SHIP HULL RESISTANCE ESTIMATION USING MULTIVARIATE STATISTICS

by

Ahmet Pala

B.S., Naval Architecture and Marine Engineering, İstanbul Technical University, 2017

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University

2020

ACKNOWLEDGEMENTS

First of all, I would like to thank my dear advisor, Assoc. Prof. Wolfgang Hörmann, who gave me all kinds of support throughout my entire study, increased my motivation to learn new topics and broadened my vision. Without his guidance, it was not possible to complete this thesis. Also I would like to thank my thesis jury for their time and valuable comments. I would also like to thank my mother Türkan Pala, my father Vahip Pala, my brother Volkan Pala, my sister Rabia Pala and my cousin Ayaz Yıldırım, who have always been with me throughout this whole process. Moreover, I would like to thank Çiğdem Karademir for her support and assistance in all matters during this difficult process. I would also like to thank Assoc. Prof. Devrim Bülent Danışman and Cihad Celik, who provided all kinds of assistance for the data I will use during my studies. In addition to all these, I would like to thank my company Sanmar Shipyards and planning manager Koray Altay, who always supported me and provided all kinds of convenience throughout my graduate period.

ABSTRACT

SHIP HULL RESISTANCE ESTIMATION USING MULTIVARIATE STATISTICS

The biggest cost in shipbuilding belongs to the main engine and propeller systems to be used. The selection of these equipment to be used is made according to the resistance values of the ship. Since these are very expensive systems, it is critical to determine the resistance values of the ships precisely. For this reason, many studies have been done on estimating the resistance values of ships. The performances of these studies are measured by comparing the estimations with the towing tank test results. Within the scope of this thesis study, the towing tank test reports of a total of 58 different cargo ships are examined in detail and the needed data are extracted from these reports. Then, the hull resistance values of these ships are estimated with different statistical approaches. First, the problem is handled using a generalized linear model, and then predictions are made with artificial neural networks, which have been used many times in the literature. After all, because of the longitudinal nature of the available data, the problem is addressed using mixed effect linear models. Finally, hull resistance values are estimated by using generalized linear mixed models, which are the combination of generalized linear and mixed effect linear models. Comparisons are made among the completed statistical models using the leave one ship out cross validation technique. In addition, the results of these models are compared with the Holtrop & Mennen method, which is widely used in this field.

ÖZET

ÇOK DEĞİŞKENLİ İSTATİSTİKLER KULLANILARAK GEMİ GÖVDE DİRENCİ TAHMİNİ

Gemi üretimindeki en büyük maliyet, kullanılacak olan ana makine ve pervane sistemlerine aittir. Kullanılacak olan bu ekipmanların seçimi, geminin direnç değerlerine göre yapılmaktadır. Bunlar çok pahalı sistemler oldukları için, gemilerin direnç değerlerinin hassas bir şekilde belirlenmesi kritik bir öneme sahiptir. Bu sebeple, gemilerin direnç değerlerini tahmin etme üzerine bir çok çalışma yapılmıştır. Yapılan bu çalışmaların performansları, gemilerin belli bir ölçeğe göre küçültülerek model havuzlarında yapılan deney sonuçları ile kıyaslanarak ölçülmektedir. Bu tez çalışması kapsamında, toplamda 58 farklı kargo gemisinin deney sonuçları incelenmiş ve veriler düzenli bir hale getirildikten sonra farklı istatistiksel yaklaşımlarla bu gemilerin gövde direnç değerleri tahmin edilmiştir. Öncelikle, genelleştirilmiş lineer modeller kullanılarak problem ele alınmış, daha sonrasında ise literatürde de bir çok kez kullanılmış olan yapay sinir ağları ile tahminler yapılmıştır. Ardından, eldeki verilerin boyuna veri yapısına sahip olması sebebiyle, lineer karma modeller kullanılarak problem ele alınmıştır. Son olarak, genelleştirilmiş lineer ve karma lineer modelin bir kombinasyonu olan genelleştirilmiş karma lineer modeller kullanılarak gövde direnç değerleri tahmin edilmiştir. Tamamlanan modeller arasında leave one ship out cross validation tekniği kullanılarak karşılaştırmalar yapılmıştır. Ayrıca, kurulan bu modellerin sonuçları, bu alanda yaygınca kullanılan Holtrop & Mennen yöntemi ile de kıyaslanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$E(.)$	Expectation operator
$f(.)$	Activation function in ANN
$g(.)$	Link function in GLM and GLMM
$g^{-1}(.)$	Inverse link function in GLM and GLMM
u_i	Random regression coefficients for subject i
w_0	Bias term in ANN
w_i	Weights in ANN
X_{ij}	Fixed effect covariate set
$V(.)$	Variance operator
y_{ij}	Response variable for subject i and measurement j
$\hat{y}_{ij}$	Predicted esponse variable for subject i and measurement j
Z_{ij}	Random effect covariate set
β	Fixed regression coefficients
Δ	Total weight of the ship (ton)
η_{ij}	Linear predictor for subject i and measurement j
ϵ_{ij}	Error term for subject i and measurement j
∇	Total volume of the ship under water (m^3)
μ	Mean response

LIST OF ACRONYMS/ABBREVIATIONS

AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BIC	Bayesian Information Criterion
CFD	Computational Fluid Dynamics
CV	Cross Validation
DNN	Deep Neural Network
DWT	Deadweight Tonnage
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
ITTC	International Towing Tank Conference
ITU	Istanbul Technical University
LM	Linear Model
MAE	Mean Absolute Error
MELM	Mixed Effect Linear Model
MSE	Mean Squared Errors
RMSE	Root Mean Squared Errors
RNN	Recurrent Neural Network
SSE	Sum of Squared Errors
TEU	Twenty Foot Equivalent Unit

1. INTRODUCTION

A typical modern ship has several significant points to be considered so that the vessel can operate with minimum cost and maximum efficiency. One of the most important phenomena for the design of a ship is resistance since it directly influences the hull-form, selection of propeller and main engine which are the most expensive parts of a ship. Resistance is the power in the opposite direction of the movement while a ship is moving through in calm water. The resistance of a vessel can be divided into two main components; Viscous Resistance *(R_V)* resulting from the viscosity of the fluid and Wave Resistance *(R_W)* caused by the movement in water also known as the Wave Making Resistance according to Hughes hypothesis [2]. There are also appendages used on the hull surface of vessels causing another type of resistance, Appendage Resistance *(R_{APP})*. Instead of estimating resistance components separately, this study uses multivariate statistics for directly forecasting the ship hull resistance which is sum of all three components.

Ship resistance estimation is one of the most important parts of the design process for Naval Architect and Marine Engineers because the most crucial and expensive parts of a vessel such as the main engine are directly dependent on the resistance. The resistance affects directly the fuel consumption of a ship; therefore, the most suitable propulsion systems should be used for maximizing efficiency and reducing the fuel consumption. At this point, since the propulsion systems are the most expensive part of a vessel, the estimation of the ship resistance must be accurate as much as possible. There are three ways to calculate ship resistance; empirical formulas, Computational Fluid Dynamics (CFD) and towing tank tests. The first one is generally preferred at the preliminary design phases for the purpose of having an approximate idea about the ship resistance. This regression technique used to measure the resistance of ships is very fast and practical; however, estimated resistance values from empirical formulas are not considered to be the precise predictions of the ship resistance generally. The most common example for this type of resistance estimation technique is the Holtrop and

Mennen method which used 334 set of model test to build up its empirical formula [3]. On the other hand, CFD takes a lot of time to estimate ship resistance since this method requires detailed modeling of the vessel and calculation of the interaction of the fluid with the ship hull. This method is more reliable than the results obtained from empirical formulas. Beside all these, the most reliable and accurate values for ship resistance are obtained from towing tank tests. The towing tank is used to measure the hydrodynamic performance of ships and marine structures. A carriage for testing the ship resistance of a scaled ship model moves along the towing tank and measures the model resistance. Towing tank tests require a certain period of time and are also more costly than other methods. However, this method gives the most accurate estimations for ship resistance and they are acceptable for classification societies which establish and maintain technical standards for the construction and operation of ships and offshore structures.

Instead of using traditional techniques for estimating ship resistance, this thesis aims at developing regression models for estimating the hull resistance of cargo ships including bulk carrier, tanker, general cargo and container ships using multivariate statistical techniques. Firstly, generalized linear models, the improved versions of the standard linear model, are used for estimating the hull resistance. After that, since there are numerous studies on estimating the ship resistance, artificial neural networks are also used for the same purpose. Additionally, since the data used in this kind of studies have the structure of repeated measurement, mixed effect linear models are also tried. Finally, some trials are performed for estimating the ship hull resistance values of the cargo ships using the generalized linear mixed models, which can be accepted as the combination of generalized linear and mixed effect linear models. Since the expression of hull geometry and the complicated structure of the fluid make the model very complex, these machine learning methods were tried to deal with this problem thanks to their very powerful learning features.

2. LITERATURE OVERVIEW

One of the most important phenomena for the design of a ship is resistance since it directly influences the hull form, selection of propeller and main engine which are the most expensive parts of a ship. Resistance is the power in the opposite direction of the movement while a ship is moving through in calm water. The resistance of a vessel can be divided into two main components; Viscous Resistance *(R_V)* resulting from the viscosity of the fluid and Wave Resistance *(R_W)* caused by the movement in water also known as the Wave Making Resistance according to Hughes hypothesis [2]. There are also Appendages used on the hull surface of vessels causing another type of resistance, Appendage Resistance *(R_{APP})*. Instead of estimating resistance components separately, this study uses multivariate statistics for directly forecasting the ship hull resistance which is sum of all three components. Since the ship resistance is one of the most critical part of the shipbuilding process, there are numerous studies on this field to estimate accurately the resistance as Holtrop and Mennen did [4]. In their study, they used 334 set of hulls' resistance experiment results in order to build a statistical simple model for resistance prediction. Their approach on this field is a well known method for estimating ship resistance especially in the preliminary design stage. Holtrop and Mennen created a method for ship resistance estimation developed through a regression analysis of random model experiments and full scale data obtained from Netherland Ship Model Basin. They mainly divide the ship total resistance into several components and created statistical models to predict them separately. After all, in their study, they reported the totals of these components, which they estimated separately, as the estimated total resistance value. In most of the studies after this study, the Holtrop and Mennen method is considered as a base model and comparisons are reported with this method. Although this method provides a very useful foresight about ship resistance, the results of the Holtrop and Mennen method are not accepted as final accurate ship resistance values at all. The reason for this is that this method used cannot make extremely accurate predictions under all conditions. Hollenbach did one of the important works on this field, and as a result of this study he proposed a

standard formula using the basic values of the ships, just like the method of Holtrop and Mennen [5]. In addition to statistical approach, there are also CFD studies on this field to estimate ship total resistance as Abdelkhalek et al. did in 2014 [6]. In their study, they model a 35000 DWT tanker ship in computer environment and estimated the resistance values at different speeds with CFD. In this method, modeling the ship hull form in computer environment and taking the results from analyzes take a lot of time.

In order to obtain more accurate results faster, Mason et al. [7] used an artificial neural network with one hidden layer and 15 hidden neurons for estimating ship hull resistance more accurately since ANN can deal with very complex nonlinear problems. They stated that ANN is very suitable for this kind of problems. Furthermore, Grabowska and Szczuko [8] also tried to deal with ship resistance prediction using artificial neural network and state that ANN perform well on the ship resistance estimation problem. According to their study, they think that in the future ship resistance estimation problem can be dealt with just statistical models such as ANN instead of towing tank tests thanks to the developments in machine learning methods. There are some suggestions about the number of hidden layers and neurons in an artificial neural network model; however, Stinchcombe and White [9] claim that one hidden layer is enough for a well performing ANN model even for complicated nonlinear models. There is a linear relation between number of hidden layers and the complexity of the ANN model; therefore, increasing number of hidden layers may lead to overfitting. This means that if a model includes a lot of covariates, ANN need to use many number of hidden neurons to learn the model well, which leads to overfitting in general. The standard training algorithm for the weights in ANN is traditional weight backpropagation which updates the numerical values of weights until the desired conditions of the model are satisfied; however, a better way for weight update, resilient backpropagation without weight backtracking was suggested by Riedmiller and Braun in 1993 [10]. One year later, Riedmiller [11] offered resilient backpropagation with weight backtracking technique in ANN since he states that it is more efficient and less time consuming for the training phase. He also says that there is a risk of missing local minimum point for

the error, and if resilient backpropagation with weight backtracking is applied, weights with a smaller total error value can be obtained.

Instead of using artificial neural networks, Skupien and Prokopowicz preferred multiple regression analysis after principle component analysis for calculating the resistance push boats and pushed barge trains using 194 measuring points [12]. In addition to the studies for estimating ship resistance with artificial neural networks using the towing tank test results, there are also some studies based on the operating ship data including wind and wave information. As an example, Wang et al [13] studied the estimation of ship hull resistance with LASSO regression using weather data in addition to main parameters of ship such as waterline length, draught and displacement. Moreover, Petersen and Jacobsen [14] developed two statistical models, Gaussian Process (GP) and ANN, for ship propulsion efficiency of a domestic ferry in 2012.

It is obvious that ship resistance estimation studies using a machine learning technique are carried out in different ways. However, some studies use less parameters than should be because addition of a new parameter increases the complexity of the model. Consequently, higher learning parameters need to be used in order to establish a proper model, which also increases the risk of overfitting. Therefore, before constructing such a problem, all possible parameters can be tried and eliminated with some particular techniques such as forward feature selection algorithm as used in this study. Dimension reduction may harm the regression models since there are nonlinear relationships between response and independent variables in this particular problem. Moreover, normalization works will also affect the model efficiency in a positive way in terms of preventing high difference between regression parameters and overfitting.

3. PROBLEM DEFINITION

3.1. Data Summary

The data used in this study are collected from Ata Nutku Ship Model Testing Laboratory of Istanbul Technical University (ITU), which consist of experimental results of 58 different cargo ship models including bulk carrier, tanker, general cargo and container ships.

Table 3.1. Cargo ship data summary

Ship Type	**Number of Hulls**	**Total Experiments**
Bulk Carrier	4	133
Container Ship	9	330
General Cargo Ship	13	428
Tanker	32	1086
Total	**58**	**1977**

In this laboratory ships are scaled according to particular ratios, λ, and ship models are produced for measuring the resistance with the help of towing tank systems. The towing tank in Ata Nutku Ship Model Testing Laboratory has 160 m length, 6 m width and 3.5 m depth. The carriage in this laboratory can take measurements up to 5.5 m/s speeds [15].

The reports attained from this laboratory include the main hydrostatic characteristics of the tested ship models as well as their speeds. For each ship, experiments are carried out for different loading conditions such as “Heavy Loaded”, “Design”, and “Ballast”. The “Heavy Loaded” is the case where the ship carries the maximum load and the “Design” is the case where it was designed to carry, only ballast water. On the other hand, the “Ballast” is the case where the ship has no cargo. For each case, results were obtained accordingly at different speeds. All cargo ship reports were examined in

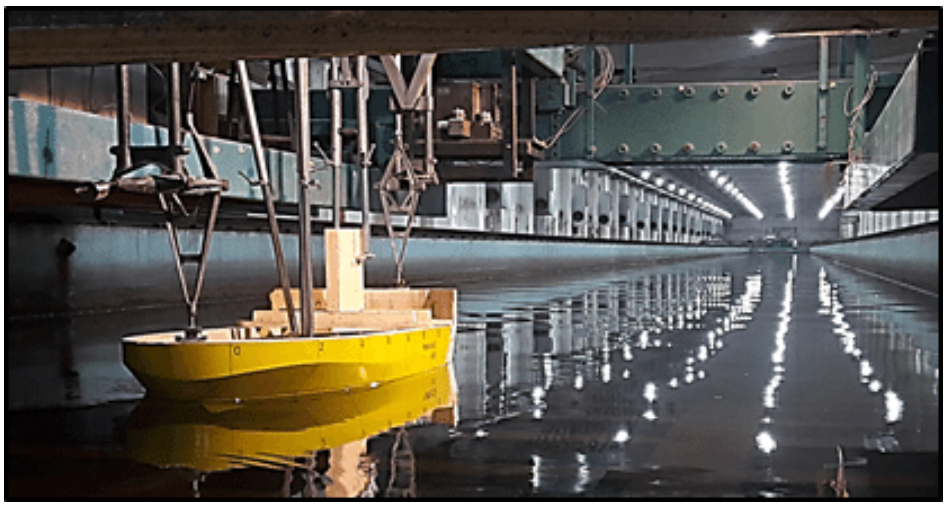

Figure 3.1. An example view from Ata Nutku Ship Model Testing Laboratory of Istanbul Technical University (ITU)

detail and the following features were extracted from each vessel model report:

- *Ship Type:* Type of the vessel.
- *Loading Condition:* Ship's loading condition for the corresponding experiment.
- *Length Overall (L_{OA}):* The overall length of the ship (m).
- *Length Between Perpendiculars (L_{BP}):* Length between the fore-side of the stem and the after side of the rudder post (m).
- *Waterline Length (L_{WL}):* The length of ship where it lay on the water surface (m).
- *Waterline Breadth (B_{WL}):* The breadth of ship where it lay on the water surface (m).
- *Draught (T_s):* The total height under water at the center of ship (m).
- *Draught AP (T_A):* The total height under water at the zeroth frame (aft) of ship (m).
- *Draught FP (T_F):* The total height under water at the last frame (forward) of ship (m).
- *Displacement Volume (∇):* Total volume of the ship under water (m^3).
- *Displacement (Δ):* Total weight of the ship (ton).

- *Wetted Surface (A_{WS}):* Total are of the ship hull under water (m^2).
- *Total Rudder Area (A_R):* Total rudder area of the ship (m^2).
- *Total Appendage Area (A_A):* Total appendage area such as bilge keel of the ship(m^2).
- *Bulb Section Area (A_B):* If any, bulbous bow section area of the ship (m^2).
- *Bulb Section Area Center (H_B):* If any, bulb section area center of the ship (m).
- *Transom Area (A_T):* If any, total transom area of the ship (m^2).
- *Transom Area Center (H_T):* If any, transom area center of the ship (m).
- *Block Coefficient (C_B):* The ratio of the volume of displacement to the rectangular which has the same length, width and draught with that ship (dimensionless).
- *Prismatic Coefficient (C_P):* The ratio of the volume of displacement to the volume of a prism having the same length and the sectional area as the ship (dimensionless).
- *Midship Section Coefficient (C_M):* The ratio of the mid-ship section of ship to a the plane with same breadth and draught as the ship (dimensionless).
- *Waterplane Area Coefficient (C_{WP}):* The ratio of the water-plane of the ship to the plane with the same length and breadth as the ship (dimensionless).
- *Longitudinal Center of Buoyancy (L_{CB}):* Longitudinal center of underwater volume from the aft perpendicular (m).
- *Longitudinal Center of Flotation (L_{CF}):* Geometric center of the ship's waterline plane from the aft perpendicular (m). The ship trims about this point.
- *Service Speed (V_s):* Service speed of the model during the experiment (m/s).
- *Temperature:* If stated, environment temperature during the experiment (Celcius) . This information is not available in all reports.
- *Water Density (ρ):* If stated, density of water in the towing tank (kg/m^3). This information is not available in all reports.
- *Form Factor (1+k):* If stated, the parameter that changes depend on the ship hull type (dimensionless). This value is same for full scale ship and the model. In some of the reports, this value was calculated according to an empirical formula, Prohaska Method [7]. However, this information was not stated in other reports.

In addition to all these characteristics, there are also experiment results:

- *Froude Number (Fr):* the ratio of a characteristic velocity to a gravitational wave velocity (dimensionless) which is calculated as $Fr = V/gL$.
- *Frictional Resistance Coefficient (C_f):* Resistance because of the friction of ship hull (dimensionless).
- *Viscous Resistance Coefficient (C_v):* Resistance because of the viscosity of the fluid (dimensionless).
- *Wave Resistance Coefficient (C_r):* Resistance because of the produces waves by the ship (dimensionless).
- *Total Resistance Coefficient (C_t):* Sum of the frictional, viscous and wave resistance coefficients (dimensionless). This coefficient will be used for the calculation of the total resistance.
- *Total Resistance (R_m):* Total resistance of the ship model (Newton).
- *Total Percentage of Appendage Resistance (Rm_{APP}):* Is stated, total percentage of appendage resistance among total resistance value (%). This information is not available in all reports.

In addition to the hydrostatic information given in reports, following parameters can be used in the statistical model since they affect the ship hull form and total resistance. Number of propeller chance the aft form of the ship because ships have as many skegs as propellers. Moreover, bulbs on ships cause a decrease in total resistance.

- *Number Of Propeller:* Total number of propeller in the ship. This information can be important since the propeller output changes the structure of the ship geometry. In general, ships have one propeller; however, some ships have two propellers.
- *Bulb:* It is a binary variable states whether there is a bulbous bow or not. It is a very critical information since bulbous bow creates a significant decrease in resistance.

This study aims to predict the hull total resistance of the cargo vessels, appendage resistance is not included since the appendage information is not available in all measurement results. Resistance is not divided into sub components, the statistical models are built for estimating directly the hull resistance of cargo ships using available information given in reports.

3.2. Holtrop and Mennen Studies

Holtrop and Mennen offered a statistical method using 334 different hulls for estimating ship resistance including also appendage resistance [3]. This method is widely used in Naval Architecture and Marine Engineering for getting an initial estimate for hull resistance, however, this method is not acceptable for final prediction of ship resistance. For the exact estimation for the ship resistance, international organizations approve towing tank test or CFD results.

Four years later, they developed this statistical approach by extending the usage area of the estimation techniques. In their second study [4], they also focused on making better estimations for ships with low L/B ratios and slender naval ships with complex appendage arrangement which are not included in this thesis. In this thesis, the recent study in [4] was used as a base method for comparison since it is widely used in this field.

In their study, Holtrop and Mennen [4] created a method for ship resistance estimation developed through a regression analysis of random model experiments and full scale data consisting of 334 hulls obtained from Netherland Ship Model Basin. They divided the ship total resistance into sub components as given in Equation 3.1.

$$R_{TOTAL} = R_F(1 + k_1) + R_{APP} + R_W + R_B + R_{TR} + R_A \tag{3.1}$$

In this expression:

- R_F: Frictional resistance according to the ITTC-1957 friction formula
- $1 + k_1$: Form factor describing the viscous resistance of the hull form in relation to R_F
- R_{APP}: Resistance of appendages
- R_W: Wave-making and wave-breaking resistance
- R_B: Additional pressure resistance of bulbous bow near the water surface
- R_{TR}: Additional pressure resistance of immersed transom stern
- R_A: Model - Ship correlation resistance

After dividing ship resistance into these parts, they developed many regression models and offered formulations for calculating each component separately. At the final stage, they obtain the ship total resistance value by adding up all the calculated values. For the each particular component, they classified the vessels based on their significant features such as L/B ratio. Then they built regression models and determine the needed regression parameters. For creating detailed prediction formulas, they also derived some coefficients which are only used for inter processes. For more detailed information, the reader may refer to the paper of Holtrop and Mennen for background of the aforementioned processes [4].

For this thesis study, Holtrop & Mennen Method is used as a benchmark method. The original cargo ship data extracted from experimental reports were used for obtaining the Holtrop & Mennen results. For this calculation, information and formulas given in the article [4] was used for only ship hull resistance prediction, appendages resistances were not included to the calculations. In order to check the accuracy of the results, some example ship models are also tested with a widely used software, Maxsurf, which is used for estimating ship resistance according to some empirical formulas such as Holtrop and Mennen method. For this purpose, ship models are modeled in 3D environment and used in this software for obtaining the results. For example, the ship model numbered 363 is modeled and used in Maxsurf. Then the hull resistance

estimations are compared with the results of calculations based on the formulas given in the article. The mean absolute relative error for this ship model (design and ballast loading conditions) is obtained as 0.052, which shows that the calculations are quite acceptable. The estimations of this ship model and the Maxsurf results are given in Figure 3.2 and Figure 3.3.

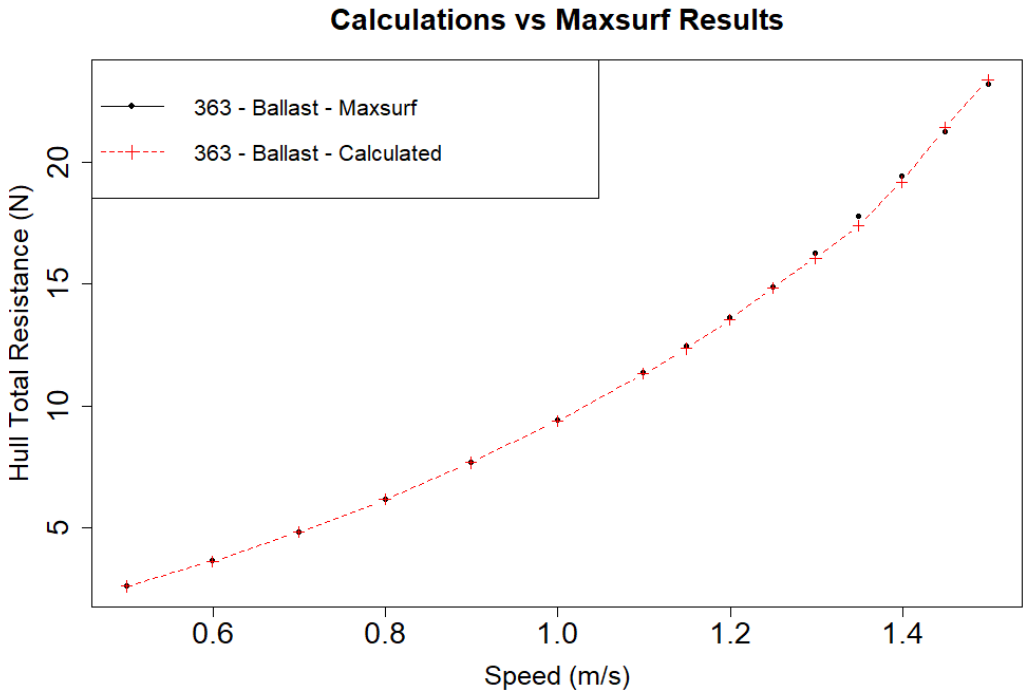

Figure 3.2. Calculated vs Maxsurf results of ship model 363, ballast loading condition

After concluded that the estimations using the formulas given in the article are accurate enough, the hull resistance values obtained as a result of this study were used as a base model to measure the performance of the created statistical models; GLM, MELM, ANN and GLMM.

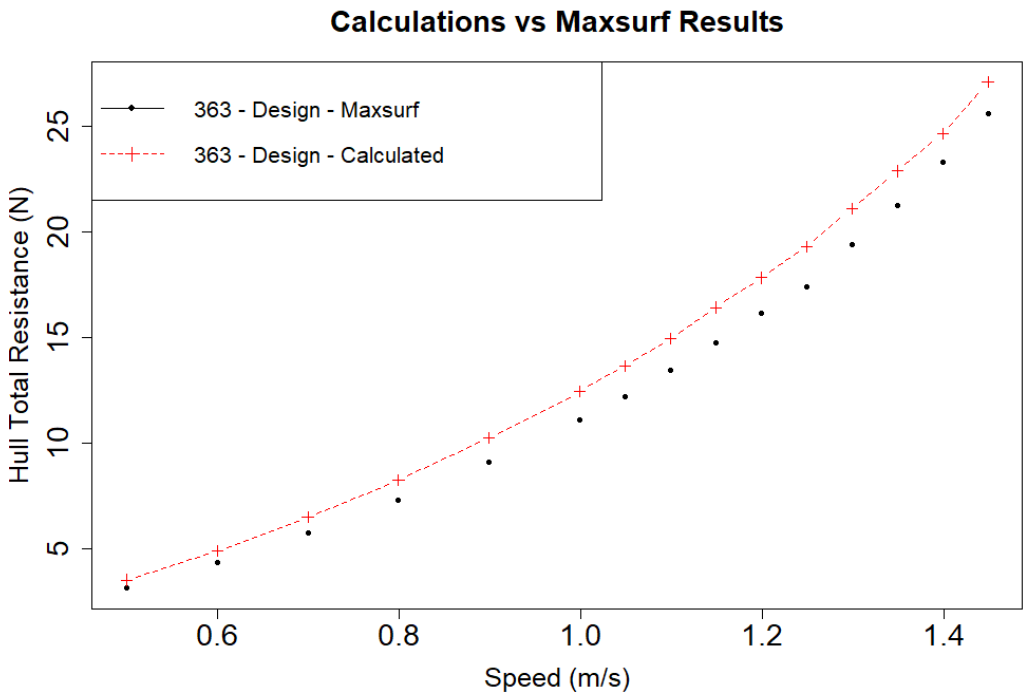

Figure 3.3. Calculated vs Maxsurf results of ship model 363, design loading condition

4. STATISTICAL METHODS

4.1. Generalized Linear Model

For the case of linear models (LM), it is assumed that the error term is distributed based on normal distribution around zero and the variance of the residuals is constant [16]. However, in some cases it may be necessary to relax these assumptions in order to establish statistically better prediction models. At this point, Nelder and Wedderburn formulated GLM for the purpose of standardizing some statistical approaches such as logistic and Poisson regressions [17]. Generalized Linear Models (GLM) are used for estimating response variables in particular cases. In GLM, response variables can follow some special distributions such as Poisson, Binomial, Gamma and so on. The standard formulation for expectation of dependent variable (Y) is given in Equation 4.1.

$$E(Y) = \mu = g^{-1}(X\beta) \tag{4.1}$$

In Equation 4.1, *E(Y)* is the expectation, μ is the mean response dependent on independent covariates *(X)* and β, the parameters of the regression model, which are generally predicted with maximum likelihood or Bayesian techniques. The function *g(.)*, in another saying "link function", is fixed and known. It is also assumed that the link function is smooth and monotone. A well known and widely used example of link function as used in this thesis study is logarithm which makes the expectation:

$$E(Y) = \mu = g^{-1}(X\beta) = e^{(X\beta)} \tag{4.2}$$

In this expression, the *exp(Xβ)* is the $X\beta$ th power of the Euler's Number (e). In generalized linear models, it is not assumed that the independent variables (X) have linear relationship with dependent variables (Y). Instead, it is assumed that the dependent variables have linear relationship with transformed response variables obtained using the inverse form of the link function. Moreover, the residuals are not needed to be

normally distributed. Dependent variables in GLM are distributed based on the distribution of GLM family such as Gamma distribution. For the variance calculation, the distribution of the dependent variable matters. For example, if there is a generalized linear model with Poisson family, it is known that the observations y_i are Poisson random variates. Therefore, the variance calculation is made according to the Poisson distribution and its formula is given in Equation 4.3.

$$Var(Y_i) = V(\mu) = V(g^{-1}(X\beta)) = \mu \tag{4.3}$$

Since the expected value and the variance are equal to the rate parameter λ (μ) in Poisson distribution, they both equal to the rate parameter in Equation 4.3. Moreover, $V(\mu)$ represents the variance function.

4.2. Artificial Neural Network

Artificial neural networks are more advantageous compared to other statistical methods (eg. mixed effect linear model and generalized linear model) since it is not necessary to specify the relationship between variables before building the model, but it is very difficult to interpret the significance of variables after building the model [18]. In this sense, it has advantages and disadvantages compared to other statistical methods. Artificial neural networks are inspired by the biological neural networks in nervous system and they are used for dealing with complex problems having deeply complicated relations between covariates and response variables [19]. In a typical human brain, there are billions of neurons, the processors, connected with a great number of synapses. In such a system, signals are transmitted via these synapses and human brain can memorize the past information, in another saying, human brain learns thanks to these actions. The synapses take signals from neurons processing the information and transfer them to the others, and an input information is turned into an output information after it is processed in a neuron.

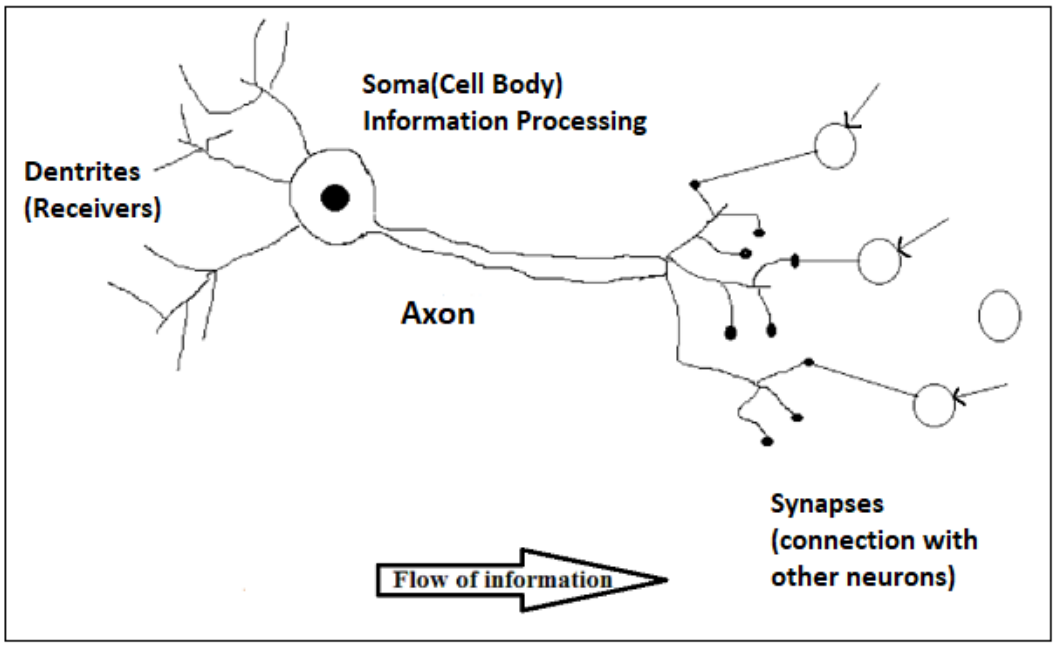

Figure 4.1. A typical neuron structure with synapses in human brain

Based on the working principle of the nervous system, in an artificial neural network (ANN), there are interconnected neurons representing by circles as in Figure 4.2 and they receive numerical data from other neurons. These numerical values are multiplied by the weights and added up in the next neurons. These summations are used in predetermined activation functions and the output values of the neurons are calculated until the output neuron [20]. The basic building block of a neural network is the perceptron. A simple perceptron consists of input neurons and an output neuron. The numerical values from the input neurons are multiplied by their respective weights and these multiplied values are added up with also bias value in output neuron. Then, the result from this addition is put into the activation function to get the final result. A typical example of a perceptron is given in Figure 4.2.

