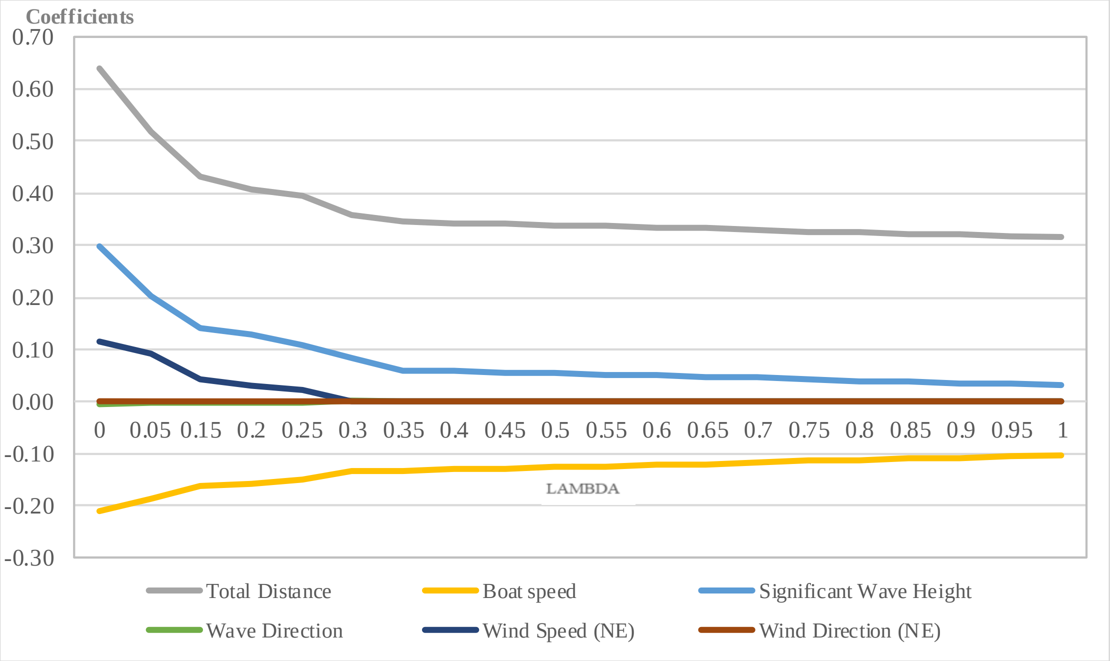


Figure 26 Lasso Significant Variable Selection for Boat B

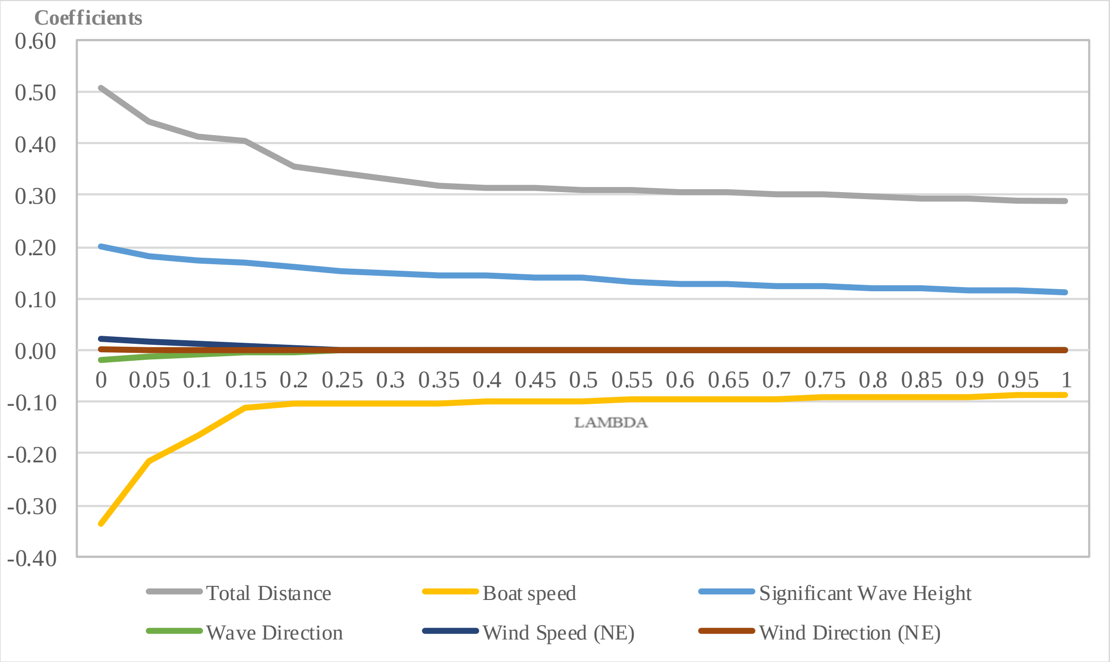


Figure 27 Lasso Significant Variable Selection for Combined boat

Long-term Prediction

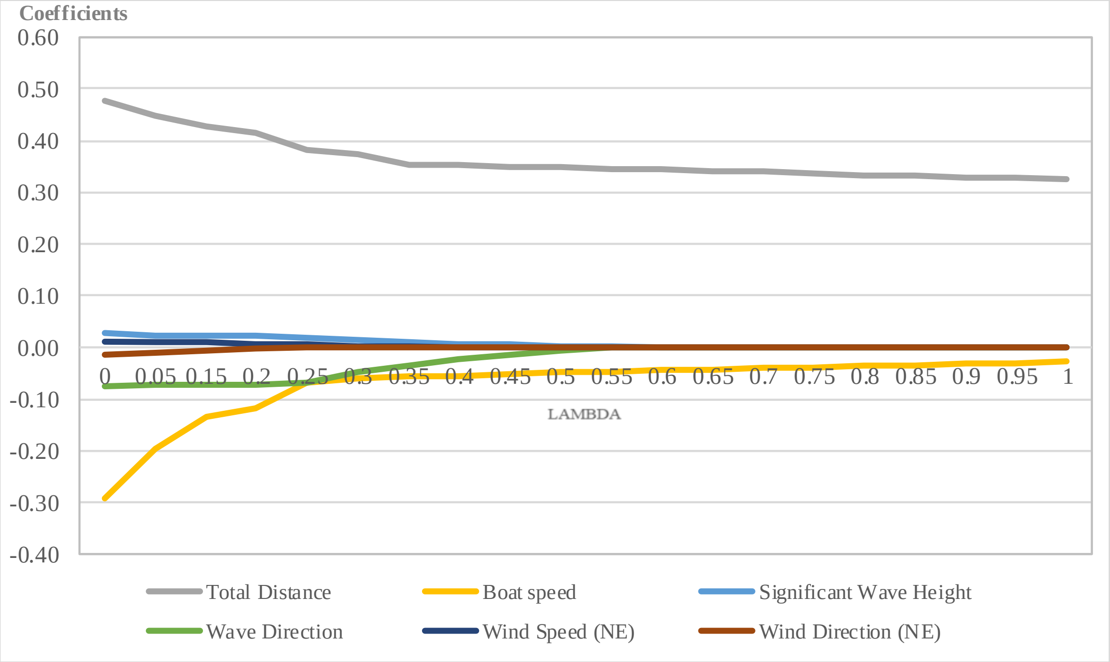


Figure 28 Lasso Significant Variable Selection for Boat A

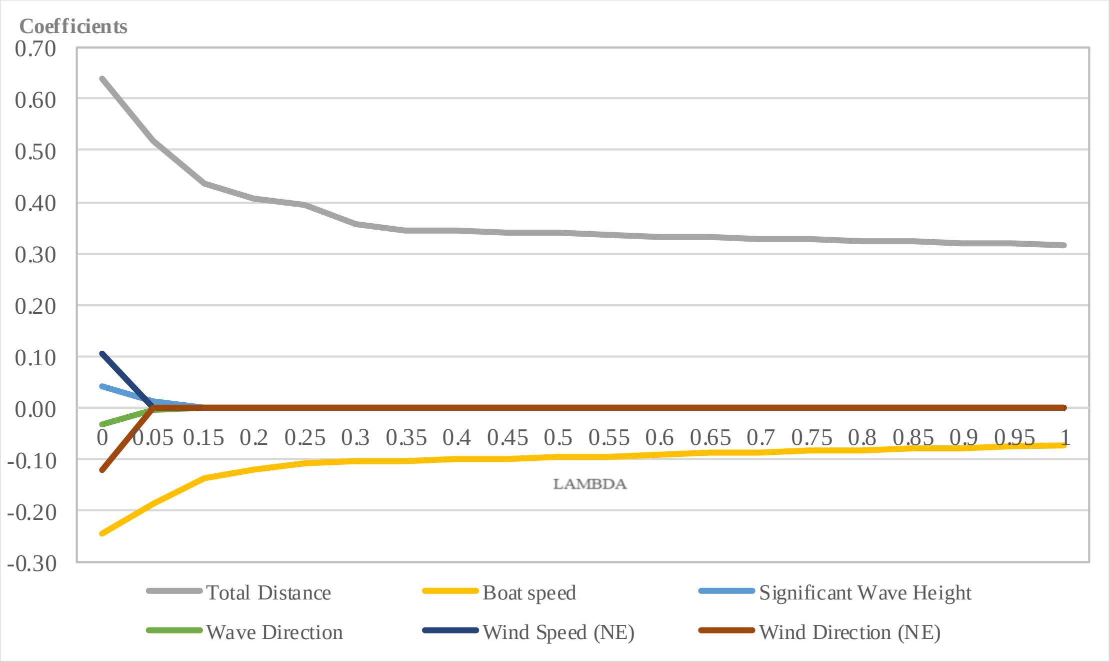


Figure 29 Lasso Significant Variable Selection for Boat B

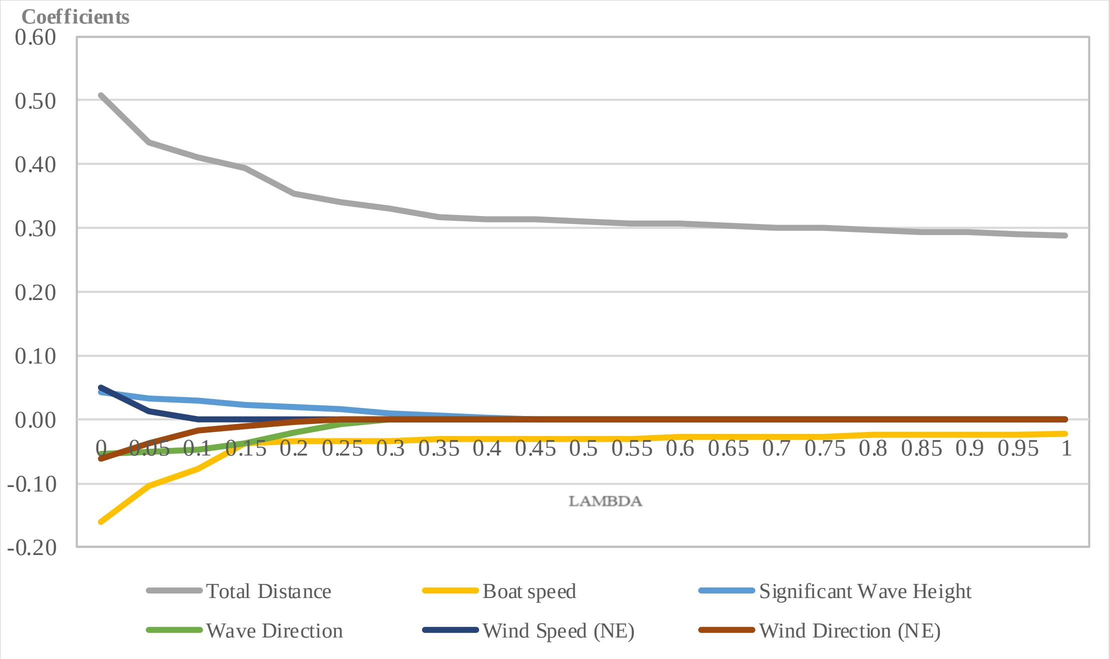


Figure 30 Lasso Significant Variable Selection for Combined boat

Table 15 Variables Selection Results (1st order dataset)

Model	Boat A	Boat B	Combined boat
Short-term	Total Distance Boat speed Significant Wave Height	Total Distance Boat speed Significant Wave Height	Total Distance Boat speed Significant Wave Height
Long-term	Total Distance Boat speed	Total Distance Boat speed	Total Distance Boat speed

Table 16 Summary Result from Lasso Variables Selection Models (1st order dataset)

Model	Boat	R-sq (pred.)	MAPE
Short-term	**A**	87.65%	8.41%
	B	87.26%	9.83%
	Combined	86.92%	8.75%
Long-term	**A**	85.84%	10.41%
	B	82.99%	12.31%
	Combined	83.41%	11.26%

The regression equations of the variable selection models were shown below.

Regression for Variable Selection Model (1st order)

Short-term Prediction

Boat A

Travelling Time = 0.2632 + 0.4586 Total Distance - 0.3376 Boat speed + 0.4493 Significant Wave Height

Boat B

Travelling Time = -0.0855 + 0.4709 Total Distance - 0.3144 Boat Speed + 0.3996 Significant Wave Height

Combined boat

Travelling Time = 0.0349 + 0.4947 Total Distance - 0.2463 Boat speed + 0.2536 Significant Wave Height - 0.1519 Boat Code

Long-term Prediction

Boat A

Travelling Time = 0.4946 + 0.4059 Total Distance -0.2376 Boat speed

Boat B

Travelling Time = 0.1937 + 0.5265 Total Distance -0.1071 Boat Speed

Combined boat

Travelling Time = 0.3059 -0.1735 Boat Code + 0.5938 Total Distance -0.0106 Boat speed

4.1.6 Elastic Net Regression Model

The optimal tuning parameter for both datasets was further analyzed with Elastic net model with 1st order and 1st order with Interaction variables with cross-validation technique with fixed L1 ratio at 0.5 and varying λ_1 & λ_2 to find the optimal tuning parameter that provided the lowest MAPE. Table 17 summarized the results for optimal dataset and tuning parameter for short-term and long-term prediction. The forecasting error slightly decreased from Lasso model for all cases. For short-term prediction, boat A had the lowest MAPE of 5.55%, boat B of 7.43%, and combined boat of 6.48% obtained from 1st order data. Moreover, MAPEs for short-term were lower than for long-term prediction, indicating similar results from previous regression models. All predictors, except wind direction were selected in the models.

On the other hand, for long-term prediction, the lowest MAPEs obtained for boat A at 8.51%, boat B of 10.09%, and combined boat of 9.34% from 1st order with interaction data. The optimal models of all boats had total distance, boat speed, and wave direction with the interaction term between boat speed and wind direction as the main variables for prediction.

Among various regression models, elastic net regression models provided the lowest MAPEs, following with Lasso regression and PLS for both short-term and long-term prediction models for boat A, boat B, and combined boat. The dominant characteristic of elastic net regression that includes Lasso and Ridge regression helps to improve prediction performance over other regression models. Elastic net methods overcome the limitations of the LASSO that tends to select one variable from a group and ignore the others. R-sq(pred.) of all models are higher than 83%, indicating high ability for prediction.

Table 17 Summary Result from Optimal Dataset and Tuning Parameter Analysis

Model	Boat	Dataset	R-sq (pred.)	Minimum MAPE	Best tuning parameter, λ_1	Best tuning parameter, λ_2
Short-term	**A**	1^{st} order	89.10%	5.55%	0.35	0.18
		1^{st} order with Interaction	90.69%	5.60%	0.16	0.08
	B	1^{st} order	87.57%	7.43%	0.40	0.20
		1^{st} order with Interaction	88.56%	7.47%	0.23	0.12
	Combined	1^{st} order	87.43%	6.48%	0.30	0.15
		1^{st} order with Interaction	88.05%	6.57%	0.25	0.12
Long-term	**A**	1^{st} order	89.29%	8.54%	0.47	0.23
		1^{st} order with Interaction	89.33%	8.51%	0.27	0.13
	B	1^{st} order	82.91%	10.17%	0.32	0.16
		1^{st} order with Interaction	83.45%	10.09%	0.21	0.11
	Combined	1^{st} order	84.86%	9.46%	0.23	0.11
		1^{st} order with Interaction	85.36%	9.34%	0.18	0.09

The regression equations of the Elastic net models were shown below.