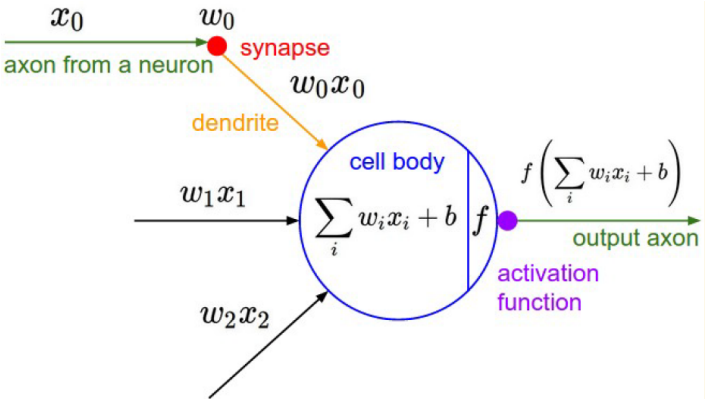

Figure 4.2. A typical perceptron [1]

In Figure 4.2, the terms x_1 and x_2 represent the input values while w_1 and w_2 indicate the respective weights of the input variables. The term w_0 stands for the bias term. In artificial neural networks, all indexed w values are defined as parameters of the ANN. The input values are multiplied by their respective weights, added up and become an output value after transformed by the activation function represented by f in the figure. In the artificial neural networks consisting of this kind of perceptrons, information is transmitted in only one direction, which is called as feed-forward neural networks. In addition to the feed-forward neural networks allowing the information transfer through the direction from input to output layer, there are recurrent neural networks (RNN) that let information circulate not only in one direction but across the network [21]. However, feed-forward neural networks are preferred in this thesis study to deal with the ship hull resistance estimation problem since it is the standard approach of the ANN. A typical neural network consists of an input layer containing input values, an output layer, and a hidden layer located between them. For the multilayered artificial neural networks, in another saying deep neural networks (DNN), there are more than one hidden layer between input and output layers [1].

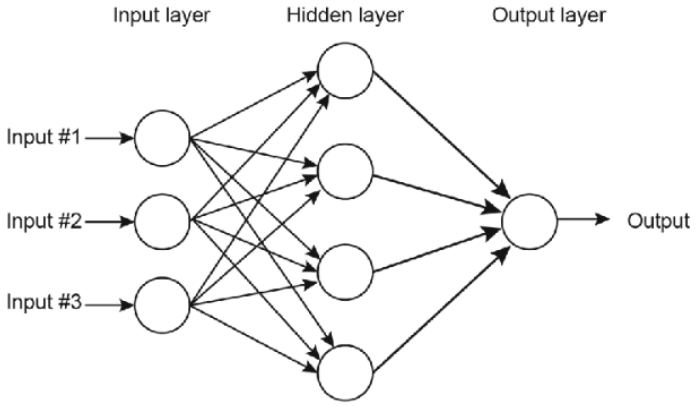

Figure 4.3. An example for feed-forward neural networks with one hidden layer

In a typical artificial neural network, there are three main layers consisting of one input, output and hidden layer. For the regression cases with only one response variables, there is only one hidden neuron in the output layer. On the other hand, for the classification problem, there can be more than one hidden neuron, one for each class in the output layer. These neurons for each class can take 1 or 0 for representing whether the corresponding observation stands for the related class. The number of neurons in the input layer is as much as the number of variables in the relevant model, so this number of neurons is directly related to the selected input variables. On the other hand, the number of neurons in hidden layers is determined by the user, and the number of neurons here affects the complexity of the model. As the number of neurons here increases, the number of parameters in the model will increase, so the model becomes more complex [18].

Activation functions that convert the weighted sums of the input values into the new outputs have different types in artificial neural networks. Thanks to these non-decreasing and differentiable functions, artificial neural networks have a nonlinear structure. Activation functions should be selected according to the structure of the output values as in GLM. For example, if a data set with binary response variable is used, the logistic function can be preferred in the ANN as the activation function

because the logistic function maps values between 0 and 1. After getting the related values, they can be rounded to 0 or 1 and the final results can be obtained [18]. There are many types of activation functions commonly used in neural networks, some of them are:

- *Linear Function:* The simplest activation function used in ANNs. This function is generally used in the output neuron for the regression problems with continuous response variables as in the ship hull resistance prediction issue.

$$f(x) = x \tag{4.4}$$

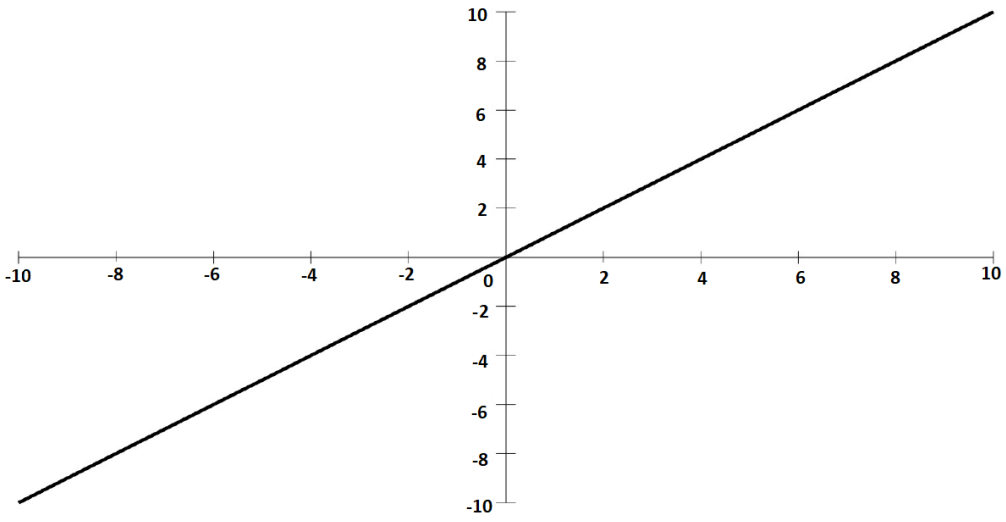

Figure 4.4. Linear function

- *Unit Step Function:* This function equals the values to 1 for positive ones and 0 for negative ones. This is widely preferred to used in binary cases.

$$f(x) = 0 \; when \; x < 0, \; f(x) = 1 \; when \; x \geq 0 \tag{4.5}$$

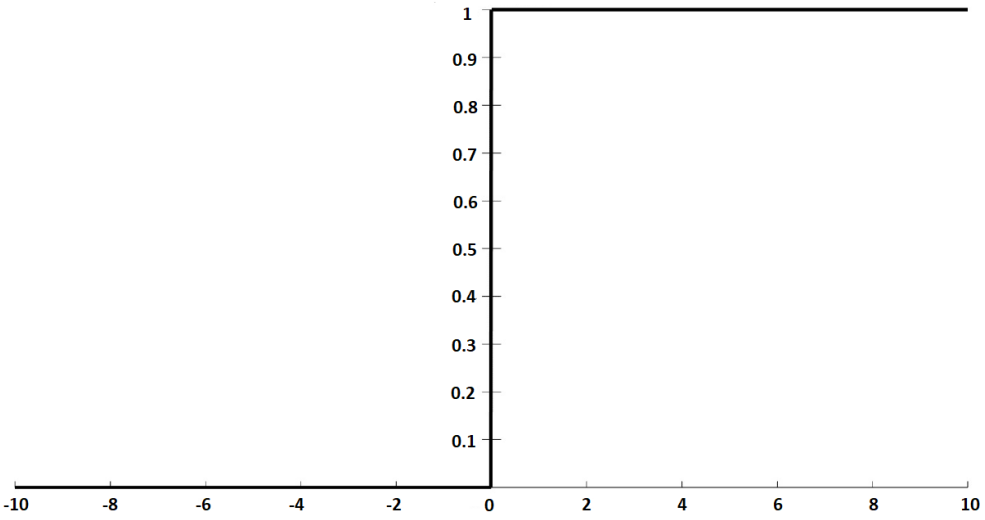

Figure 4.5. Unit step function

- *Sigmoid Function:* The sigmoid function having an S shape in its nature produces the values mapped between 0 and 1. This specific function represents a special form of the logistic function.

$$f(x) = \frac{1}{(1 + e^{-x})} \tag{4.6}$$

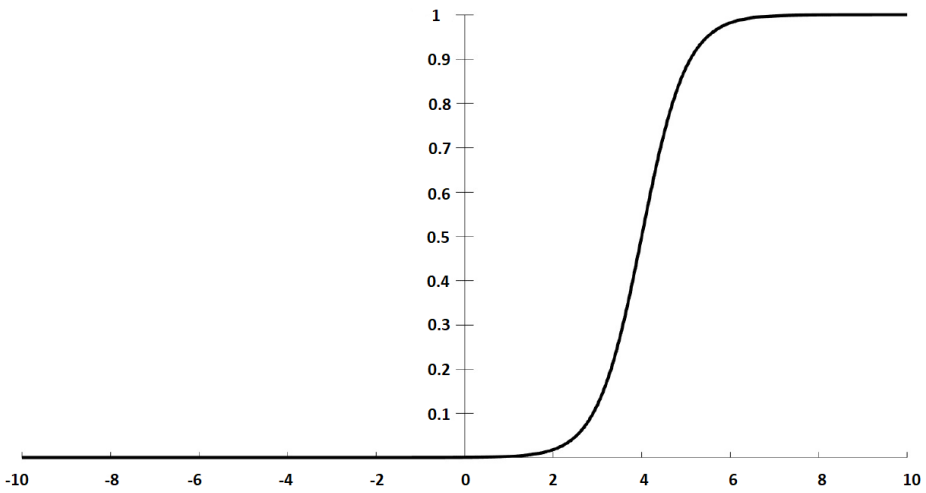

Figure 4.6. Sigmoid function

- *Hyperbolic Tangent (tanh):* This nonlinear function can be accepted as the scaled form of the sigmoid function. Instead of giving results between 0 and 1, hyperbolic tangent function produces outputs between -1 and 1.

$$f(x) = tanh(x) \tag{4.7}$$

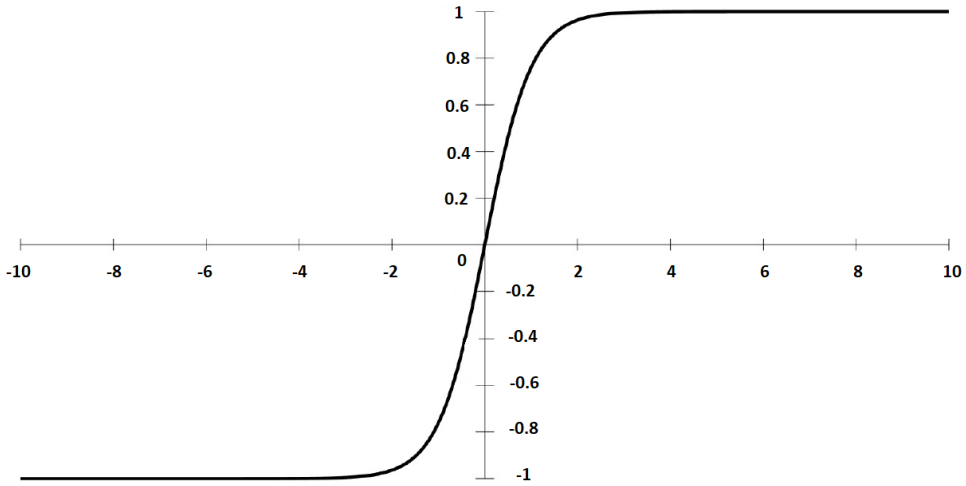

Figure 4.7. Hyperbolic tangent function

Learning procedure consists of two main parts in artificial neural networks; calculating cost function (also called as loss function) measuring how well the predictions are according to the original results and revising the weights as well as the bias terms. These two processes in ANNs are called as forward and backpropagation, respectively. In the forward propagation, the weighted sums are calculated and used as inputs in activation functions in each neuron to transfer to the next layer in the network. After completing the forward propagation, weights are updated in the opposite direction of their partial derivatives in the backpropagation phase. For the deep neural network cases (DNN), there are more than one hidden layer. After reaching the output neuron, cost function is calculated and the error values are obtained. For the cost function in regression models, a particular type of the sum of squared errors (SSE) multiplied by 1/2 is used since taking derivative is easier for this function than the standard SSE

function [22].

$$SSE = \frac{1}{2}\sum(y_i - \hat{y}_i)^2 \tag{4.8}$$

In Equation 4.8, y_i and $\hat{y}_i$ refer to the real and the predicted values, respectively. Based on the calculated error values with cost function, weight and biases are corrected after calculating the derivatives of the activation functions in each neuron so that the model can minimize the total error value. A first order iterative optimization algorithm, the gradient descent, is used to decide how much to change the weight and bias values at each stage. Since the gradient descent algorithm is kind of a heuristic method, the main problem in ANN is to find an acceptable local optimum [23].

4.3. Longitudinal Data Approach

4.3.1. Longitudinal Data

Simple linear models assume that each observation is independent of all other observations for continuous random variables [16]. However in some cases, there are observations that highly dependent on each other such as repeated measurements. The measurements made at certain intervals for each person, for example in hospitals, are also of this type [24]. As in the field of health, such data are available in the financial and economic sectors, and observations are not independent from each other. The general name of this type of data is longitudinal data and different approaches have been developed to establish a statistical model with this kind of data. Longitudinal data consist of repeated measurements on the same sample over time according to Liang and Zeger [25], which means that measurements are carried out for samples more than once while some individual features remain same and some other variables change over measurements such as time or velocity as in ship resistance estimation problem. For longitudinal data, each scalar measurement of a continuous random variable for subject i and measurement j can be symbolized as y_{ij} [26]. As can be understood from this data structure, all response variables are indexed in certain measurements and

this information is stored. This information is then used in the statistical model, and unlike standard linear models, each measurement is not considered to be independent of each other.

Various methods have been developed to establish regression problems with longitudinal data. For example, Liang and Zeger [25] developed a special generalized linear model in 1986 and conducted a longitudinal data analysis study. Moreover, Laird and Ware [27] studied on longitudinal data with Gaussian outcome in 1982. For the data set with non-Gausian response variable, there was a study carried out by Ochi and Prencite [28] in which they used probit link function for the binary response variable.

In this particular case of ship resistance, the hydrostatic values of vessel such as draft, width and length remain the same for each loading case, while the speed value varies for each measurement. From this point of view, it can be suggested that this cargo ship data set is a type of longitudinal data since it contains repeated measurement results. For example, for the ship model numbered 409, 15 different speed values were tested in 3 different loading conditions (design, ballast and heavy loaded). The characteristics are all the same except for the speed in each loading condition. For this reason, each loading case for a particular ship can be considered as an individual. Due to this particular feature of this data set, in addition to generalized linear model and artificial neural network, a longitudinal data approach can be used for ship hull resistance estimation. For this purpose, Mixed Effect Linear Model (MELM) and Generalized Linear Mixed Model (GLMM) are tried for dealing with this regression case.

4.3.2. Mixed Effect Linear Model

Since the simple linear model assumes that the observations are independent from each other, there are numerous studies for finding an alternative way to deal with the data with dependent observations, for a particular case, longitudinal data. First of all, the independence assumption must be relaxed and statistical models should be estab-

lished in line with this approach. At this point, mixed effect linear model, in another saying mixed error component model is a very useful technique for building regression models for longitudinal data. In a standard mixed effect linear model, there are random effects that vary on an individual basis as well as fixed effects considered constant for each observation [29]. For typical longitudinal data with repeated measurements in the form of (y_{ij},X_{ij}^T) for j from 1 to n_i and i from 1 to N, where X_{ij}^T represents the covariates for jth scalar measurement of a continuous random variable for subject i. Accordingly, a basic mixed effect linear model can be expressed as follows:

$$y_{ij} = X_{ij}^T\beta + Z_{ij}^T u_i + \epsilon_{ij} \tag{4.9}$$

In Equation 4.9, Z_{ij} indicates a subset of covariate set and has random effects, the term u_i represents the random regression coefficients for ith subject. Finally, ϵ_{ij} represents the error terms that are assumed to follow normal distribution with mean zero [26].

In mixed effect linear models, the terms u_i indicating the random regression coefficients are considered to be independent, and normally distributed with mean zero. There is also another parameter, the variance of the random regression coefficients to be estimated in mixed effect linear models. Note that the likelihood function of the mixed effect linear model is optimized with respect to the random effects. Therefore, estimates of fixed regression coefficients is dependent on the random effects. Moreover, the random regression coefficients u_i and the error term ϵ_{ij} are also considered to be independent from each other [30]. In a mixed effect linear model, both fixed and random regression coefficients are calculated and they are used to predict y for new covariate values. For each individual, random regression coefficients are estimated for the required random effects. For example, if there are 5 different groups with several observations in the data set of a regression problem, and this problem is set up to include a random intercept value on group basis, a total of 5 different random intercept values are estimated, one for each group. If a prediction is made for an individual, for which observations exist, the random regression coefficient u_i estimated for that individual is used for that prediction. However, if a prediction is required for

the response of an individual for which no observations exist in the data, the prediction is made by using only the fixed regression coefficients, since there is no estimated random regression coefficient for this new group and the average of the random effects is zero [31, 32].

In this thesis, R-package "lme4" is used for the mixed effect linear models. Mixed effect linear models are created with the "lmer" function in this package. Within this function, as in known standard linear model function "lm", the response variable and covariates are specified in the function and separated by "∼". However, unlike this standard function, the random regression coefficients can be included by defining them with the aid of the symbol (|) in parentheses. This function estimates model parameters optimizing Restricted Maximum Likelihood (REML) criterion by default. However, the user may prefer to use Maximum Likelihood instead of REML by defining "REML = FALSE" inside of the function. For more detailed information, the reader can view the documentation of the R-package "lme4" [32].

$$melm = lmer(Y \sim X_1 + (1|X_2), REML = TRUE, data) \tag{4.10}$$

In the example given in Equation 4.10, there are 2 covariates for the response variable Y; covariate 1 (X_1) and covariate 2 (X_2). In this example, covariate 1 has the fixed effect, that is, the regression coefficient belonging to this covariate does not vary, like the regression coefficients in standard linear models. However, for covariate 2 (usually a grouping variable) , different intercept values are estimated for each group represented in covariate 2 (X_2). Since the fixed intercept value is already estimated in the model, when the new sample belonging to that group is estimated, the final intercept value is determined by adding these two values. For example, if we build a mixed effect linear model with ship speed (V_s) as fixed effect and random intercept values on the basis of ship type to estimate total hull resistance in a data set containing 4 different ship types (bulk carrier, container ship, general cargo ship and tanker), the summary function gives an output as in Figure 4.8.

```
> summary(lmer_exp)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: Hull_Total_Res ~ Vs + (1 | Ship_Type)
   Data: exp_lmer

REML criterion at convergence: 577

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.5625 -0.6268 -0.0510  0.4776  4.5838

Random effects:
 Groups    Name        Variance Std.Dev.
 Ship_Type (Intercept) 7.674    2.770
 Residual              4.305    2.075
Number of obs: 132, groups:  Ship_Type, 4

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  -0.5505     1.4693   3.5258  -0.375    0.729
Vs           30.4472     1.0054 128.2314  30.282   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
   (Intr)
Vs -0.305
```

Figure 4.8. Example “lmer” output

As can be seen from Figure 4.8, the formulation written into the function is specified in the “Formula” section. Statistical summary of residuals is also provided in the relevant output. In addition, 4 different random intercept values were estimated for 4 different ship types in the model, and their variance and standard deviation are shown in the “Random Effects” section. At the end of this section, total observation and number of different groups (ship type in this particular example) are also given. The estimated fixed regression coefficients and p-values of the variables can be seen in the “Fixed Effects” section. This section is quite similar to the output of the standard linear models in R. In addition to all these, the correlation between variables having fixed effects can be viewed in the “Correlation of Fixed Effects” section.

4.3.3. Generalized Linear Mixed Model

Mixed effect linear models assume that the response variable follow standard normal distribution. This assumption may not be valid for some cases because of

the restriction of the distribution of response variables. At this point, Generalized Linear Mixed Models (GLMM) provide a flexible way to build statistical models for longitudinal data by allowing the response variables to have different distributions [33]. This statistical method, as in the Mixed Effect Linear Model, relaxes the independence assumption between observations by defining fixed and random effects on an individual basis as discussed in Subsection 4.3.2, while also allowing the response variables to be distributed with a specified distribution as in Generalized Linear Models. In essence, they can be accepted as a combination of generalized linear models and mixed effect linear models [34]. In this regard, this statistical method provides various useful models for explaining grouped data including longitudinal data by assigning random effects for each group [35].

As with generalized linear models, there are link functions in generalized linear mixed models and these functions are applied to the response variable. While linear models assume that the responses are normally distributed, generalized models including generalized linear mixed models allow for different response distributions such as Gamma distribution. Therefore, the distribution best suited for the data can be used. In a typical generalized linear mixed model, with continuous response variable y, the linear predictor η is explained by the combination of fixed and random regression coefficients, β and u_i, respectively. These specified regression coefficients are multiplied by the relevant subsets of the covariates in the notation given as follows:

$$\eta_{ij} = X_{ij}^T \beta + Z_{ij}^T u_i \tag{4.11}$$

In Equation 4.11, Z_{ij} and X_{ij} represent the subsets of covariates with random and fixed effects, respectively for ith subject and jth scalar measurement. The linear predictor η given in the previous formula is expressed with the generic link function $g(.)$ and response variable y is given in Equation 4.12.

$$g(E(y_{ij})) = \eta_{ij} \tag{4.12}$$

Additionally, the expectation of y can be expressed as:

$$E(y_{ij}) = g^{-1}(\eta_{ij}) = \mu \tag{4.13}$$

In Equation 4.13, $g(.)^{-1}$ and μ represent the inverse of the link function $g(.)$ and mean value of the continuous response variable y, respectively. Finally, the response variable y can be individually expressed as:

$$y_{ij} = g^{-1}(\eta_{ij}) + \epsilon_{ij} \tag{4.14}$$

In Equation 4.14, the term ϵ_{ij} stands for the error for subject i and jth scalar measurement [36]. Unlike MELMs, the response variables do not have to be normally distributed in generalized linear mixed models. Instead, the distribution of the response variables is specified before the model is created. The appropriate distribution for the available data can be found by comparing the AIC values in models with the same response variables, and the distribution that gives the lowest AIC value among the models can be selected. Random regression coefficients are assumed to have normal distribution with mean zero. The variance of the random effects is a parameter to be estimated in the model [37].

GLMM is used to find appropriate error structures while explaining correlated data unlike the normality assumption of response variables in MELM. To estimate the parameters of GLMM, the likelihood function of the model should be derived on the assumptions. Maximum likelihood and the restricted maximum likelihood methods are used to estimate the unknown parameters by numerical optimization using integrations over random effects. According to Harville [38], the restrictred maximum likelihood method is more reliable in estimation of unbiased variance parameters for some problems. Likelihood functions can become difficult to derive, complicated and high dimensional due to the specified distribution and the size of the random effects. It is difficult to estimate the parameters by numerical search methods for such complex equations. Several numerical integration techniques are used to fit a GLMM by

using numerical approximations. Laplace, numerical quadrature, non-adaptive Gauss-Hermite quadrature, adaptive Gauss-Hermite quadrature, and Bayesian approaches are the methods for approximate integrations to predict GLMM parameters. Laplace method uses great approximations on variables and produces poor estimations. However, it is easiest to solve the problem by Laplace method. Adaptive Gauss-Hermite quadrature is significantly better but restrict the number of random effects in integrations. Bayesian method can be used for more complicated real data but requires extensive computing power used in simulations within the framework. Tuerlinckx et al. discusses these methods in addition to likelihood approaches in a broad survey on GLMM. The reader may refer to this paper for mathematical and statistical background of the aforementioned approaches [38, 39].

In this thesis, R-package "lme4" , which is also used in the mixed effect linear model, is used for the generalized linear mixed models. The "glmer" function in this package estimates the parameters of generalized linear mixed models. Within this function, as in known generalized linear models, the response variable and covariates are specified in the function and separated by $\sim$. The distribution of response variable and the link function are likewise referred to as "family". In contrast, the random effects are specified by separating them with the symbol (|) in parentheses. This function estimates model parameters for non-gaussian families using Maximum Likelihood (Laplace Approximation). For more detailed information, the reader can view the documentation of the R-package "lme4" [32].

$$glmm = glmer(Y \sim X_1 + (1|X_2), family = Gamma(link = \text{“}log\text{”}), data) \quad (4.15)$$

For the example given in Equation 4.15, there are 2 covariates for the response variable Y; covariate 1 (X_1) and covariate 2 (X_2). In this example, covariate 1 has the fixed effect, that is, the regression parameter belonging to this variable does not vary as in generalized linear models. However, for covariate 2 (usually a grouping variable), different intercept values are estimated for each group represented in covariate 2 (X_2). Since fixed intercept value is already estimated in the model, when the new

sample belonging to that group is estimated, the final intercept value is determined by adding these two values. For example, if we build a generalized linear mixed model with Gamma family, identity link function, ship speed (V_s) as fixed effect and random intercept values on the basis of ship type to estimate total hull resistance in a data set containing 4 different ship types (bulk carrier, container ship, general cargo ship and tanker), the summary function gives an output as in Figure 4.9.

```
> summary(glmer_exp)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Gamma  ( inverse )
Formula: Hull_Total_Res ~ Vs + (1 | Ship_Type)
   Data: exp_glmer

     AIC      BIC   logLik deviance df.resid
   647.7    659.3   -319.9    639.7      128

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.40282 -0.64716  0.08365  0.62688  2.47736

Random effects:
 Groups    Name        Variance  Std.Dev.
 Ship_Type (Intercept) 2.357e-05 0.004855
 Residual              7.586e-02 0.275428
Number of obs: 132, groups:  Ship_Type, 4

Fixed effects:
             Estimate Std. Error t value Pr(>|z|)
(Intercept)  0.203134   0.009023   22.51   <2e-16 ***
Vs          -0.217762   0.013472  -16.16   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
   (Intr)
Vs -0.829
```

Figure 4.9. Example “glmer” output

The information in this output is mostly similar to the information in the example output for the “lmer” function in Figure 4.8. Differently, Akaike information criterion (AIC), Bayesian information criterion (BIC), family (Gamma in this example), link function and the log likelihood (logLik) are reported in this output.

4.4. Model Comparisons

4.4.1. Leave One Out Cross Validation

In a k-fold cross validation, the data are divided into k equal slices and each slice is once used as the test set, while the remaining k-1 slices are use as training sets [40]. In such a validation technique, the maximum value of k is the number of observation. As the value of k increases, the reliability of validation increases in parallel. However, in any case, keeping the value of k too high is not practical because as the value of k increases, the computational load also increases. The most desired cross validation in appropriate conditions is leave one out cross validation.

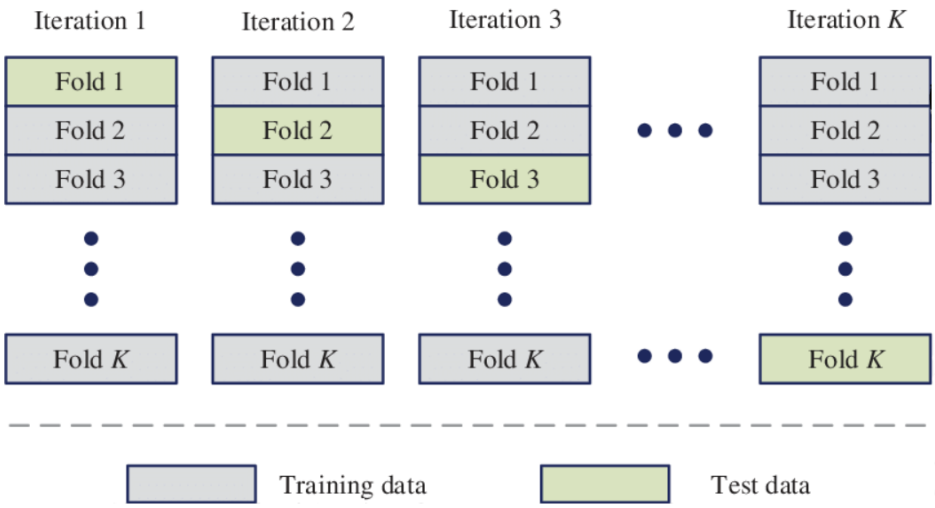

Figure 4.10. K - Fold cross validation

Since the data we have are kind of longitudinal data, it is not possible to handle each observation separately; there are different loading states for each ship and many speed measurements are made for each loading condition. For example, if we use some of the 15 different test results in the design loading state of a ship in train data and the other in test data, the statistical model we have will have already learned the characteristics of the ship in the train data. For this reason, a special method of leave

one out cross validation is used in this particular problem. At each stage, all loading conditions and test results (at each speed value) of a ship were used as the test set, while the information of all remaining ships was used as train data. This application has been applied for all established statistical models (generalized linear model, artificial neural networks, mixed effect linear models and generalized linear mixed models), and in each step the mean absolute errors are calculated. Finally, the overall mean absolute error values for each model are reported separately. This special type of the leave one out cross validation method is called as "Leave one ship out cross validation" in the rest of the thesis.

4.4.2. 4-Fold Cross Validation

In artificial neural networks built in this study, for selecting the most convenient variable and parameter sets to use in the model, 4-Fold cross validation is applied for determining the model performance. The reason for preferring the stratified 4-Fold cross validation technique is that as stated in the Table 3.1, there are only 4 bulk carriers in the data set. For the generalized linear model and mixed effect linear model, 4-Fold cross validation was not performed since in these statistical models, parameter selection can be made according to the significance levels of the variables and AIC values, but unfortunately, this is not possible in ANN. Since many calculations will be carried out for each parameter set, 4-Fold cross validation was preferred instead of leave one ship out cross validation to prevent the computational cost from being too high. In each train and test split, all the ship types are divided homogeneously. Applying the 4-Fold cross validation technique also reduced the risk of errors caused by train test split. For each trial with the potential parameter sets, the data are split into 4 folds; 3 for training and 1 for validation. Then, the MSE values are calculated for each fold and the MSE for the addition of each potential input trail was reported. The visual representation of 4-Fold cross validation is given in Figure 4.11.

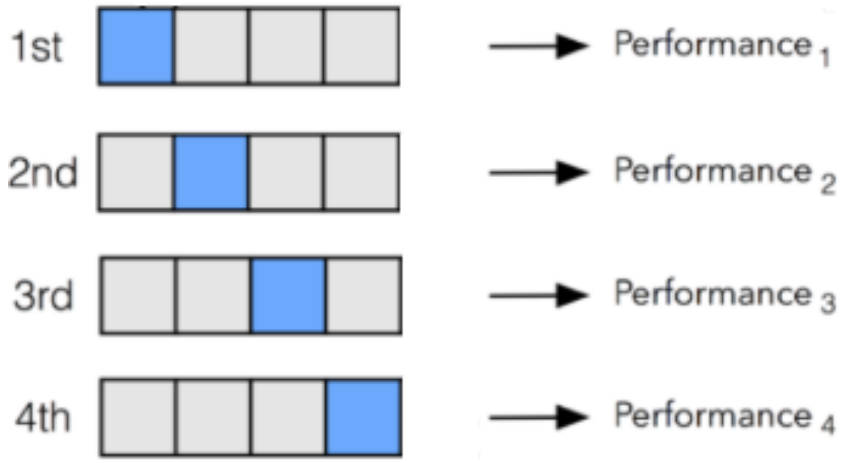

Figure 4.11. 4 - Fold cross validation

5. MODELING THE RESISTANCE DATA

At the first part, all numerical variables are mapped between 0 and 1 to reduce the error rate that the variables with high range difference will cause. Especially artificial neural networks are very sensitive in this regard and such errors cause unfavorable estimations. Since the normalized data are used in ANN, it has also been used in other statistical methods to be a fair comparison. Variables with a high range difference cause serious differences in regression parameters, which leads to unstable regression models. For example, in a regression problem, variables with very large values become much more important than they should have with larger regression parameters [41]. In such cases, the fact that the statistical model has variables with very different ranges directly affects the performance of the model negatively. To prevent such a risk of error, numerical values are mapped to a specified range, which is known as standardization or normalization [42]. It can be claimed that the normalization procedure is needed in this study since the numerical covariates have different ranges. For instance, the mean value of the total height (draught) under water at the center of the ship (T_s) is 0.2155 while the mean value of the total wetted surface area of the ship (A_{WS}) is 3.697, which means that there is approximately 94 % difference between them. The summary of the other numerical values of these two features are as follows. For the very detailed summary of all covariates, see Appendix B.

Table 5.1. Statistical summary of T_s and A_{WS} variables

	T_s	A_{WS}
Min.	0.095	2.168
1st Qu.	0.169	3.147
Median	0.22	3.715
Mean	0.2155	3.697
3rd Qu.	0.264	4.183
Max.	0.311	5.208

Normalizing the numerical variables scales the data into a common interval and ensures that the range difference between covariates does not lead to unusable regression coefficients. At the same time, this approach does not violate the proportional distinction of the variables since the order of numbers from large to small does not change and the difference between them is preserved proportionally. The normalization function given in Equation 5.1 is applied all numerical columns, and for a fair comparison, these normalized data are used in all statistical models.

$$X_{normalized} = \frac{X - X_{minimum}}{X_{maximum} - X_{minimum}} \tag{5.1}$$

After normalizing the numerical variables, the cargo ship data are divided into train (60%) and test data (40%) homogeneously so that an initial statistical model can be obtained for GLM, MELM and GLMM, and the significant features can be detected. For the train and test split phase, stratified sampling was used. In stratified sampling, train and test data contain the same proportion of all types of samples. In this particular problem, train and test data should contain the specified proportion of data for each cargo ship type. For example, if we are going to use 60 % of the data for the train set, we should use 60 % of the container ships in the train data. This should also be valid for all other cargo ship types. Additionally, the values of the respective ships at each speed value should be considered as a whole. For example, if a ship has 45 lines of data in total in three different loading conditions (design, ballast and heavy loaded), these values cannot be in both train and test data at the same time, either in the train or the test data. In line with these rules, the stratified sampling method was used for data split. Finally, not only the train data include 60% of the whole data set, but also the 60 % of the each ship type (tanker, bulk carrier, container and general cargo ships) separately.

5.1. Generalized Linear Model

Classical Linear models assume that the data must have constant variance; however, in real life the data generally do not have constant variance. Besides, linear

models assume that the response variable is distributed normally, and estimations are made according to this assumption. To avoid these constraints, instead of linear models, the more advanced version, generalized linear model that relaxes both response distribution and constant variance assumption is tried to be used as an estimation technique in this study. The generalized linear models are more flexible than the standard linear models since their distributions do not have to be normal distribution, it can be Poisson, Gamma and so on. This makes it easier to solve different types of problems having response variables that are not normally distributed thanks to the generalized linear models. For example, for the case of non-negative and continuous response variable regression problems, the models built with Gamma family provide very good results according to Chen et al. [43]. However, it is not possible to specify the distribution of response variable in standard linear model. If we used a standard linear model specifically for this problem, using log(y) transformation would be useful. Because it helps in providing constant variance assumption, it can also be useful for linearity assumption. However, GLM is prefered as the "LM for log(y)" has predictions that are not equal to the mean of the Y-values (but the log-predictions are equal to the log-Y values).

After determining the train and test data in the data pre-processing phase, all numeric and grouping variables are tried to predict the hull total resistance. For the initial model including all potential covariates, some families and link functions were tested. Since the Gamma distribution is commonly used for non-negative continuous response variables [43], it was preferred for this ship hull resistance estimation problem which is conducted for predicting non-negative continuous response variable, ship hull resistance. To test the accuracy, two different models, with Gamma and Gaussian family (log link function), were established with all possible covariates. While the AIC value of the Gaussian model is 4282.6, the AIC value of the model created with Gamma family is 3863.6, which show that the Gamma family suits better on the data. The main reason that gamma is better than normal is the variance function of the glm model with gamma family. This means for gamma family the variance increases with the mean response and this is the case for our data as we can see in the Figure 6.3, in

Chapter 6. For the GLM with gaussian family the variance is constant. That is the reason, why the AIC is so much higher for the gaussian family for our data.

At the final stage, the most appropriate and well performing family and link function are Gamma family and logarithm link function, respectively. After determining distribution family and the link function, quadratic terms are added to the model by considering the relations between covariates and response variables. The relations of the covariates with the logarithm of response variable, hull total resistance, were inspected visually. As an example, some of these relations can be seen in the Figure 5.1. For a more detailed plot, see Appendix B.

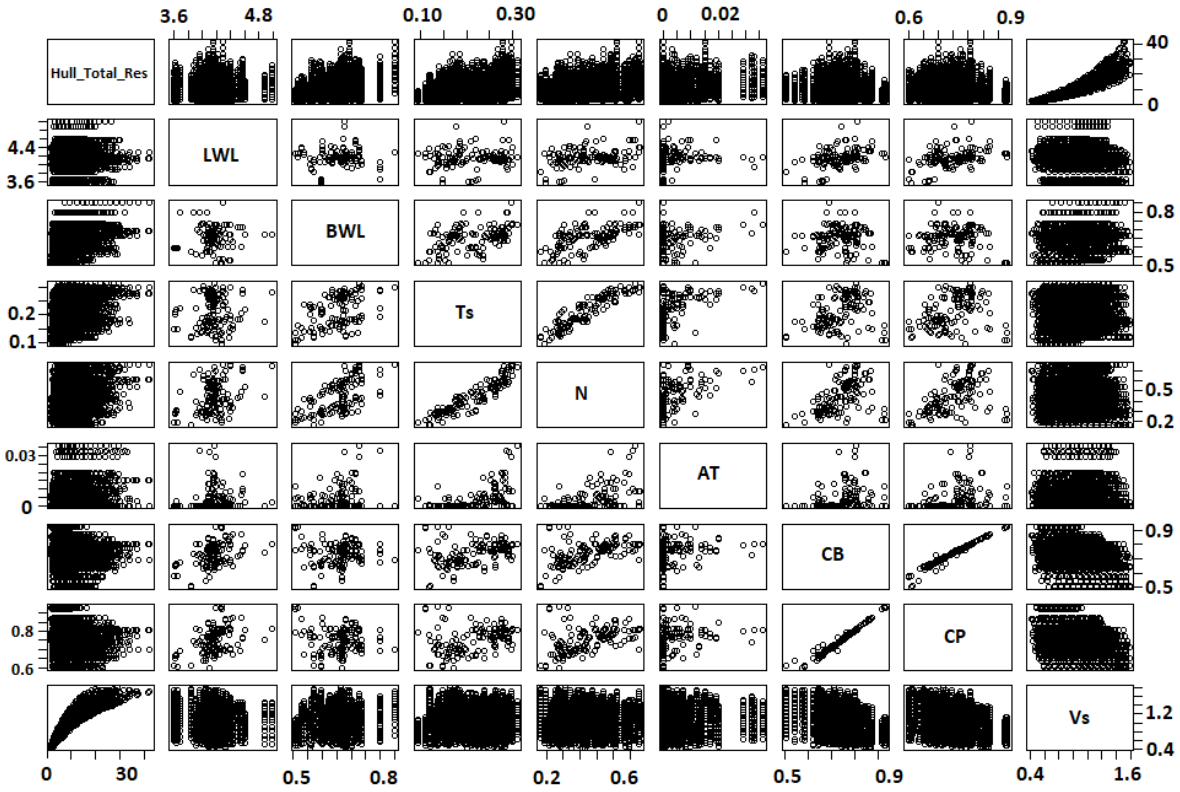

Figure 5.1. Relations of some features between the hull resistance

As it can be seen from Figure 5.1, there should be a somehow nonlinear relationship between ship speed (V_s) and the hull resistance. Therefore, different forms of the

nonlinear version of the ship speed such as square of speed were tried. Within the same consideration, these nonlinear terms also tried in mixed effect linear and generalized linear mixed models.

During the trials of these quadratic terms, the linearity assumptions (between transformed response and covariates) and the Akaike Information Criterion (AIC) are tested. If there is an improvement in linearity assumption and the AIC value is decreased, the recently added quadratic term was included to the model. After all potential quadratic terms have been tried, the ones that lead to improvement in the model have been identified and then the process of removing variables from the model has begun. Variables with low significance levels (according to 95% confidence interval) were excluded from the model considering also the AIC values. After completing all phases, the final GLM model for the prediction of hull total resistance is given in Equation 5.2:

$$glm = glm(Hull\_Total\_Res \sim L_{WL} + B_{WL} + T_A + A_{WS} + A_B + A_T + C_P + C_M + C_{WP} + V_s + V_s^2 + V_s^3, family = Gamma(link = \text{“}log\text{”})) \quad (5.2)$$

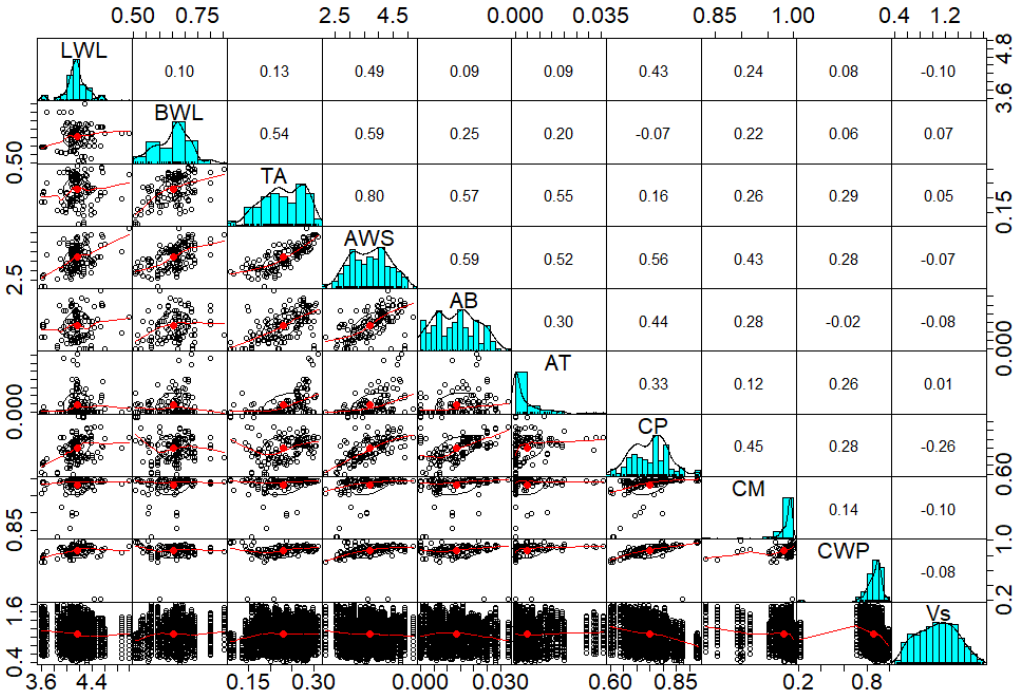

Figure 5.2. Relations between selected GLM variables

For this final GLM model created by using train data, the residual plot and the summary information are also given in Figure 5.3 and Figure 5.4, respectively.

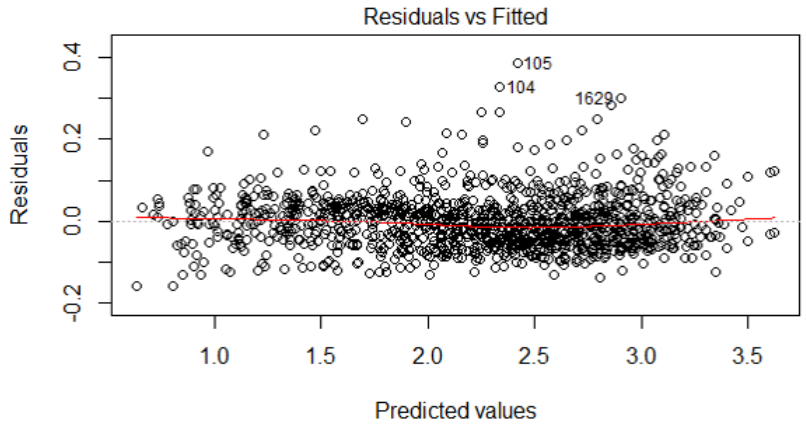

Figure 5.3. Residual plot of the final GLM model

```
> summary(glm1)

Call:
glm(formula = Hull_Total_Res ~ LWL + BWL + TA + AWS + AB + AT +
    CP + CM + CWP + Vs + I(Vs^2) + I(Vs^3), family = Gamma(link = "log"),
    data = train.data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.15902  -0.04669  -0.00638   0.03345   0.38363

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.13391    0.02176   6.155 1.02e-09 ***
LWL         -0.15247    0.01869  -8.158 8.42e-16 ***
BWL          0.31719    0.01721  18.426  < 2e-16 ***
TA           0.12100    0.02335   5.182 2.57e-07 ***
AWS          0.43030    0.03679  11.697  < 2e-16 ***
AB          -0.14002    0.01241 -11.279  < 2e-16 ***
AT           0.04046    0.01132   3.575 0.000365 ***
CP           0.42729    0.01980  21.577  < 2e-16 ***
CM           0.11234    0.01194   9.408  < 2e-16 ***
CWP         -0.18540    0.01575 -11.771  < 2e-16 ***
Vs           4.96935    0.09785  50.787  < 2e-16 ***
I(Vs^2)     -3.56728    0.21692 -16.445  < 2e-16 ***
I(Vs^3)      1.54289    0.14300  10.789  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 0.004423396)

    Null deviance: 456.5389  on 1226  degrees of freedom
Residual deviance:   5.1214  on 1214  degrees of freedom
AIC: 2308

Number of Fisher Scoring iterations: 4
```

Figure 5.4. Summary of the final GLM model

It can be said that the linearity assumption is satisfied according to the residual plot. Moreover, all the covariates have P values level smaller than 0.05 and the dispersion parameter is 0.00442. Accordingly, the shape parameter of the Gamma distribution is 226 which is simply the reciprocal of the estimate of the dispersion parameter.

After all these processes, final GLM model was reached; however, the exact comparison of the statistical models in this study is carried out based on leave one ship out cross validation technique since the size of the data set is relatively small. For the leave one ship out cross validation procedure, these selected features, family and the link function is used and the results are reported in the Section 6.

5.2. Artificial Neural Networks

In this ship resistance estimation problem, these normalized cargo ship data are used as well as in other statistical methods. First of all, some studies for determining the significant features to use in the model are carried out. In GLM, MELM and GLMM, feature selection is made based on the significance levels of the variables and the AIC values; however, feature selection is more complicated in ANN models. Moreover, there are several features in the data set which can be potential input variables in the ANN model. For the purpose of determining the eligible features to use in the ANN model, forward feature selection algorithm is used, which adds the variables one by one and decides whether to include them according to the model performance. After selecting the significant variables to use in the model, parameter tuning is carried out for selecting the suitable parameter set and improving the model performance.

5.2.1. Feature Selection

For choosing an appropriate ANN model, forward feature selection algorithm which inserts the most suitable variable to the input set is preferred in this study as stated before. In forward feature selection algorithm an initial base statistical model is created in the first step. Then, all possible variables are used separately for predicting the outcome, and the variable giving the least error value is selected as an input variable for the model. The selected variable is not in the potential variable set anymore, instead, it is in the input variable set. This process is carried out iteratively until there is no improvement in the error function after adding a potential input variable [44]. The sequential forward feature selection pseudo code is as follows [45].

(i) Create an empty set: $Y_k = \emptyset, k = 0$,

(ii) Select best remaining feature: $x^+ = argmax_{x^+ \in Y^k}[J(Y_k + x^+)]$

(iii) If $J(Y_k + x^+) > J(Y_k)$;

  (a) Upgrade $Y_{k+1} = Y_k + x^+$

  (b) k=k+1

(c) Go back to step (ii)

In this pseudo code, Y_k represents the input variable set while x^+ indicates the potential input variable. Moreover, the function J calculates the model performance and k stands for time.

Instead of restricting the feature selection process by the improvement in the error value, in this study the maximum size of input variable set is set as 9 because of the high computational load. Increasing the number of input variables also increases the complexity of the model and the risk of overfitting. Moreover, for this particular problem, too many input variables are not necessary. For example, Mason et al. built artificial neural networks for predicting ship hull resistance with only 4 input variables [7].

Before starting the forward feature selection process, 2 hidden neurons and 1 hidden layer, logistic activation function, SSE loss function, resilient backpropagation with weight backtracking algorithm and 0.5 threshold value were chosen as the parameter set of a simple ANN model. Also note that, for the artificial neural networks in R ("neuralnet" package) initial weights are randomly selected from standard normal distribution [18]. Therefore, there is a randomness caused by this situation. From this point of view, 5 replications were performed for each specified artificial neural network parameter set and the model with the lowest mean squared error value was selected. This process was applied for each parameter set separately.

Applying all these processes with 4-Fold cross validation, the result matrix is produced and given in Table 5.2.

Table 5.2. 4 - Fold cross validation results for variable set in ANN

N	**MSE**	**Formula**
1	8.1883096	Hull_Res $\sim Vs$
2	1.5380961	Hull_Res $\sim Vs + A_{WS}$
3	1.4369872	Hull_Res $\sim Vs + A_{WS} + C_M$
4	1.2121215	Hull_Res $\sim Vs + A_{WS} + C_M + C_B$
5	0.9966383	Hull_Res $\sim Vs + A_{WS} + C_M + C_B + L_{WL}$
6	1.0930449	Hull_Res $\sim Vs + A_{WS} + C_M + C_B + L_{WL} + H_T$
7	0.9788777	Hull_Res $\sim Vs + A_{WS} + C_M + C_B + L_{WL} + H_T + D$
8	1.1607413	Hull_Res $\sim Vs + A_{WS} + C_M + C_B + L_{WL} + H_T + D + H_B$
9	1.384635	Hull_Res $\sim Vs + A_{WS} + C_M + C_B + L_{WL} + H_T + D + H_B + N$

As it can be seen from the Table 5.2, the parameter set giving the least MSE value is the input set consisting of 7 input variables including ship speed (V_s), wetted surface area (A_{WS}), mid-ship section coefficient (C_M), block coefficient (C_B), waterline length (L_{WL}), transom area center (H_T) and displacement of ship (D). In addition, after 7 variables, it is seen that mean CV value increases for the addition of a new variable. For the parameter tuning phase, this set of input parameters is used. The relations between these covariates can be seen in the Figure 5.5.

5.2.2. Parameter Tuning

After selecting the input variable set, parameter tuning study for selecting the suitable number of hidden neurons and the algorithm is carried out. In artificial neural networks, increasing the number of hidden layers allows the model to fit better on the data; however, this results in the risk of overfitting since it makes the model more complicated. For this regard, according to their study, Hornik et al. claim that the ANNs with only one hidden layer are sufficient for most of the cases with piece wise continuous functions [9]. In addition, Mason et al. [7] built artificial neural networks

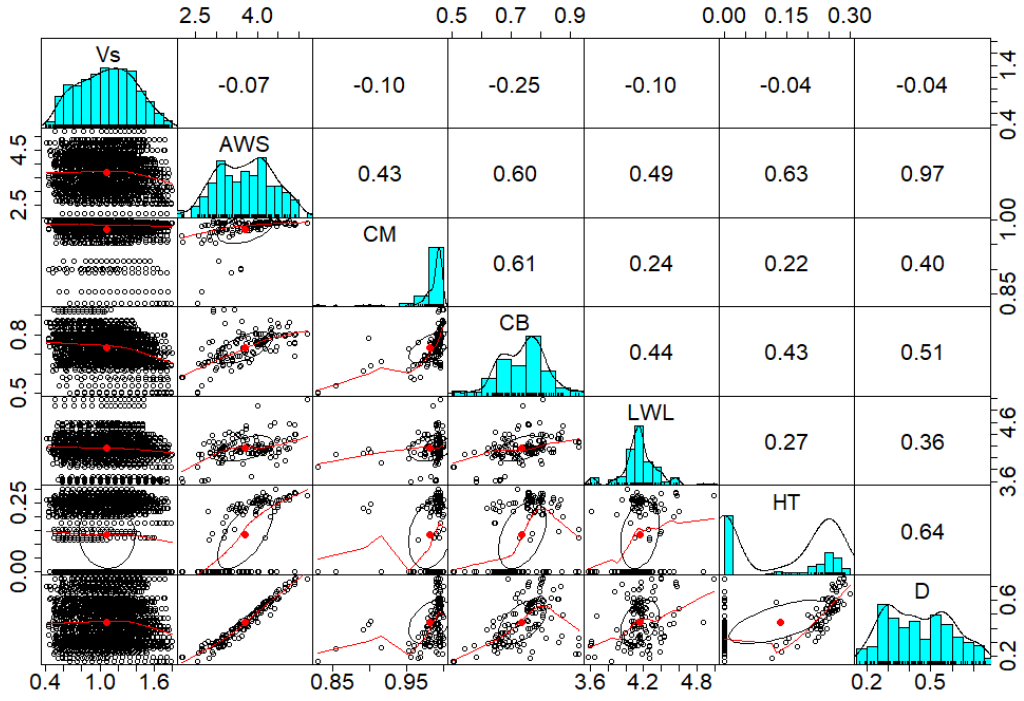

Figure 5.5. Relations between selected ANN variables

with a single hidden layer to predict the ship hull resistance values. Based on this information, a parameter tuning study is not performed for the number of hidden layers and trials are carried out with 1 hidden layer in this thesis. In addition to the number of hidden layer, there is a tradeoff between model complexity and the number of hidden neurons. The more hidden neurons an ANN has, the more complex structure it has. Moreover, as the number of hidden neuron increases, the probability of overfitting and the computational load increases in parallel. As an initial idea for the number of needed hidden neurons, it must be between the number of input and the output neurons [46]. For this particular problem, it should be between 1 and 7. In the parameter tuning study, the number of hidden neurons from 1 to 9 are tried.

In addition to the number of hidden neurons, the algorithm for learning in the ANN model should be selected. Since there will be several trials because of the 4-fold cross validation algorithm for determining the best parameter set, a fast and robust learning algorithm should be selected. In the literature, the resilient backpropagation

algorithm using traditional backpropagation technique and updating weights in the opposite direction of the partial derivatives is accepted as a well performing and fast learning technique in artificial neural networks [19, 47]. In this regard, two types of resilient backpropagation: resilient backpropagation with and without weight backtracking are tried in the parameter tuning study for ANN. For the activation function, a parameter tuning is not performed and logistic function is preferred. As in the feature selection study, 4-fold cross validation is applied for each parameter set, and the results are reported in Table 5.3.

Table 5.3. Parameter tuning results

Neuron	**Algorithm**	**ActivationFunc**	**MSEcrossTest**
1	rprop+	logistic	1.384555
1	rprop-	logistic	1.502684
2	rprop+	logistic	1.226831
2	rprop-	logistic	1.54056
3	rprop+	logistic	1.288606
3	rprop-	logistic	1.486358
4	rprop+	logistic	1.525687
4	rprop-	logistic	2.137777
5	rprop+	logistic	2.110248
5	rprop-	logistic	125.230066
6	rprop+	logistic	2.184847
6	rprop-	logistic	67.932753
7	rprop+	logistic	3.215091
7	rprop-	logistic	2.607394
8	rprop+	logistic	130.129618
8	rprop-	logistic	2.327957
9	rprop+	logistic	3.536267
9	rprop-	logistic	5.232552

As it can be seen from the Table 5.3, the parameter set giving the least 4-fold cross validation error includes 2 hidden neurons and resilient backpropagation with weight backtracking. In some cases, the ANN models suffer from giving very bad results because ANN is unable to estimate the resistance of tanker ships with different characteristics compared to others. This is maybe originated from the low number of inputs for this case. However, all the statistical techniques use this data set and thus the comparison will be fair. After all these studies, the final ANN model for the ship hull resistance estimation problem is shown in Figure 5.6.

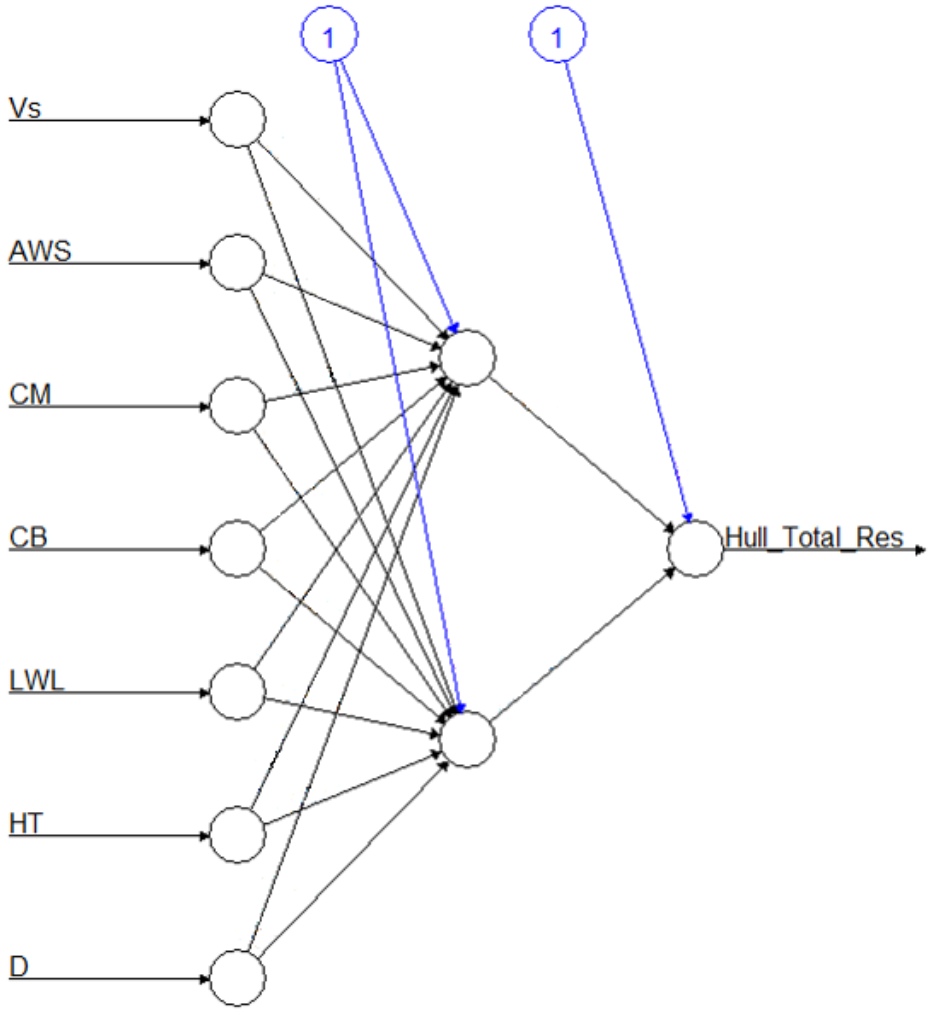

Figure 5.6. Final ANN model

5.3. Mixed Effect Linear Model

The cargo ship data obtained from Ata Nutku Ship Model Testing Laboratory of Istanbul Technical University contain the results of the experiments conducted for each scaled ship hull model at many different speeds, which means that this data are kind of repeated measurements and can be modeled as longitudinal data. The reason for this is that many different experiments are carried out in different loading situations of each ship model, only by changing the speed values. Therefore, in this thesis, the mixed effect linear model, as a longitudinal data approach, is tried to estimate the hull total resistance as well as the generalized linear model and artificial neural networks.

First of all, the train and test split is made as stated in the beginning of Chapter 5. Within the mixed effect linear model, after the regression coefficients of all covariates are set as fixed regression coefficients, it is assumed that there are random intercept values as random effects on each individual basis. In mixed effect linear models, random intercepts are assumed to be normally distributed, and random intercepts and error terms are independent from each other [48]. In this particular ship hull resistance estimation problem, each loading condition of each ship model is considered as independent individual, and the random regression coefficients are determined accordingly, since all data except the speed value remain the same in the individual loading conditions (design, ballast and heavy loaded) of each ship model.

After all, log transformation is determined as the most suitable transformation providing constant variance assumption among all transformation attempts for the relevant model. After determining the most suitable transformation for this mixed effect linear model, it is tested whether the AIC value of the model is decreased by adding the nonlinear terms of the ship speed to the model since there should be a nonlinear relation between ship speed and the hull resistance according to the visual inspection as stated in Section 5.1. As mentioned before, all variables are used as fixed effects, and it is stated that there are random intercept values on the basis of each group. Therefore, these tried covariates are evaluated as fixed effects, which have the

same value for each sample and do not change on a group basis. If this added quadratic term decreased the value of AIC, it was included in the model, if it did not, it was not added to the model. During this stage, significance levels of the covariates are also considered as a covariate selection parameter. Variables having p-values greater than 0.05 are not added to the model. All possible nonlinear terms are tried and finally this stage is completed.

In the next step, variables are removed according to their significance levels. Variables having p-values greater than 0.05 are removed from the model, and AIC values are checked simultaneously. Final mixed effect linear model for this ship hull resistance estimation problem is given in Equation 5.3.

$$lmer.cargoships = lmer(log(Hull\_Total\_Res) \sim L_{WL} + B_{WL} + A_{WS} + C_P + C_M + Bulb + V_s + V_s^2 + \sqrt{V_s} + (1|Ind), data = train.data) \quad (5.3)$$

In this final mixed effect linear model stated in Equation 5.3, the response variable is the logarithm of the hull total resistance value. In this model, the feature designated "Ind" represents the loading states of each ship individually, and random intercept values are estimated on the basis of this feature. The grouping variable "Ind" is created by combining ship model number and loading condition for each line. As an example, for the ship 424 and "Design" loading condition, the "Ind" variable for corresponding lines is "424,Design". This operation allows us to find out which ship model and loading condition the related row belongs to via a single variable. In summary, each individual ship model has intercept values assigned for separate loading cases. As stated in the Subsection 4.3.2, the random effect ("Ind" here) is specified with the aid of the symbol "|". On the other hand, all variables specified in the model are used as fixed effects and these values do not differ on the basis of "Ind", they remain the same. This model is called "mixed effect" model because it has both fixed and random effects.

After this model is established, for example, a ship included in the train data set has an estimated special intercept value for a certain loading condition. If an estimation is made at another speed value for the respective loading condition of this ship, this specified intercept value is used, all other regression coefficients are already the same, and they do not change on a group basis. However, if estimation is to be made for a new ship and loading condition set that is not included in the train data with this model, each estimated random intercept value is averaged and used as the random intercept value of the new ship and loading condition to be estimated. Since the mean value of the random regression coefficients is zero in mixed effect linear models, only the fixed regression coefficients are used for this kind of estimations.

Additionally, the relations between selected variables are also given in Figure 5.7.

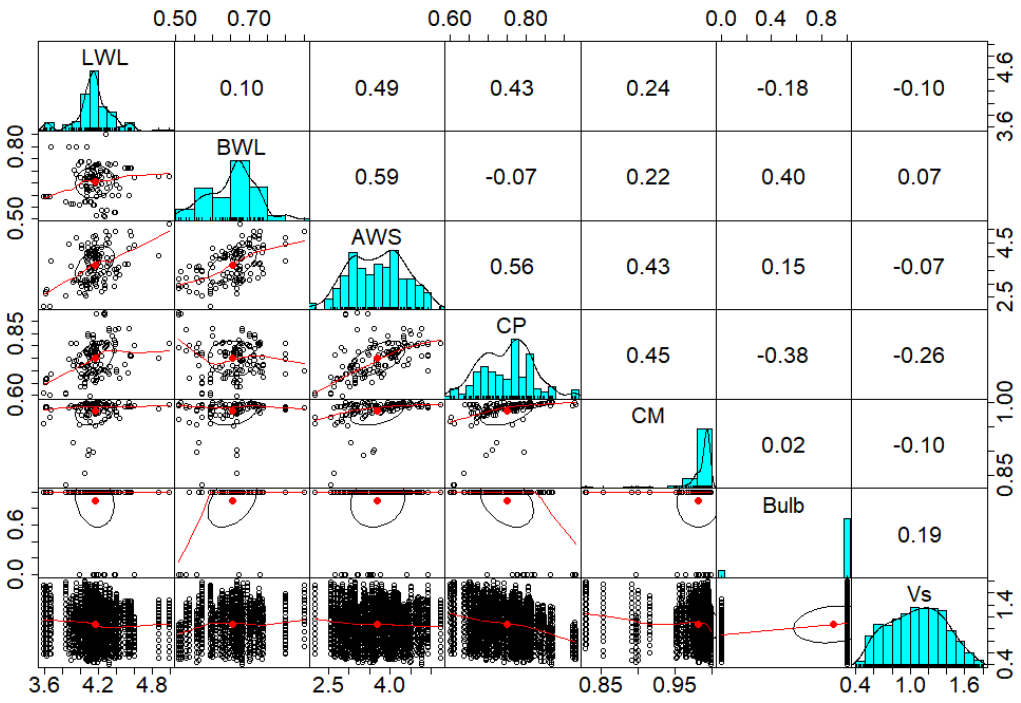

Figure 5.7. Relations between selected MELM variables

For the final mixed effect linear model, waterline length (L_{WL}), waterline breadth (B_{WL}), wetted surface area (A_{WS}), prismatic coefficient (C_P), mid-ship section coef-

ficient (C_M), bulb (whether there is a bulb or not, binary variable), ship speed (V_s), square of the ship speed (V_s^2) and the square root of the ship speed ($\sqrt{V_s}$) are selected as the significant variables. The summary of the final mixed effect linear model is given in Figure 5.8.

```
> summary(lmer.cargoships)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: log(Hull_Total_Res) ~ LWL + BWL + AWS + CP + CM + Bulb + Vs +      I(Vs^2) + sqrt(Vs) + (1 | Ind)
   Data: train.data

REML criterion at convergence: -4033.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.2329 -0.5846 -0.0608  0.4365  5.6344

Random effects:
 Groups   Name        Variance Std.Dev.
 Ind      (Intercept) 0.003057 0.05529
 Residual             0.001689 0.04110
Number of obs: 1227, groups:  Ind, 82

Fixed effects:
             Estimate Std. Error         df t value             Pr(>|t|)
(Intercept)   -0.18663    0.04928  130.33196  -3.787             0.000232 ***
LWL           -0.16627    0.04778   75.02450  -3.480             0.000840 ***
BWL            0.26454    0.05143   74.96389   5.143   0.0000020895968662 ***
AWS            0.58478    0.06061   75.05433   9.649   0.0000000000000087 ***
CP             0.25365    0.04923   75.27644   5.152   0.0000020036712210 ***
CM             0.09940    0.03939   74.79910   2.524             0.013741 *
Bulb          -0.07557    0.02501   75.13472  -3.022             0.003433 **
Vs             1.58841    0.14026 1147.95705  11.325 < 0.0000000000000002 ***
I(Vs^2)       -0.30790    0.05911 1149.37192  -5.209   0.0000002248271796 ***
sqrt(Vs)       1.91945    0.11067 1147.58786  17.344 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) LWL    BWL    AWS    CP     CM     Bulb   Vs     I(V^2)
LWL         -0.360
BWL         -0.325  0.278
AWS          0.396 -0.399 -0.764
CP          -0.318  0.035  0.543 -0.656
CM          -0.444 -0.031 -0.082 -0.076 -0.309
Bulb        -0.542  0.178  0.119 -0.367  0.504 -0.059
Vs           0.469 -0.011 -0.008  0.006 -0.014 -0.007 -0.016
I(Vs^2)     -0.427  0.013  0.012 -0.008  0.018  0.011  0.016 -0.970
sqrt(Vs)    -0.487  0.010  0.006 -0.005  0.013  0.005  0.015 -0.987  0.922
```

Figure 5.8. Summary of the final MELM model

As it can be seen from the Figure 5.8, there are no variables having a P-value greater than 0.05 at all, which means that all variables are significant according to the 95 % significance level. Moreover, the residual plot of the final mixed effect linear model created by using train data is given in Figure 5.9.