<u>Best Elastic Net Model</u>

<u>Short-term Prediction</u>

Boat A

Travelling Time (1^{st} order)	=	1.2894 + 0.4522 Total Distance -0.2536 Boat speed + 0.3577 Significant Wave Height -0.1211 Wave Direction +0.1180 Wind Speed (E1) +0.0012 Wind Direction (E1)

Boat B

Travelling Time (1^{st} order)	=	1.3424 + 0.5193 Total Distance -0.1074 Boat speed + 0.1962 Significant Wave Height -0.0612 Wave Direction +0.0711 Wind Speed (E1)

Combined boat

Travelling Time (1^{st} order)	=	1.3741 + 0.5056 Total Distance -0.2215 Boat speed + 0.1811 Significant Wave Height -0.0216 Wave Direction +0.0198 Wind Speed (E1) -0.2175 Boat code

<u>Long-term Prediction</u>

Boat A

Travelling Time (1^{st} order with Interaction)	=	1.2894 + 0.3935 Total Distance -0.0788 Boat speed + 0.0719 Significant Wave Height -0.0892 Wave Direction -0.0022 Distance*Wave Direction +0.0007 Speed*Wind speed -0.0094 Speed*Wind direction

Boat B

Travelling Time (1^{st} order with Interaction)	=	1.3424 + 0.4446 Total Distance -0.0029 Boat speed -0.1397 Wave Direction +0.0102 Distance*Sig Wave -0.0365 Distance*Wind direction +0.0395 Speed*Wind direction

Combined boat

Travelling Time (1^{st} order with Interaction)	=	1.3990 + 0.4899 Total Distance -0.0822 Boat speed +0.0542 Significant Wave Height -0.0409 Wave Direction +0.0012 Wind Speed (E1) -0.0293 Distance*Wave Direction - 0.0466 Distance*Wind direction -0.0287 Speed*Sig Wave -0.0119 Speed*Wind direction -0.1683 Boat code

Furthermore, PLS, Lasso, and Elastic net methods prove to be good for prediction. In conclusion, the Elastic net regression is considered the best model.

After that, residual plots and prediction intervals for the best model were generated using validate data to verify validity of Elastic net regression models as shown in Figures 31-33 for short-term and Figures 34-36 for long-term prediction. The residual plots illustrated that there were only some peaks from random error and no significant peak in residual versus observation order graphs, indicating that residuals are independent. Most of residual values lined in the middle in histogram meaning that most of them fitted very well with the validate data. From AD test, it was found that boats A and B had p-value > 0.05, indicating normal distribution for residuals of short-term prediction. However, the errors for combined boat did not fit the normal distribution (p-value < 0.05). For long-term prediction, the residuals for all boats fitted the normal distribution. Furthermore, most of predicted values are in the middle line in plots of prediction interval, implying that the obtained models are suitable with the data.

The equations of the optimal models provided the information that align with the theoretical background [5]. Positive sign of distance's coefficient illustrates that the longer distance leads to longer travelling time and negative sign for boat speed means that the higher boat speed decreases total travelling time. Positive sign for wave height indicates that the higher wave height makes travelling time of boats to be longer and positive sign of wind speed means that the higher wind speed makes the travelling time to be longer. However, wind speed relates to wind direction whether it is the same direction as the boat (support sailing) or opposite direction (obstruct sailing). There was no clear indication for wave direction and wind direction since they changed over time.

Short-term Prediction

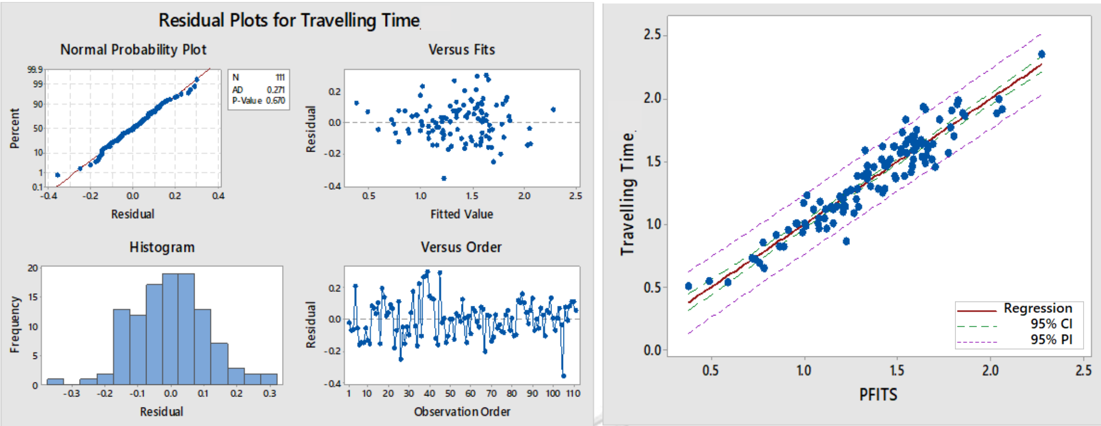


Figure 31 Residual Plot and Prediction Interval for Boat A

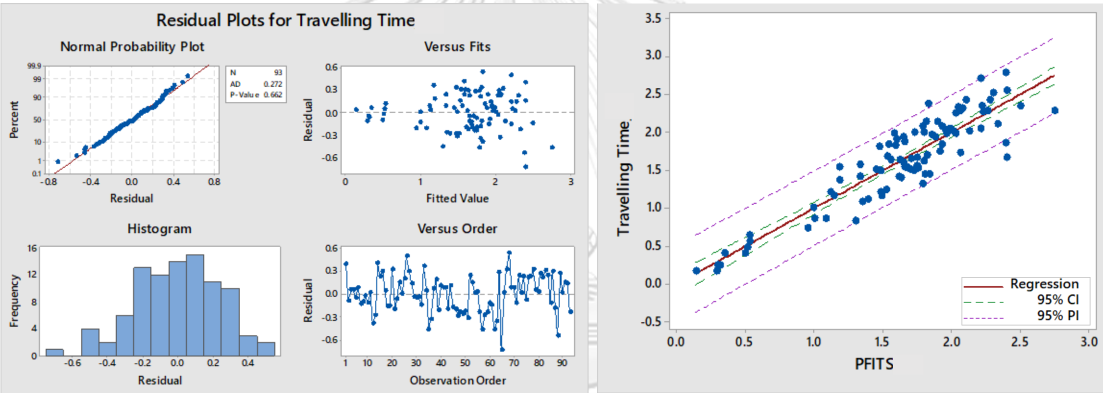


Figure 32 Residual Plot and Prediction Interval for Boat B

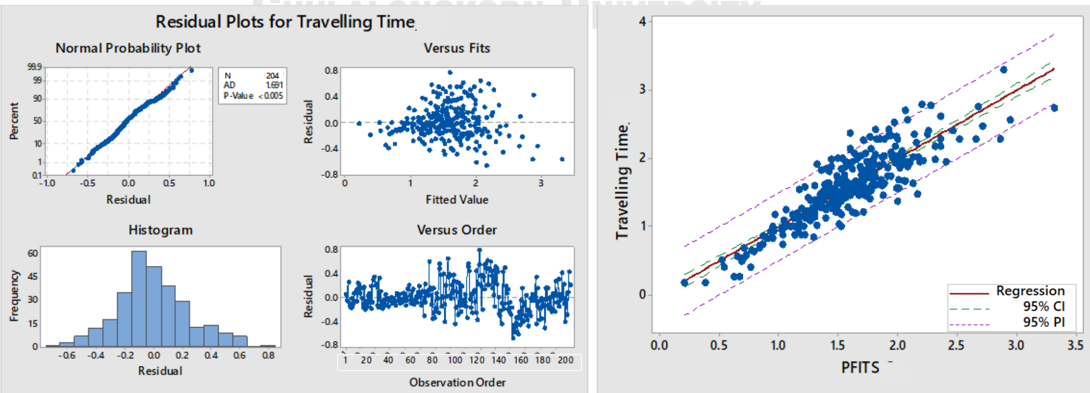


Figure 33 Residual Plot and Prediction Interval for Combined boat

Long-term Prediction

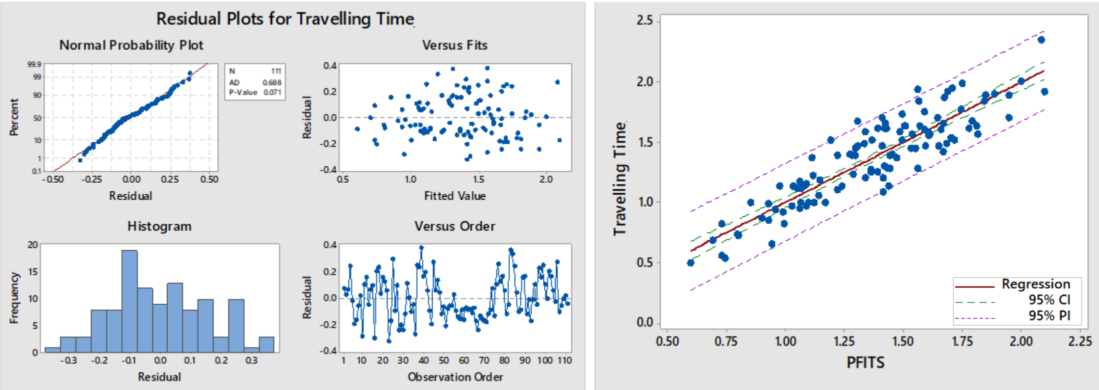


Figure 34 Residual Plot and Prediction Interval for Boat A

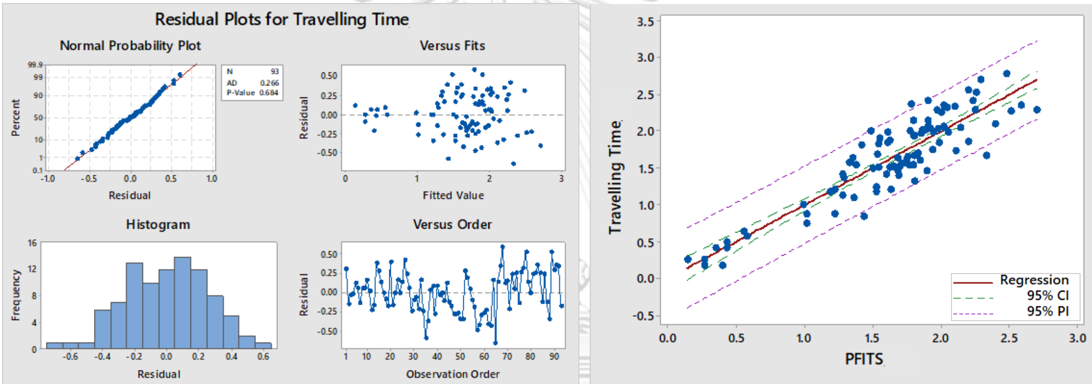


Figure 35 Residual Plot and Prediction Interval for Boat B

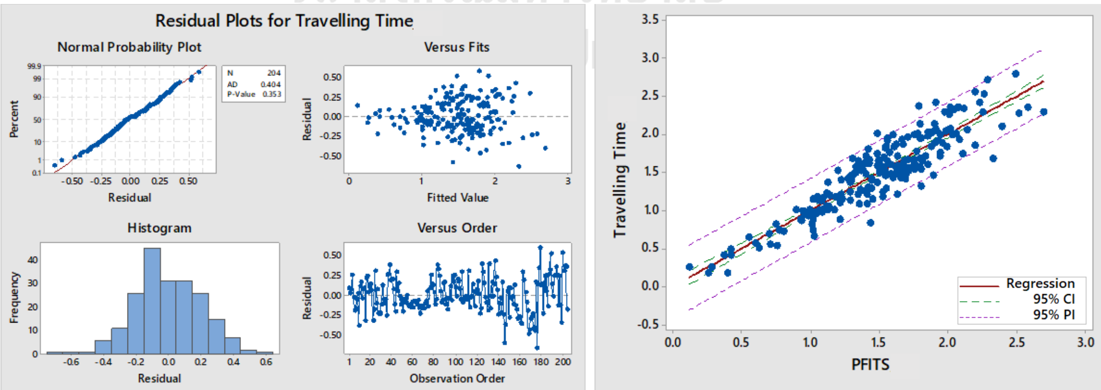


Figure 36 Residual Plot and Prediction Interval for Combined boat

4.2 Model Evaluation

For short-term prediction, the best model was Elastic net regression with 1st order dataset while Elastic net regression with 1st order with interaction terms was the optimal model for long-term prediction. The evaluation approach was analyzed using 2019 data with the input from the optimal model resulted by calculating MAPEs and comparing between the actual values of travelling time and the prediction values from the models for both short-term and long-term prediction.

The model evaluation results were shown in Table 18. There are small differences of MAPEs between validate data and this test data, indicating that obtained elastic net regression did not overfit. The residual plots as shown in Figures 37-42 provided the positive outlook for the models, same as ones using validate data that residuals are independent and most of residual values lined in the middle in histogram, meaning that most of them fitted very well with the test data. From AD test, it was found that boat A had p-value > 0.05, indicating normal distribution for residuals of short-term prediction. However, the errors for boat B and combined boat did not fit the normal distribution (p-value < 0.05). For long-term prediction, the residuals for all boats were not normally distributed. Moreover, the prediction intervals illustrated that the final optimal models are suitable with the data since most of predicted values are in the middle line in the plots. To conclude with the final point, Elastic net model is suitable for crew boat travelling time prediction with high accuracy.