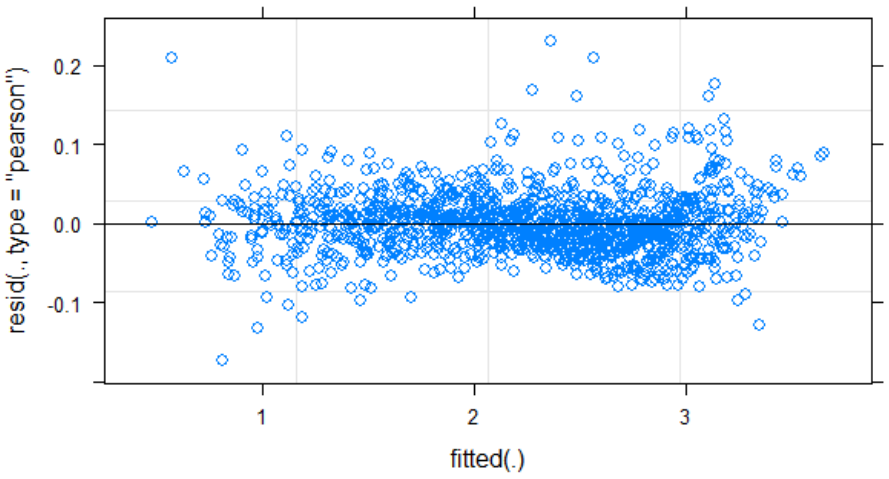

Figure 5.9. Residual plot of the final MELM model

As it can be seen from the Figure 5.9, the residual plot does not clearly contradict the constant variance and linearity assumption. Compared to the final GLM residual plot given in Figure 5.3, the variance of residuals in final MELM is relatively smaller. Also, according to the visual inspection, the final MELM can be said to be more successful in linearity assumption than the final GLM.

As a result of all these stages, the final mixed effect linear model is established. However, leave one ship out cross validation will be used for the final comparison with other statistical techniques.

5.4. Generalized Linear Mixed Model

Generalized linear mixed model is tried to deal with ship hull resistance estimation problem in this study, both because this model is a kind of longitudinal data approach and does not limit the distribution of response variables only to normal distribution. In essence, it can be sait that generalized linear mixed models are the combination of generalized linear and mixed effect linear models. For the beginning, the normalized

cargo ship data are split into train and test set as stated in the beginning of Chapter 5. Within the generalized linear mixed models, after the regression coefficients of all covariates are set as fixed effects, it is assumed that there are random intercept values as random effects on each individual basis. The term "Ind" is created as stated in Subsection 5.3. In this particular problem, each loading condition of each ship is treated as an individual. Therefore, in GLMM, random intercept values are estimated separately for each loading condition of each ship.

Since the Gamma distribution performed well on the cargo ship data in generalized linear models built in this study before, the generalized linear mixed models are also established with Gamma distribution with log link function. Consequently, log link function satisfied the linearity assumption. Moreover, for the case of non-negative and continuous response variable regression problems, the models built with Gamma family provide very good results according to Chen et al. [43]. For the comparison, generalized linear mixed models are created with Gamma and Gaussian families and log link functions using all possible covariates in the train data. Moreover, for the generalized linear model, these families are also tried with same data set and the AIC values are obtained as in Table 5.4.

Table 5.4. AIC comparison of initial GLMs and GLMMs

Model	**Family**	**AIC**
GLMM	Gaussian	3489.2
GLMM	Gamma	3264.4
GLM	Gaussian	4282.6
GLM	Gamma	3863.6

As it can be seen from the Table 5.4, the GLMM with Gamma family has the lowest AIC value; therefore, the Gamma family is selected. In addition, more sophisticated models were also tried before deciding on the random intercept model. Comparisons were made on which values can be changed and standardized on an individual basis. When comparing models, AIC values (for models with the same response variables)

and the model assumptions were checked. For example, the model including random intercepts on ship type and loading condition basis is tested, and the results are reported at the end of this section. Finally, as an appropriate approach to the problem at hand, it is concluded that it is useful to assign random intercept values on the basis of each particular ship and loading condition in the generalized linear mixed effect model.

In generalized linear mixed models, random intercepts are assumed to be normally distributed. Moreover, random intercepts and error terms are assumed to be independent from each other [37]. In this particular ship hull resistance estimation problem, each loading condition of each ship model was considered as independent individual, and the random regression coefficients were determined accordingly, since all data except the speed value remained the same in the individual loading conditions (design, ballast and heavy loaded) of each ship model.

After setting the distribution, link function, fixed and random effects (intercept for this problem) for this generalized linear mixed model, a GLMM is created with all potential numeric and grouping variables. Then, it is tested whether the AIC value of the model is decreased by adding the nonlinear terms of covariates, especially those with high t value (eg. ship speed), to the model. If this added quadratic term decreases the value of AIC, it is included in the model, if it does not, it is not added to the model. During this stage, significance levels of the covariates are also considered as a covariate selection parameter. Variables having more than 5 % p-value are not added to the model. All possible nonlinear terms are tried and finally this stage is completed. Note that all variables in the model are set as fixed effect. In addition, random intercept values for each individual are estimated in the model.

In the next step, variables are removed according to their significance levels. Variables having more than 0.05 p-values are removed from the model, and AIC values are checked simultaneously. Final generalized linear mixed model for this ship hull

resistance estimation problem is given in Equation 5.4.

$$glmm = glmer(Hull\_Total\_Res \sim L_{WL} + B_{WL} + A_{WS} + C_B + C_{WP} + Bulb + Vs + \sqrt{V_s} + (1|Ind), family = Gamma(link = \text{“}log\text{”})) \quad (5.4)$$

In this final generalized linear mixed model stated in Equation 5.4, the feature designated “Ind” represents the loading states of each ship individually, and random intercept values are estimated on the basis of this feature. In summary, each individual ship model has intercept values assigned for separate loading cases. Except for the intercept, all variables specified in the model are used as fixed effects and these values do not differ on the basis of “Ind”, they remain the same. This model is defined as a “mixed effect” model because it has both fixed and random effects. As a result of feature selection process in generalized linear mixed model for this particular ship hull resistance estimation problem, waterline length (L_{WL}), waterline beam (B_{WL}), wetted surface area (A_{WS}), block coefficient (C_B), water plane area coefficient (C_{WP}), bulb (binary), vessel speed (V_s) and the square root of vessel speed are selected as significant covariates to build the final statistical model. Note that the tried grouping variables (ship type and loading condition) are not statistically significant for the model according to their P-values, they do not improve the model performance. The reason for this may be that the data set we have consists of ship types with similar characteristics (cargo ships). Since the distinction between these ships is provided by other covariates (eg. length, depth, width), it can be considered acceptable that these grouping variables are not significant. If there were other ships with very different characteristics in our data set (eg. yachts), then information such as ship type would be meaningful. In addition, one of the reasons why loading condition is not significant may be that the resistance variations between loading conditions differ depending on the size of each ship. For example, the relationships between the design and ballast conditions of a large-sized tanker ship and a container ship with relatively small dimensions may differ. While the resistance difference between the loading states of larger sizes can be higher both proportionally and quantitatively, this difference can be considered less in

small sizes. It may also not be true to say that every ship has standard characteristics, for example, for the heavy loaded case. Considering all these, it is understandable that the loading situation is not significant for this model. Although ship type and loading condition are not significant as fixed effects in this model structure, the model containing random intercept values on the basis of ship type and loading condition will also be tested in this section. The relations between the selected variables for final GLMM are given in Figure 5.10.

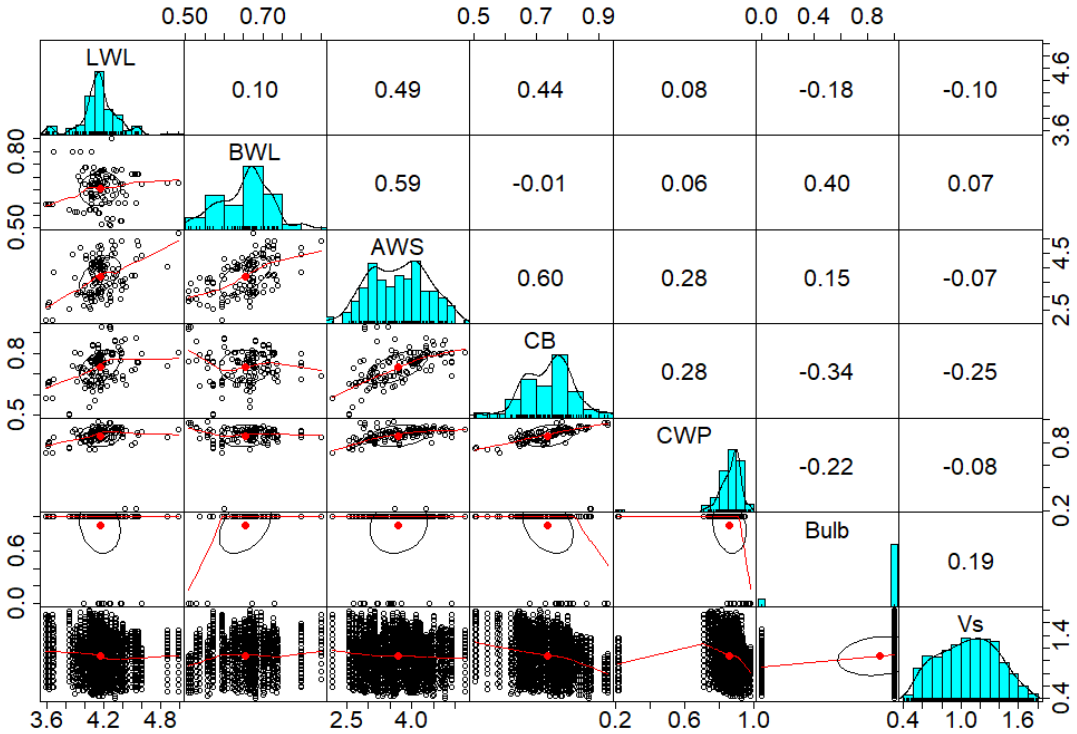

Figure 5.10. Relations between selected GLMM variables

Compared to the final MELM having 9 features, generalized linear mixed model have 8 covariates to estimate the hull total resistance. Both these methods have waterline length (L_{WL}), water plane surface area (A_{WS}) and vessel speed (V_s), which are also used in all built statistical methods in this thesis. Moreover, waterline breadth (B_{WL}), binary bulb info and the square root of the vessel speed ($\sqrt{V_s}$) are also common features used in these two statistical models. Instead of using square of the vessel

speed (V_s^2), prismatic coefficient (C_P) and mid-ship section coefficient (C_M), generalized linear mixed model uses block coefficient (C_B) and water plane are coefficient (C_{WP}) as input variable to predict the hull resistance. On the other hand, final GLM model have 12 covariates including V_s^2 and V_s^3. Instead of using the binary bulb variable representing whether a vessel has bulb or not as in MELM and GLMM, the final generalized linear model has the bulb section area (A_B). It also has aft draught (T_A) and transom area (A_T) which are not used in ANN, MELM and GLMM. As opposed to GLM, MELM and GLMM, ANN does not use any type of bulb information.

For the final GLMM model created by using train data, the residual plot and the summary information are also given in Figure 5.11 and Figure 5.12, respectively.

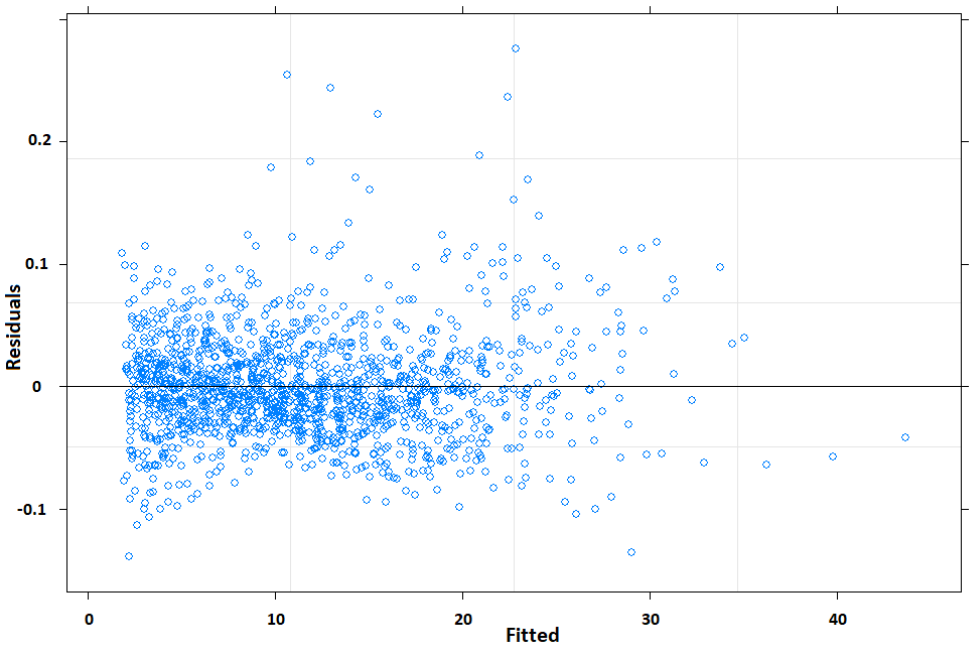

Figure 5.11. Residual plot of the final GLMM model

```
> summary(glmer.cargoships)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Gamma  ( log )
Formula: Hull_Total_Res ~ LWL + BWL + AWS + CB + CWP + Bulb + Vs + sqrt(Vs) +      (1 | Ind)
   Data: train.data

     AIC      BIC   logLik deviance df.resid
  1828.3   1886.4   -903.1   1806.3     1444

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.8369 -0.5594 -0.0799  0.4594  5.6506

Random effects:
 Groups   Name        Variance Std.Dev.
 Ind      (Intercept) 0.000979 0.03129
 Residual             0.002382 0.04880
Number of obs: 1455, groups:  Ind, 97

Fixed effects:
            Estimate Std. Error t value Pr(>|z|)
(Intercept) -0.13772    0.10131  -1.359 0.174012
LWL         -0.15387    0.06286  -2.448 0.014382 *
BWL          0.26734    0.06526   4.097 4.19e-05 ***
AWS          0.63387    0.08816   7.190 6.49e-13 ***
CB           0.29364    0.07806   3.762 0.000169 ***
CWP         -0.15535    0.06709  -2.316 0.020585 *
Bulb        -0.11459    0.05411  -2.117 0.034223 *
Vs           0.86328    0.03506  24.625  < 2e-16 ***
sqrt(Vs)     2.50815    0.04421  56.727  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) LWL    BWL    AWS    CB     CWP    Bulb   Vs
LWL         -0.440
BWL         -0.153  0.185
AWS          0.622 -0.497 -0.581
CB          -0.650  0.150  0.349 -0.746
CWP         -0.757  0.242 -0.001 -0.341  0.194
Bulb        -0.805  0.317 -0.131 -0.532  0.618  0.423
Vs           0.109  0.010  0.009 -0.010  0.017  0.005  0.007
sqrt(Vs) -0.115 -0.009 -0.009  0.008 -0.014 -0.005 -0.008 -0.987
```

Figure 5.12. Summary of the final GLMM model

It can be said that the linearity assumption is satisfied according to the residual plot given in Figure 5.11. Moreover, all the covariates have P values level less than 0.05, which means that they can be accepted as significant variables to estimate the hull resistance. Both final MELM and GLMM satisfy this assumption. Compared to MELM residual plot, error variance of the GLMM is relatively higher, however, this is not a problem in fitting the model and providing assumptions. Note that the data used for this residual plot are the train data consisting of 60 % of the entire data, and the final comparison will be carried on with all data using this built model in Chapter 6. In addition, in GLMM, as in MELM, residuals are dispersed around zero. Note that the shape parameter of the final generalized linear mixed model with Gamma family is determined as 108.71 which is a relatively large value and the distribution is close

to normal distribution. However, the final generalized linear mixed model is also tried with Gaussian family and the AIC value is determined as 2720.5 while the AIC value of the final generalized linear mixed model with Gamma family is 1828.3, which means that the Gamma family suits better on these data. Note that these AIC values are different than the values given in Table 5.4 because they are the AIC values of the initial models with all possible (original) variables. In other words, there is no variable selection process for these models and they are initial models.

In addition to the generalized linear mixed model with random intercept for all 97 groups of grouping variable "Ind", another model that includes random intercept values on ship type and loading condition basis was established within the train data. In this model, a new grouping variable named "*Ship_Loading*" has been created, which contains ship type and loading condition information for ease of operation (for example "Tanker,Design", "Tanker,Ballast" etc.). Thanks to creating this variable, ship type and loading condition information could be obtained from a single variable, and a model containing random intercepts on the basis of this variable is established. Since there are four different ship types (bulk carrier, container ship, general cargo ship and tanker) and three different loading conditions (design, ballast and loading condition) in total, this grouping variable contains 12 different categories in total. Initially, two different models, with Gamma and Gaussian family, were set up with all the variables available, and AIC values were compared. While the AIC value of the model established with the Gaussian family was 4154.5, the AIC value of the model built with the Gamma family was 3795.2. This shows that the Gamma family suits better in this particular problem. Then, as mentioned in the previous model, the model building stages were completed. Finally, the model containing random intercept values on the basis of ship type and loading condition is completed with log link function, the summary of this model is given in Figure 5.13.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Gamma  ( log )
Formula: Hull_Total_Res ~ LWL + BWL + AWS + AT + HT + CB + LCB + Number_of_Propeller +
    Vs + sqrt(Vs) + (1 | Ship_Loading)
   Data: new_train.data

     AIC      BIC   logLik deviance df.resid
  2901.1   2969.8  -1437.6   2875.1     1442

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6166 -0.7098 -0.1010  0.5731  5.1982

Random effects:
 Groups       Name        Variance  Std.Dev.
 Ship_Loading (Intercept) 0.0009769 0.03125
 Residual                 0.0048707 0.06979
Number of obs: 1455, groups:  Ship_Loading, 12

Fixed effects:
                    Estimate Std. Error t value Pr(>|z|)
(Intercept)         -0.504303   0.039619 -12.729  < 2e-16 ***
LWL                 -0.166557   0.019362  -8.602  < 2e-16 ***
BWL                  0.180892   0.023561   7.677 1.62e-14 ***
AWS                  0.643286   0.032971  19.511  < 2e-16 ***
AT                   0.089226   0.013250   6.734 1.65e-11 ***
HT                   0.031935   0.008465   3.773 0.000162 ***
CB                   0.477681   0.016316  29.277  < 2e-16 ***
LCB                  0.102012   0.011649   8.757  < 2e-16 ***
Number_of_Propeller -0.040172   0.006128  -6.555 5.56e-11 ***
Vs                   0.852211   0.052430  16.254  < 2e-16 ***
sqrt(Vs)             2.524627   0.066375  38.036  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5.13. Summary of the GLMM model with random intercepts on the ship type and loading condition basis

As it can be seen from the Figure 5.13, all the covariates have P values level less than 0.05, which means that they can be accepted as significant variables to estimate the hull resistance. However, the AIC value of this model is 2901.1, while the AIC value of the previous model with random intercept values on the basis of each “Ind” (each particular ship and loading condition) is 1828.3 as can be seen from the Figure 5.12. This situation shows us that the model containing random intercept values on the basis of “Ind” fits better on the train data. Therefore, for comparisons with other models, the generalized linear mixed model with random intercepts on “Ind” basis is used.

After all these processes, final GLMM model is reached; however, the exact comparison of the statistical model in this study is carried out based on leave one ship out cross validation technique instead of using other types of k - fold cross validation since the size of the data set is relatively small. The details of the specific type of leave one ship out cross validation technique is discussed in Subsection 4.4.1. For the leave

one ship out cross validation procedure, these selected features, family and the link function is used and the results are reported in the Section 6.

6. RESULTS AND COMPARISONS

After obtaining the final versions of all the statistical models established, mean absolute errors are calculated for each model (generalized linear model, artificial neural networks, mixed effect linear model and generalized linear mixed models) by applying leave one ship out cross validation separately. In all these calculations, as mentioned earlier, all experiment data of each ship are used separately as test data, and the experiment values of all remaining ship models are used as train data. Finally, the leave one ship out cross validation results containing mean absolute and mean absolute relative errors are given in Table 6.1.

Table 6.1. Leave one ship out cross validation results

	GLM	**ANN**	**MELM**	**GLMM**
Mean Absolute Error	0.7713	0.7990	0.7300	0.7024
Mean Absolute Relative Error	0.0630	0.0687	0.0594	0.0568

While MELM and GLMM perform close to each other on this data set, the method with the lowest mean absolute error among these established statistical models is the generalized linear mixed model with 0.7024. At the same time, GLMM has the lowest mean absolute relative error, 0.0568, which means that hull resistance values are estimated with approximately 5.68 percent error margin. In addition, the mixed effect linear model, another longitudinal data approach, also performed well and had 0.73 mean absolute error and 0.0594 mean absolute relative error. On the other hand, GLM and ANN also made close predictions and had mean absolute error values of 0.7713 and 0.7790, respectively. However, these two statistical methods yielded less accurate results than the models established using the longitudinal data approach. The histogram of the leave one ship out CV forecasting errors and relative errors using GLMM are given in Figure 6.1 and Figure 6.2, respectively (the errors are calculated by subtracting the test results from the estimated values).

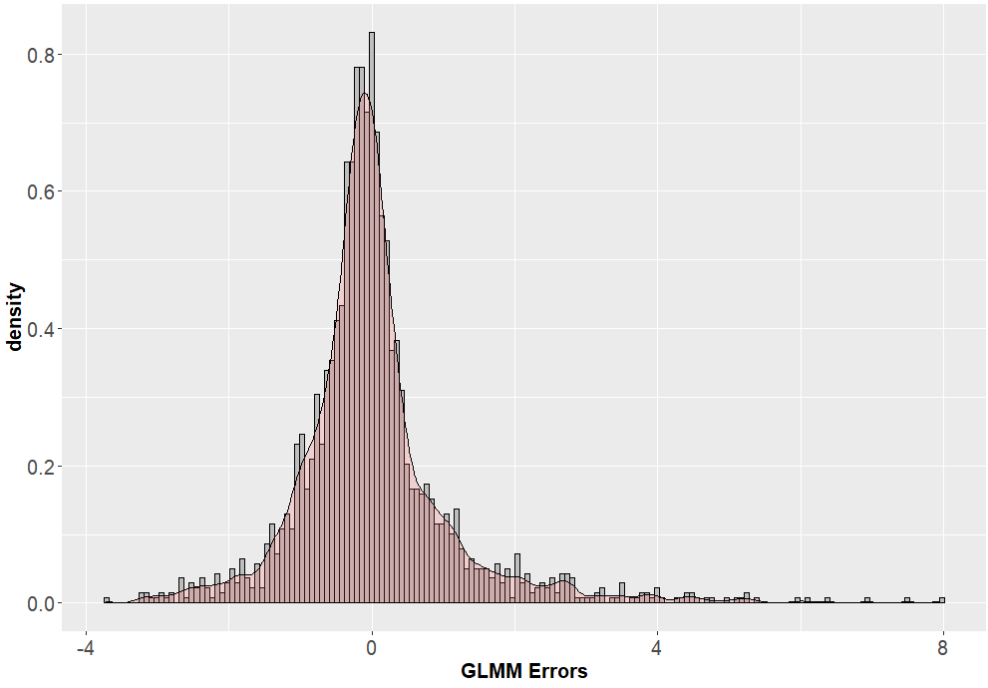

Figure 6.1. Leave one ship out CV forecasting errors using GLMM

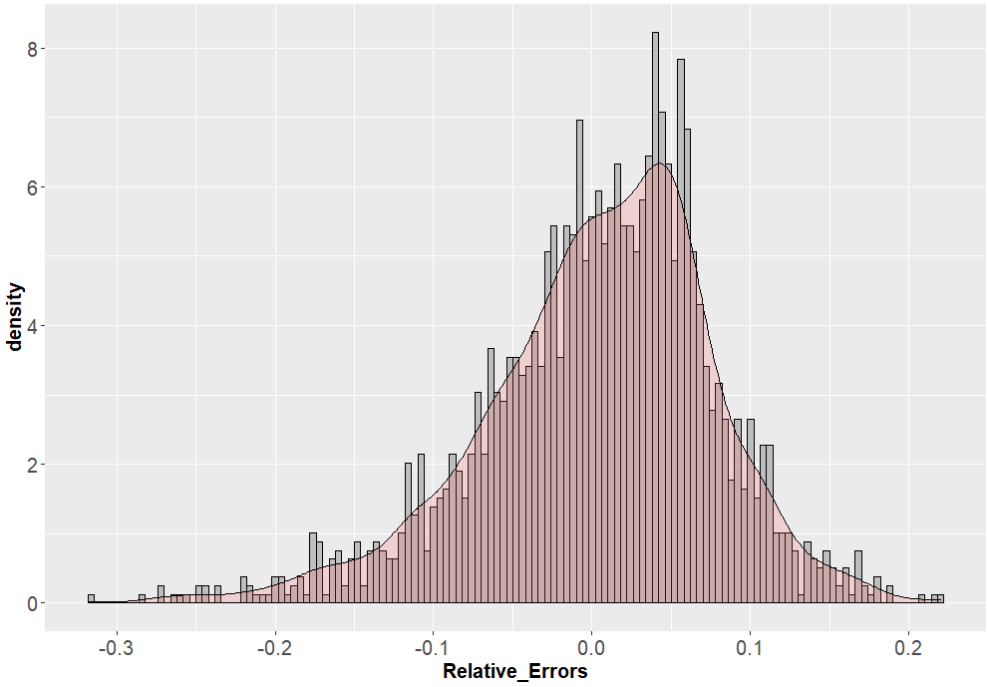

Figure 6.2. Leave one ship out CV forecasting errors using GLMM

As it can be seen from the Figure 6.1, errors are distributed around zero. On the other hand, the relative error histogram given in Figure 6.2 represents a kind of right skewed distribution. In these estimates, it can be said that it is relatively more frequent to estimate lower values than the experimental results. On the other hand, the variance of positive relative error values is smaller than the negative ones. The reason why the histograms of errors and relative errors differ slightly from each other is that the ratios of the estimated values to the experiment result value on the basis of each sample differ. In addition to the error histograms, the graph of GLMM predictions vs hull resistance values obtained from ITU Ata Nutku Ship Model Testing Laboratory experiment results is given in Figure 6.3.

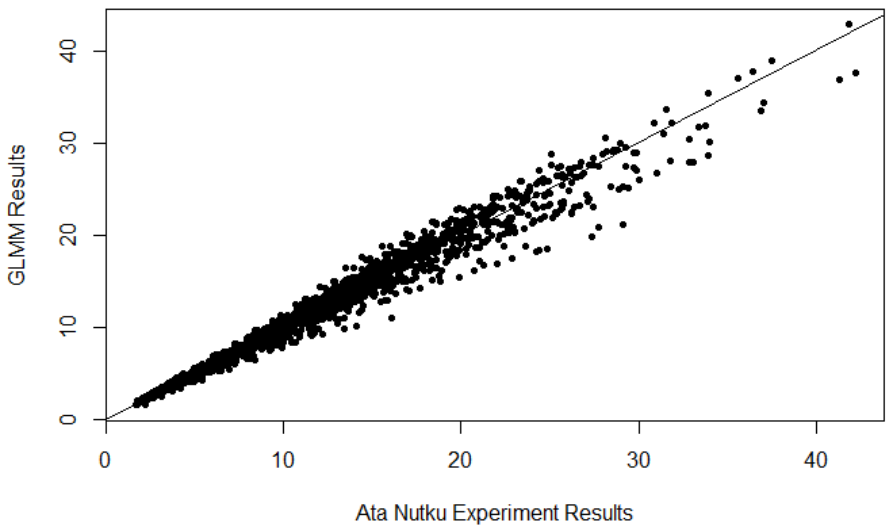

Figure 6.3. GLMM predictions vs experimental results

A very detailed results summary for Holtrop and Mennen method and built statistical models according to the ship type and loading conditions is given in Table 6.2.

Table 6.2. Ship type and loading condition based MAE comparisons

Ship Type	**Loading**	**Mean Absolute Error**				
		GLM	ANN	MELM	GLMM	Holtrop
Bulk Carrier	Design	0.5818	0.6748	0.5302	0.4889	4.6721
	Ballast	0.3235	0.4683	0.328	0.3002	1.3352
	Heavy Loaded	0.5116	0.6116	0.4175	0.3968	4.2661
General Cargo	Design	0.5741	0.7657	0.6795	0.5849	3.0846
	Ballast	0.5575	0.6456	0.5734	0.4826	1.0529
	Heavy Loaded	0.4303	0.483	0.4191	0.3813	3.6868
Container Ship	Design	0.9711	1.0787	0.961	0.9779	2.2029
	Ballast	0.6561	0.7026	0.6924	0.6816	2.3748
	Heavy Loaded	0.8073	0.6988	0.7795	0.7704	2.5627
Tanker	Design	0.8877	0.8625	0.8222	0.8054	3.9895
	Ballast	0.8549	0.8558	0.7926	0.8003	1.7123
	Heavy Loaded	0.9144	0.8164	0.6938	0.6374	3.5449

It can be seen from the Table 6.2, GLM, ANN, MELM and GLMM has made much better predictions in all loading conditions and ship types than the Holtrop & Mennen method. Among all these methods, GLMM is the one with the lowest MAE values. This method gave its best estimations in bulk carrier's ballast and heavy loaded state and general cargo ships in heavy loaded case. For general cargo ships and tankers, GLMM predicts the hull total resistance value more accurate in heavy loaded state. On the other hand, for bulk carriers and container ships, this method gives better results in ballast condition than other loading cases. For all ship types and loading conditions except container ships, generalized linear mixed models gave the lowest MAE values. For the container ships, mixed effect linear models, generalized linear models and artificial neural networks performed better than GLMM in design, ballast and heavy loaded cases, respectively.

In the light of all this information and the results table, it can be said that longitudinal data approaches have generally made good predictions in this particular problem. A reason for the mixed effect linear model and generalized linear mixed model to perform very well on this data set may be that the structure of our data seems to make the "longitudinal" approach most successful, and the usage of such special methods for this type of data has caused the error rate to be low. This study can be considered as a different approach since there are not enough studies on the use of this method to predict ship resistance. At the same time, this thesis proposed a useful model for predicting ship resistance. Among the longitudinal data approaches, GLMM made better predictions than MELM, although there was no critical difference between their MAE values. For the very detailed comparison among all ship types and loading conditions, see the row based heat map and ship type and loading condition based mean absolute relative errors given in Appendix C. There are also the mean absolute errors of the statistical techniques for each individual hull model and loading condition.

As can be seen from the Table 6.2, the Holtrop & Mennen method generally gives relatively more accurate results in ballast loading conditions compared to other loading conditions. In addition, for bulk carriers and tankers heavy loaded conditions, it gives more adequate predictions than in design loading condition, this is the opposite in general cargo and container ships. The mean absolute error value of the Holtrop & Mennen method is 2.74, which is much larger than the results of built statistical methods given in Table 6.1. The ship type based MAE values of the Holtrop & Mennen method is given in Table 6.3.

Table 6.3. Ship type based Holtrop and Mennen method results

	MAE of Holtrop & Mennen
Bulk Carrier	3.45
General Cargo Ship	2.23
Container Ship	2.36
Tanker	2.96
OVERALL	**2.74**

As it can be seen from the Table 6.3, Holtrop and Mennen method performs relatively well on general cargo and container ships compared to bulk carriers and tankers. However, in general, Holtrop and Mennen technique established on the basis of very old ship models (until 1982) is not sufficient to predict hull total resistance values of the models in the data accurately. Despite this, the Holtrop and Mennen method catches the trend in estimation of the hull resistance of many ship models, but suffers from overestimate or under estimate.

The results of GLMM, MELM, ANN and Holtrop & Mennen applications and the experimental results are drawn on the same graphs for each loading condition of ships with the mean absolute error of GLMMs. The following four charts show the average performance of generalized linear mixed models selecting one ship and loading condition for each of the 4 different ship types.

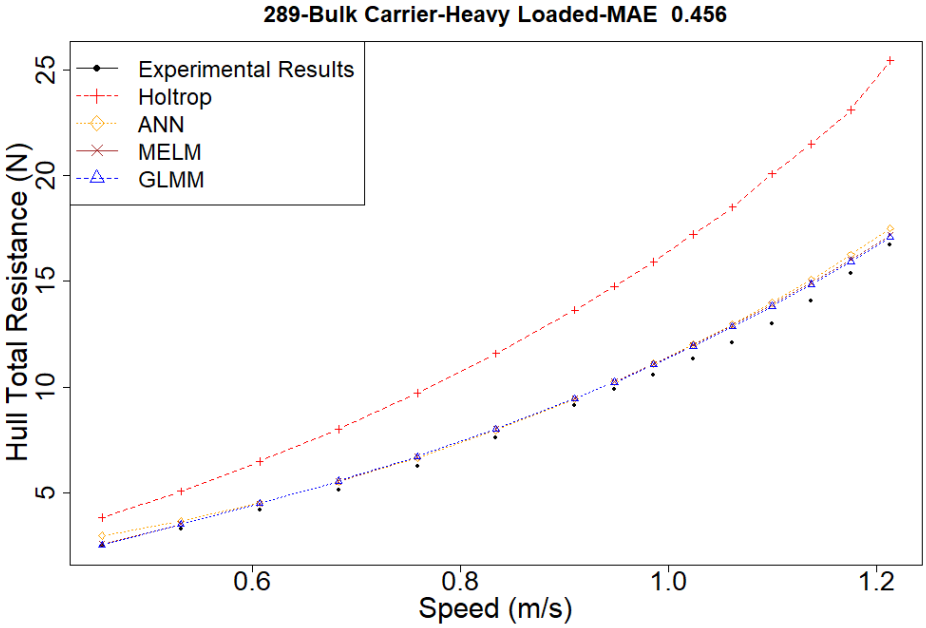

Figure 6.4. Resistance Estimation of Model No: 289, Bulk Carrier, Loading Condition: Heavy Loaded

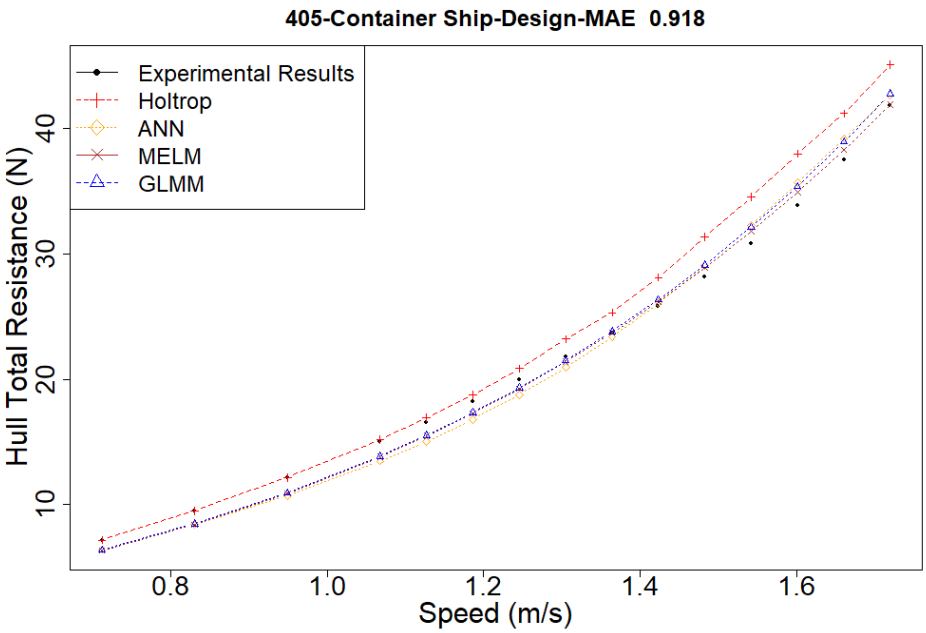

Figure 6.5. Resistance Estimation of Model No: 405, Container Ship, Loading Condition: Design

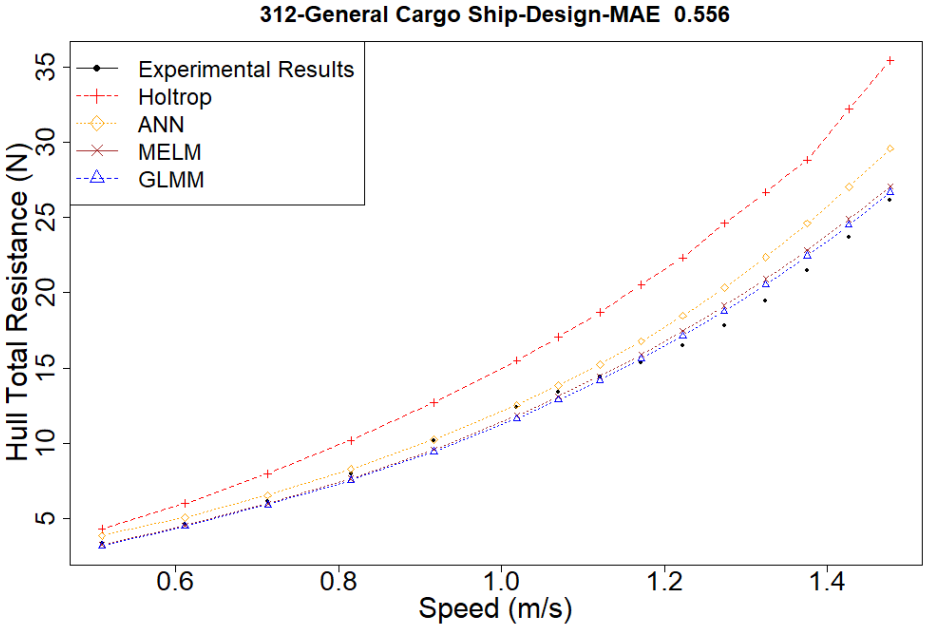

Figure 6.6. Resistance Estimation of Model No: 312, General Cargo Ship, Loading Condition: Design

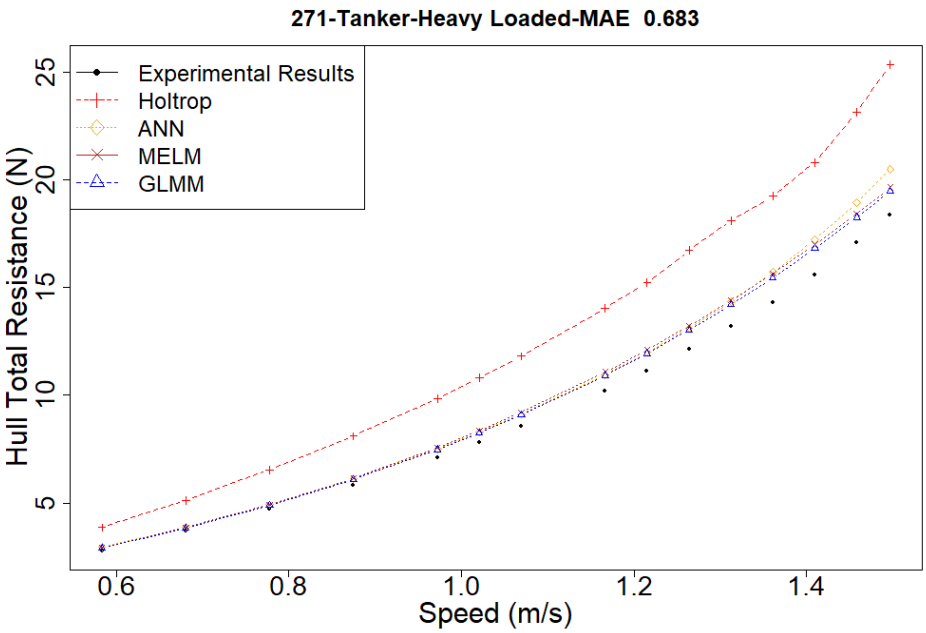

Figure 6.7. Resistance Estimation of Model No: 271, Tanker, Loading Condition: Heavy Loaded

Additionally, the prediction results of the tanker ship with model no 363 which is one of the best predicted models of GLMM and container ship numbered 278, which is one of the worst predictions, are as follows:

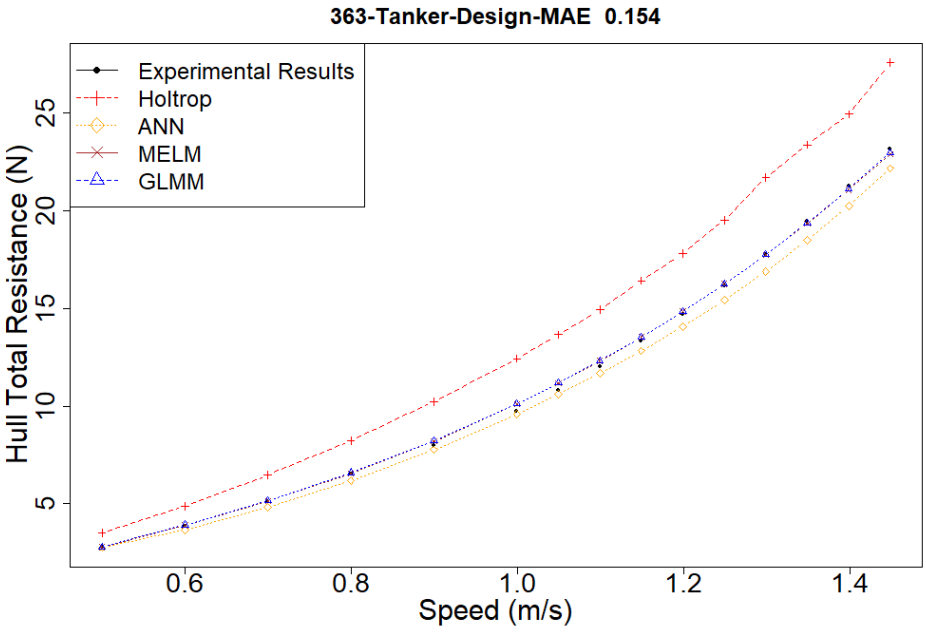

Figure 6.8. Resistance Estimation of Model No: 363, Tanker, Loading Condition: Design

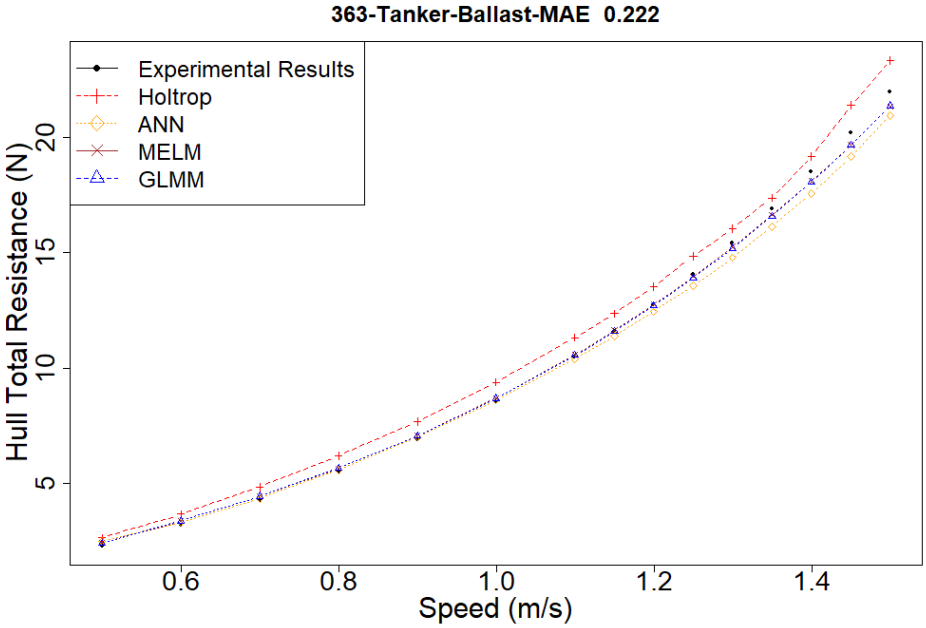

Figure 6.9. Resistance Estimation of Model No: 363, Tanker, Loading Condition: Ballast

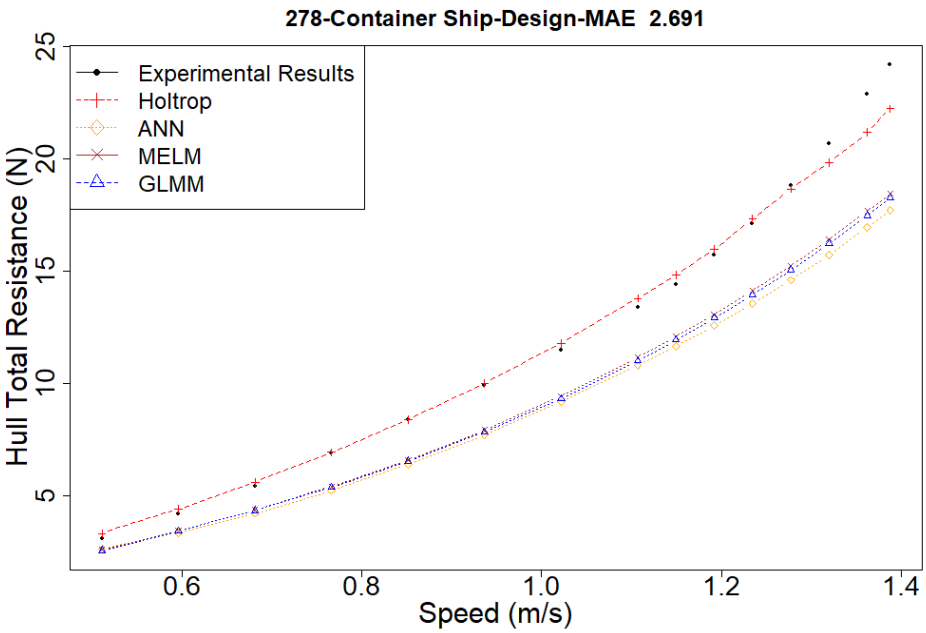

Figure 6.10. Resistance Estimation of Model No: 278, Container Ship, Loading Condition: Design

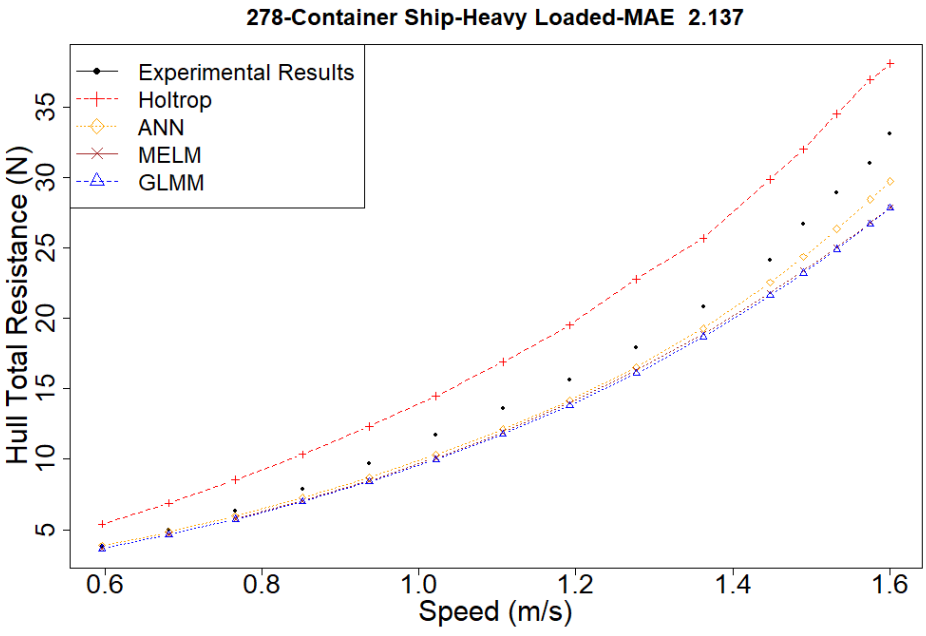

Figure 6.11. Resistance Estimation of Model No: 278, Container Ship, Loading Condition: Heavy Loaded

7. CONCLUSIONS AND FUTURE WORK

Within the scope of this study, the test experimental results obtained from the ITU Ata Nutku Ship Model Testing Laboratory were examined in detail, and the results of the tests together with the hydrostatic values of the ships were regularly drawn from the reports related to cargo ships containing bulk carriers, tankers, container and general cargo ships. Then, on this regularly prepared cargo ship data, the results were obtained by applying the method of Holtrop & Mennen, which is widely used and accepted in the field of naval architecture and marine engineering for the initial design stage. By establishing statistical models on this data set, it was tried to estimate the hull resistance of the ship directly. To solve this problem, firstly the improved version of the standard linear model, generalized linear models were used, and then artificial neural networks, which have been used for this problem recently, have been established. After that, because the data in the resistance measurement reports of the ships are longitudinal data, the problem was addressed with a kind of longitudinal data approach; mixed effect linear model. Finally, the combination of generalized linear and mixed effect linear models, generalized linear mixed models were tried to deal with this particular forecasting problem. Once all the statistical models have been built, the one that will best address this problem has been identified with leave one ship out cross validation. As a result of all calculations, it was found that the generalized linear mixed model, which was not widely used for this problem before, was quite successful on this particular problem. In addition, another tried longitudinal data approach, mixed effect linear models, also gave well estimations close to GLMMs. A reason for this situation may be that each observation is not completely independent from each other in the data in this type of problem. Considering the connection of the variables with each other, statistical models should be established. In another saying, independence assumption should be relaxed since this data set has the structure of repeated measurements and observations are not totally independent.

In order to make other statistical models, especially ANN, perform better on this problem, studies with data sets containing more observations can be done. In this thesis, many studies have been done for improving the model, and finding a well performed parameter sets for ANN by using various validation types. The same care has been exhibited for other statistical models, but in this particular problem, longitudinal data approaches have given the most successful results. While ANNs are expected to give more accurate results on this problem, the situation developed differently. Therefore, it would not be correct to say that ANN is more successful than other statistical problems in all complex problems. Especially in ship resistance prediction problems, ANN is widely used, but preferring longitudinal data approaches can give useful results for this particular problem because of its longitudinal nature. One of the reasons for the ANN model to perform relatively poorly may be that the data set at hand is not very large. In addition to the comparison with ANNs, longitudinal data approaches predicted ship resistances much better than the Holtrop & Mennen method commonly used in the initial design phase.

As a result of this study, it can be said that it is a useful way to use a longitudinal data approach such as mixed effect linear model and generalized linear mixed model to estimate the resistance values of ships with repeated measurements data structure. While conducting resistance tests of ships, the ships are tested many times at different speeds under certain loading conditions and resistance measurements are carried out, which means that except speed value, all other variables remain same for several experiments and observations are not completely independent from each other. Therefore, a longitudinal data approach handling each ship with certain loading condition as an individual and assuming that they are dependent is needed. This approach can be useful not only for ship problems but also for problems with similar data structures.

For the future work, based on this particular problem, each resistance component can be handled separately and estimation models can be established to obtain the resistance values. In addition, resistance values of appendages, which are not taken into account in this study and cannot be made because of insufficient data, can also

be calculated and included in the total resistance.

REFERENCES

1. Arnold, T. B., "Yale University Data Mining and Machine Learning, Lecture Notes 12 - 15: Neural Networks", `http://euler.stat.yale.edu/ tba3/stat665/`, accessed in September 2019.

2. Marón, D. and M. Santos, "Aplicación de Redes Neuronales para la Estimación de la Resistencia al Avance de Buques", *Actas de las XXXVIII Jornadas de Automática*, Vol. 38, pp. 393–400, 2017.

3. Holtrop, J. and G. Mennen, "A statistical power prediction method", *International Shipbuilding Progress*, Vol. 25, No. 290, pp. 253–256, 1978.

4. Holtrop, J. and G. Mennen, "An approximate power prediction method", *International Shipbuilding Progress*, Vol. 29, No. 335, pp. 166–170, 1982.

5. Hollenbach, K., "Estimating Resistance and Propulsion for Single Screw and Twin Screw Ships", *Ship Technology Research*, Vol. 45, No. 2, pp. 237–250, 1998.

6. Abdelkhalek, H., D. F. Han, L. T. Gao and Q. Wang, "Numerical estimation of ship resistance using CFD with different turblence model", *Advanced Materials Research*, Vol. 1021, pp. 209–213, 2014.

7. Mason, G., C. R. Smith and B. R. von Konsky, "Artificial neural networks for hull resistance prediction", *3rd Int. Conf. on Computer Applications and Information Technology in the Maritime Industries (COMPIT)*, 2004.

8. Grabowska, K. and P. Szczuko, "Ship resistance prediction with Artificial Neural Networks", *2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 168–173, IEEE, 2015.

9. Hornik, K., M. Stinchcombe, H. White *et al.*, "Multilayer feedforward networks

are universal approximators.", *Neural Networks*, Vol. 2, No. 5, pp. 359–366, 1989.

10. Riedmiller, M. and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm", *IEEE International Conference on Neural Networks*, pp. 586–591, IEEE, 1993.

11. Riedmiller, M., "Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms", *Computer Standards & Interfaces*, Vol. 16, No. 3, pp. 265–278, 1994.

12. Skupień, E. and J. Prokopowicz, "Methods of Calculating Ship Resistance on Limited Waterways", *Polish Maritime Research*, Vol. 21, pp. 12–17, 2015.

13. Wang, S., B. Ji, J. Zhao, W. Liu and T. Xu, "Predicting ship fuel consumption based on LASSO regression", *Transportation Research Part D: Transport and Environment*, Vol. 65, pp. 817–824, 2017.

14. Petersen, J. P., D. J. Jacobsen and O. Winther, "Statistical modelling for ship propulsion efficiency", *Journal of Marine Science and Technology*, Vol. 17, No. 1, pp. 30–39, 2012.

15. Delen, C. and B. Şakir, "Uncertainty analysis of resistance tests in Ata Nutku Ship model testing Laboratory of Istanbul Technical University", *Türk Denizcilik ve Deniz Bilimleri Dergisi*, Vol. 1, No. 2, pp. 69–88, 2015.

16. Walpole, R., *Essentials of Probability & Statistics for Engineers & Scientists*, Pearson, 2013.

17. Nelder, J. A. and R. W. Wedderburn, "Generalized linear models", *Journal of the Royal Statistical Society: Series A (General)*, Vol. 135, No. 3, pp. 370–384, 1972.

18. Günther, F. and S. Fritsch, "neuralnet: Training of neural networks", *The R Journal*, Vol. 2, No. 1, pp. 30–38, 2010.

19. Kumar, A. and D. Zhang, "Personal recognition using hand shape and texture", *IEEE Transactions on Image Processing*, Vol. 15, No. 8, pp. 2454–2461, 2006.

20. Rojas, R., *Neural Networks: A Systematic Introduction*, Springer Science & Business Media, 2013.

21. Chen, W., K. Chiu and M. Fuge, "Aerodynamic design optimization and shape exploration using generative adversarial networks", *AIAA Scitech 2019 Forum*, p. 2351, 2019.

22. Ciaburro, G. and B. Venkateswaran, *Neural Networks with R: Smart Models Using CNN, RNN, Deep Learning, and Artificial Intelligence Principles*, Packt Publishing Ltd, 2017.

23. Hinton, G., N. Srivastava and K. Swersky, "Neural Networks for Machine Learning Lecture Notes: Overview of mini-batch gradient descent", `https://www.cs.toronto.edu/tijmen/csc321/slides/lecture_slides_lec6`, accessed in February 2020.

24. Schober, P. and T. Vetter, "Repeated Measures Designs and Analysis of Longitudinal Data: If at First You Do Not Succeed—Try, Try Again", *Anesthesia & Analgesia*, Vol. 127, pp. 569–575, 2018.

25. Liang, K.-Y. and S. L. Zeger, "Longitudinal data analysis using generalized linear models", *Biometrika*, Vol. 73, No. 1, pp. 13–22, 1986.

26. Liu, Y. and M. Bottai, "Mixed-ffects models for conditional quantiles with longitudinal data", *International Journal of Biostatistics*, Vol. 5, No. 1, pp. 1–24, 2009.

27. Laird, N. M. and J. H. Ware, "Random-effects models for longitudinal data", *Biometrics*, Vol. 38, No. 4, pp. 963–974, 1982.

28. Ochi, Y. e. and R. L. Prentice, "Likelihood inference in a correlated probit regression model", *Biometrika*, Vol. 71, No. 3, pp. 531–543, 1984.

29. Baltagi, B., *Econometric Analysis of Panel Data*, John Wiley & Sons, 2008.

30. Pinheiro, J. and D. Bates, *Mixed-Effects Models in S and S-PLUS*, Springer Science & Business Media, 2006.

31. Bates, D., M. Mächler, B. Bolker and S. Walker, "Fitting linear mixed-effects models using lme4", *arXiv preprint arXiv:1406.5823*, 2014.

32. Bates, D., M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, F. Scheipl, G. Grothendieck, P. Green *et al.*, "Package 'lme4"', *Version*, Vol. 1, p. 17, 2018.

33. Kachman, S. D., "An introduction to generalized linear mixed models", *Proceedings of a Symposium at the Organizational Meeting for a NCR Coordinating Committee on "Implementation Strategies for National Beef Cattle Evaluation"*, pp. 59–73, Athens, 2000.

34. Gbur, E. E., W. W. Stroup, K. S. McCarter, S. Durham, L. J. Young, M. Christman, M. West and M. Kramer, *Generalized linear mixed models*, ASA, CSSA, and SSSA, Madison, WI, 2012.

35. Fitzmaurice, G. M., N. M. Laird and J. H. Ware, *Applied Longitudinal Analysis*, John Wiley & Sons, 2012.

36. Pinheiro, J. C. and E. C. Chao, "Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models", *Journal of Computational and Graphical Statistics*, Vol. 15, No. 1, pp. 58–81, 2006.

37. Liu, X., *Methods and Applications of Longitudinal Data Analysis*, Elsevier, 2015.

38. Harville, D. A., "Maximum likelihood approaches to variance component estimation and to related problems", *Journal of the American Statistical Association*, Vol. 72, No. 358, pp. 320–338, 1977.

39. Tuerlinckx, F., F. Rijmen, G. Verbeke and P. De Boeck, "Statistical inference in generalized linear mixed models: A review", *British Journal of Mathematical and Statistical Psychology*, Vol. 59, No. 2, pp. 225–255, 2006.

40. Stone, M., "Cross-validatory choice and assessment of statistical predictions", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 36, No. 2, pp. 111–133, 1974.

41. Theodoridis, S. and K. Koutroumbas, *Pattern Recognition & Matlab Intro*, Academic Press, Inc., 2010.

42. Akbulut, O., "Feature Normalization Effect in Emotion Classification based on EEG Signals", *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, Vol. 24, No. 1, pp. 60–66, 2020.

43. Chen, X., A. Y. Aravkin and R. D. Martin, "Generalized linear model for gamma distributed variables via elastic net regularization", *arXiv preprint arXiv:1804.07780*, 2018.

44. Mao, K., "Fast orthogonal forward selection algorithm for feature subset selection", *IEEE Transactions on Neural Networks*, Vol. 13, No. 5, pp. 1218–1224, 2002.

45. Smith, A., O. Mendoza-Schrock, S. Kangas, M. Dierking and A. Shaw, "An end-to-end vechicle classification pipeline using vibrometry data", *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR V*, Vol. 9079, p. 90790O, International Society for Optics and Photonics, 2014.

46. Hornik, K., "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, Vol. 4, No. 2, pp. 251–257, 1991.

47. Almeida, C., C. Baugh, C. Lacey, C. Frenk, G. Granato, L. Silva and A. Bressan, "Modelling the dusty universe–I. Introducing the artificial neural network and first applications to luminosity and colour distributions", *Monthly Notices of the Royal Astronomical Society*, Vol. 402, No. 1, pp. 544–564, 2010.

48. Helwig, N. E., "Efficient estimation of variance components in nonparametric mixed-effects models with large samples", *Statistics and Computing*, Vol. 26, No. 6, pp. 1319–1336, 2016.

APPENDIX A: BASIC INFORMATION FOR CARGO SHIPS

A.1. Bulk Carriers

Bulk carriers are vessels that can carry dry bulk cargoes such as unpackaged ore, scrap metal, and grain.

Figure A.1. Bulk Carrier, Tigris

A.2. Tankers

Tankers are special cargo ships designed to transport liquid or gaseous cargoes. In particular, oil transportation in the world is carried out with these ships.

Figure A.2. Tanker, New Medal

A.3. General Cargo Ships

General cargo ships are capable of transporting a variety of goods in different forms such as boxed, palletized, refrigerated and also capable of carrying bulk materials such as grain.

Figure A.3. General Cargo Ship, Melody

A.4. Container Ships

A container ship is a freight ship designed to transport boxes called containers in international standards. It is mostly preferred in international commercial freight transportation because it reduces costs compared to other transportation vehicles. Their carriage capacities are expressed by TEU, which means that how many containers they carry with international standards. As an example, 1500 TEU container ship can carry up to 1500 standard containers.

Figure A.4. Container Ship, CMA CGM Marko Polo

APPENDIX B: SUMMARY STATISTICS

```
 Ship_Type         Loading_Condition       LBP             LWL             BWL              Ts              TA
Length:1977        Length:1977        Min.   :3.533   Min.   :3.586   Min.   :0.5090   Min.   :0.0950   Min.   :0.1090
Class :character   Class :character   1st Qu.:4.000   1st Qu.:4.072   1st Qu.:0.6000   1st Qu.:0.1690   1st Qu.:0.1930
Mode  :character   Mode  :character   Median :4.068   Median :4.153   Median :0.6650   Median :0.2200   Median :0.2320
                                      Mean   :4.085   Mean   :4.159   Mean   :0.6538   Mean   :0.2155   Mean   :0.2289
                                      3rd Qu.:4.187   3rd Qu.:4.261   3rd Qu.:0.7000   3rd Qu.:0.2640   3rd Qu.:0.2710
                                      Max.   :4.871   Max.   :4.988   Max.   :0.8510   Max.   :0.3110   Max.   :0.3110
      TF                N                D               AWS               AR               AA               AB
Min.   :0.0760   Min.   :0.1675   Min.   :0.1675   Min.   :2.168   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
1st Qu.:0.1370   1st Qu.:0.3060   1st Qu.:0.3060   1st Qu.:3.147   1st Qu.:0.04400   1st Qu.:0.0770   1st Qu.:0.00700
Median :0.2140   Median :0.4280   Median :0.4280   Median :3.715   Median :0.05280   Median :0.1040   Median :0.01400
Mean   :0.2021   Mean   :0.4457   Mean   :0.4457   Mean   :3.697   Mean   :0.05312   Mean   :0.1025   Mean   :0.01352
3rd Qu.:0.2640   3rd Qu.:0.5610   3rd Qu.:0.5610   3rd Qu.:4.183   3rd Qu.:0.06000   3rd Qu.:0.1300   3rd Qu.:0.01900
Max.   :0.3110   Max.   :0.7560   Max.   :0.7560   Max.   :5.208   Max.   :0.12600   Max.   :0.1870   Max.   :0.03300
      HB                AT                HT               CB               CP               CM              CWP
Min.   :0.0000   Min.   :0.000000   Min.   :0.0000   Min.   :0.5040   Min.   :0.5990   Min.   :0.8300   Min.   :0.2150
1st Qu.:0.0980   1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.:0.6790   1st Qu.:0.6950   1st Qu.:0.9810   1st Qu.:0.8310
Median :0.1220   Median :0.001000   Median :0.1750   Median :0.7520   Median :0.7610   Median :0.9910   Median :0.8810
Mean   :0.1185   Mean   :0.004634   Mean   :0.1336   Mean   :0.7372   Mean   :0.7492   Mean   :0.9816   Mean   :0.8595
3rd Qu.:0.1470   3rd Qu.:0.006537   3rd Qu.:0.2520   3rd Qu.:0.7940   3rd Qu.:0.8010   3rd Qu.:0.9950   3rd Qu.:0.9070
Max.   :0.2670   Max.   :0.036000   Max.   :0.3000   Max.   :0.9280   Max.   :0.9320   Max.   :0.9990   Max.   :0.9840
      LCB               LCF                Fr              Bulb        Number_of_Propeller       Vs          Hull_Total_Res
Min.   :-0.8400   Min.   :-0.2660   Min.   :0.06324   Min.   :0.0000   Min.   :1.000       Min.   :0.4116   Min.   : 1.61
1st Qu.:-0.0720   1st Qu.:-0.1560   1st Qu.:0.13062   1st Qu.:1.0000   1st Qu.:1.000       1st Qu.:0.8290   1st Qu.: 6.11
Median :-0.0110   Median :-0.0960   Median :0.16942   Median :1.0000   Median :1.000       Median :1.0848   Median :10.34
Mean   : 0.2875   Mean   : 0.2238   Mean   :0.16866   Mean   :0.8897   Mean   :1.235       Mean   :1.0754   Mean   :11.52
3rd Qu.: 0.0510   3rd Qu.: 0.0060   3rd Qu.:0.20512   3rd Qu.:1.0000   3rd Qu.:1.000       3rd Qu.:1.3124   3rd Qu.:15.56
Max.   : 2.1790   Max.   : 2.1220   Max.   :0.29264   Max.   :1.0000   Max.   :2.000       Max.   :1.7939   Max.   :42.24
```

Figure B.1. The summary of the features in cargo ship data

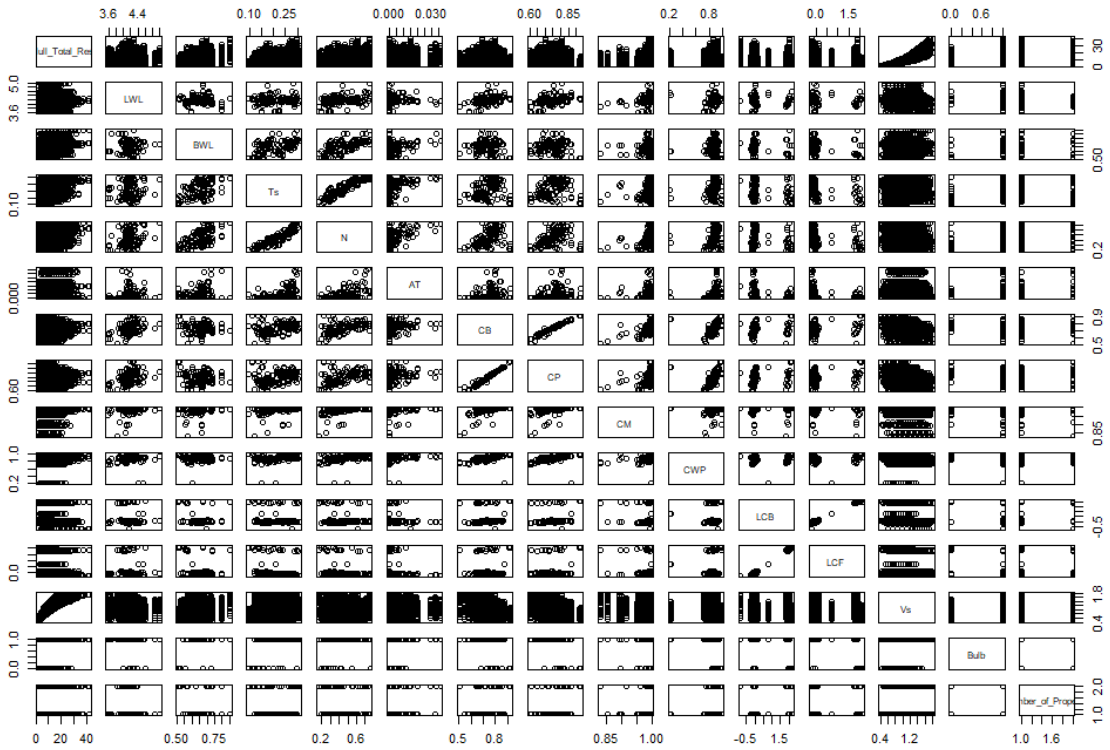

Figure B.2. The relations of the features with the logarithm of hull resistance

APPENDIX C: DETAILED RESULTS

Ship Type	Loading Condition	Mean Absolute Relative Error				
		GLM	ANN	MELM	GLMM	Holtrop
Bulk Carrier	Design	0.0618	0.0785	0.0565	0.0527	0.5759
	Ballast	0.0585	0.0778	0.0577	0.0551	0.2445
	Heavy Loaded	0.0616	0.0831	0.054	0.0522	0.5216
General Cargo Ships	Design	0.0589	0.0704	0.0678	0.0601	0.3086
	Ballast	0.0604	0.0629	0.0619	0.0535	0.1466
	Heavy Loaded	0.0377	0.0477	0.0346	0.0336	0.4316
Container Ship	Design	0.0724	0.0854	0.07	0.0705	0.1846
	Ballast	0.0533	0.0722	0.059	0.0552	0.2497
	Heavy Loaded	0.0544	0.0577	0.0548	0.0548	0.2024
Tanker	Design	0.0609	0.0658	0.0559	0.0546	0.3028
	Ballast	0.0685	0.0701	0.0617	0.0623	0.1737
	Heavy Loaded	0.0726	0.066	0.0541	0.0488	0.2858

Figure C.1. Row based detailed heatmap of final predictions

Table C.1. Ship type and loading condition based mean absolute relative error comparisons

Ship Type	**Loading**	**Mean Absolute Relative Error**				
		GLM	ANN	MELM	GLMM	Holtrop
Bulk Carrier	Design	0.0618	0.0785	0.0565	0.0527	0.5759
	Ballast	0.0585	0.0778	0.0577	0.0551	0.2445
	Heavy Loaded	0.0616	0.0831	0.0540	0.0522	0.5216
General Cargo	Design	0.0589	0.0704	0.0678	0.0601	0.3086
	Ballast	0.0604	0.0629	0.0619	0.0535	0.1466
	Heavy Loaded	0.0377	0.0477	0.0346	0.0336	0.4316
Container Ship	Design	0.0724	0.0854	0.0700	0.0705	0.1846
	Ballast	0.0533	0.0722	0.0590	0.0552	0.2497
	Heavy Loaded	0.0544	0.0577	0.0548	0.0548	0.2024
Tanker	Design	0.0609	0.0658	0.0559	0.0546	0.3028
	Ballast	0.0685	0.0701	0.0617	0.0623	0.1737
	Heavy Loaded	0.0726	0.0660	0.0541	0.0488	0.2858

Table C.2. All predictions for each model no and loading condition

No-Type-Loading	**Mean Absolute Error**				
	GLM	**ANN**	**MELM**	**GLMM**	**Holtrop**
424-General Cargo-Design	0.431	0.814	1.445	0.465	0.592
424-General Cargo-Ballast	0.48	0.987	1.692	0.494	1.307
420-General Cargo-Design	1.574	1.63	1.44	1.46	2.328
420-General Cargo-Ballast	0.811	0.925	0.752	0.76	2.868
416-General Cargo-Design	0.133	0.323	0.331	0.375	1.027
416-General Cargo-Ballast	0.328	0.204	0.653	0.7	1.088
409-Tanker-Design	0.451	0.436	0.373	0.36	1.314
409-Tanker-H.Loaded	0.421	0.908	0.358	0.337	1.31
409-Tanker-Ballast	0.537	0.637	0.524	0.537	0.969
405-Container-Design	0.698	1.134	0.763	0.918	1.688
398-General Cargo-H.Loaded	0.096	0.286	0.093	0.118	3.25
398-General Cargo-Design	0.323	0.233	0.359	0.354	3.084
398-General Cargo-Ballast	0.351	0.213	0.094	0.142	0.088
397-Tanker-H.Loaded	1.182	0.82	0.549	0.249	1.793
397-Tanker-Design	0.429	0.266	0.233	0.266	1.742
397-Tanker-Ballast	0.546	0.436	0.087	0.109	2.772
397-Tanker-Ballast (Lighter)	0.781	0.46	0.314	0.321	0.449
385-Tanker-Design	1.092	1.378	1.205	1.267	0.74
385-Tanker-Ballast	1.423	1.339	1.151	1.569	0.572
385M-Tanker-Design	1.122	1.469	1.221	1.319	0.851
385M-Tanker-Ballast	1.314	1.153	1.046	1.456	0.712
382-Tanker-Design	1.565	1.443	1.639	1.698	1.381
382-Tanker-Ballast	1.814	1.438	1.848	1.835	1.48
255-Container-Design	0.772	0.705	0.591	0.625	2.921
255-Container-H.Loaded	0.489	0.846	0.39	0.392	3.104
255-Container-Ballast	0.687	0.395	0.534	0.685	5.841
257-Container-Design	1.111	0.572	1.158	1.162	2.918

Table C.2. All predictions for each model no and loading condition (cont.)