Table 18 Best Regression Model's Evaluation for Short-term & Long-term Prediction

Prediction	**Boat**	**Regression Model**	**MAPE (Validate data)**	**MAPE (Test data)**
Short-term	**A**	Elastic net (1^{st} order)	5.55%	5.83%
	B	Elastic net (1^{st} order)	7.43%	7.95%
	Combined	Elastic net (1^{st} order)	6.48%	6.93%
Long-term	**A**	Elastic net (1^{st} order with Interaction)	8.51%	9.11%
	B	Elastic net (1^{st} order with Interaction)	10.09%	11.09%
	Combined	Elastic net (1^{st} order with Interaction)	9.34%	10.28%

Short-term Prediction

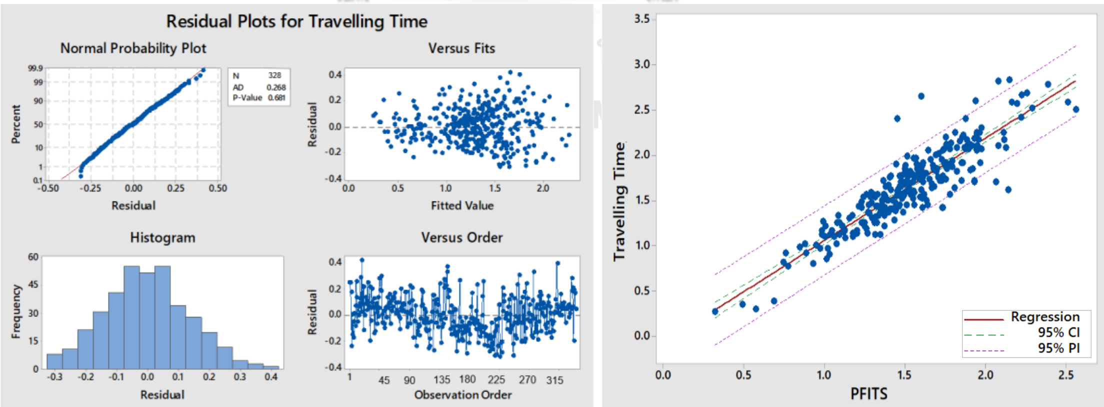


Figure 37 Residual Plot and Prediction Interval for Boat A

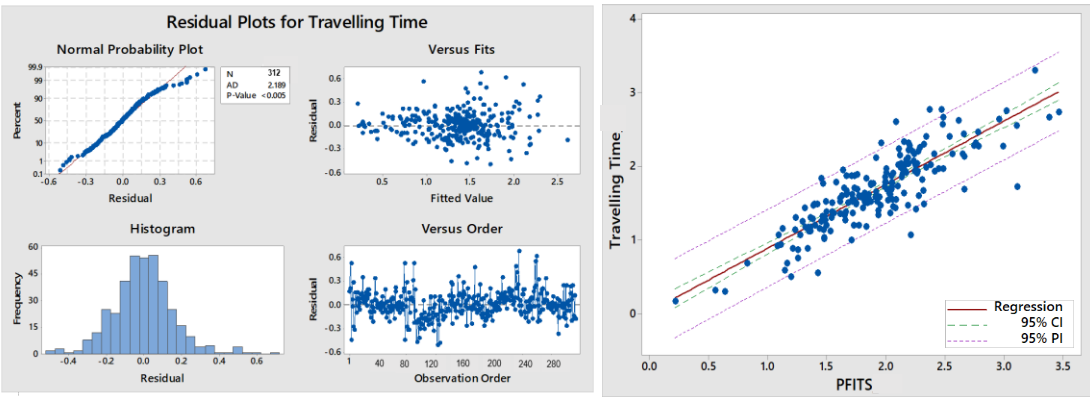


Figure 38 Residual Plot and Prediction Interval for Boat B

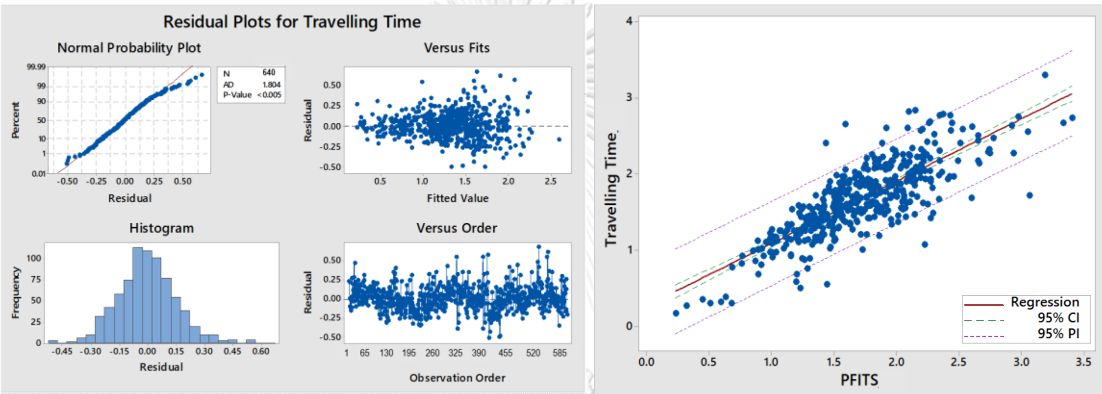


Figure 39 Residual Plot and Prediction Interval for Combined boat

Long-term Prediction

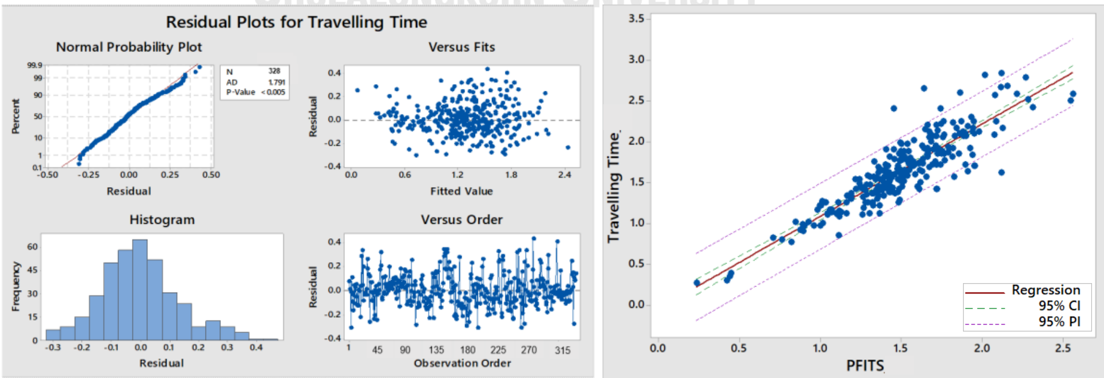


Figure 40 Residual Plot and Prediction Interval for Boat A

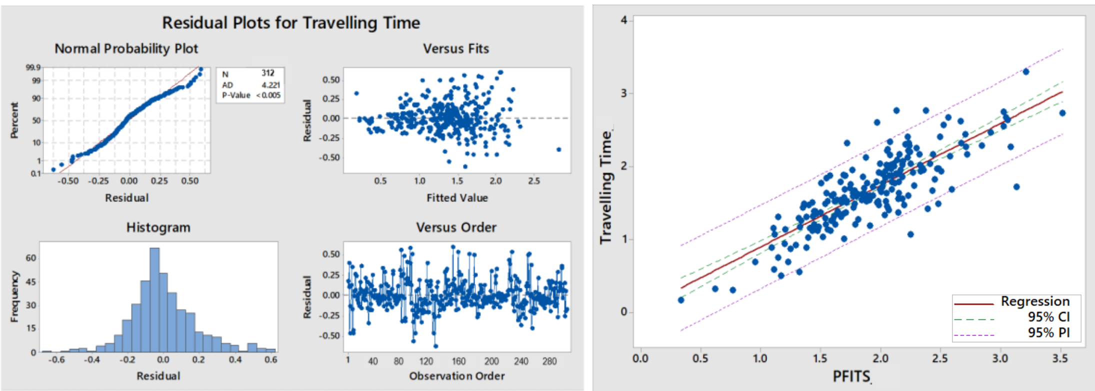


Figure 41 Residual Plot and Prediction Interval for Boat B

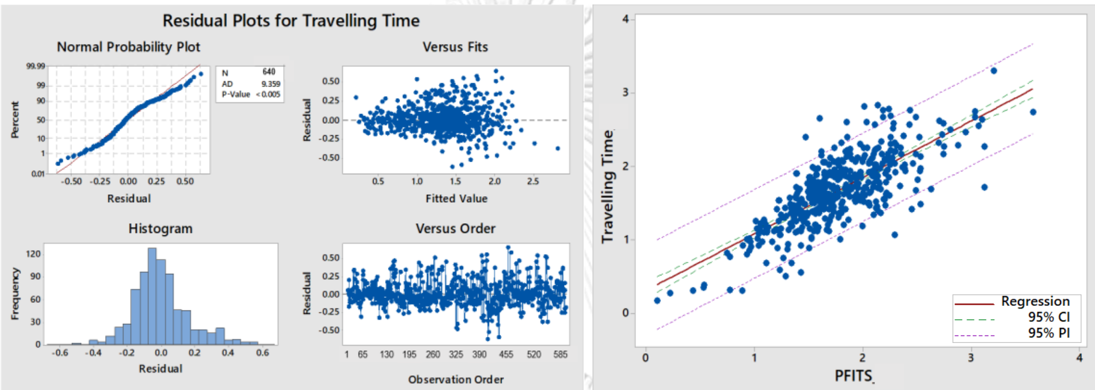


Figure 42 Residual Plot and Prediction Interval for Combined boat

CHAPTER V
CONCLUSION AND RECOMMENDATION

The objective of this research is to develop the prediction models for crew boat's sailing time for both short-term and long-term periods to further apply for crew boat operation in the future. Short-term model is used for operational level adjustment of boat speed to achieve travelling time at 1.22 hrs as per annual goal. Long-term model is for planning 1 month in advance for crew boats.

The morning route (6-10 A.M.) for crew boat operation was studied for travelling time prediction of boat A, boat B, and combined boat. The 2017-2019 data were collected as hourly basis for internal and external factors that are potentially affected crew boat sailing time from literature review and consulting SME. The outliers and influential data points using cook's distance analysis were eliminated since they might have significant effect on output before plugging in the models.

Data analysis was performed and the final variables to be included in the model were total distance & boat speed as internal factors and significant wave height, wave direction, wind speed, and wind direction as external factors to forecast travelling time. The relationship between selected 6 input parameters both 1st order and 1st order with Interaction dataset was considered for further analysis.

5.1 Short-term prediction

Short-term models aim to use for boat speed adjustment to meet the required travelling time of 1.22 hrs. The input data selection for this model built-up can be divided into two groups, group A and group B. Group A was selected based on the exact operating time of daily crew boats which was for travelling time, distance, and 4-hr average (6-10 A.M.) for boat speed. Moreover, Group B was for environmental data (significant wave height, wave direction, wind speed, and wind direction). These factors cannot be directly used in the model in the real application since there is no environmental data at that time (6-10 A.M.) when crew boats sail for prediction.

Hence, the optimal time frame for environmental data was identified based on the lowest MAPE. Data for 4-hr average (before sailing: 2-6 A.M.) were used for wave data while 5 A.M. data were applied for wind data. 1st order and 1st order with Interaction dataset was used for further analysis to compare MAPEs from various regression models.

The dataset which provided the lowest MAPE was selected as representative for each regression as summarized in Table 19. R-sq(pred.) for each boat is over 87%, implying high prediction ability for selected models. Elastic net regression models are the best short-term prediction models, providing the lowest MAPEs at 5.55%, 7.43%, and 6.48% for boat A, boat B, and combined boat, respectively.

Equations for the optimal models were summarized in Table 20. 1st order was the optimal dataset. The parameters that affect crew boat sailing time for all boats are total distance, boat speed, significant wave height, wave direction, and wind speed which aligned with the theoretical background. The longer distance, higher significant wave height, and higher wind speed lead to longer travelling time. However, wind speed is dependent to wind direction whether it is the same direction (support sailing) or opposite direction (obstruct sailing) for the boats. The higher boat speed results in decreasing sailing time. Short-term travelling time prediction's performance significantly increases with high accuracy and will help AA company to efficiently manage crew boats through operational adjustment, meet annual KPI and reduce operating cost.

Table 19 R-sq (pred.) and MAPE Results for Best Model Selected from Various Regression Models for Short-term Prediction

Boat Name	**Regression Model**	**Dataset**	**Rsq. (pred)**	**MAPE**
A	Multiple Regression	1^{st} order	88.69%	7.90%
	Stepwise Regression	1^{st} order with Interaction	90.30%	7.45%
	PLS	1^{st} order	88.69%	6.75%
	Lasso Regression	1^{st} order with Interaction	90.69%	5.64%
	Elastic net Regression	1^{st} order	89.10%	5.55%
B	Multiple Regression	1^{st} order	87.20%	10.13%
	Stepwise Regression	1^{st} order	87.30%	9.76%
	PLS	1^{st} order	87.20%	9.01%
	Lasso Regression	1^{st} order	87.57%	7.54%
	Elastic net Regression	1^{st} order	87.57%	7.43%
Combined	Multiple Regression	1^{st} order	87.19%	8.96%
	Stepwise Regression	1^{st} order	87.22%	8.76%
	PLS	1^{st} order	87.19%	8.26%
	Lasso Regression	1^{st} order with Interaction	88.04%	6.64%
	Elastic net Regression	1^{st} order	87.43%	6.48%

Optimal Model for Prediction (Elastic Net Model)

Table 20 Optimal Models for Short-term Prediction

Boat	Equation		
A	Travelling Time (1st order)	=	1.2894 + 0.4522 Total Distance -0.2536 Boat speed + 0.3577 Significant Wave Height -0.1211 Wave Direction +0.1180 Wind Speed (E1) +0.0012 Wind Direction (E1)
B	Travelling Time (1st order)	=	1.3424 + 0.5193 Total Distance -0.1074 Boat speed + 0.1962 Significant Wave Height -0.0612 Wave Direction +0.0711 Wind Speed (E1)
Combined	Travelling Time (1st order)	=	1.3741 + 0.5056 Total Distance -0.2215 Boat speed + 0.1811 Significant Wave Height -0.0216 Wave Direction +0.0198 Wind Speed (E1) -0.2175 Boat code

5.2 Long-term prediction

This type of prediction applied for future planning for crew boats. The same approach as short-term was employed to select proper time interval for input data. The exact operating time of daily crew boats was used for travelling time & distance and 4-hr average (6-10 A.M.) for boat speed. The best time interval for environmental data that provided the lowest MAPE was quarter data so they were selected to use for model built-up.

The prediction's performance in terms of R-sq(pred.) and MAPEs were summarized in Table 21. Elastic net regression with 1st order with Interaction dataset was the best regression models for all boats, showing the lowest MAPE among various models. Compared with short-term prediction model, MAPEs for long-term prediction were slightly higher according to longer period of time use for prediction and same environmental data as quarter was applied in the model built-up.

Table 21 R-sq (pred.) and MAPE Results for Best Model Selected from Various Regression Models for Long-term Prediction

Boat Name	Regression Model	Dataset	Rsq.(pred)	New MAPE
A	1st order (w/o Interaction)	1st order	88.92%	10.50%
	1st order (stepwise)	1st order	88.97%	10.51%
	PLS with Interaction	1st order with Interaction	88.60%	9.43%
	Lasso Regression (with Interaction)	1st order with Interaction	89.33%	8.55%
	Elastic net Regression (with Interaction)	1st order with Interaction	89.33%	8.51%
B	1st order	1st order	82.97%	13.10%
	1st order (stepwise)	1st order	82.99%	12.31%
	PLS	1st order	83.04%	11.86%
	Lasso Regression (with Interaction)	1st order with Interaction	83.19%	10.23%
	Elastic net Regression (with Interaction)	1st order with Interaction	83.45%	10.09%
Combined	1st order	1st order	84.60%	11.86%
	1st order + Interaction (Stepwise)	1st order with Interaction	84.82%	11.21%
	PLS with Interaction	1st order with Interaction	84.78%	9.96%
	Lasso Regression	1st order	84.87%	9.85%
	Elastic net Regression (with Interaction)	1st order with Interaction	85.36%	9.34%

The equations of the best models were shown in Table 22. The input parameters that have effect for sailing time for all boats were total distance, boat speed, and wave direction with interaction between boat speed and wind direction. Regarding to the optimal models, the longer total distance and higher significant wave height (+ sign in equations) increase travelling time while the higher boat speed decreases the total travelling time (-sign in equation).

After that, the best prediction models were re-evaluated with different dataset in 2019. Table 23 summarized the evaluation results between validate and test data. It was found that there is small change in MAPEs from using new evaluation data, indicating that models are not overfitted and suitable for future use. In conclusion, elastic net model is the best model for both short-term and long-term prediction for crew boat's travelling time. AA company can further utilize the obtained models for operational adjustment and future planning to increase efficiency in forecasting and optimize boat route to meet annual KPI and reduce overall operating cost.

Table 22 Optimal Models for Long-term Prediction

Boat	Equation		
A	Travelling Time (1^{st} order with Interaction)	=	1.2894 + 0.3935 Total Distance -0.0788 Boat speed + 0.0719 Significant Wave Height -0.0892 Wave Direction -0.0022 Distance*Wave Direction +0.0007 Speed*Wind speed -0.0094 Speed*Wind direction
B	Travelling Time (1^{st} order with Interaction)	=	1.3424 + 0.4446 Total Distance -0.0029 Boat speed -0.1397 Wave Direction +0.0102 Distance*Sig Wave -0.0365 Distance*Wind direction +0.0395 Speed*Wind direction
Combined	Travelling Time (1^{st} order with Interaction)	=	1.3990 + 0.4899 Total Distance -0.0822 Boat speed +0.0542 Significant Wave Height -0.0409 Wave Direction +0.0012 Wind Speed (E1) -0.0293 Distance*Wave Direction - 0.0466 Distance*Wind direction -0.0287 Speed*Sig Wave -0.0119 Speed*Wind direction -0.1683 Boat code

Table 23 Optimal Regression Model's Evaluation for Short-term & Long-term Prediction

Prediction	Boat	Regression Model	MAPE (Validate data)	MAPE (Test data)
Short-term	A	Elastic net (1st order)	5.55%	5.83%
	B	Elastic net (1st order)	7.43%	7.95%
	Combined	Elastic net (1st order)	6.48%	6.93%
Long-term	A	Elastic net (1st order with Interaction)	8.51%	9.11%
	B	Elastic net (1st order with Interaction)	10.09%	11.09%
	Combined	Elastic net (1st order with Interaction)	9.34%	10.28%

From previous study, the main factors that used for travelling time prediction were total distance, boat speed, wind speed, and wave height. However, it was found that the additional parameters that also had effect on sailing time were wind direction and wave direction from this study, resulting in improvement in accuracy for prediction. The company will gain additional travelling time 37 hrs. from boat A and 28 hrs. from boat B in 2019 for short-term prediction from applying the optimal models. On the other hand, additional travelling time from 34 hrs. for boat A and 26 hrs. from boat B for long-term prediction.

Although, elastic net regression is more complex model for prediction than multiple regression model, it is worthy to apply for model built-up since it proved to enhance better ability for prediction as shown from lower forecast error and higher accuracy. Moreover, this study can help planning team to get insight of operational parameters that can be adjusted and finally annual key performance indicator will be achieved from applying the optimal models.

5.3 Model Application

5.3.1 Short-term prediction

The internal factors which are daily total distance and travelling time of 1.22 hrs. (company target) and external factors comprising of significant wave height & wave direction use 4-hr average data (before sailing at 2-6 A.M.) and wind speed & wind direction use data at 5 A.M. to plug in the optimal short-term models obtained from the previous steps. Then, the recommended boat speeds for boat A, boat B, and combined boat are obtained for daily operational adjustment. See example below for boat A. Based on below operating conditions in Table 24, it is recommended for boat A to operate at boat speed of 19.78 knot for that day.

<u>Optimal Model for Boat A</u>

Travelling Time (1st order) = 1.2894 + 0.4522 Total Distance -0.2536 Boat speed + 0.3577 Significant Wave Height -0.1211 Wave Direction +0.1180 Wind Speed (E1) +0.0012 Wind Direction (E1)

Table 24 Example for Short-term Application

Travelling time (hrs.)	Total distance (mile)	Recommended Boat speed (kt)	Significant wave height (m)	Wave direction (°)	Wind speed (kt)	Wind direction (°)
1.22	20	19.78	1	50	10	350

5.3.2 Long-term prediction

AA company aims to use the long-term prediction to plan crew boat operation 1 month in advance. This type of prediction can be applied in two scenarios. One application is to calculate the proper boat speed by fixing travelling time at 1.22 hrs. Then, the company will gain insight about ranges of boat speeds that need to be maintained and operated in monthly basis to achieve the annual target of travelling time.

The other scenario is to calculate travelling time for each boat by using constant boat speed at 16 knots (company's KPI). Then, company will recognize whether there is an opportunity to accelerate some works to perform earlier in that particular month or not. This determination is based on travelling time must be less than 1.22 hrs. The input data is daily total distance and boat speed as internal factors while the external factors comprising of significant wave height, wave direction, wind speed, and wind direction using quarter data of the same period from previous year to apply in the optimal long-term prediction models to obtain either recommended both speed or travelling time for crew boat planning 1 month in advance. See example for travelling time calculation for boat A in Table 25 below. The travelling time is 0.99 hrs. which is still below target, indicating opportunity to visit the nearby platform to accelerate some work.

<u>Optimal Model for Boat A</u>

Travelling Time (1^{st} order with Interaction)	=	1.2894 + 0.3935 Total Distance -0.0788 Boat speed + 0.0719 Significant Wave Height -0.0892 Wave Direction -0.0022 Distance*Wave Direction +0.0007 Speed*Wind speed -0.0094 Speed*Wind direction

Table 25 Example for Long-term Application

Calculated Travelling Time (hrs.)	Total distance (mile)	Boat speed (kt)	Significant wave height (m)	Wave direction (°)	Distance* Wave Direction	Speed* Wind speed	Speed* Wind direction
0.99	31	14	1	50	1550	240	400

5.4 Recommendation

The further improvement can be made by including boat direction in the model built-up or predict as single hop instead of total distance to gain higher accuracy. Furthermore, other algorithms for elastic net regression i.e. Caret, etc. should be trialed.

Moreover, new adaptive techniques i.e. adaptive lasso, adaptive elastic net, and pairwise elastic net or using other regression techniques i.e. robust regression and principal components regression should be applied to compare with this optimal elastic net regression model following the same step as this study.

In addition, current models should be re-evaluated annually with data of current year operation to ensure that there is no significant change for all predictors that might lead to change in overall processes, resulting in increase in forecasting error.

Appendix A

Appendix 1 Data collection results

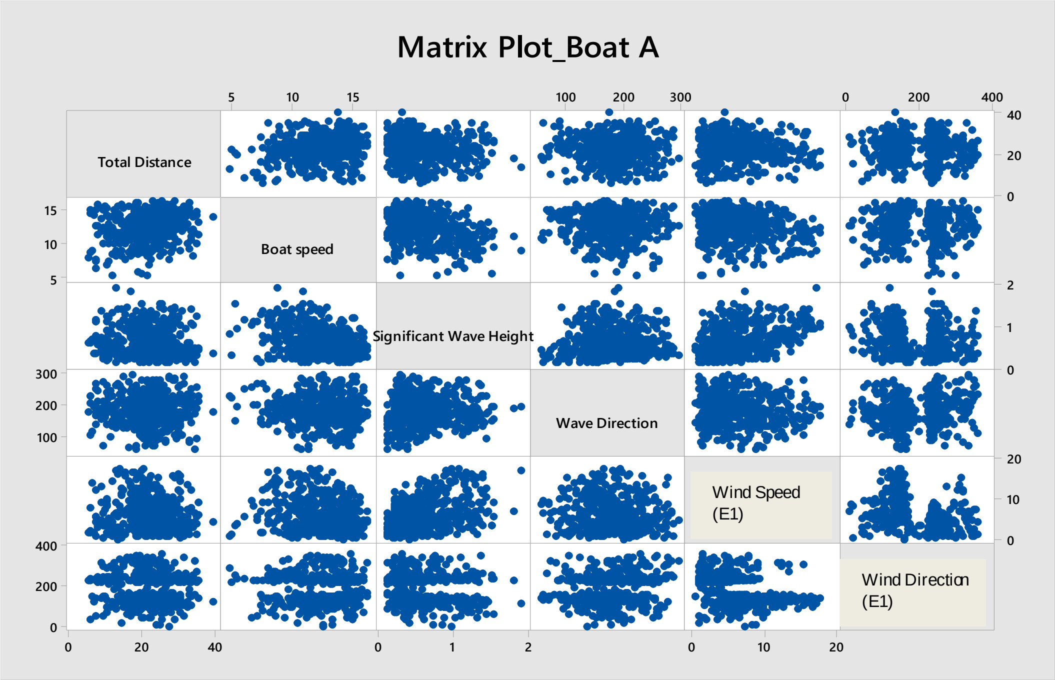


Figure 43 Matrix Plot for Boat A

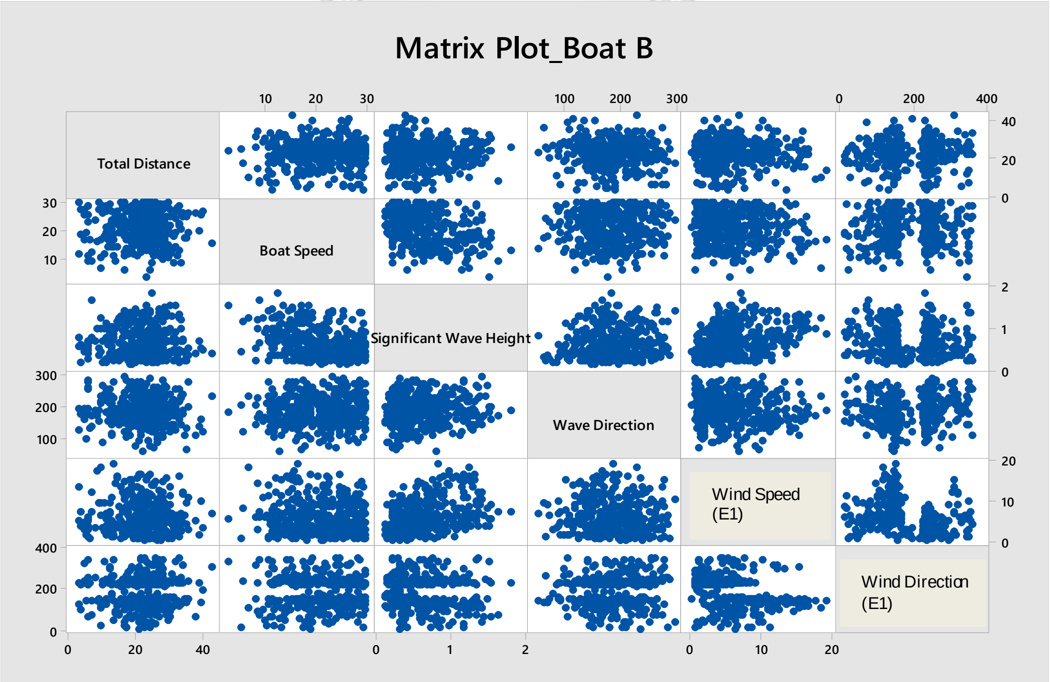


Figure 44 Matrix Plot for Boat B

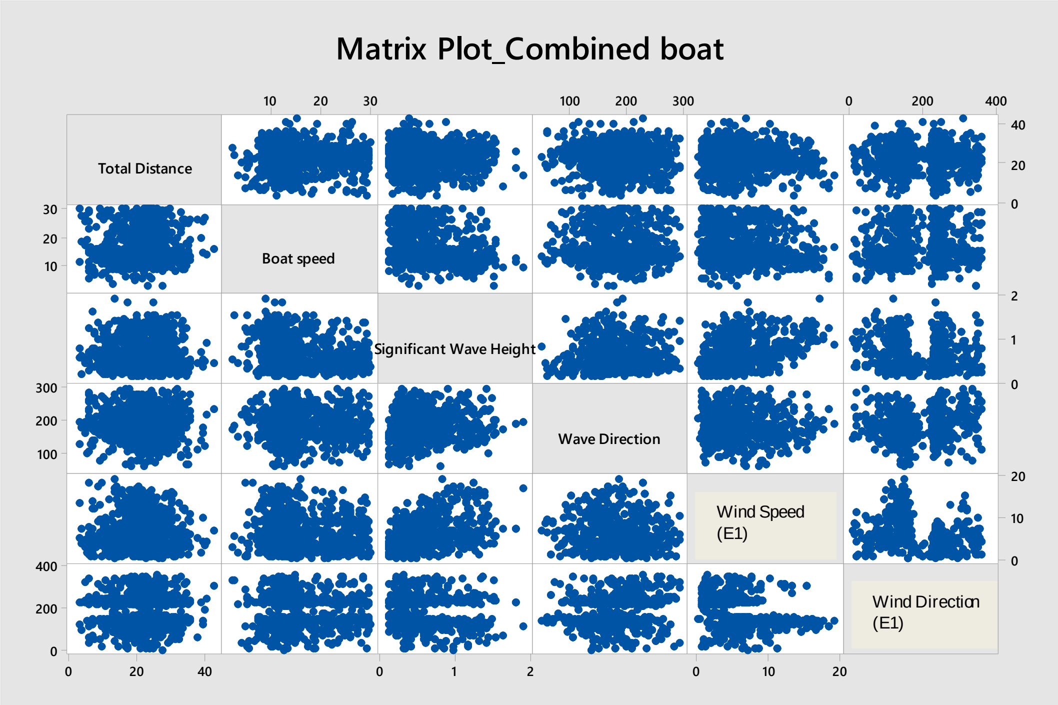


Figure 45 Matrix Plot for Combined boat

Appendix 2 Model Selection Regression Result

A2.1.1 Multiple regression with all variable input analysis

Short-term Prediction

A2.1.1.1 Multiple regression with all variable input analysis: Boat A, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	58.6388	9.7731	506.52	0.000
Total Distance	1	56.6175	56.6175	2934.35	0.000
Boat speed	1	0.9112	0.9112	47.23	0.000
Significant Wave Height	1	0.8760	0.8760	45.40	0.000
Wave Direction	1	0.6072	0.6072	31.47	0.000
Wind Speed (E1)	1	0.0753	0.0753	3.90	0.049
Wind Direction (E1)	1	0.0005	0.0005	0.03	0.869
Error	371	7.1583	0.0193		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.138905	89.12%	88.94%	88.69%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.3777	0.00714	180.48	0.000	
Total Distance	0.4933	0.00726	54.17	0.000	1.03
Boat speed	-0.2639	0.00784	-6.87	0.000	1.20
Significant Wave Height	0.3679	0.00860	6.74	0.000	1.44
Wave Direction	-0.0412	0.00735	-5.61	0.000	1.06
Wind Speed (E1)	0.0188	0.00911	1.98	0.049	1.62
Wind Direction (E1)	0.0013	0.00790	0.17	0.869	1.22

Regression Equation

Travelling Time = 0.3777 + 0.4933 Total Distance - 0.2639 Boat speed
+ 0.3679 Significant Wave Height - 0.0412 Wave Direction
+ 0.0188 Wind Speed (E1) + 0.0013 Wind Direction (E1)

A2.1.1.2 Multiple regression with all variable input analysis: Boat A, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	10	55.8087	5.58087	205.06	0.000
Boat speed	1	0.7464	0.74636	27.42	0.000
Distance*Sig Wave	1	0.0209	0.02090	0.77	0.381
Distance*Wave Direction	1	2.4581	2.45811	90.32	0.000
Distance*Wind Speed	1	1.5802	1.58022	58.06	0.000
Distance*Wind direction	1	5.9043	5.90432	216.94	0.000
Sig Wave*Wind speed	1	0.1444	0.14444	5.31	0.022
Sig Wave*Wind direction	1	0.0000	0.00000	0.00	1.000
Wave direction*Wind speed	1	0.5555	0.55553	20.41	0.000

Wave direction*Wind direction	1	2.9969	2.99692	110.11	0.000
Wind speed*Wind direction	1	0.0749	0.07495	2.75	0.098
Error	367	9.9884	0.02722		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.164974	84.82%	84.41%	83.67%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.9465	0.00849	151.96	0.000	
Boat speed	-0.2429	0.00948	-5.24	0.000	1.24
Distance*Sig Wave	0.0272	0.0261	0.88	0.381	9.41
Distance*Wave Direction	0.00110	0.0189	9.50	0.000	4.96
Distance*Wind Speed	0.001959	0.0243	7.62	0.000	8.21
Distance*Wind direction	0.000122	0.0162	14.73	0.000	3.65
Sig Wave*Wind speed	0.01096	0.0227	2.30	0.022	7.16
Sig Wave*Wind direction	0.00002	0.0230	0.00	1.000	7.30
Wave direction*Wind speed	-0.000158	0.0271	-4.52	0.000	10.15
Wave direction*Wind direction	-0.00011	0.0177	-10.49	0.000	4.35
Wind speed*Wind direction	-0.00061	0.0244	-1.66	0.098	8.21