No-Type-Loading	**Mean Absolute Error**				
	GLM	**ANN**	**MELM**	**GLMM**	**Holtrop**
257-Container-H.Loaded	0.472	0.381	0.514	0.495	3.397
257-Container-Ballast	1.042	0.828	1.038	1.22	1.549
260-Container-Design	0.768	0.635	0.796	0.696	0.932
260-Container-H.Loaded	0.481	0.38	0.481	0.414	0.631
260-Container-Ballast	0.588	0.645	0.809	0.583	1.205
260B-Container-Design	0.627	0.638	0.641	0.623	0.952
260B-Container-H.Loaded	0.416	0.321	0.381	0.344	0.686
260B-Container-Ballast	0.842	0.877	1.031	0.844	0.916
265-Tanker-Design	0.507	1.169	0.415	0.418	6.277
265-Tanker-H.Loaded	0.776	0.987	0.676	0.681	5.817
265-Tanker-Ballast	0.962	0.364	0.836	0.829	3.046
266B-Tanker-Design	0.327	0.308	0.206	0.171	3.325
266B-Tanker-Ballast	0.476	0.445	0.243	0.247	1.417
269B-Tanker-H.Loaded	2.131	1.665	1.617	1.552	2.047
269B-Tanker-Design	2.294	1.888	1.728	1.656	3.153
269B-Tanker-Ballast	3.373	2.977	3.249	2.92	1.274
269-Tanker-Design	0.722	0.973	1.107	1.036	4.25
269-Tanker-Ballast	0.937	1.014	0.858	0.833	3.316
271-Tanker-Design	0.797	0.844	0.696	0.583	2.9
271-Tanker-H.Loaded	0.993	0.878	0.811	0.683	3.753
271-Tanker-Ballast	0.869	0.961	0.744	0.688	1.21
273-Tanker-Design	0.226	0.283	0.24	0.201	2.481
273-Tanker-H.Loaded	0.172	0.163	0.156	0.122	2.606
273-Tanker-Ballast	0.239	0.24	0.353	0.356	0.239
274N-Tanker-Design	1.742	2.4	1.52	1.543	9.107
274-Tanker-Design	1.385	1.188	1.477	1.396	6.431

Table C.2. All predictions for each model no and loading condition (cont.)

No-Type-Loading	**Mean Absolute Error**				
	GLM	**ANN**	**MELM**	**GLMM**	**Holtrop**
274-Tanker-Ballast	0.289	0.261	0.39	0.252	1.386
276-Tanker-Design	1.696	1.814	1.506	1.496	5.615
276-Tanker-Ballast	1.001	0.913	1.214	0.953	1.97
277-Tanker-Design	0.725	0.609	0.751	0.743	5.741
277-Tanker-Ballast	0.486	0.4	0.467	0.463	0.281
278-Container-H.Loaded	2.232	1.452	2.024	2.137	3.873
278-Container-Design	2.69	2.955	2.577	2.691	0.47
279-Tanker-Design	1.667	1.391	1.78	1.765	2.132
279-Tanker-Ballast	0.689	0.521	0.848	0.866	0.34
280-Tanker-Design	0.96	0.635	0.946	0.842	4.878
280-Tanker-Ballast	0.886	1.232	0.791	0.921	3.678
281-Tanker-H.Loaded	0.984	0.926	0.569	0.561	8.364
281-Tanker-Design	0.681	0.628	0.427	0.405	7.892
281-Tanker-Ballast	0.512	0.955	0.566	0.62	2.016
284-Tanker-Design	0.562	0.325	0.466	0.4	3.954
284-Tanker-Ballast	0.63	0.33	0.325	0.239	1.127
289-Bulk Carrier-H.Loaded	0.672	0.582	0.495	0.456	4.911
289-Bulk Carrier-Design	0.823	0.7	0.759	0.703	4.93
289-Bulk Carrier-Ballast	0.309	0.246	0.295	0.296	2.012
291-Tanker-Design	1.197	0.144	1.137	1.142	2.266
291-Tanker-H.Loaded	1.619	1.404	1.553	1.576	1.582
291-Tanker-Ballast	1.836	1.843	2.018	2.13	0.995
293-Container-Design	0.672	0.979	0.424	0.318	2.948
293-Container-Ballast	0.206	0.706	0.152	0.136	0.585
294-Tanker-Design	0.369	0.211	0.253	0.222	3.495
294-Tanker-H.Loaded	0.3	0.236	0.23	0.213	3.685

Table C.2. All predictions for each model no and loading condition (cont.)

No-Type-Loading	**Mean Absolute Error**				
	GLM	**ANN**	**MELM**	**GLMM**	**Holtrop**
294-Tanker-Ballast	0.436	0.442	0.446	0.474	1.231
297-Tanker-Design	0.781	0.382	0.476	0.71	3.387
297-Tanker-H.Loaded	0.534	0.439	0.397	0.38	4.343
301-Bulk Carrier-Design	0.347	0.481	0.357	0.352	5.055
301-Bulk Carrier-H.Loaded	0.352	0.585	0.34	0.338	3.621
301-Bulk Carrier-Ballast	0.399	0.533	0.401	0.367	1.079
301M-Bulk Carrier-Design	0.321	0.5	0.362	0.376	5.314
302-Tanker-Design	0.774	2.191	0.894	0.825	5.444
302-Tanker-Ballast	0.504	0.708	0.491	0.404	1.987
303-Container-Design	0.735	1.298	0.865	0.918	2.704
303-Container-H.Loaded	0.754	0.829	0.887	0.841	3.686
303-Container-Ballast	0.396	0.35	0.406	0.382	4.542
304-Tanker-Design	1.094	0.916	0.867	0.777	6.029
304-Tanker-Ballast	1.122	1.279	0.97	0.811	2.906
306-General Cargo-Design	0.393	0.315	0.276	0.281	4.196
306-General Cargo-Ballast	0.464	0.377	0.196	0.196	0.555
309-Tanker-Design	0.604	0.618	0.684	0.599	3.862
309-Tanker-Ballast	0.261	0.256	0.29	0.303	2.446
311-General Cargo-H.Loaded	0.509	0.539	0.531	0.471	3.869
311-General Cargo-Design	0.779	0.765	0.83	0.737	3.771
311-General Cargo-Ballast	0.241	0.376	0.234	0.134	0.668
312-General Cargo-Design	0.515	1.466	0.655	0.556	4.682
312-General Cargo-Ballast	0.515	0.44	0.597	0.672	2.093
313-General Cargo-Design	0.478	0.399	0.499	0.437	3.507
313-General Cargo-H.Loaded	0.618	0.579	0.568	0.501	3.855
313-General Cargo-Ballast	0.362	0.335	0.232	0.193	1.316

Table C.2. All predictions for each model no and loading condition (cont.)

No-Type-Loading	**Mean Absolute Error**				
	GLM	**ANN**	**MELM**	**GLMM**	**Holtrop**
316-Tanker-Design	1.118	0.716	0.761	0.66	6.601
316-Tanker-Ballast	0.766	0.336	0.535	0.473	2.525
320-Tanker-Design	0.852	0.615	0.899	0.876	3.063
320-Tanker-Ballast	0.769	0.743	0.827	0.788	3.845
322-Bulk Carrier-Design	0.801	0.97	0.62	0.51	3.475
322-Bulk Carrier-Ballast	0.262	0.602	0.288	0.237	0.915
338-Tanker-Design	0.386	0.644	0.414	0.439	7.148
338-Tanker-Ballast	0.763	1.056	0.852	0.882	4.039
345-General Cargo-Design	0.366	0.676	0.647	0.648	3.103
345-General Cargo-Ballast	0.884	1.275	1.059	1.112	0.369
346-General Cargo-Design	0.378	0.806	0.272	0.329	3.554
346-General Cargo-Ballast	1.518	1.844	0.778	0.674	1.008
357-General Cargo-Design	0.437	0.309	0.389	0.336	3.552
357-General Cargo-Ballast	0.296	0.339	0.269	0.29	0.498
363-Tanker-Design	0.321	0.5	0.155	0.154	2.735
363-Tanker-Ballast	0.275	0.408	0.215	0.222	0.75
256-Tanker-Design	0.27	0.311	0.193	0.253	2.718
305-Tanker-Design	0.795	0.642	0.768	0.734	3.891
305-Tanker-Ballast	0.553	0.862	0.416	0.343	0.825
321-Tanker-Design	0.781	0.855	0.69	0.592	3.962
321-Tanker-Ballast	0.396	0.455	0.301	0.521	0.469
324-General Cargo-Design	0.9	0.802	0.877	0.861	2.428
324-General Cargo-Ballast	0.732	0.674	0.635	0.648	0.522
334-Container-Design	0.665	0.817	0.833	0.85	4.292
334-Container-Ballast	0.833	1.133	0.878	0.921	1.986
352-General Cargo-Design	0.723	1.352	0.77	0.735	4.276

Table C.2. All predictions for each model no and loading condition (cont.)

No-Type-Loading	**Mean Absolute Error**				
	GLM	**ANN**	**MELM**	**GLMM**	**Holtrop**
352-General Cargo-Ballast	0.239	0.353	0.199	0.212	1.178

APPENDIX D: EXAMPLE PREDICTIONS AND HISTOGRAMS

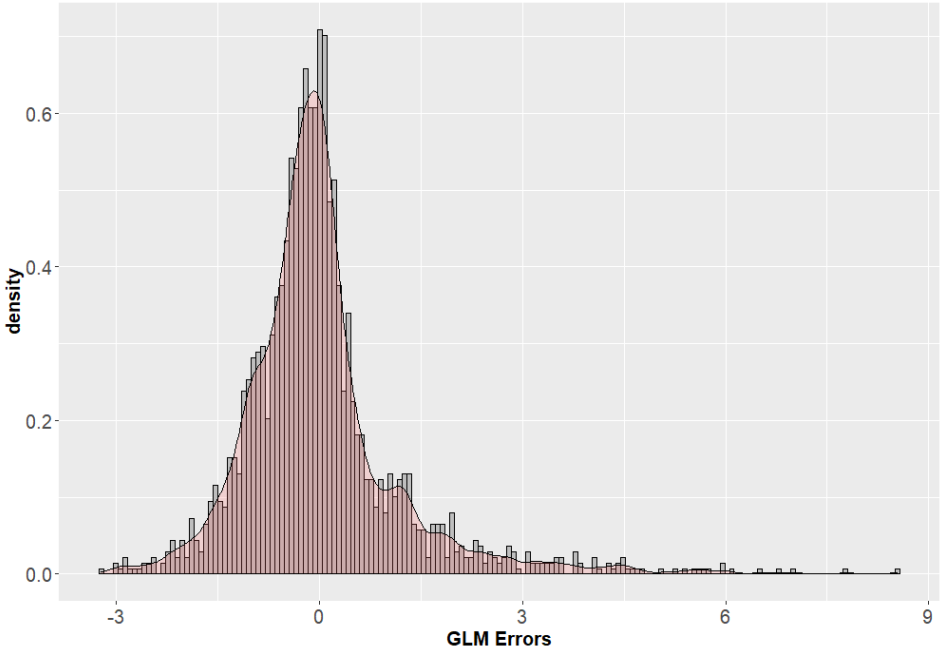

Figure D.1. Histogram of GLM errors

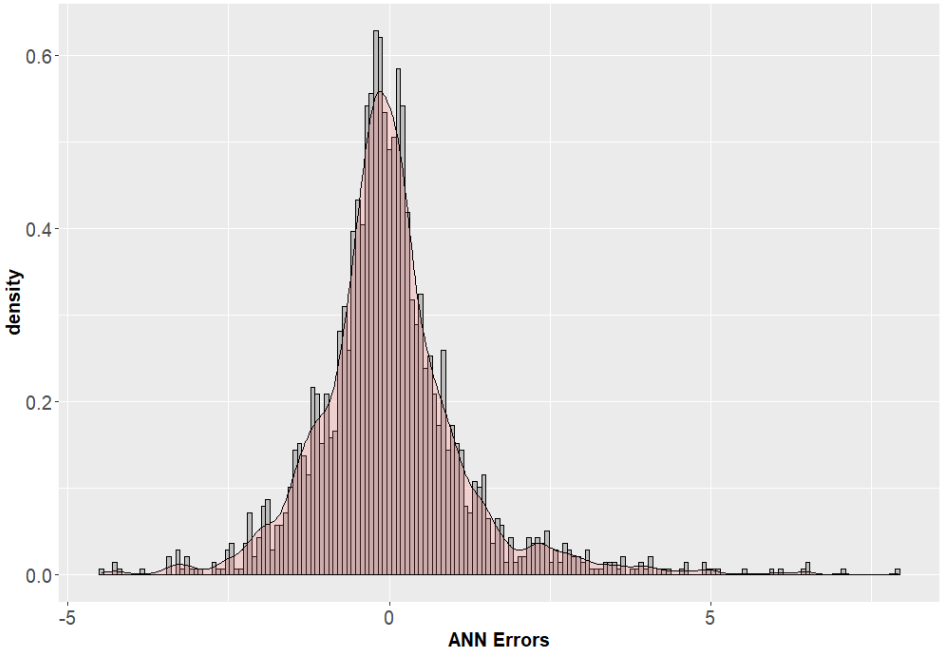

Figure D.2. Histogram of ANN errors

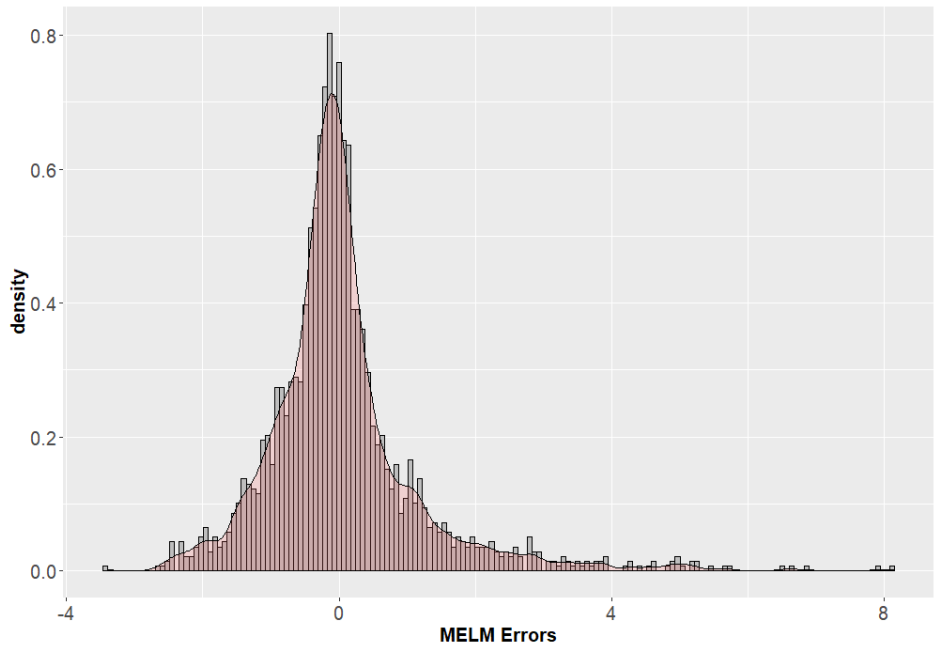

Figure D.3. Histogram of MELM errors

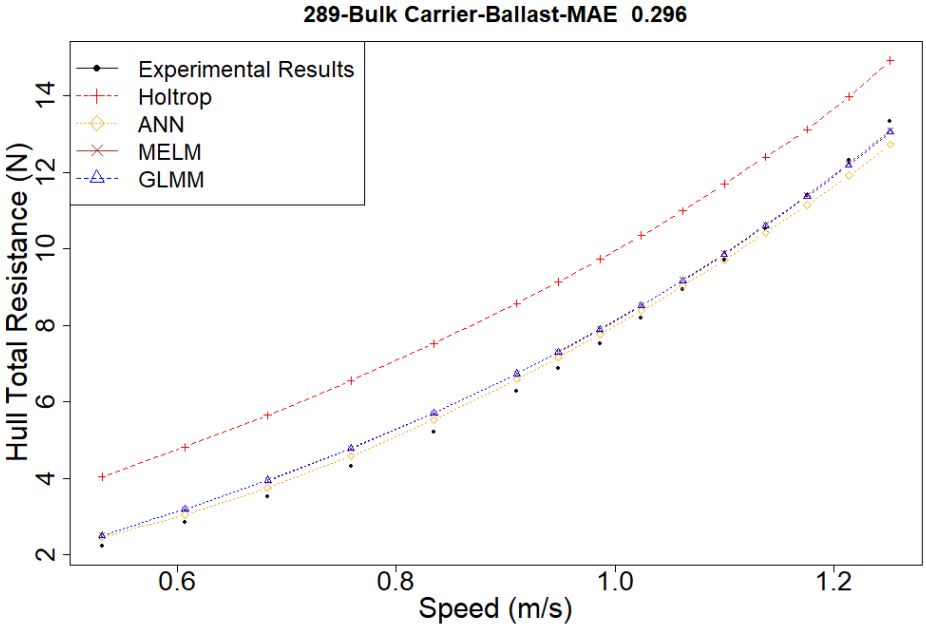

Figure D.4. Resistance Estimation of Model No: 289, Bulk Carrier, Loading Condition: Ballast

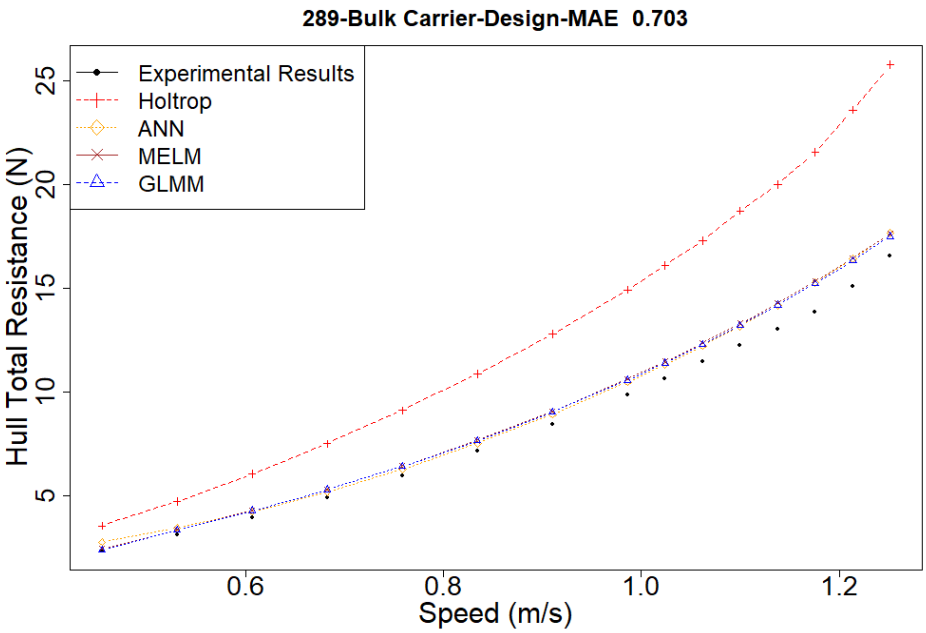

Figure D.5. Resistance Estimation of Model No: 289, Bulk Carrier, Loading Condition: Design

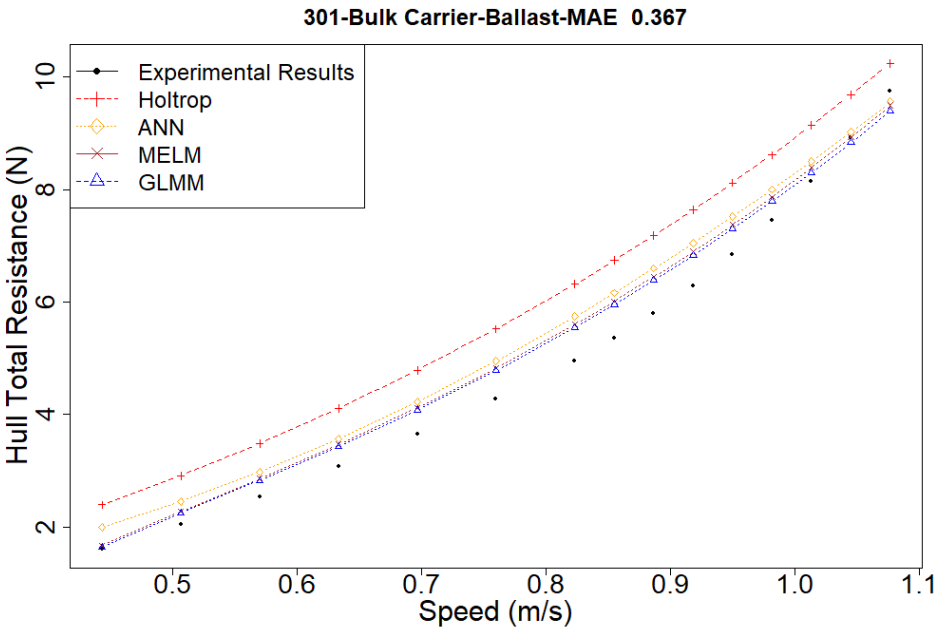

Figure D.6. Resistance Estimation of Model No: 301, Bulk Carrier, Loading Condition: Ballast

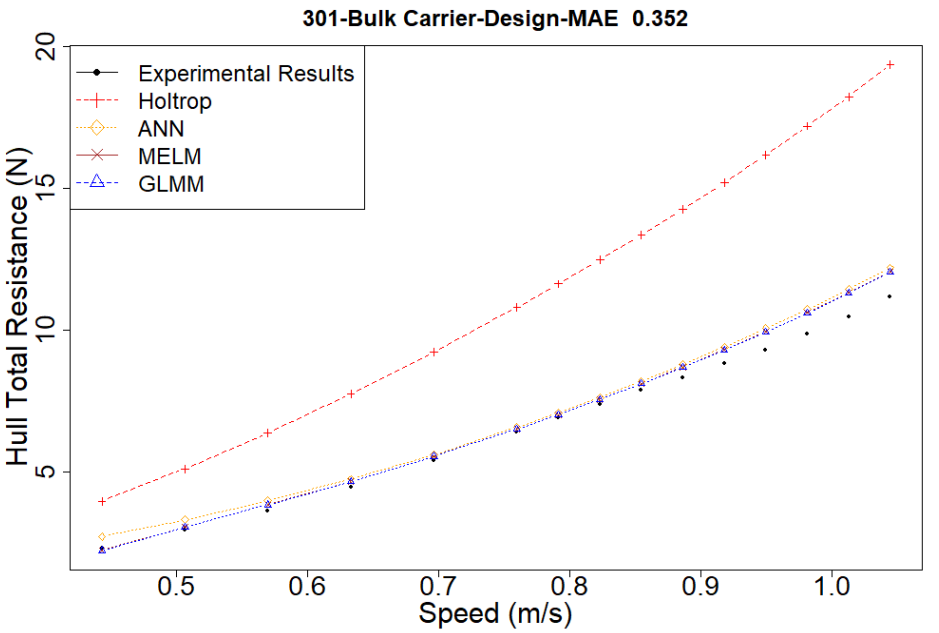

Figure D.7. Resistance Estimation of Model No: 301, Bulk Carrier, Loading Condition: Design

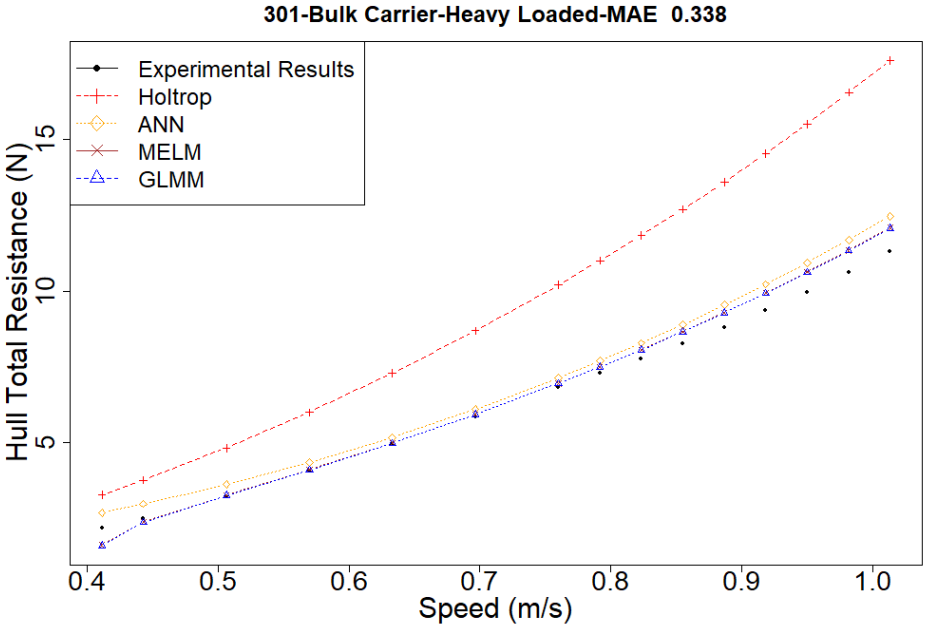

Figure D.8. Resistance Estimation of Model No: 301, Bulk Carrier, Loading Condition: Heavy Loaded

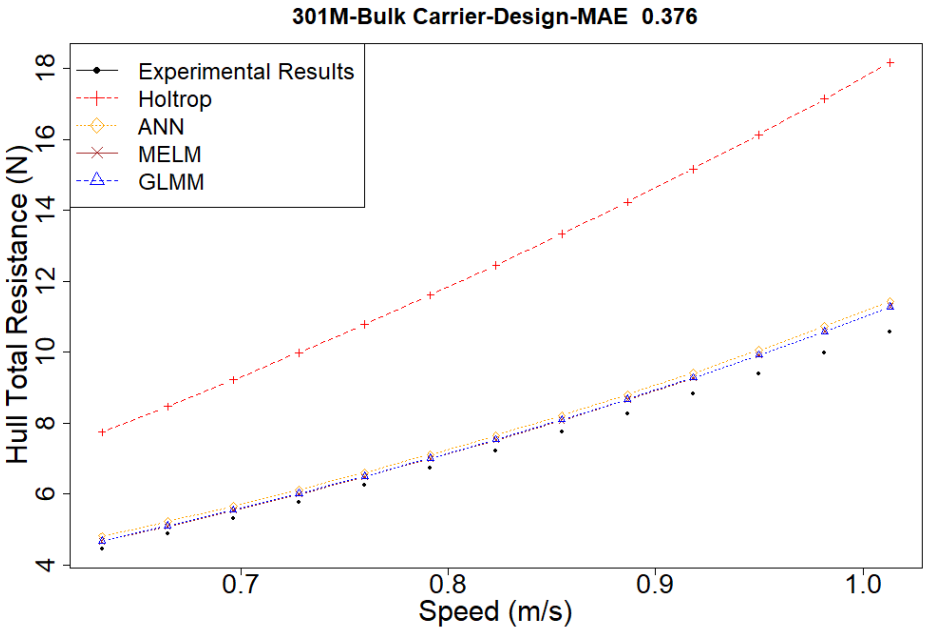

Figure D.9. Resistance Estimation of Model No: 301M, Bulk Carrier, Loading Condition: Design

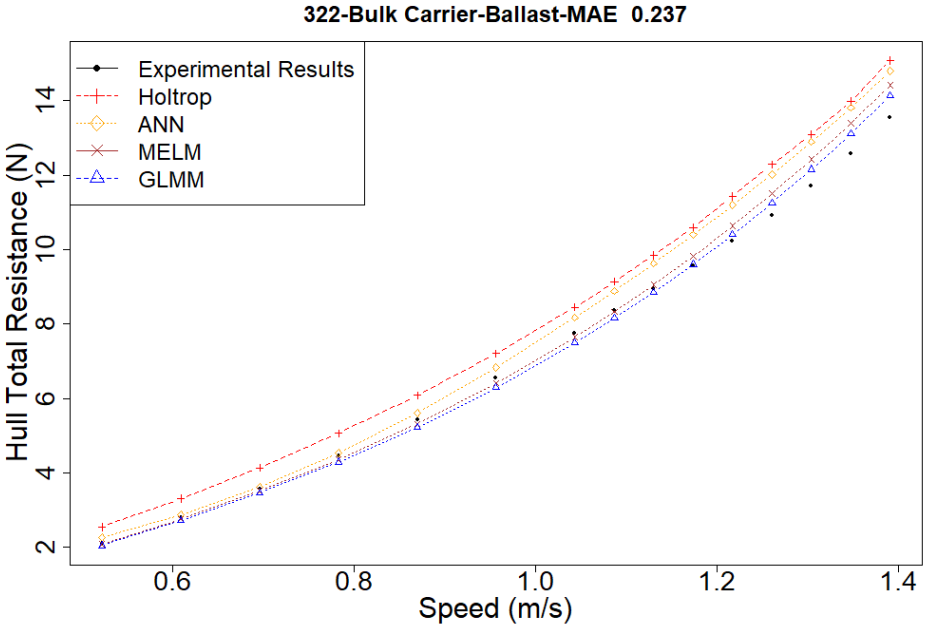

Figure D.10. Resistance Estimation of Model No: 322, Bulk Carrier, Loading Condition: Ballast