Regression Equation

Travelling Time = 0.9465 - 0.2429 Boat speed + 0.0272 Distance*Sig Wave
+ 0.00110 Distance*Wave Direction + 0.001959 Distance*Wind Speed
+ 0.000122 Distance*Wind direction + 0.01096 Sig Wave*Wind speed
+ 0.00002 Sig Wave*Wind direction
- 0.000158 Wave direction*Wind speed

- 0.00011 Wave direction*Wind direction
- 0.00061 Wind speed*Wind direction

A2.1.1.3 Multiple regression with all variable input analysis: Boat B, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	77.3531	12.8922	453.36	0.000
Total Distance	1	69.9818	69.9818	2460.92	0.000
Boat Speed	1	0.0400	0.0400	1.41	0.237
Significant Wave Height	1	2.8523	2.8523	100.30	0.000
Wave Direction	1	0.0119	0.0119	0.42	0.519
Wind Speed (E1)	1	0.0624	0.0624	2.19	0.139
Wind Direction (E1)	1	0.0000	0.0000	0.00	0.990
Error	385	10.9483	0.0284		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.168633	87.60%	87.41%	87.20%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.1106	0.00852	157.61	0.000	
Total Distance	0.5646	0.00856	49.61	0.000	1.01
Boat Speed	-0.2109	0.00918	-1.19	0.237	1.16
Significant Wave Height	0.2977	0.00975	10.02	0.000	1.31

Wave Direction	-0.0056	0.00861	-0.65	0.519	1.02
Wind Speed (E1)	0.0148	0.0100	1.48	0.139	1.38
Wind Direction (E1)	0.0001	0.00940	0.01	0.990	1.21

Regression Equation

Travelling Time = -0.1106 + 0.5646 Total Distance - 0.2109 Boat Speed
+ 0.2977 Significant Wave Height – 0.0056 Wave Direction
+ 0.0148 Wind Speed (E1) + 0.0001 Wind Direction (E1)

A2.1.1.4 Multiple regression with all variable input analysis: Boat B, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	11	75.7944	6.89040	209.35	0.000
Distance*Speed	1	3.0540	3.05400	92.79	0.000
Distance*Sig Wave	1	1.3587	1.35870	41.28	0.000
Distance*Wave Direction	1	1.9896	1.98955	60.45	0.000
Speed*Sig Wave	1	0.4303	0.43028	13.07	0.000
Speed*Wave direction	1	2.5074	2.50743	76.18	0.000
Speed*Wind speed	1	0.1170	0.11698	3.55	0.060
Speed*Wind direction	1	0.0005	0.00054	0.02	0.898
Sig Wave*Wind speed	1	0.0940	0.09399	2.86	0.092
Sig Wave*Wind direction	1	0.0115	0.01150	0.35	0.555
Wave direction*Wind direction	1	0.0186	0.01856	0.56	0.453
Wind speed*Wind direction	1	0.2208	0.22078	6.71	0.010
Error	380	12.5070	0.03291		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.181420	85.84%	85.43%	84.98%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.6264	0.00916	146.50	0.000	
Distance*Speed	-0.01257	0.0231	-9.63	0.000	6.36
Distance*Sig Wave	0.2052	0.0293	6.43	0.000	10.23
Distance*Wave Direction	0.00124	0.0258	7.77	0.000	7.94
Speed*Sig Wave	-0.01435	0.0253	-3.62	0.000	7.60
Speed*Wave direction	-0.000123	0.0206	-8.73	0.000	5.03
Speed*Wind speed	0.000564	0.0250	1.89	0.060	7.40
Speed*Wind direction	-0.000002	0.0243	-0.13	0.898	7.01
Sig Wave*Wind speed	0.00870	0.0217	1.69	0.092	5.62
Sig Wave*Wind direction	0.000199	0.0283	0.59	0.555	9.49
Wave direction*Wind direction	0.000001	0.0280	0.75	0.453	9.32
Wind speed*Wind direction	-0.000088	0.0226	-2.59	0.010	6.05

Regression Equation

Travelling Time = 0.6264 - 0.01257 Distance*Speed + 0.2052 Distance*Sig Wave
+ 0.00124 Distance*Wave Direction - 0.01435 Speed*Sig Wave
- 0.000123 Speed*Wave direction + 0.000564 Speed*Wind speed
- 0.000002 Speed*Wind direction + 0.00870 Sig Wave*Wind speed
+ 0.000199 Sig Wave*Wind direction
+ 0.000001 Wave direction*Wind direction
- 0.000088 Wind speed*Wind direction

A2.1.1.5 Multiple regression with all variable input analysis: Combined boat, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	135.227	19.318	758.32	0.000
Total Distance	1	127.153	127.153	4991.24	0.000
Boat speed	1	0.292	0.292	11.47	0.001
Significant Wave Height	1	3.659	3.659	143.63	0.000
Wave Direction	1	0.218	0.218	8.55	0.004
Wind Speed (E1)	1	0.251	0.251	9.84	0.002
Wind Direction (E1)	1	0.002	0.002	0.06	0.805
Boat Code	1	1.841	1.841	72.27	0.000
Error	762	19.412	0.025		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.159609	87.45%	87.33%	87.19%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0827	0.00575	228.86	0.000	
Total Distance	0.5075	0.00577	70.65	0.000	1.00
Boat speed	-0.3361	0.00771	-3.39	0.001	1.79
Significant Wave Height	0.2006	0.00672	11.98	0.000	1.36
Wave Direction	-0.0192	0.00587	-2.92	0.004	1.04
Wind Speed (E1)	0.0218	0.00696	3.14	0.002	1.46

Wind Direction (E1)	0.0016	0.00636	0.25	0.805	1.22
Boat Code	-0.1794	0.00761	-8.50	0.000	1.75

Regression Equation

Travelling Time = 0.0827 + 0.5075 Total Distance - 0.3361 Boat speed + 0.2006 Significant Wave Height - 0.0192 Wave Direction + 0.0218 Wind Speed (E1) + 0.0016 Wind Direction (E1) - 0.1794 Boat Code

A2.1.1.6 Multiple regression with all variable input analysis: Combined boat, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	13	131.540	10.1184	331.15	0.000
Distance*Speed	1	2.325	2.3254	76.10	0.000
Distance*Wave Direction	1	1.515	1.5150	49.58	0.000
Distance*Wind Speed	1	2.442	2.4419	79.92	0.000
Distance*Wind direction	1	3.279	3.2789	107.31	0.000
Speed*Sig Wave	1	0.058	0.0582	1.91	0.168
Speed*Wave direction	1	0.329	0.3286	10.75	0.001
Speed*Wind speed	1	0.318	0.3176	10.39	0.001
Speed*Wind direction	1	0.509	0.5090	16.66	0.000
Sig Wave*Wind speed	1	0.459	0.4593	15.03	0.000
Sig Wave*Wind direction	1	0.604	0.6038	19.76	0.000
Wave direction*Wind direction	1	0.645	0.6448	21.10	0.000
Wind speed*Wind direction	1	1.121	1.1215	36.70	0.000
Boat Code	1	1.591	1.5907	52.06	0.000

Error	756	23.100	0.0306
Total	769	154.639	

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.174801	85.06%	84.81%	84.37%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.7151	0.00630	208.97	0.000	
Distance*Speed	-0.00993	0.0180	-8.72	0.000	8.15
Distance*Wave Direction	0.00076	0.0175	7.04	0.000	7.73
Distance*Wind Speed	0.01796	0.0188	8.94	0.000	8.89
Distance*Wind direction	0.000093	0.0191	10.36	0.000	9.13
Speed*Sig Wave	-0.0403	0.0167	-1.38	0.168	7.03
Speed*Wave direction	-0.000049	0.0201	-3.28	0.001	10.20
Speed*Wind speed	-0.000838	0.0183	-3.22	0.001	8.40
Speed*Wind direction	-0.000049	0.0193	-4.08	0.000	9.35
Sig Wave*Wind speed	0.01257	0.0146	3.88	0.000	5.37
Sig Wave*Wind direction	0.000947	0.0168	4.45	0.000	7.14
Wave direction*Wind direction	-0.000005	0.0192	-4.59	0.000	9.27
Wind speed*Wind direction	-0.000148	0.0163	-6.06	0.000	6.66
Boat Code	-0.1213	0.00841	-7.22	0.000	1.78

Regression Equation

Travelling Time = 0.7151 - 0.00993 Distance*Speed + 0.00076 Distance*Wave Direction
+ 0.01796 Distance*Wind Speed + 0.000093 Distance*Wind direction
- 0.0403 Speed*Sig Wave - 0.000049 Speed*Wave direction
- 0.000838 Speed*Wind speed - 0.000049 Speed*Wind direction

+ 0.01257 Sig Wave*Wind speed + 0.000947 Sig Wave*Wind direction
- 0.000005 Wave direction*Wind direction
- 0.000148 Wind speed*Wind direction - 0.1213 Boat Code

Long-term Prediction

A2.1.2.1 Multiple regression with all variable input analysis: Boat A, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	58.7600	11.7520	621.24	0.000
Total Distance	1	57.0919	57.0919	3018.01	0.000
Boat speed	1	2.4531	2.4531	129.68	0.000
Significant Wave Height	1	0.1919	0.1919	10.15	0.002
Wave Direction	1	1.8508	1.8508	97.84	0.000
Wind Direction (E1)	1	0.0139	0.0139	0.73	0.393
Error	372	7.0372	0.0189		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.137539	89.30%	89.16%	88.92%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.0001	0.00707	182.27	0.000	
Total Distance	0.5923	0.00723	54.94	0.000	1.04
Boat speed	-0.4044	0.00725	-11.39	0.000	1.05

Significant Wave Height	0.2185	0.00741	3.19	0.002	1.09
Wave Direction	-0.00323	0.00745	-9.89	0.000	1.11
Wind Direction (E1)	-0.000181	0.00739	-0.86	0.393	1.09

Regression Equation

Travelling Time = 1.0001 + 0.5923 Total Distance - 0.4044 Boat speed
+ 0.2185 Significant Wave Height - 0.003231 Wave Direction
- 0.000181 Wind Direction (E1)

A2.1.2.2 Multiple regression with all variable input analysis: Boat A, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	58.7836	7.3480	386.60	0.000
Total Distance	1	56.0142	56.0142	2947.06	0.000
Boat speed	1	2.3555	2.3555	123.93	0.000
Significant Wave Height	1	0.0911	0.0911	4.80	0.029
Wave Direction	1	0.8138	0.8138	42.82	0.000
Wind Direction (E1)	1	0.0209	0.0209	1.10	0.295
Distance*Wave Direction	1	0.0055	0.0055	0.29	0.589
Speed*Wave direction	1	0.0077	0.0077	0.41	0.524
Wind speed*Wind direction	1	0.0065	0.0065	0.34	0.559
Error	369	7.0135	0.0190		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.137865	89.34%	89.11%	88.79%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.082	0.00709	181.84	0.000	
Total Distance	0.5939	0.00734	54.29	0.000	1.07
Boat speed	-0.3995	0.00733	-11.13	0.000	1.07
Significant Wave Height	0.269	0.0133	2.19	0.029	3.50
Wave Direction	-0.003131	0.0109	-6.54	0.000	2.36
Wind Direction (E1)	-0.000328	0.0109	-1.05	0.295	2.36
Distance*Wave Direction	-0.00003	0.00798	-0.54	0.589	1.26
Speed*Wave direction	-0.00013	0.00881	-0.64	0.524	1.54
Wind speed*Wind direction	-0.00064	0.0167	-0.59	0.559	5.55

Regression Equation

Travelling Time = 1.082 + 0.5939 Total Distance - 0.3995 Boat speed
+ 0.269 Significant Wave Height - 0.003131 Wave Direction
- 0.000328 Wind Direction (E1) - 0.00003 Distance*Wave Direction
- 0.00013 Speed*Wave direction
- 0.00064 Wind speed*Wind direction

A2.1.2.3 Multiple regression with all variable input analysis: Boat B, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	73.7051	14.7410	389.83	0.000
Total Distance	1	71.2239	71.2239	1883.51	0.000
Boat Speed	1	0.6794	0.6794	17.97	0.000
Significant Wave Height	1	0.0839	0.0839	2.22	0.137

Wave Direction	1	0.1381	0.1381	3.65	0.057
Wind Speed (E1)	1	0.0445	0.0445	1.18	0.278
Error	386	14.5963	0.0378		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.194459	83.47%	83.26%	82.97%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.288	0.00982	136.68	0.000	
Total Distance	0.6145	0.00987	43.40	0.000	1.01
Boat Speed	-0.0746	0.0102	-4.24	0.000	1.09
Significant Wave Height	0.169	0.0103	1.49	0.137	1.10
Wave Direction	-0.01199	0.0104	-1.91	0.057	1.12
Wind Speed (E1)	0.00646	0.0102	1.09	0.278	1.09

Regression Equation

Travelling Time = 0.288 + 0.6145 Total Distance - 0.0746 Boat Speed
+ 0.169 Significant Wave Height - 0.01199 Wave Direction
+ 0.00646 Wind Speed (E1)

A2.1.2.4 Multiple regression with all variable input analysis: Boat B, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	73.4825	10.4975	272.02	0.000
Significant Wave Height	1	0.0020	0.0020	0.05	0.821
Wave Direction	1	0.0272	0.0272	0.71	0.401

Distance*Wind Speed	1	2.1580	2.1580	55.92	0.000
Distance*Wind direction	1	8.3863	8.3863	217.31	0.000
Speed*Wind speed	1	0.0420	0.0420	1.09	0.298
Speed*Wind direction	1	0.4727	0.4727	12.25	0.001
Wind speed*Wind direction	1	0.0484	0.0484	1.25	0.263
Error	384	14.8190	0.0386		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.196446	83.22%	82.91%	82.51%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.583	0.00992	135.30	0.000	
Significant Wave Height	0.32	0.0129	0.23	0.821	1.68
Wave Direction	-0.00542	0.0107	-0.84	0.401	1.16
Distance*Wind Speed	0.03835	0.0274	7.48	0.000	7.62
Distance*Wind direction	0.000204	0.0222	14.74	0.000	5.00
Speed*Wind speed	0.000558	0.0281	1.04	0.298	8.01
Speed*Wind direction	-0.000054	0.0185	-3.50	0.001	3.45
Wind speed*Wind direction	-0.000252	0.0297	-1.12	0.263	8.95

Regression Equation

Travelling Time = 0.583 + 0.32 Significant Wave Height - 0.00542 Wave Direction
+ 0.03835 Distance*Wind Speed + 0.000204 Distance*Wind direction
+ 0.000558 Speed*Wind speed - 0.000054 Speed*Wind direction
- 0.000252 Wind speed*Wind direction

A2.1.2.5 Multiple regression with all variable input analysis: Combined boat, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	131.259	21.876	713.92	0.000
Total Distance	1	126.976	126.976	4143.73	0.000
Boat speed	1	1.723	1.723	56.23	0.000
Significant Wave Height	1	0.486	0.486	15.86	0.000
Wave Direction	1	1.570	1.570	51.23	0.000
Wind Direction (E1)	1	0.189	0.189	6.18	0.013
Boat Code	1	3.951	3.951	128.94	0.000
Error	763	23.381	0.031		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.175051	84.88%	84.76%	84.60%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.6748	0.00631	208.68	0.000	
Total Distance	0.59600	0.00634	64.37	0.000	1.01
Boat speed	-0.1098	0.00815	-7.50	0.000	1.67
Significant Wave Height	0.2591	0.00665	3.98	0.000	1.11
Wave Direction	-0.002335	0.00665	-7.16	0.000	1.11
Wind Direction (E1)	-0.000446	0.00650	-2.49	0.013	1.06
Boat Code	-0.1904	0.00839	-11.35	0.000	1.77

Regression Equation

Travelling Time = 0.6748 + 0.59600 Total Distance - 0.1098 Boat speed
+ 0.2591 Significant Wave Height - 0.002335 Wave Direction
- 0.000446 Wind Direction (E1) - 0.1904 Boat Code

A2.1.2.6 Multiple regression with all variable input analysis: Combined boat, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	130.636	16.3295	517.71	0.000
Significant Wave Height	1	0.011	0.0110	0.35	0.556
Wave Direction	1	0.868	0.8684	27.53	0.000
Distance*Wind Speed	1	6.703	6.7027	212.50	0.000
Distance*Wind direction	1	15.410	15.4103	488.57	0.000
Speed*Wind speed	1	0.021	0.0208	0.66	0.418
Speed*Wind direction	1	1.034	1.0343	32.79	0.000
Wind speed*Wind direction	1	0.380	0.3797	12.04	0.001
Boat Code	1	3.173	3.1726	100.59	0.000
Error	761	24.003	0.0315		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.177599	84.48%	84.31%	84.10%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.095	0.00640	205.68	0.000	
Significant Wave Height	0.577	0.0100	0.59	0.556	2.44
Wave Direction	-0.01896	0.00737	-5.25	0.000	1.32
Distance*Wind Speed	0.04219	0.0160	14.58	0.000	6.26
Distance*Wind direction	0.00183	0.0128	22.10	0.000	4.01
Speed*Wind speed	0.000316	0.0176	0.81	0.418	7.53
Speed*Wind direction	-0.000068	0.0140	-5.73	0.000	4.75
Wind speed*Wind direction	-0.000444	0.0186	-3.47	0.001	8.44
Boat Code	-0.1869	0.00932	-10.03	0.000	2.12

Regression Equation

Travelling Time = 1.095 + 0.577 Significant Wave Height - 0.01896 Wave Direction
+ 0.04219 Distance*Wind Speed + 0.00183 Distance*Wind direction
+ 0.000316 Speed*Wind speed - 0.000068 Speed*Wind direction
- 0.000444 Wind speed*Wind direction - 0.1869 Boat Code

A2.2.1 Stepwise regression analysis

Short-term Prediction

A2.2.1.1 Stepwise regression analysis: Boat A, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	58.6382	11.7276	609.41	0.000
Total Distance	1	56.7567	56.7567	2949.28	0.000
Boat speed	1	0.9155	0.9155	47.57	0.000

Significant Wave Height	1	0.8816	0.8816	45.81	0.000
Wave Direction	1	0.6086	0.6086	31.63	0.000
Wind Speed (E1)	1	0.0827	0.0827	4.30	0.039
Error	372	7.1589	0.0192		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.138724	89.12%	88.97%	88.76%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.3822	0.00714	180.71	0.000	
Total Distance	0. 4931	0.00724	54.31	0.000	1.03
Boat speed	-0.02640	0.00782	-6.90	0.000	1.20
Significant Wave Height	0.1669	0.00857	6.77	0.000	1.44
Wave Direction	-0.000857	0.00732	-5.62	0.000	1.05
Wind Speed (E1)	0.00424	0.00840	2.07	0.039	1.38

Regression Equation

Travelling Time = 0.3822 + 0. 4931 Total Distance - 0.02640 Boat speed
+ 0.1669 Significant Wave Height - 0.000857 Wave Direction
+ 0.00424 Wind Speed (E1)

A2.2.1.2 Stepwise regression analysis: Boat A, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	59.7816	6.6424	406.35	0.000

Total Distance	1	55.4722	55.4722	3393.54	0.000
Boat speed	1	0.8554	0.8554	52.33	0.000
Significant Wave Height	1	0.3146	0.3146	19.25	0.000
Wave Direction	1	0.4222	0.4222	25.83	0.000
Wind Speed (E1)	1	0.0388	0.0388	2.37	0.124
Total Distance*Boat speed	1	0.1870	0.1870	11.44	0.001
Total Distance*Wave Direction	1	0.1830	0.1830	11.20	0.001
Boat speed*Significant Wave Height	1	0.4095	0.4095	25.05	0.000
Significant Wave Height*Wind Speed (E1)	1	0.0930	0.0930	5.69	0.018
Error	368	6.0155	0.0163		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.127853	90.86%	90.63%	90.30%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.614	0.00745	170.79	0.000	
Total Distance	0.5506	0.00676	58.25	0.000	1.05
Boat speed	-0.0331	0.00742	-7.23	0.000	1.27
Significant Wave Height	0.2905	0.00851	4.39	0.000	1.67
Wave Direction	0.000831	0.00685	5.08	0.000	1.08
Wind Speed (E1)	-0.00347	0.00800	-1.54	0.124	1.48

Total Distance*Boat speed	-0.001455	0.00589	-3.38	0.001	1.08
Total Distance*Wave Direction	-0.000072	0.00688	-3.35	0.001	1.03
Boat speed*Significant Wave Height	-0.04999	0.00709	-5.01	0.000	1.48
Significant Wave Height*Wind Speed (E1)	0.01165	0.00697	2.38	0.018	1.48

Regression Equation

Travelling Time = 0.614 + 0.5506 Total Distance - 0.0331 Boat speed
+ 0.2905 Significant Wave Height + 0.000831 Wave Direction
- 0.00347 Wind Speed (E1) - 0.001455 Total Distance*Boat speed
- 0.000072 Total Distance*Wave Direction
- 0.04999 Boat speed*Significant Wave Height
+ 0.01165 Significant Wave Height*Wind Speed (E1)

A2.2.1.3 Stepwise regression analysis: Boat B, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	87.5755	38.6235	1359.14	0.000
Total Distance	1	70.3677	70.3677	2476.21	0.000
Boat speed	1	12.7893	12.7893	450.33	0.000
Significant Wave Height	1	4.4185	4.4185	155.48	0.000
Error	389	11.0544	0.0284		
Total	392	98.6299			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.168575	87.48%	87.42%	87.30%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.1195	0.00851	157.67	0.000	
Total Distance	0.5036	0.00854	49.76	0.000	1.00
Boat speed	-0.0097	0.00854	-24.89	0.000	1.00
Significant Wave Height	0.3073	0.00854	12.47	0.000	1.00

Regression Equation

Travelling Time = 0.1195 + 0.5036 Total Distance - 0.0097 Boat speed + 0.3073 Significant Wave Height

A2.2.1.4 Stepwise regression analysis: Boat B, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	77.9370	11.1339	412.51	0.000
Total Distance	1	70.5158	70.5158	2612.59	0.000
Boat Speed	1	0.0416	0.0416	1.54	0.215
Significant Wave Height	1	2.2416	2.2416	83.05	0.000
Wind Speed (E1)	1	0.0051	0.0051	0.19	0.664
Wind Direction (E1)	1	0.0126	0.0126	0.47	0.495
Boat Speed*Significant Wave Height	1	0.3297	0.3297	12.21	0.001
Wind Speed (E1)*Wind Direction (E1)	1	0.2524	0.2524	9.35	0.002
Error	384	10.3645	0.0270		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.164289	88.26%	88.05%	87.84%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.3722	0.00967	136.52	0.000	
Total Distance	0.4467	0.00836	51.11	0.000	1.01
Boat Speed	-0.00780	0.00891	-1.24	0.215	1.15
Significant Wave Height	0.0809	0.00978	9.11	0.000	1.38
Wind Speed (E1)	0.02110	0.0103	0.43	0.664	1.53
Wind Direction (E1)	0.000487	0.00925	0.68	0.495	1.24
Boat Speed*Significant Wave Height	-0.01617	0.00933	-3.49	0.001	1.09
Wind Speed (E1)*Wind Direction (E1)	-0.000103	0.0101	-3.06	0.002	1.14

Regression Equation

Travelling Time = 0.3722 + 0.4467 Total Distance - 0.00780 Boat Speed
+ 0.0809 Significant Wave Height + 0.02110 Wind Speed (E1)
+ 0.000487 Wind Direction (E1)
- 0.01617 Boat Speed*Significant Wave Height
- 0.000103 Wind Speed (E1)*Wind Direction (E1)

A2.2.1.5 Stepwise regression analysis: Combined boat, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	135.226	22.538	885.78	0.000
Boat Code	1	1.856	1.856	72.94	0.000
Total Distance	1	127.163	127.163	4997.79	0.000

Boat speed	1	0.297	0.297	11.66	0.001
Significant Wave Height	1	3.684	3.684	144.80	0.000
Wave Direction	1	0.216	0.216	8.50	0.004
Wind Speed (E1)	1	0.278	0.278	10.93	0.001
Error	763	19.414	0.025		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.159511	87.45%	87.35%	87.22%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0877	0.00575	229.00	0.000	
Boat Code	-0.1296	0.00759	-8.54	0.000	1.74
Total Distance	0.4079	0.00576	70.70	0.000	1.00
Boat speed	-0.00471	0.00768	-3.41	0.001	1.78
Significant Wave Height	0.2321	0.00670	12.03	0.000	1.36
Wave Direction	-0.000365	0.00585	-2.91	0.004	1.03
Wind Speed (E1)	0.00526	0.00640	3.31	0.001	1.24

Regression Equation

Travelling Time = 0.0877 + 0.4079 Total Distance - 0.00471 Boat speed
+ 0.2321 Significant Wave Height - 0.000365 Wave Direction
+ 0.00526 Wind Speed (E1) - 0.1296 Boat Code

A2.2.1.6 Stepwise regression analysis: Combined boat, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	11	136.126	12.375	506.69	0.000
Total Distance	1	127.576	127.576	5223.46	0.000
Boat speed	1	0.244	0.244	10.01	0.002
Boat Code	1	1.694	1.694	69.38	0.000
Significant Wave Height	1	2.863	2.863	117.23	0.000
Wave Direction	1	0.123	0.123	5.02	0.025
Wind Speed (E1)	1	0.027	0.027	1.10	0.295
Wind Direction (E1)	1	0.021	0.021	0.86	0.353
Boat speed*Significant Wave Height	1	0.115	0.115	4.69	0.031
Significant Wave Height*Wind Speed (E1)	1	0.301	0.301	12.34	0.000
Significant Wave Height*Wind Direction (E1)	1	0.227	0.227	9.30	0.002
Wind Speed (E1)*Wind Direction (E1)	1	0.291	0.291	11.93	0.001
Error	758	18.513	0.024		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.156281	88.03%	87.85%	87.66%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0614	0.00669	193.75	0.000	
Total Distance	0.4498	0.00566	72.27	0.000	1.01
Boat speed	-0.00023	0.00762	-3.16	0.002	1.83
Boat Code	-0.1249	0.00750	-8.33	0.000	1.77
Significant Wave Height	0.0786	0.00679	10.83	0.000	1.45
Wave Direction	-0.000278	0.00580	-2.24	0.025	1.06
Wind Speed (E1)	0.00952	0.00730	1.05	0.295	1.68
Wind Direction (E1)	-0.000023	0.00642	-0.93	0.353	1.30
Boat speed*Significant Wave Height	-0.00709	0.00634	-2.17	0.031	1.13
Significant Wave Height*Wind Speed (E1)	0.01622	0.00645	3.51	0.000	1.58
Significant Wave Height*Wind Direction (E1)	0.000809	0.00672	3.05	0.002	1.41
Wind Speed (E1)*Wind Direction (E1)	-0.000090	0.00763	-3.45	0.001	1.33

Regression Equation

Travelling Time = 0.0614 + 0.4498 Total Distance - 0.00023 Boat speed
+ 0.0786 Significant Wave Height - 0.000278 Wave Direction
+ 0.00952 Wind Speed (E1) - 0.000023 Wind Direction (E1)
- 0.00709 Boat speed*Significant Wave Height
+ 0.01622 Significant Wave Height*Wind Speed (E1)
+ 0.000809 Significant Wave Height*Wind Direction (E1)
- 0.000090 Wind Speed (E1)*Wind Direction (E1) - 0.1249 Boat Code

Long-term Prediction

A2.2.2.1 Stepwise regression analysis: Boat A, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	58.7461	14.6865	776.92	0.000
Total Distance	1	57.2803	57.2803	3030.14	0.000
Boat speed	1	2.4487	2.4487	129.54	0.000
Significant Wave Height	1	0.2201	0.2201	11.64	0.001
Wave Direction	1	2.0532	2.0532	108.61	0.000
Error	373	7.0510	0.0189		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.137490	89.28%	89.17%	88.97%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.4714	0.00707	182.33	0.000	
Total Distance	0.4927	0.00722	55.05	0.000	1.04
Boat speed	-0.04040	0.00725	-11.38	0.000	1.05
Significant Wave Height	0.2297	0.00727	3.41	0.001	1.05
Wave Direction	-0.003300	0.00722	-10.42	0.000	1.04

Regression Equation

Travelling Time = 0.4714 + 0.4927 Total Distance - 0.04040 Boat speed + 0.2297 Significant Wave Height - 0.003300 Wave Direction

A2.2.2.2 Stepwise regression analysis: Boat A, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	59.3023	9.8837	564.58	0.000
Total Distance	1	55.6372	55.6372	3178.13	0.000
Boat speed	1	2.8390	2.8390	162.17	0.000
Significant Wave Height	1	0.2194	0.2194	12.53	0.000
Wave Direction	1	1.9698	1.9698	112.52	0.000
Total Distance*Boat speed	1	0.2851	0.2851	16.28	0.000
Total Distance*Wave Direction	1	0.3175	0.3175	18.14	0.000
Error	371	6.4948	0.0175		
Total	377	65.7971			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.132311	90.13%	89.97%	89.75%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.123	0.00688	188.09	0.000	
Total Distance	0.4121	0.00707	56.37	0.000	1.08
Boat speed	-0.00566	0.00719	-12.73	0.000	1.11
Significant Wave Height	0.2294	0.00700	3.54	0.000	1.05
Wave Direction	0.000380	0.00703	10.61	0.000	1.06
Total Distance*Boat speed	-0.001800	0.00611	-4.04	0.000	1.09
Total Distance*Wave Direction	-0.000168	0.00602	-4.26	0.000	1.06

Regression Equation

Travelling Time = 0.123 + 0.4121 Total Distance - 0.00566 Boat speed
+ 0.2294 Significant Wave Height + 0.000380 Wave Direction
- 0.001800 Total Distance*Boat speed
- 0.000168 Total Distance*Wave Direction

A2.2.2.3 Stepwise regression analysis: Boat B, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	73.5007	36.7504	965.89	0.000
Total Distance	1	71.5001	71.5001	1879.20	0.000
Boat Speed	1	0.6722	0.6722	17.67	0.000
Error	389	14.8007	0.0380		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.195059	83.24%	83.15%	82.99%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.1937	0.00985	136.26	0.000	
Total Distance	0.6396	0.00989	43.35	0.000	1.00
Boat Speed	-0.00714	0.00989	-4.20	0.000	1.00

Regression Equation

Travelling Time = 0.1937 + 0.6396 Total Distance - 0.00714 Boat Speed

A2.2.2.4 Stepwise regression analysis: Boat B, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	73.7473	18.4368	490.24	0.000
Total Distance	1	71.7361	71.7361	1907.49	0.000
Boat Speed	1	0.7192	0.7192	19.12	0.000
Significant Wave Height	1	0.0457	0.0457	1.22	0.271
Boat Speed*Significant Wave Height	1	0.2100	0.2100	5.58	0.019
Error	387	14.5541	0.0376		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.193927	83.52%	83.35%	83.15%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.400	0.00982	136.90	0.000	
Total Distance	0.4106	0.00984	43.67	0.000	1.01
Boat Speed	-0.0195	0.00986	-4.37	0.000	1.01
Significant Wave Height	0.2996	0.00984	1.10	0.271	1.01
Boat Speed*Significant Wave Height	-0.0448	0.0101	-2.36	0.019	1.01

Regression Equation

Travelling Time = 0.400 + 0.4106 Total Distance - 0.0195 Boat Speed + 0.2996 Significant Wave Height - 0.0448 Boat Speed*Significant Wave Height