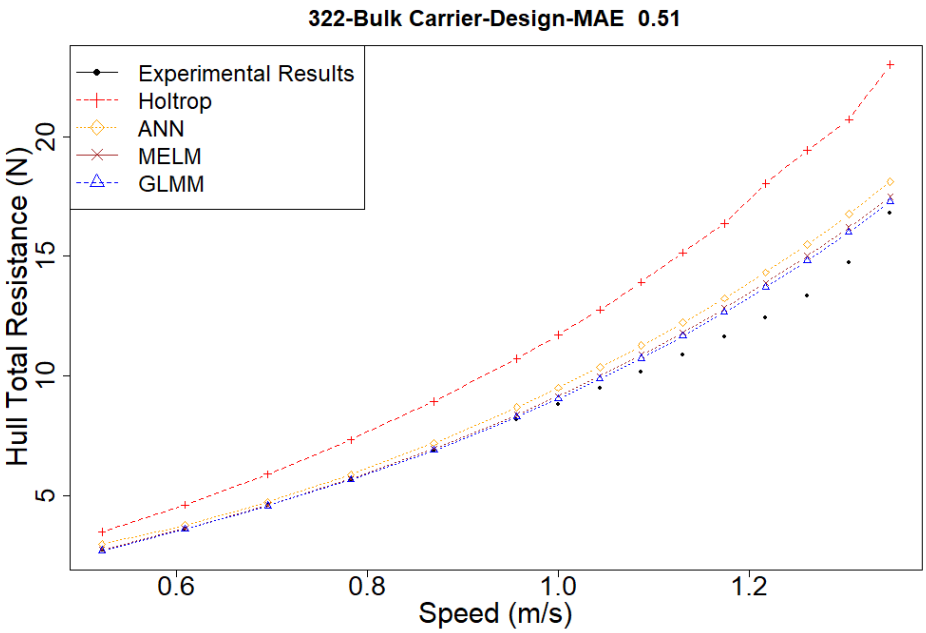

Figure D.11. Resistance Estimation of Model No: 322, Bulk Carrier, Loading Condition: Design

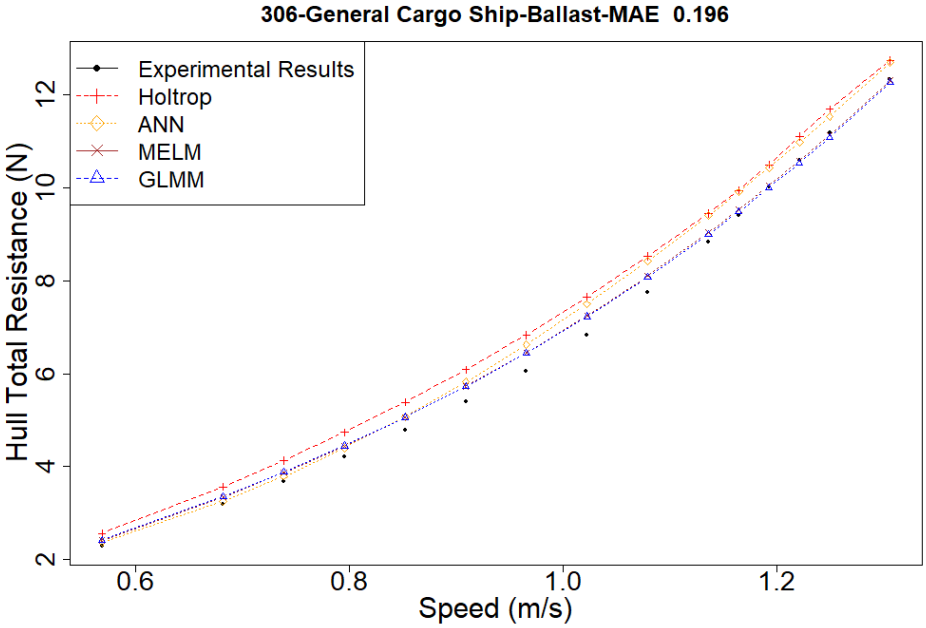

Figure D.12. Resistance Estimation of Model No: 306, General Cargo Ship, Loading Condition: Ballast

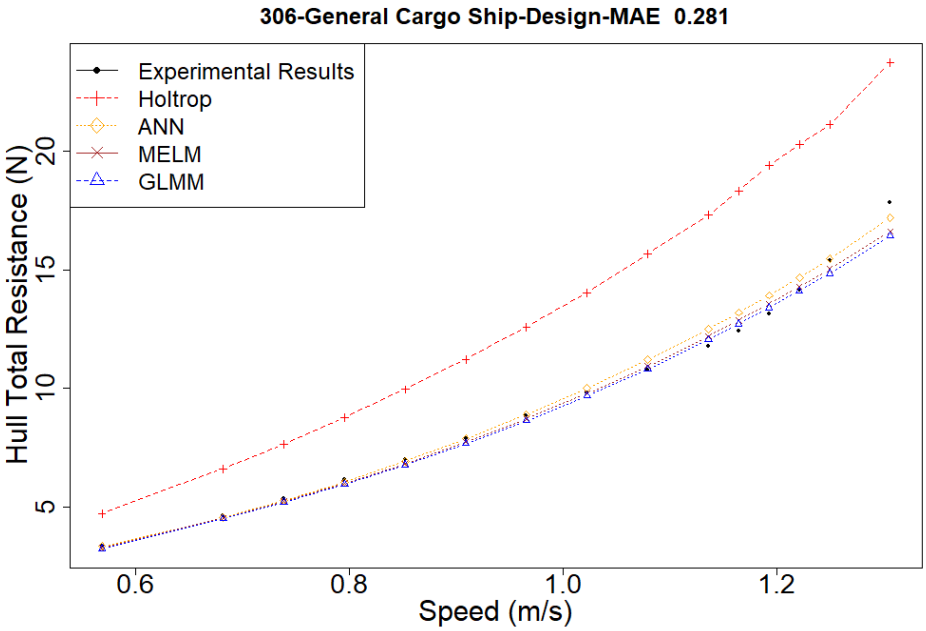

Figure D.13. Resistance Estimation of Model No: 306, General Cargo Ship, Loading Condition: Design

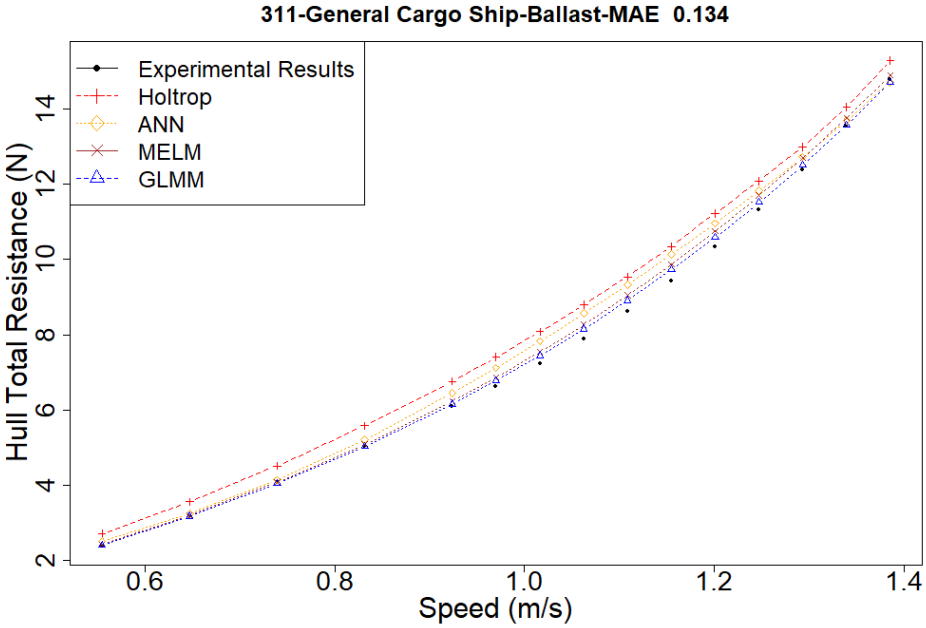

Figure D.14. Resistance Estimation of Model No: 311, General Cargo Ship, Loading Condition: Ballast

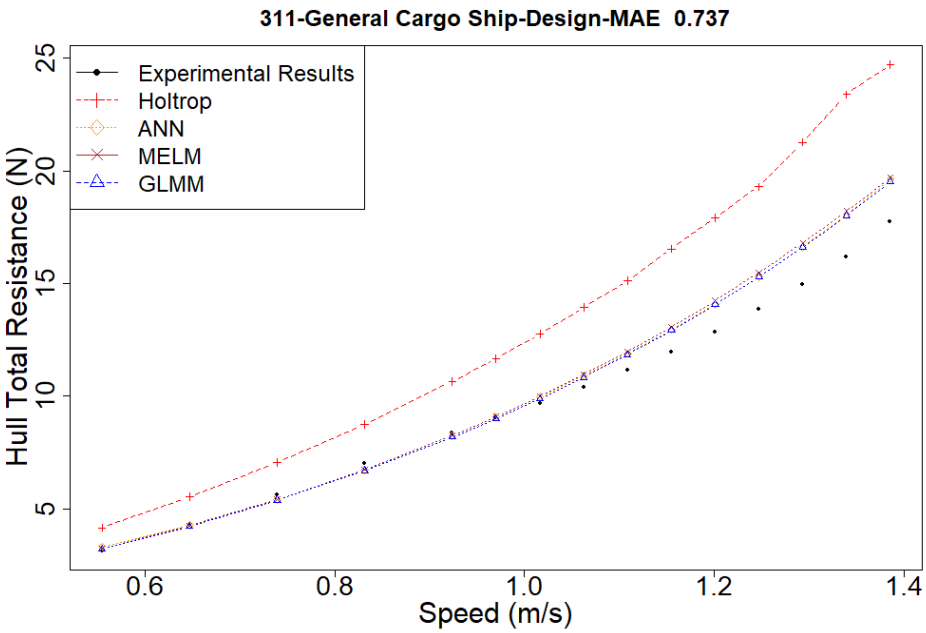

Figure D.15. Resistance Estimation of Model No: 311, General Cargo Ship, Loading Condition: Design

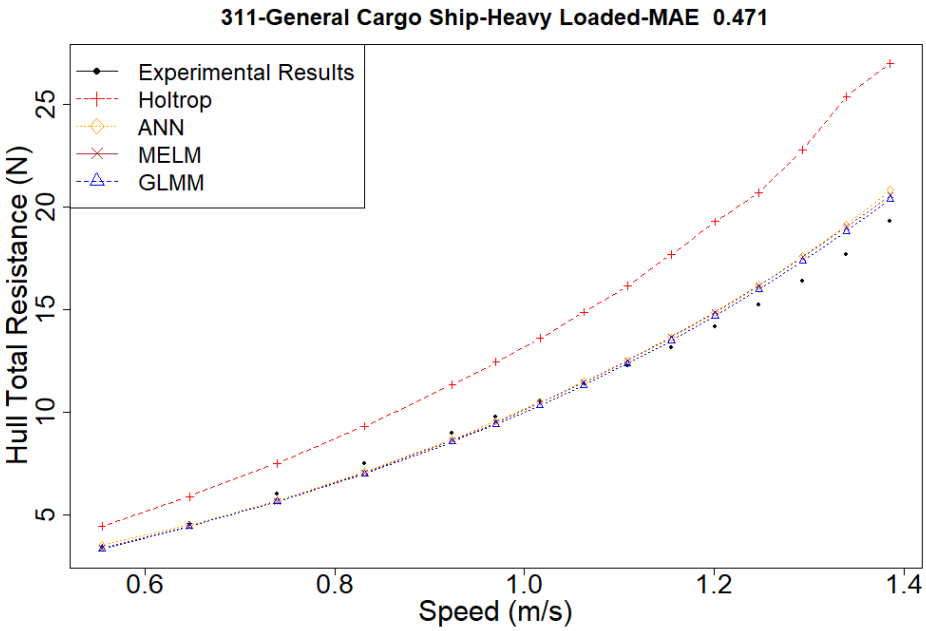

Figure D.16. Resistance Estimation of Model No: 311, General Cargo Ship, Loading Condition: Heavy Loaded

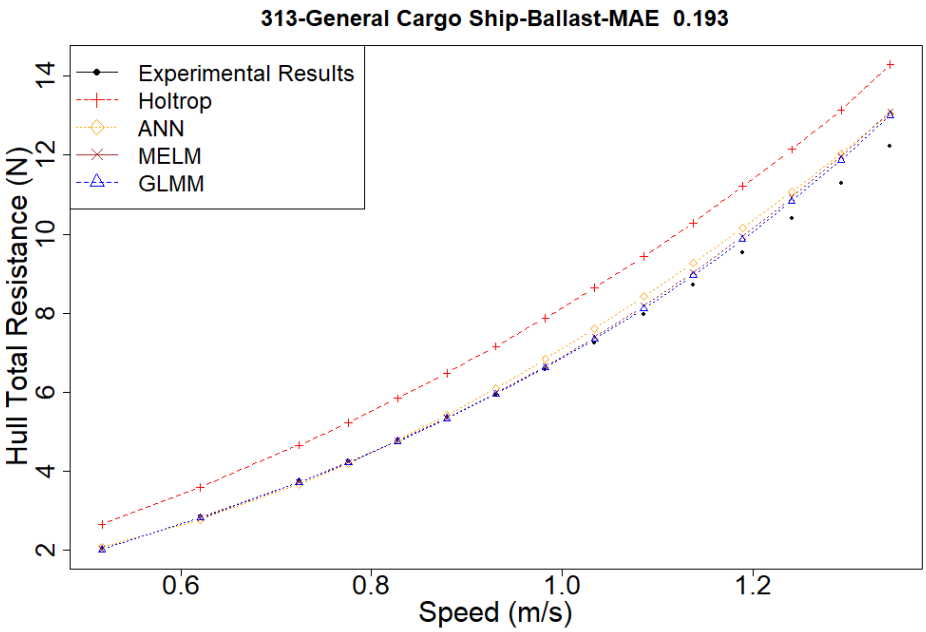

Figure D.17. Resistance Estimation of Model No: 313, General Cargo Ship, Loading Condition: Ballast

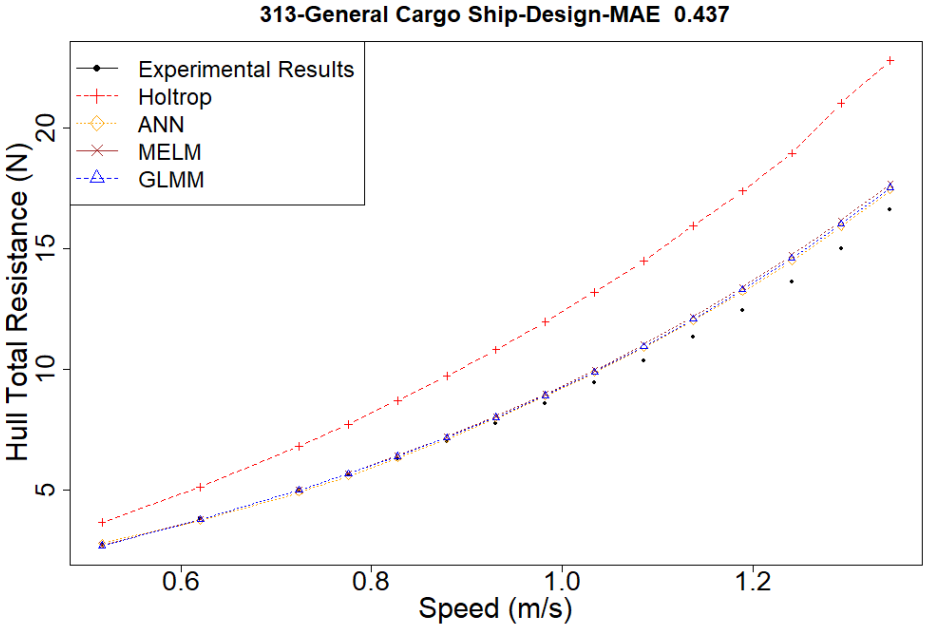

Figure D.18. Resistance Estimation of Model No: 313, General Cargo Ship, Loading Condition: Design

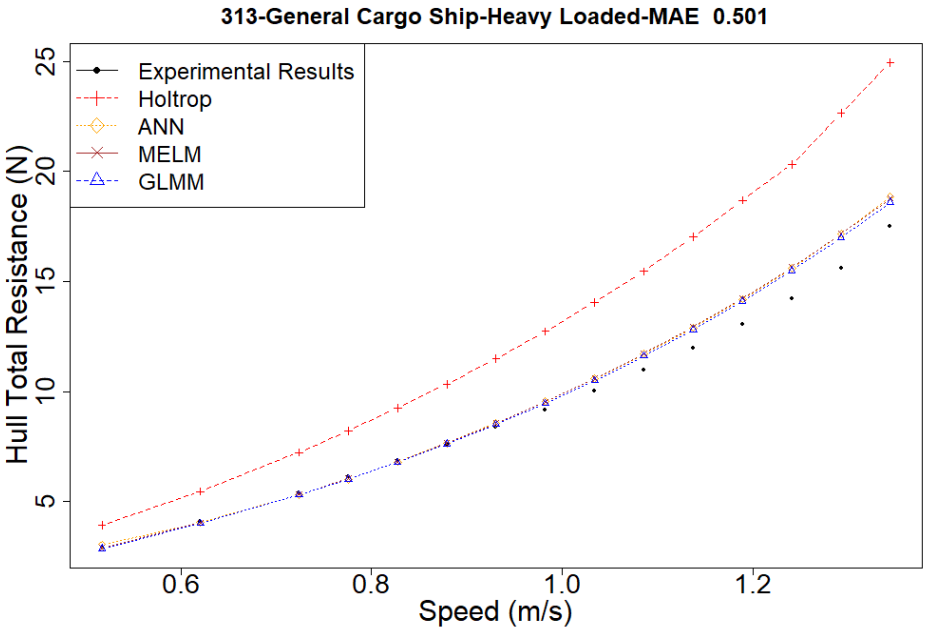

Figure D.19. Resistance Estimation of Model No: 313, General Cargo Ship, Loading Condition: Heavy Loaded

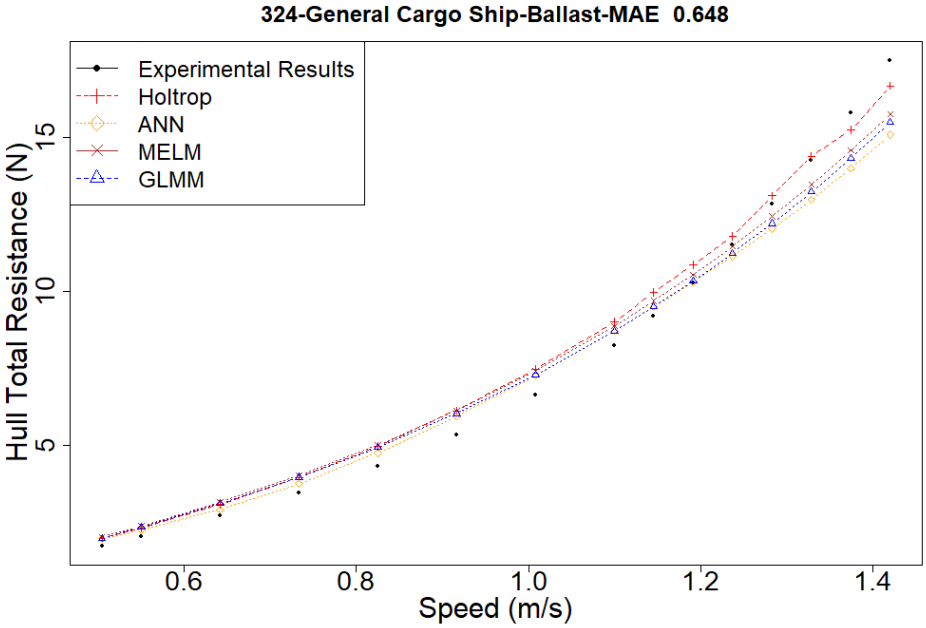

Figure D.20. Resistance Estimation of Model No: 324, General Cargo Ship, Loading Condition: Ballast

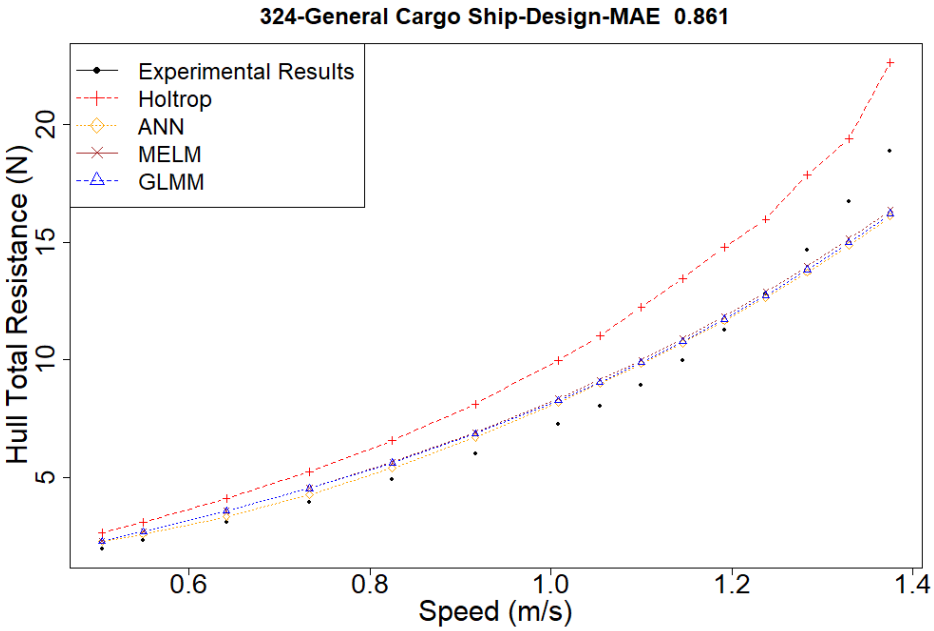

Figure D.21. Resistance Estimation of Model No: 324, General Cargo Ship, Loading Condition: Design

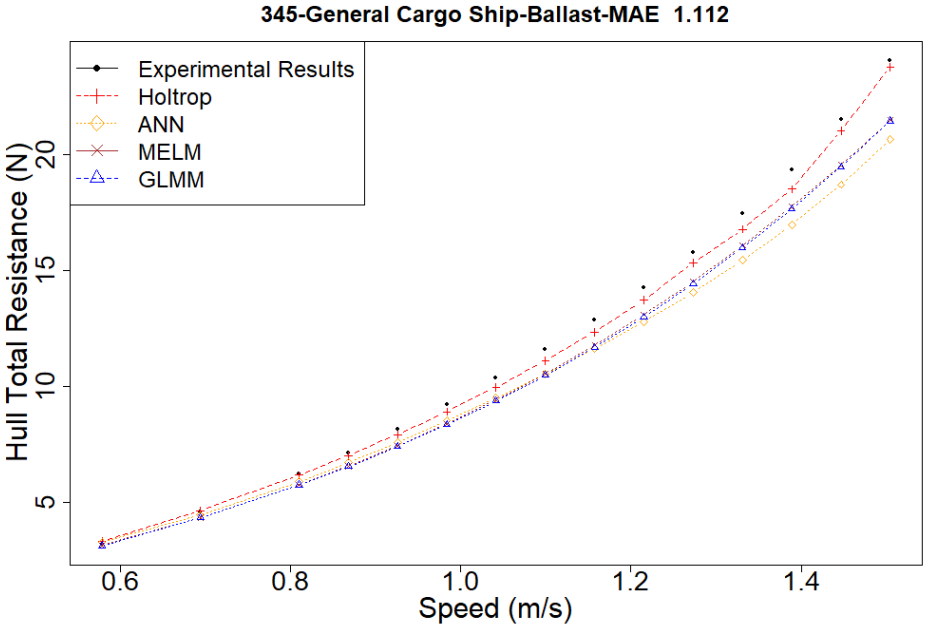

Figure D.22. Resistance Estimation of Model No: 345, General Cargo Ship, Loading Condition: Ballast

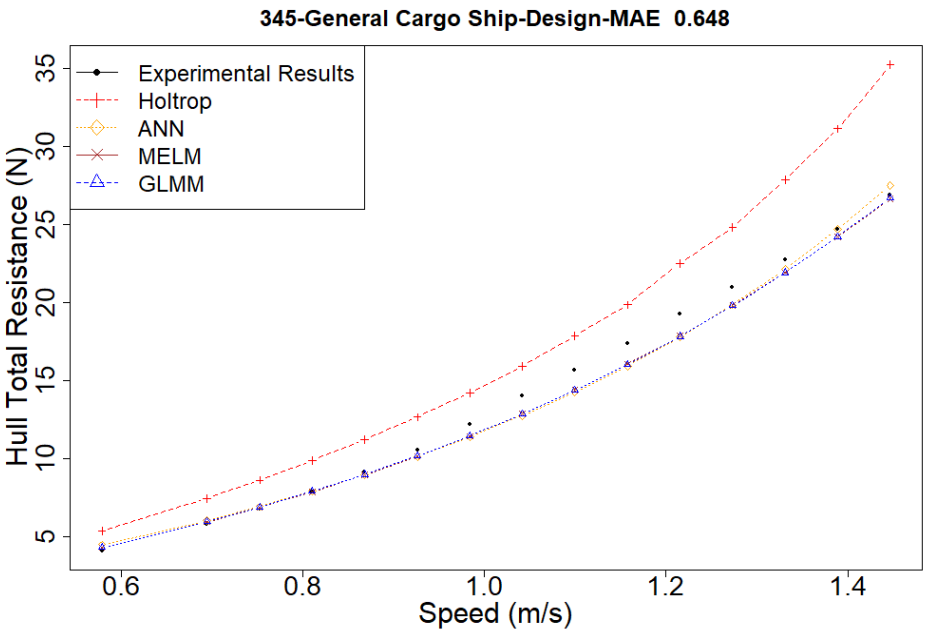

Figure D.23. Resistance Estimation of Model No: 345, General Cargo Ship, Loading Condition: Design

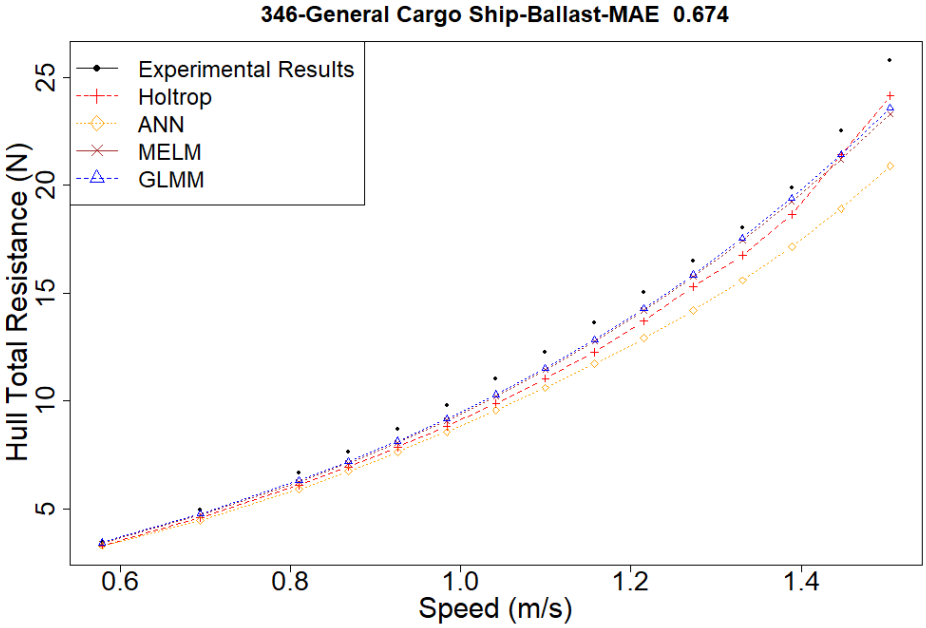

Figure D.24. Resistance Estimation of Model No: 346, General Cargo Ship, Loading Condition: Ballast

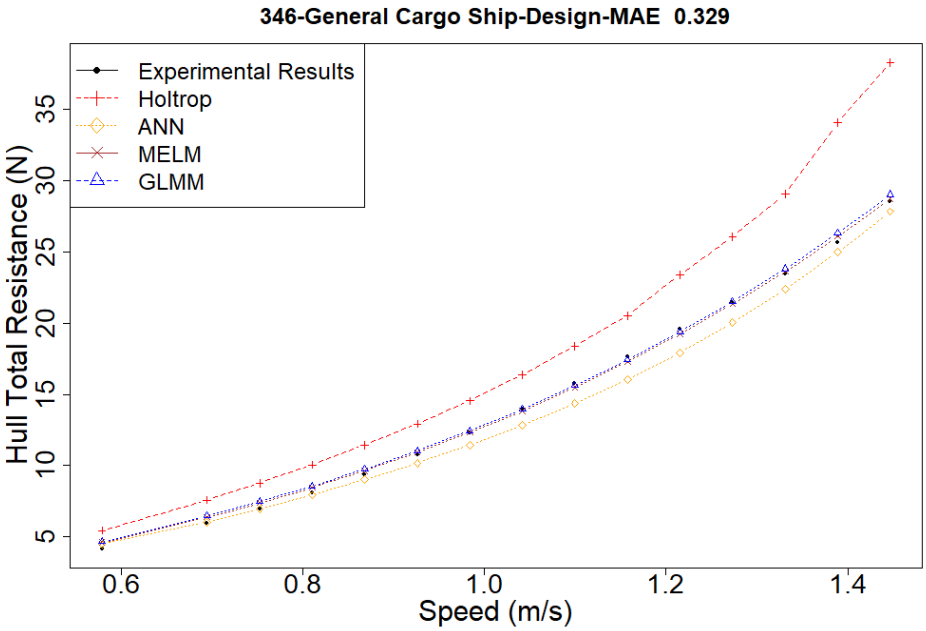

Figure D.25. Resistance Estimation of Model No: 346, General Cargo Ship, Loading Condition: Design

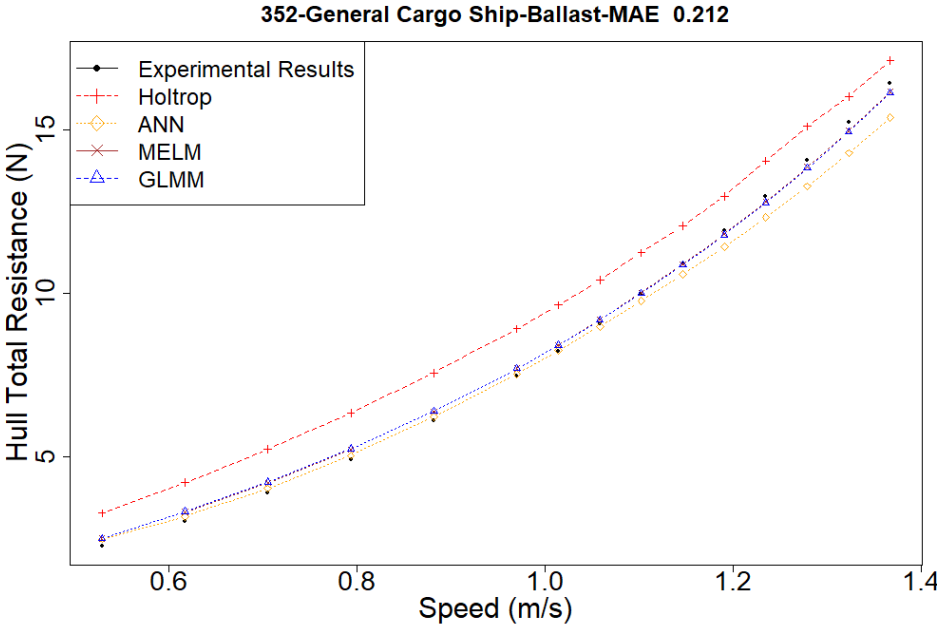

Figure D.26. Resistance Estimation of Model No: 352, General Cargo Ship, Loading Condition: Ballast

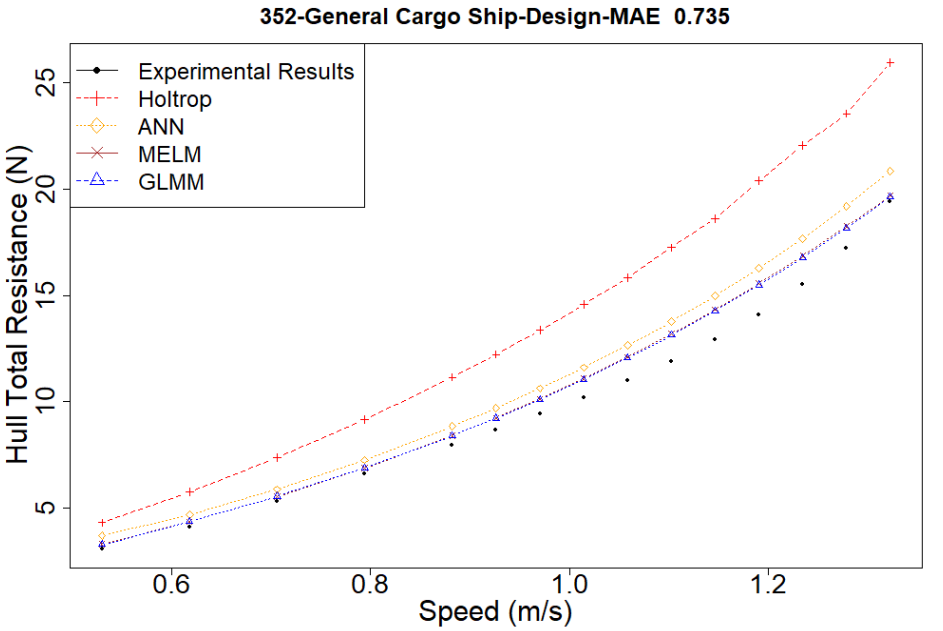

Figure D.27. Resistance Estimation of Model No: 352, General Cargo Ship, Loading Condition: Design