A2.2.2.5 Stepwise regression analysis: Combined boat, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	131.259	21.876	713.92	0.000
Boat Code	1	3.951	3.951	128.94	0.000
Total Distance	1	126.976	126.976	4143.73	0.000
Boat speed	1	1.723	1.723	56.23	0.000
Significant Wave Height	1	0.486	0.486	15.86	0.000
Wave Direction	1	1.570	1.570	51.23	0.000
Wind Direction (E1)	1	0.189	0.189	6.18	0.013
Error	763	23.381	0.031		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.175051	84.88%	84.76%	84.60%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.6748	0.00631	208.68	0.000	
Boat Code	-0.1904	0.00839	-11.35	0.000	1.77
Total Distance	0.5080	0.00634	64.37	0.000	1.01
Boat speed	-0.01098	0.00815	-7.50	0.000	1.67

Significant Wave Height	0.2591	0.00665	3.98	0.000	1.11
Wave Direction	-0.002335	0.00665	-7.16	0.000	1.11
Wind Direction (E1)	-0.000446	0.00650	-2.49	0.013	1.06

Regression Equation

Travelling Time = 0.6748 + 0.5080 Total Distance - 0.01098 Boat speed
+ 0.2591 Significant Wave Height - 0.002335 Wave Direction
- 0.000446 Wind Direction (E1) - 0.1904 Boat Code

A2.2.2.6 Stepwise regression analysis: Combined boat, 1st order with Interaction

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	131.769	14.641	486.53	0.000
Total Distance	1	126.138	126.138	4191.66	0.000
Boat speed	1	1.763	1.763	58.59	0.000
Boat Code	1	3.952	3.952	131.34	0.000
Wind Direction (E1)	1	0.124	0.124	4.13	0.042
Significant Wave Height	1	0.377	0.377	12.54	0.000
Wave Direction	1	1.288	1.288	42.79	0.000
Total Distance*Wave Direction	1	0.168	0.168	5.58	0.018
Boat speed*Significant Wave Height	1	0.124	0.124	4.12	0.043
Boat speed*Wave Direction	1	0.264	0.264	8.76	0.003
Error	760	22.870	0.030		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.173472	85.21%	85.04%	84.82%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.549	0.00637	206.44	0.000	
Total Distance	0.4985	0.00635	64.74	0.000	1.03
Boat speed	-0.0283	0.00832	-7.65	0.000	1.77
Boat Code	-0.1912	0.00834	-11.46	0.000	1.78
Wind Direction (E1)	-0.000366	0.00652	-2.03	0.042	1.09
Significant Wave Height	0.1645	0.00666	3.54	0.000	1.13
Wave Direction	-0.00297	0.00667	-6.54	0.000	1.14
Total Distance*Wave Direction	-0.000093	0.00550	-2.36	0.018	1.03
Boat speed*Significant Wave Height	-0.0255	0.00716	-2.03	0.043	1.16
Boat speed*Wave Direction	0.000174	0.00666	2.96	0.003	1.20

Regression Equation

Travelling Time = 0.549 + 0.4985 Total Distance - 0.0283 Boat speed
+ 0.1645 Significant Wave Height

- 0.000366 Wind Direction (E1)
- 0.00297 Wave Direction - 0.000093 Total Distance*Wave Direction
- 0.0255 Boat speed*Significant Wave Height
+ 0.000174 Boat speed*Wave Direction - 0.1912 Boat Code

A2.3.1 Regression with Lasso Variable Selection Analysis

Short-term Prediction

2.3.1.1 Regression with Lasso Variable Selection Analysis: Boat A, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	57.845	19.2818	906.90	0.000
Total Distance	1	56.870	56.8701	2674.82	0.000
Boat speed	1	1.009	1.0088	47.45	0.000
Significant Wave Height	1	1.210	1.2097	56.90	0.000
Error	374	7.952	0.0213		
Total	377	65.797			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.145813	87.91%	87.82%	87.65%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.2632	0.00750	171.93	0.000	
Total Distance	0.4586	0.00761	51.72	0.000	1.03
Boat speed	-0.3376	0.00811	-6.89	0.000	1.17
Significant Wave Height	0.4493	0.00802	7.54	0.000	1.14

Regression Equation

Travelling Time = 0.2632 + 0.4586 Total Distance - 0.3376 Boat speed + 0.4493 Significant Wave Height

2.3.1.2 Regression with Lasso Variable Selection Analysis: Boat B, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	77.2715	25.7572	906.06	0.000
Total Distance	1	70.0311	70.0311	2463.48	0.000
Boat Speed	1	0.0245	0.0245	0.86	0.354
Significant Wave Height	1	3.7708	3.7708	132.64	0.000
Error	388	11.0300	0.0284		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.168605	87.51%	87.41%	87.26%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.0855	0.00852	157.64	0.000	
Total Distance	0.4709	0.00855	49.63	0.000	1.01
Boat Speed	-0.3144	0.00902	-0.93	0.354	1.12
Significant Wave Height	0.3996	0.00902	11.52	0.000	1.12

Regression Equation

Travelling Time = -0.0855 + 0.4709 Total Distance - 0.3144 Boat Speed
+ 0.3996 Significant Wave Height

2.3.1.3 Regression with Lasso Variable Selection Analysis: Combined boat, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	134.670	33.668	1289.77	0.000
Total Distance	1	127.105	127.105	4869.28	0.000
Boat speed	1	0.289	0.289	11.07	0.001
Significant Wave Height	1	5.421	5.421	207.67	0.000
Boat Code	1	1.658	1.658	63.53	0.000
Error	765	19.969	0.026		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.161566	87.09%	87.02%	86.92%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0349	0.00582	226.09	0.000	
Total Distance	0.4947	0.00584	69.78	0.000	1.00
Boat speed	-0.2463	0.00775	-3.33	0.001	1.77
Significant Wave Height	0.2536	0.00612	14.41	0.000	1.10
Boat Code	-0.1519	0.00762	-7.97	0.000	1.71

Regression Equation

Travelling Time = 0.0349 + 0.4947 Total Distance - 0.2463 Boat speed + 0.2536 Significant Wave Height - 0.1519 Boat Code

Long-term Prediction

2.3.2.1 Regression with Lasso Variable Selection Analysis: Boat A, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	56.636	28.3178	1159.11	0.000
Total Distance	1	56.570	56.5700	2315.54	0.000
Boat speed	1	2.170	2.1702	88.83	0.000
Error	375	9.161	0.0244		
Total	377	65.797			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.156303	86.08%	86.00%	85.84%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.4946	0.00804	160.39	0.000	
Total Distance	0.4059	0.00816	48.12	0.000	1.03
Boat speed	-0.2376	0.00816	-9.43	0.000	1.03

Regression Equation

Travelling Time = 0.4946 + 0.4059 Total Distance -0.2376 Boat speed

2.3.2.2 Regression with Lasso Variable Selection Analysis: Boat B, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	73.5007	36.7504	965.89	0.000
Total Distance	1	71.5001	71.5001	1879.20	0.000
Boat Speed	1	0.6722	0.6722	17.67	0.000
Error	389	14.8007	0.0380		
Total	391	88.3014			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.195059	83.24%	83.15%	82.99%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.1937	0.00985	136.26	0.000	
Total Distance	0.5265	0.00989	43.35	0.000	1.00
Boat Speed	-0.1071	0.00989	-4.20	0.000	1.00

Regression Equation

Travelling Time = 0.1937 + 0.5265 Total Distance -0.1071 Boat Speed

2.3.2.3 Regression with Lasso Variable Selection Analysis: Combined boat, 1st order

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	129.249	43.083	1299.79	0.000

Boat Code	1	3.590	3.590	108.30	0.000
Total Distance	1	126.690	126.690	3822.14	0.000
Boat speed	1	1.660	1.660	50.09	0.000
Error	766	25.390	0.033		
Total	769	154.639			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.182061	83.58%	83.52%	83.41%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.31641	0.00656	200.64	0.000	
Boat Code	-0.1735	0.00834	-10.41	0.000	1.61
Total Distance	0.5938	0.00658	61.82	0.000	1.00
Boat speed	-0.0106	0.00833	-7.08	0.000	1.61

Regression Equation

Travelling Time = 0.3059 -0.1735 Boat Code + 0.5938 Total Distance -0.0106 Boat speed

Appendix 3 Model Selection results

Table 26 Lasso Best Tuning Parameter Analysis for Short-term Prediction

Short-term	**Boat A**		**Boat B**		**Combined boat**	
Lambda	**1st order (%)**	**1st order with Interaction (%)**	**1st order (%)**	**1st order with Interaction (%)**	**1st order (%)**	**1st order with Interaction (%)**
0.00	7.8992	8.4287	10.1276	11.4825	8.9545	9.7595
0.05	6.8187	7.2029	9.2873	10.5140	7.7930	8.4645
0.10	6.3075	6.5038	8.8818	9.9873	7.0872	7.1930
0.15	6.0463	5.8362	8.4764	9.0932	6.9833	6.9383
0.20	5.9685	6.6417	8.0709	8.2893	7.3888	7.5438
0.25	6.3740	7.0691	7.8654	8.0899	7.8162	7.9712
0.30	6.8014	7.2295	7.6600	8.6954	7.9766	8.1316
0.35	6.9618	7.3676	7.8622	9.1228	8.1147	8.2697
0.40	7.0999	7.4890	8.2677	9.2832	8.2361	8.3911
0.45	7.2213	7.5972	8.6951	9.4213	8.3443	8.4993
0.50	7.3295	7.6948	8.8555	9.5427	8.4419	8.5969
0.55	7.4271	7.7838	8.9936	9.6509	8.5309	8.6859
0.60	7.5161	7.8654	9.1150	9.7485	8.6125	8.7675
0.65	7.5977	7.9409	9.2232	9.8375	8.6880	8.8430
0.70	7.6732	8.0112	9.3208	9.9191	8.7583	8.9133
0.75	7.7435	8.0767	9.4098	9.9946	8.8238	8.9788
0.80	7.8090	8.1383	9.4914	10.0649	8.8854	9.0404
0.85	7.8706	8.1963	9.5669	10.1304	8.9434	9.0984
0.90	7.9286	8.2511	9.6372	10.1920	8.9982	9.1532
0.95	7.9834	8.3031	9.7027	10.2500	9.0502	9.2052
1.00	8.0354	8.3525	9.7643	10.3048	9.0996	9.2546

Table 27 Lasso Best Tuning Parameter Analysis for Long-term Prediction

Long-term	**Boat A**		**Boat B**		**Combined boat**	
Lambda	**1st order (%)**	**1st order with Interaction (%)**	**1st order (%)**	**1st order with Interaction (%)**	**1st order (%)**	**1st order with Interaction (%)**
0.00	10.5076	10.6863	13.1072	13.9716	11.8619	12.1205
0.05	10.0289	10.1969	12.2875	12.7919	11.0476	11.8935
0.10	9.8672	9.9568	11.4768	11.8488	10.4366	10.5179
0.15	9.4664	9.6346	10.8568	11.0272	10.5954	10.8234
0.20	9.0356	9.1356	10.5462	10.9182	10.7009	11.2508
0.25	9.2659	9.0645	10.9517	11.4237	11.1283	11.4112
0.30	9.6714	9.7700	11.3791	11.8511	11.2887	11.5493
0.35	10.0988	10.1974	11.5395	12.0115	11.4268	11.6707
0.40	10.2592	10.3578	11.6776	12.1496	11.5482	11.7789
0.45	10.3973	10.4959	11.799	12.2710	11.6564	11.8765
0.50	10.5187	10.6173	11.9072	12.3792	11.754	11.9655
0.55	10.6269	10.7255	12.0048	12.4768	11.843	12.0471
0.60	10.7245	10.8231	12.0938	12.5658	11.9246	12.1226
0.65	10.8135	10.9121	12.1754	12.6474	12.0001	12.1929
0.70	10.8951	10.9937	12.2509	12.7229	12.0704	12.2584
0.75	10.9706	11.0692	12.3212	12.7932	12.1359	12.3200
0.80	11.0409	11.1395	12.3867	12.8587	12.1975	12.3780
0.85	11.1064	11.2050	12.4483	12.9203	12.2555	12.4328
0.90	11.1680	11.2666	12.5063	12.9783	12.3103	12.4848
0.95	11.2260	11.3246	12.5611	13.0331	12.3623	12.5342
1.00	11.2808	11.3794	12.6131	13.0851	12.4117	12.5812

Table 28 Lasso Model Coefficient of Boat A for Short-term Prediction

Lambda	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
0	0.4933	-0.2639	0.3679	-0.1412	0.0188	0.0013
0.05	0.4783	-0.2427	0.3485	-0.1300	0.0132	0.0000
0.1	0.4532	-0.2315	0.3391	-0.1165	0.0091	0.0000
0.15	0.4022	-0.2202	0.3296	-0.0503	0.0280	0.0000
0.2	0.3798	-0.1410	0.1453	-0.0278	0.0055	0.0000
0.25	0.3775	-0.1387	0.1430	-0.0135	0.0000	0.0000
0.3	0.3752	-0.1364	0.1408	-0.0046	0.0000	0.0000
0.35	0.3729	-0.1341	0.1385	0.0000	0.0000	0.0000
0.4	0.3706	-0.1318	0.1362	0.0000	0.0000	0.0000
0.45	0.3683	-0.1295	0.1339	0.0000	0.0000	0.0000
0.5	0.3660	-0.1273	0.1316	0.0000	0.0000	0.0000
0.55	0.3638	-0.1250	0.1293	0.0000	0.0000	0.0000
0.6	0.3615	-0.1227	0.1271	0.0000	0.0000	0.0000
0.65	0.3592	-0.1205	0.1248	0.0000	0.0000	0.0000
0.7	0.3570	-0.1182	0.1226	0.0000	0.0000	0.0000
0.75	0.3547	-0.1160	0.1203	0.0000	0.0000	0.0000
0.8	0.3525	-0.1137	0.1181	0.0000	0.0000	0.0000
0.85	0.3503	-0.1115	0.1158	0.0000	0.0000	0.0000
0.9	0.3480	-0.1092	0.1136	0.0000	0.0000	0.0000
0.95	0.3458	-0.1070	0.1114	0.0000	0.0000	0.0000
1	0.3436	-0.1048	0.1091	0.0000	0.0000	0.0000

Table 29 Lasso Model Coefficient of Boat B for Short-term Prediction

Lambda	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
0	0.5646	-0.2109	0.2977	-0.0056	0.1148	0.0001
0.05	0.5421	-0.1883	0.2035	-0.0035	0.0923	0.0000
0.1	0.5150	-0.1771	0.1829	-0.0030	0.0652	0.0000
0.15	0.4834	-0.1635	0.1411	-0.0028	0.0415	0.0000
0.2	0.4471	-0.1577	0.1293	-0.0016	0.0280	0.0000
0.25	0.4267	-0.1505	0.1076	-0.0014	0.0219	0.0000
0.3	0.4094	-0.1356	0.0853	-0.0004	0.0000	0.0000
0.35	0.3979	-0.1333	0.0603	0.0000	0.0000	0.0000
0.4	0.3864	-0.1310	0.0580	0.0000	0.0000	0.0000
0.45	0.3749	-0.1287	0.0558	0.0000	0.0000	0.0000
0.5	0.3633	-0.1265	0.0535	0.0000	0.0000	0.0000
0.55	0.3518	-0.1242	0.0513	0.0000	0.0000	0.0000
0.6	0.3403	-0.1219	0.0490	0.0000	0.0000	0.0000
0.65	0.3287	-0.1196	0.0468	0.0000	0.0000	0.0000
0.7	0.3172	-0.1174	0.0445	0.0000	0.0000	0.0000
0.75	0.3057	-0.1151	0.0423	0.0000	0.0000	0.0000
0.8	0.2941	-0.1128	0.0401	0.0000	0.0000	0.0000
0.85	0.2826	-0.1106	0.0378	0.0000	0.0000	0.0000
0.9	0.2711	-0.1084	0.0356	0.0000	0.0000	0.0000
0.95	0.2596	-0.1061	0.0334	0.0000	0.0000	0.0000
1	0.2480	-0.1039	0.0312	0.0000	0.0000	0.0000

Table 30 Lasso Model Coefficient of Combined boat for Short-term Prediction

Lambda	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
0	0.5075	-0.3361	0.2006	-0.0192	0.0218	0.0016
0.05	0.4436	-0.2161	0.1798	-0.0123	0.0153	0.0000
0.1	0.4137	-0.1661	0.1720	-0.0084	0.0126	0.0000
0.15	0.4049	-0.1106	0.1698	-0.0046	0.0086	0.0000
0.2	0.3545	-0.1050	0.1620	-0.0025	0.0034	0.0000
0.25	0.3419	-0.1038	0.1508	-0.0012	0.0000	0.0000
0.3	0.3293	-0.1027	0.1485	-0.0003	0.0000	0.0000
0.35	0.3180	-0.1015	0.1462	0.0000	0.0000	0.0000
0.4	0.3157	-0.1004	0.1439	0.0000	0.0000	0.0000
0.45	0.3134	-0.0992	0.1416	0.0000	0.0000	0.0000
0.5	0.3111	-0.0981	0.1393	0.0000	0.0000	0.0000
0.55	0.3088	-0.0970	0.1321	0.0000	0.0000	0.0000
0.6	0.3066	-0.0958	0.1298	0.0000	0.0000	0.0000
0.65	0.3043	-0.0947	0.1275	0.0000	0.0000	0.0000
0.7	0.3020	-0.0936	0.1253	0.0000	0.0000	0.0000
0.75	0.2997	-0.0924	0.1230	0.0000	0.0000	0.0000
0.8	0.2975	-0.0913	0.1208	0.0000	0.0000	0.0000
0.85	0.2952	-0.0902	0.1185	0.0000	0.0000	0.0000
0.9	0.2930	-0.0891	0.1163	0.0000	0.0000	0.0000
0.95	0.2907	-0.0880	0.1141	0.0000	0.0000	0.0000
1	0.2885	-0.0869	0.1118	0.0000	0.0000	0.0000

Table 31 Lasso Model Coefficient of Boat A for Long-term Prediction

Lambda	**Total Distance**	**Boat speed**	**Significant Wave Height**	**Wave Direction**	**Wind Speed (E1)**	**Wind Direction (E1)**
0	0.4772	-0.2923	0.0278	-0.0753	0.0109	-0.0142
0.05	0.4502	-0.1976	0.0237	-0.0739	0.0098	-0.0094
0.1	0.4440	-0.1617	0.0229	-0.0735	0.0091	-0.0065
0.15	0.4285	-0.1351	0.0216	-0.0726	0.0086	-0.0043
0.2	0.4131	-0.1185	0.0208	-0.0723	0.0080	-0.0024
0.25	0.3841	-0.0700	0.0181	-0.0683	0.0048	0.0000
0.3	0.3729	-0.0587	0.0123	-0.0463	0.0033	0.0000
0.35	0.3547	-0.0564	0.0094	-0.0367	0.0014	0.0000
0.4	0.3525	-0.0541	0.0062	-0.0218	0.0000	0.0000
0.45	0.3502	-0.0518	0.0043	-0.0142	0.0000	0.0000
0.5	0.3479	-0.0496	0.0029	-0.0045	0.0000	0.0000
0.55	0.3456	-0.0473	0.0005	0.0000	0.0000	0.0000
0.6	0.3434	-0.0450	0.0000	0.0000	0.0000	0.0000
0.65	0.3411	-0.0427	0.0000	0.0000	0.0000	0.0000
0.7	0.3389	-0.0405	0.0000	0.0000	0.0000	0.0000
0.75	0.3366	-0.0382	0.0000	0.0000	0.0000	0.0000
0.8	0.3344	-0.0360	0.0000	0.0000	0.0000	0.0000
0.85	0.3321	-0.0337	0.0000	0.0000	0.0000	0.0000
0.9	0.3299	-0.0315	0.0000	0.0000	0.0000	0.0000
0.95	0.3277	-0.0292	0.0000	0.0000	0.0000	0.0000
1	0.3254	-0.0270	0.0000	0.0000	0.0000	0.0000

Table 32 Lasso Model Coefficient of Boat B for Long-term Prediction

Lambda	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
0	0.6395	-0.2447	0.0416	-0.0325	0.1051	-0.1210
0.05	0.5169	-0.1849	0.0118	-0.0029	0.0000	0.0000
0.1	0.4653	-0.1693	0.0054	0.0000	0.0000	0.0000
0.15	0.4335	-0.1364	0.0000	0.0000	0.0000	0.0000
0.2	0.4053	-0.1185	0.0000	0.0000	0.0000	0.0000
0.25	0.3941	-0.1073	0.0000	0.0000	0.0000	0.0000
0.3	0.3564	-0.1050	0.0000	0.0000	0.0000	0.0000
0.35	0.3452	-0.1027	0.0000	0.0000	0.0000	0.0000
0.4	0.3429	-0.1004	0.0000	0.0000	0.0000	0.0000
0.45	0.3406	-0.0981	0.0000	0.0000	0.0000	0.0000
0.5	0.3383	-0.0958	0.0000	0.0000	0.0000	0.0000
0.55	0.3360	-0.0935	0.0000	0.0000	0.0000	0.0000
0.6	0.3337	-0.0913	0.0000	0.0000	0.0000	0.0000
0.65	0.3314	-0.0890	0.0000	0.0000	0.0000	0.0000
0.7	0.3292	-0.0867	0.0000	0.0000	0.0000	0.0000
0.75	0.3269	-0.0845	0.0000	0.0000	0.0000	0.0000
0.8	0.3246	-0.0822	0.0000	0.0000	0.0000	0.0000
0.85	0.3224	-0.0800	0.0000	0.0000	0.0000	0.0000
0.9	0.3201	-0.0778	0.0000	0.0000	0.0000	0.0000
0.95	0.3179	-0.0755	0.0000	0.0000	0.0000	0.0000
1	0.3156	-0.0733	0.0000	0.0000	0.0000	0.0000

Table 33 Lasso Model Coefficient of Combined boat for Long-term Prediction

Lambda	Total Distance	Boat speed	Significant Wave Height	Wave Direction	Wind Speed (E1)	Wind Direction (E1)
0	0.5084	-0.1609	0.0427	-0.0542	0.0498	-0.0618
0.05	0.4339	-0.1033	0.0338	-0.0506	0.0127	-0.0382
0.1	0.4124	-0.0785	0.0286	-0.0475	0.0000	-0.0184
0.15	0.3956	-0.0384	0.0238	-0.0386	0.0000	-0.0109
0.2	0.3541	-0.0347	0.0206	-0.0205	0.0000	-0.0051
0.25	0.3416	-0.0339	0.0153	-0.0065	0.0000	0.0000
0.3	0.3291	-0.0331	0.0101	0.0000	0.0000	0.0000
0.35	0.3178	-0.0324	0.0063	0.0000	0.0000	0.0000
0.4	0.3155	-0.0316	0.0012	0.0000	0.0000	0.0000
0.45	0.3132	-0.0308	0.0000	0.0000	0.0000	0.0000
0.5	0.3109	-0.0301	0.0000	0.0000	0.0000	0.0000
0.55	0.3087	-0.0293	0.0000	0.0000	0.0000	0.0000
0.6	0.3064	-0.0286	0.0000	0.0000	0.0000	0.0000
0.65	0.3041	-0.0278	0.0000	0.0000	0.0000	0.0000
0.7	0.3018	-0.0271	0.0000	0.0000	0.0000	0.0000
0.75	0.2996	-0.0263	0.0000	0.0000	0.0000	0.0000
0.8	0.2973	-0.0256	0.0000	0.0000	0.0000	0.0000
0.85	0.2951	-0.0248	0.0000	0.0000	0.0000	0.0000
0.9	0.2928	-0.0241	0.0000	0.0000	0.0000	0.0000
0.95	0.2906	-0.0233	0.0000	0.0000	0.0000	0.0000
1	0.2883	-0.0226	0.0000	0.0000	0.0000	0.0000

REFERENCES

1. Banister, D., *The Trilogy of Distance, Speed, and Time.* Journal of Transport Geography, 2011. **19**: p. 950-959.
2. R. Lu, O.T.a.E.B., *Voyage Optimisation: Prediction of Ship Specific Fuel Consumption for Energy Efficient Shipping*, in *Low Carbon Shipping Conference*. 2013.
3. D. Zissis, E.K.X., and D. Lekkas, *Real-time Vessel Behavior Prediction.* Evolving Systems, 2016.
4. A. Mancuso, V.N., A. Saporito, and D. Tumino, *Yacht Performance Monitoring in Real Sailing Conditions.* Ocean Engineering, 2019. **188**.
5. El-Hawary, F., *The Ocean Engineering Handbook*. 2000: CRC Press. Taylor & Francis group.
6. C. H. H. J. J. Myers, a.R.F.M., *Handbook of Ocean and Underwater Engineering*. 1969: United States: N. p.
7. I. L. Y. Dalgic, a.O.T., *Investigation of optimum crew transfer vessel fleet for offshore wind farm maintenance operations.* J. Wind Engineering, 2015. **39**: p. 31-52.
8. Milgram, J.H., *Fluid Mechanics for Sailing Vessel Design (*Annu. Rev. Fluid Mech. 1998.
9. Anderson, B.D., *The Physics of Sailing.* Feature article.
10. *http://thailandupstream.chevron.com/OPL/default.asp*.
11. Marrocu, G.A.a.E., *Forecasting inflation: A comparison of linear Phillips curve models and nonlinear time series models.* 2003.
12. Y. K.-l. W. Tao, G.X.-l., and F. Hui, *Comparative study of ANFIS and ANN applied to freeze-up water temperature forecasting.* J. of Hydraulic Engineering, 2013. **07**.
13. S. L. Q. Wang, a.R.L., *Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques.* J. Energy, 2018. **161**: p. 821-831.
14. Jong, J.D.S.a.G.C., *An international meta-analysis of values of travel time savings.* J. Evaluation and Program Planning, 2009. **32**: p. 315-325.
15. Kleinbaum, C.V.A.a.D.G., *Regression models for ordinal responses: A review of methods and applications.* International Journal of Epifemiology, 1997. **26**: p. 1323-1333.
16. Heragu, B.Y.E.a.S.S., *Simulation based regression analysis for the rack configuration of an autonomous vehicle storage and retrieval system.* Internation Journal of Production Research, 2010. **48**: p. 6257-6274.
17. C. M. B. Keshtegar, a.O.K., *Comparison of four heuristic regression techniques in solar radiation modelling: Kriging method VS RSM, MARS and M5 model tree*. Renewable and Sustainable Energy Reviews, 2018. **81**: p. 330-341.
18. J. A. C. S. Diaz, a.J.M.M., *Comparison of several measure-correlate-predict models using support vector regression techniques to estimate wind power densities. A case study*. J. Energy Conversion and Management, 2017. **140**: p. 334-354.
19. Aggarwal, C.C., *Outlier Analysis*. Springer. 2017.

20. Glen, S. *Cook's distance/ Cook's D: Definition, Interpretation*. Statistics How To 2016.
21. Halloran, S., *Di: Cook's distance for Identifying Influential Cases*.
22. S. Reid, R.T., and J. H. Firedman, *A study of Error Variance Estimation in Lasso Regression*, in *Statistica Sinica*. 2013.
23. Brownlee, J., *A Gentle Introduction to k-fold Cross-validation*, in *Machine Learning Mastery*. 2018.
24. *https://rapidminer.com/blog/validate-models-cross-validation/*.
25. Lee, G.A.F.S.a.A.J., *Linear Regression Analysis*. 2nd ed. Wiley Series in Probability and Statistics. 2003.
26. Efroymson, M.A., *Multiple regression analysis*. Mathematical methods for digital computers. 1960: New York: Wiley.
27. Smith, G., *Step away from stepwise.* J. Big Data, 2018. **5**.
28. Kennard, A.H.a.R., *Ridge Regression: Biased Estimation for Nonorthogonal Problems.* Technometric, 1970.
29. Kashid, A.V.D.a.D.N., *Alternative Method for Choosing Ridge Parameter for Regression.* Applied Mathematical Sciences, 2010. **4**: p. 447-456.
30. A. K. M. E. Saleh, M.A., and B. M. G. Kibria, *Theory of Ridge Regression Estimation with Applications*. Wiley Series in Probability and Statistics. 2019.
31. Fonti, V., *Feature Selection using Lasso.* Research Paper in Business Analytics, 2017.
32. Su, X.Y.a.X.G., *Linear Regression Analysis: Theory and Computing*. World Scientific. 2009.
33. Oleszak, M. *Regularization: Rigde, Lasso, and Elastic net.* https://www.datacamp.com/community/tutorials/tutorial-ridge-lasso-elastic-net.
34. *https://scikitlearn.org/stable/auto_examples/linear_model/plot_lasso_model_selection.html?highlight=lasso%20error*.
35. *https://scikit-learn.org/stable/auto_examples/linear_model/plot_lasso_and_elasticnet.html#sphx-glr-auto-examples-linear-model-plot-lasso-and-elasticnet-py*.
36. Ng, K.S., *A Simple Explanation of Partial Least Squares*. 201-: http://users.cecs.anu.edu.au/~kee/pls.pdf.
37. Gregoria, M.A., *Partial Least Squares (PLS): Origins, Evolution, and Application to Social Sciences.* Communications in Statistics -Theory and Methods, 2011. **40**: p. 2305-2317.
38. Henry, N.W. *R-square and Standardization in Regression*. Linear Models in Social Research, 2001.
39. Frost, J., *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*. 2017.
40. Armstrong, J.S., *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer Sciences + Business Media. 2001.
41. B. G. A. Myttenaere, B.L.G., and F. Rossi, *Mean absolute percentage error for regression models.* J. Neurocomputing, 2016. **192**: p. 38-48.
42. Athanasopoulos, R.J.H.a.G., *Forecasting: Principles and practice*, in *Otexts*. 2018.

43. S. Patrick, B.C., and L. A. Schwarte, *Correlation Coefficients: Appropriate Use and Interpretation.* Anesthesia and Analgesia, 2018. **126**: p. 1763-1768.

VITA

NAME Atiporn Chongterdtoonskul

DATE OF BIRTH 14 December 1985

PLACE OF BIRTH Bangkok

INSTITUTIONS ATTENDED
2018-Present:
- Sasin School of Management, Chulalongkorn University
- Department of Industrial Engineering, Chulalongkorn university

2008-2012: The Petroleum and Petrochemical College, Chulalongkorn university
2004-2008: Thammasat English Program of Engineering, Thammasat University.

HOME ADDRESS 502 Soi Onnuch 36 Suanluang district, Bangkok, 10250

PUBLICATION
1) Effects of Oxide Supports on Ethylene Epoxidation Activity over Ag-based Catalysts, Journal: Molecular Catalysis A: Chemical, February 25, 2012.
2) Ethylene Epoxidation Activity Over Ag-Based Catalysts on Different Nanocrystalline Perovskite Titanate Supports, Journal: Catalysis Letters, June 19, 2012,
3) Comparative Study on the Influence of Second Metals on Ag-loaded Mesoporous SrTiO3 Catalysts for Ethylene Oxide Evolution, Journal: Molecular Catalysis A: Chemical, March 2013.
4) Effect of Diluent Gas on Ethylene Epoxidation Activity over Various Ag-based Catalysts on Selective Oxide Supports, Journal: Molecular Catalysis A: Chemical, February 15, 2014.
5) Offshore Crew boat Sailing Time Forecast using Regression Models, ICIEA 2020 conference's proceeding, 2020.

AWARD RECEIVED The king scholarship, Faculty of Engineering, Thammasat University, 2008.