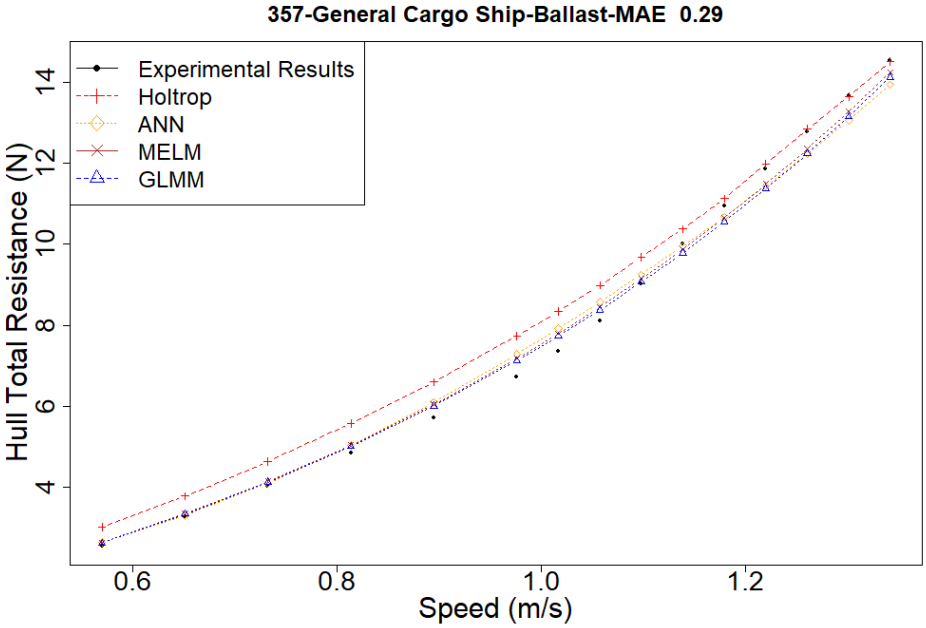

Figure D.28. Resistance Estimation of Model No: 357, General Cargo Ship, Loading Condition: Ballast

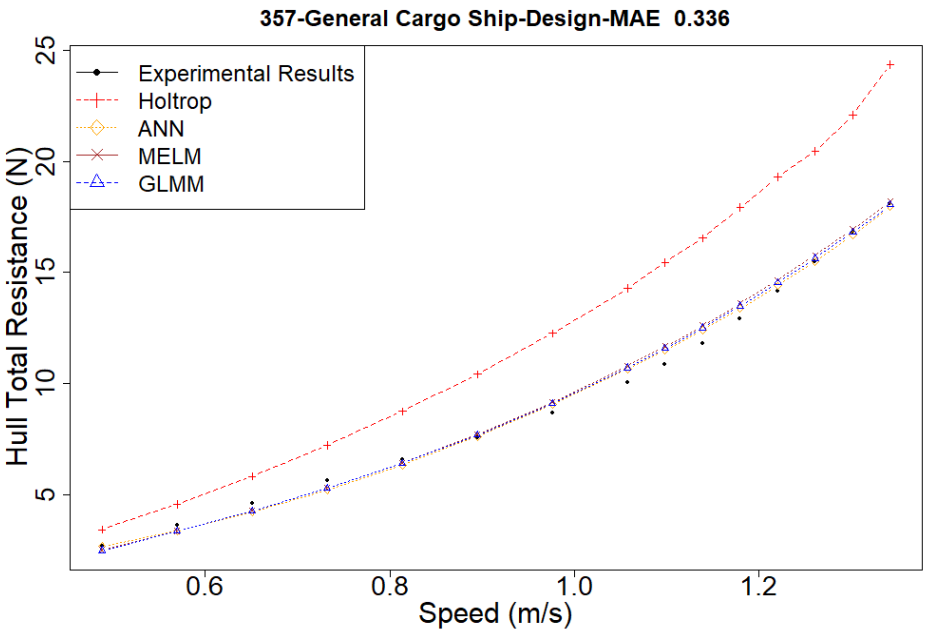

Figure D.29. Resistance Estimation of Model No: 357, General Cargo Ship, Loading Condition: Design

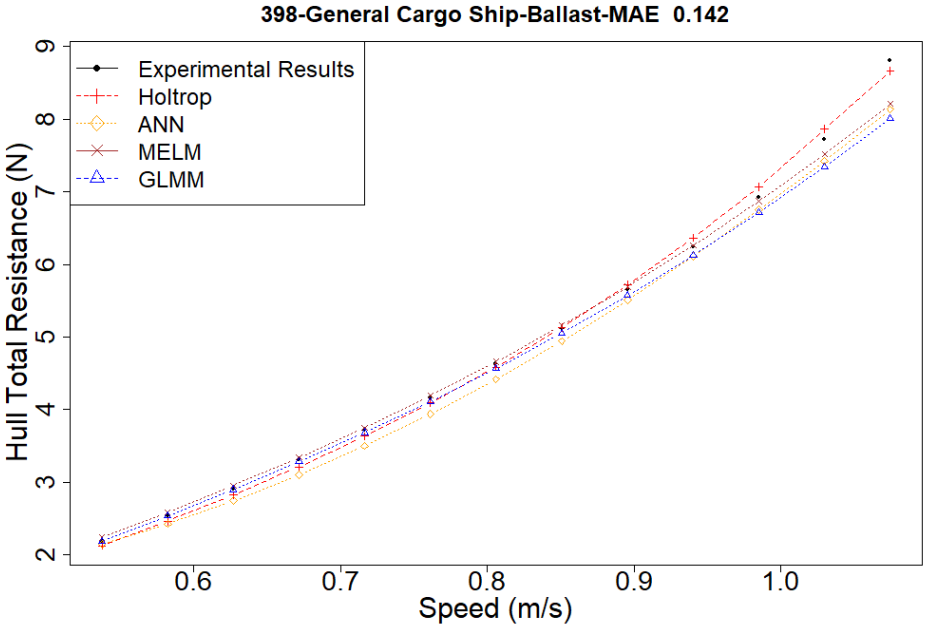

Figure D.30. Resistance Estimation of Model No: 398, General Cargo Ship, Loading Condition: Ballast

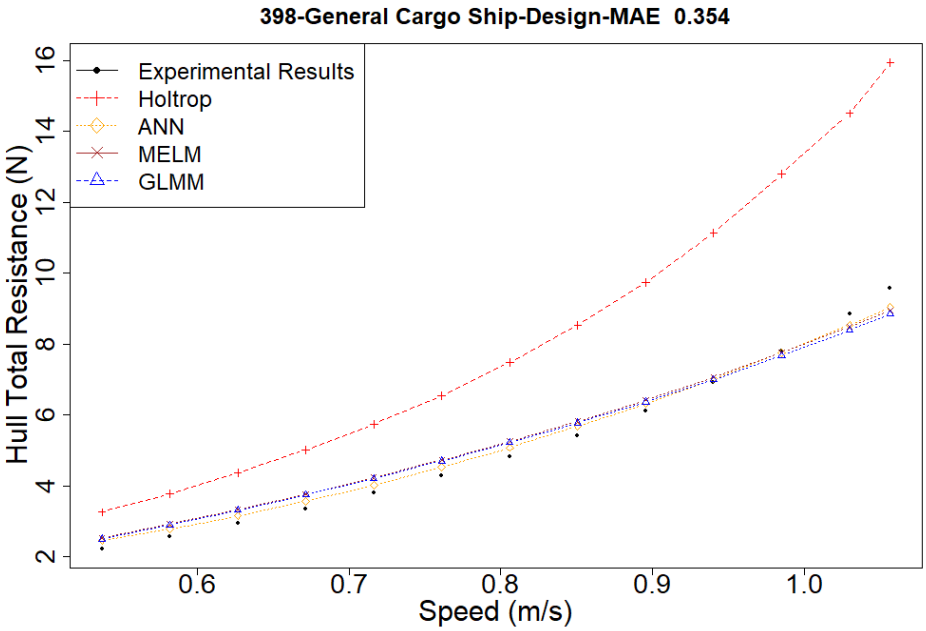

Figure D.31. Resistance Estimation of Model No: 398, General Cargo Ship, Loading Condition: Design

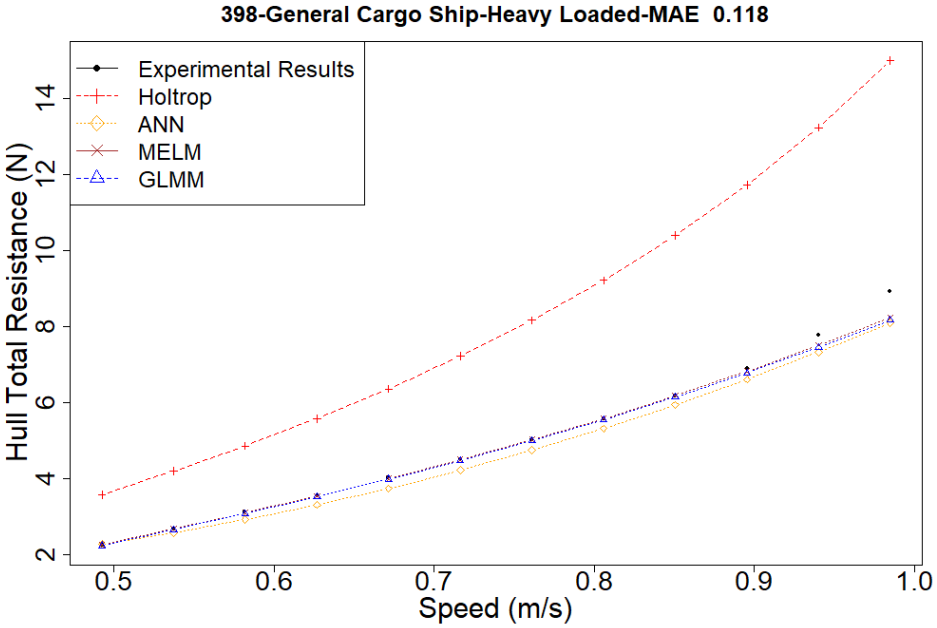

Figure D.32. Resistance Estimation of Model No: 398, General Cargo Ship, Loading Condition: Heavy Loaded

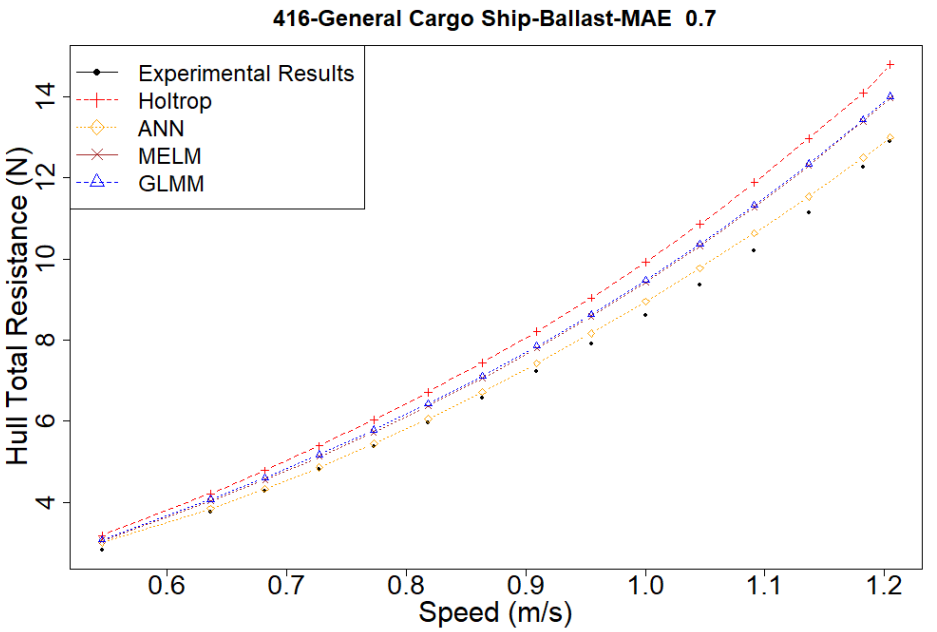

Figure D.33. Resistance Estimation of Model No: 416, General Cargo Ship, Loading Condition: Ballast

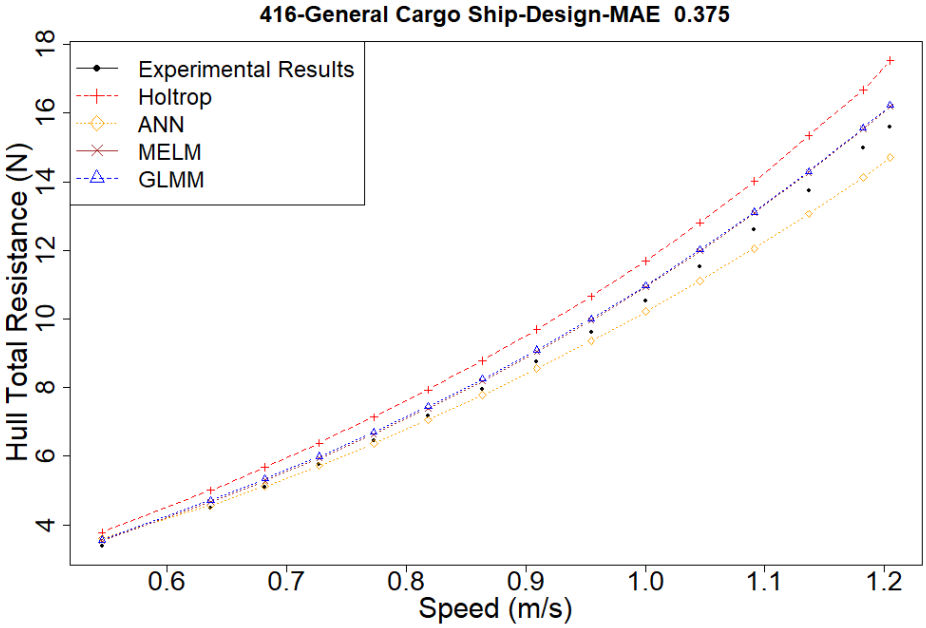

Figure D.34. Resistance Estimation of Model No: 416, General Cargo Ship, Loading Condition: Design

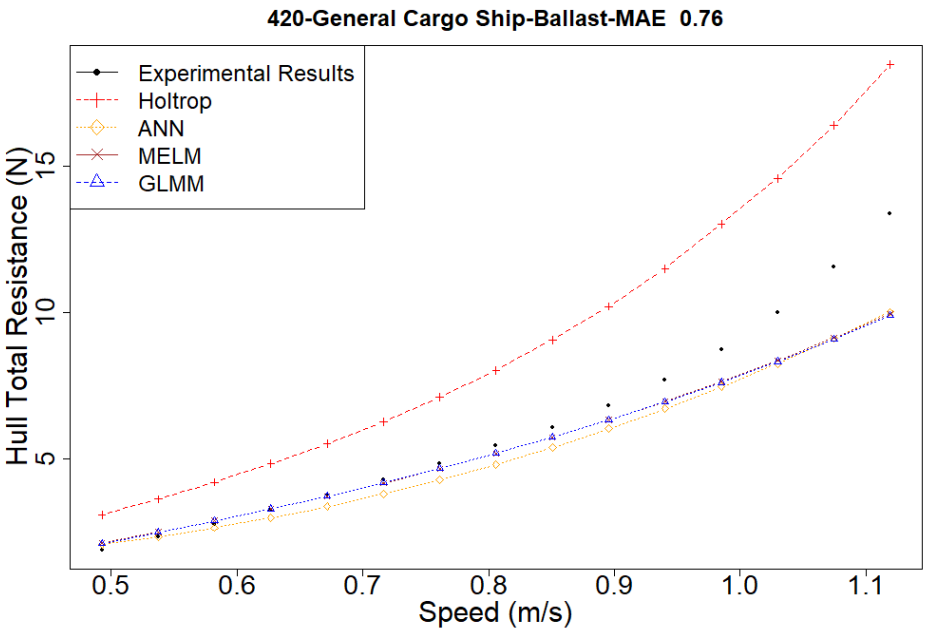

Figure D.35. Resistance Estimation of Model No: 420, General Cargo Ship, Loading Condition: Ballast

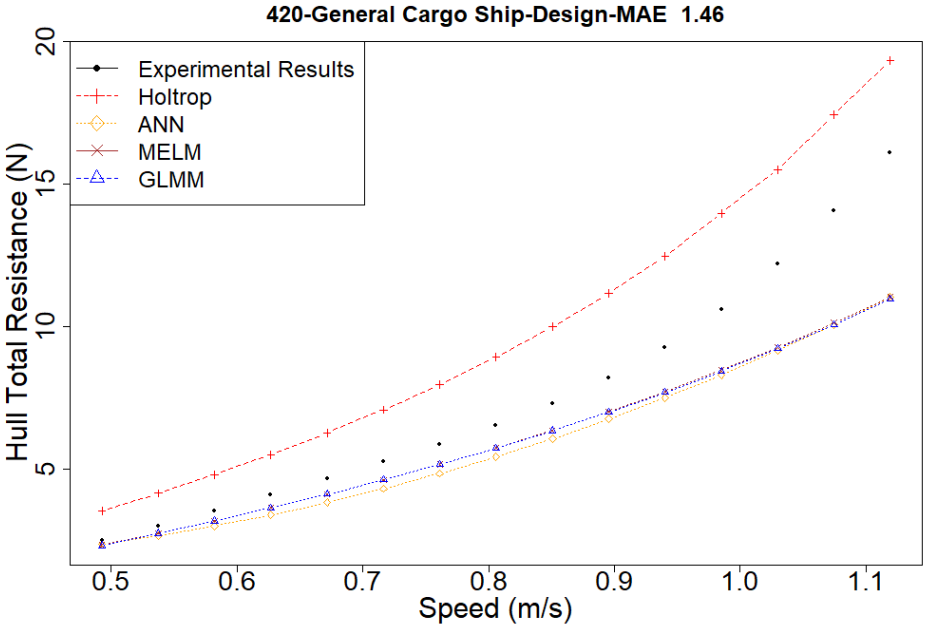

Figure D.36. Resistance Estimation of Model No: 420, General Cargo Ship, Loading Condition: Design

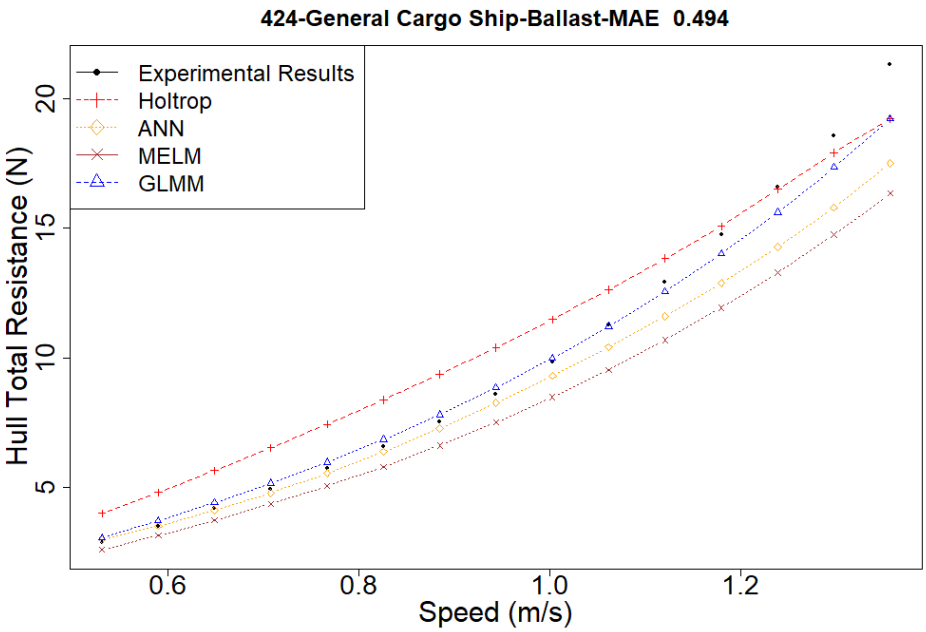

Figure D.37. Resistance Estimation of Model No: 424, General Cargo Ship, Loading Condition: Ballast

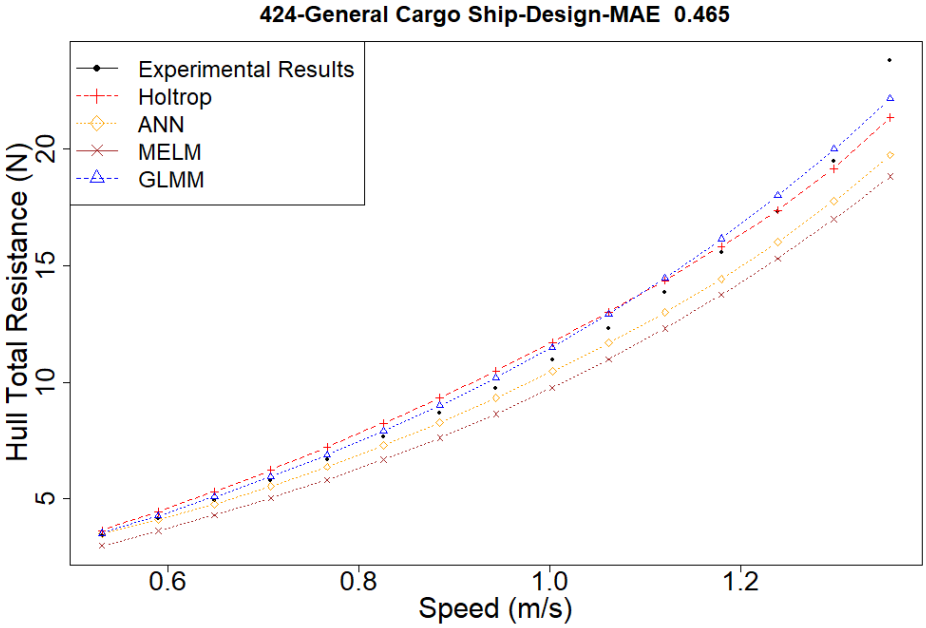

Figure D.38. Resistance Estimation of Model No: 424, General Cargo Ship, Loading Condition: Design

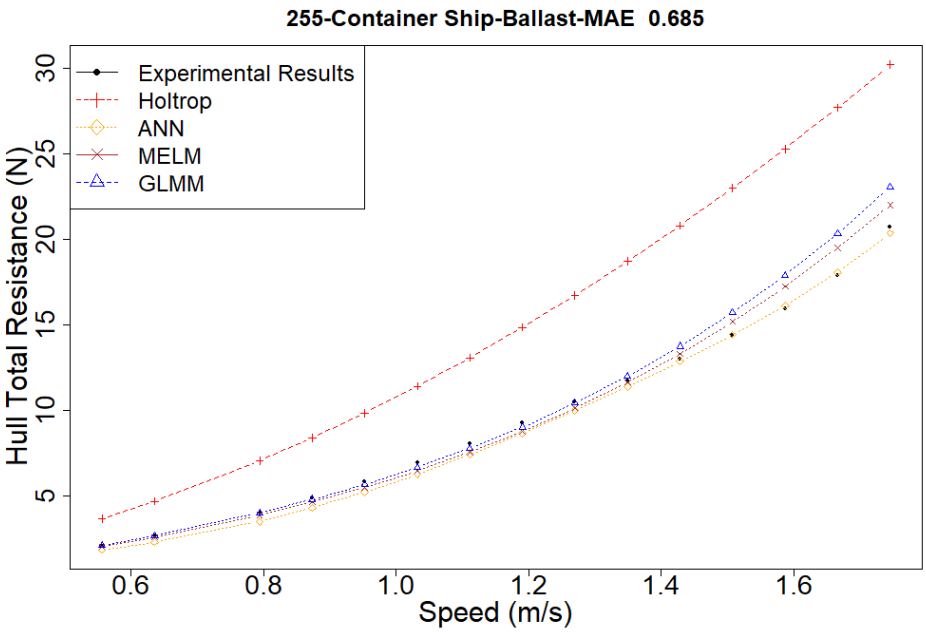

Figure D.39. Resistance Estimation of Model No: 255, Container Ship, Loading Condition: Ballast

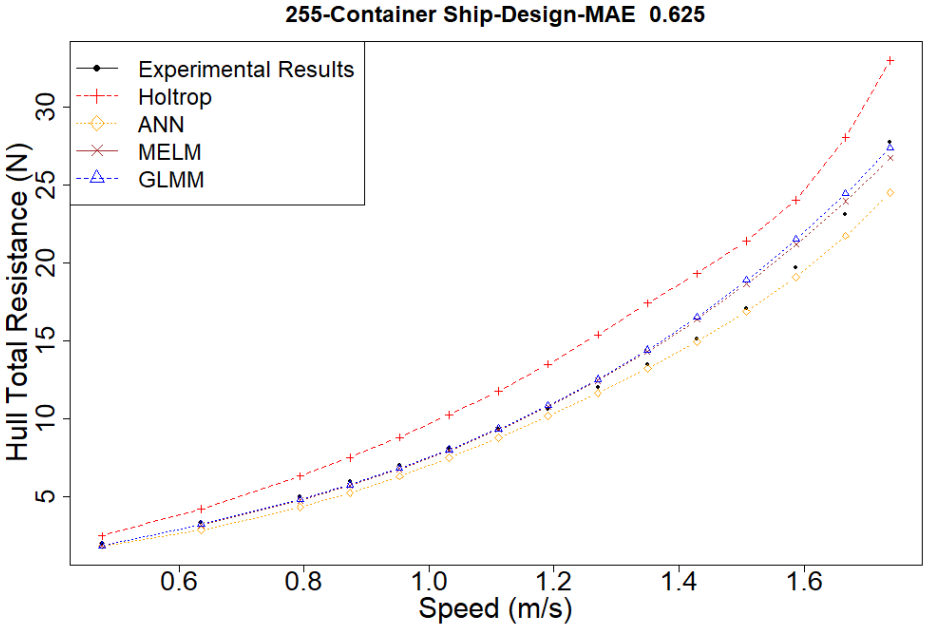

Figure D.40. Resistance Estimation of Model No: 255, Container Ship, Loading Condition: Design

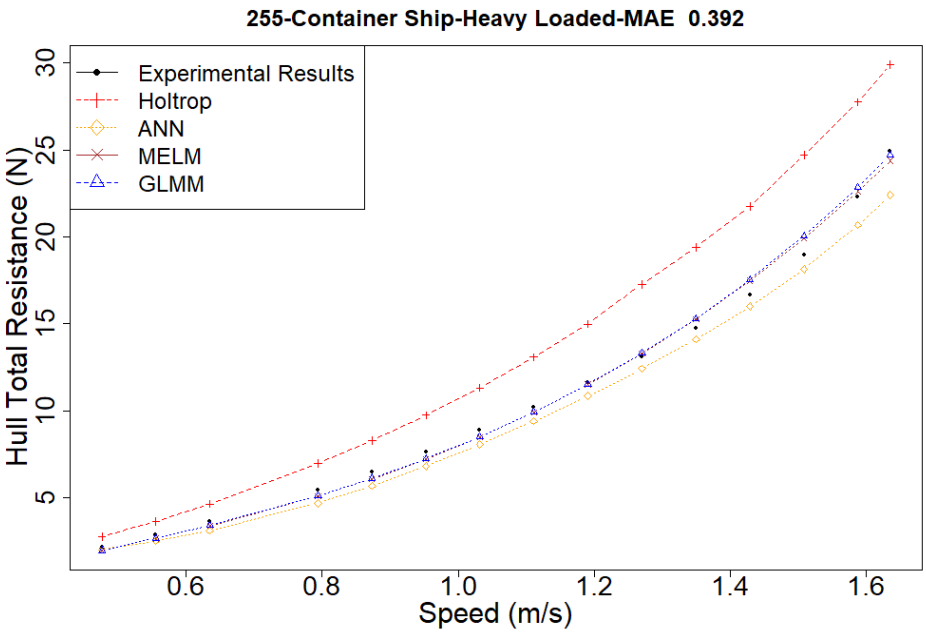

Figure D.41. Resistance Estimation of Model No: 255, Container Ship, Loading Condition: Heavy Loaded

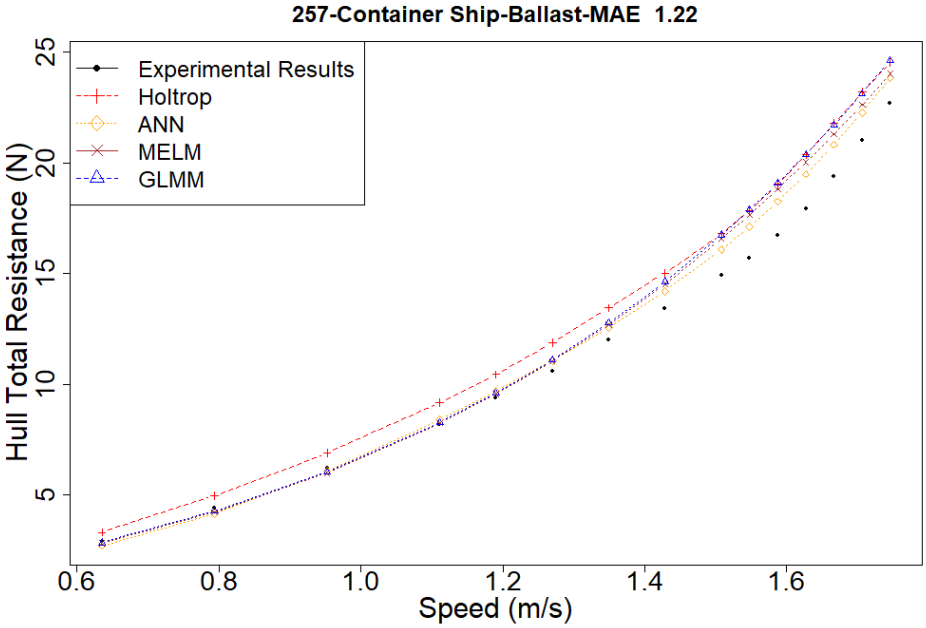

Figure D.42. Resistance Estimation of Model No: 257, Container Ship, Loading Condition: Ballast

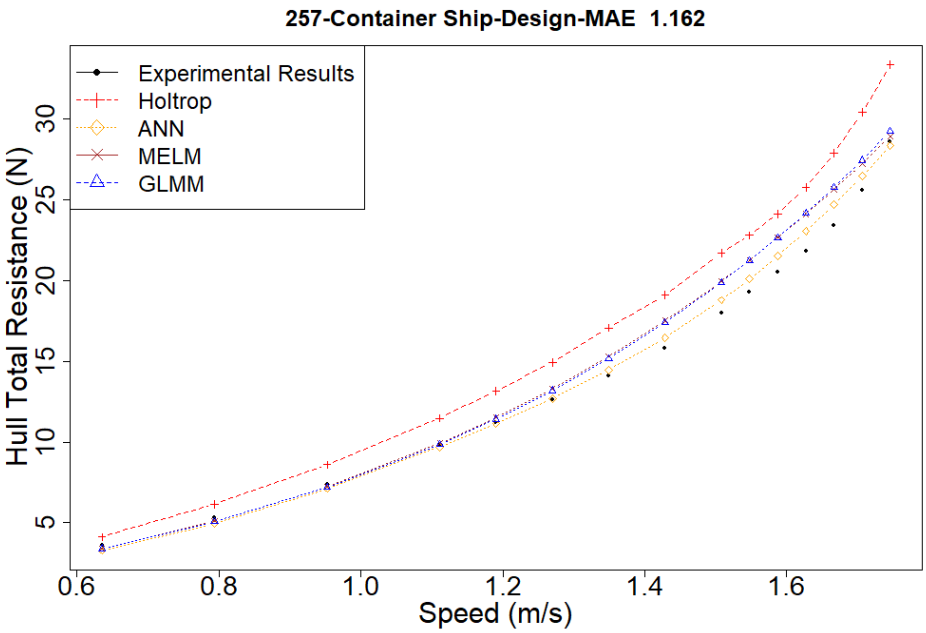

Figure D.43. Resistance Estimation of Model No: 257, Container Ship, Loading Condition: Design

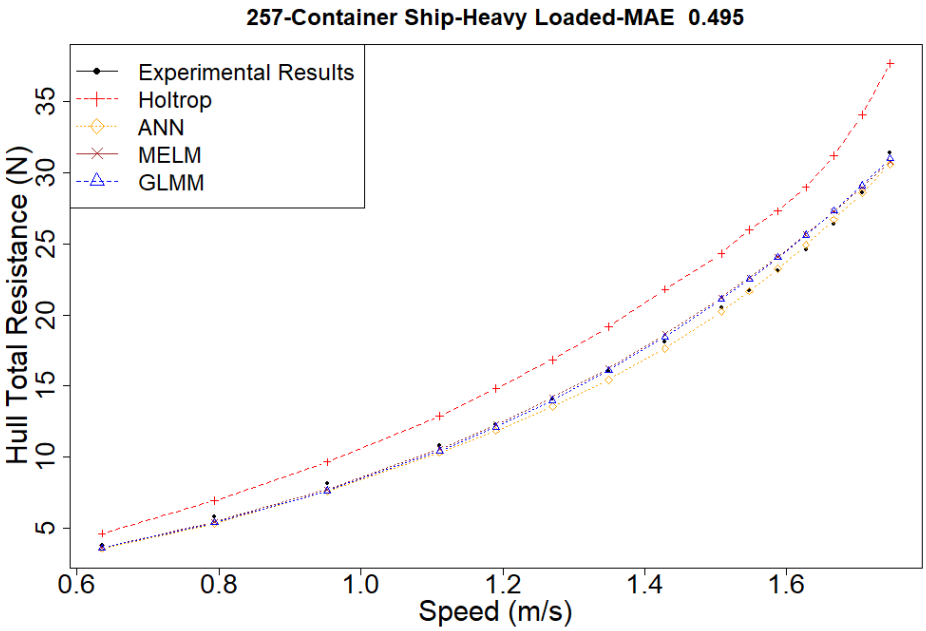

Figure D.44. Resistance Estimation of Model No: 257, Container Ship, Loading Condition: Heavy Loaded

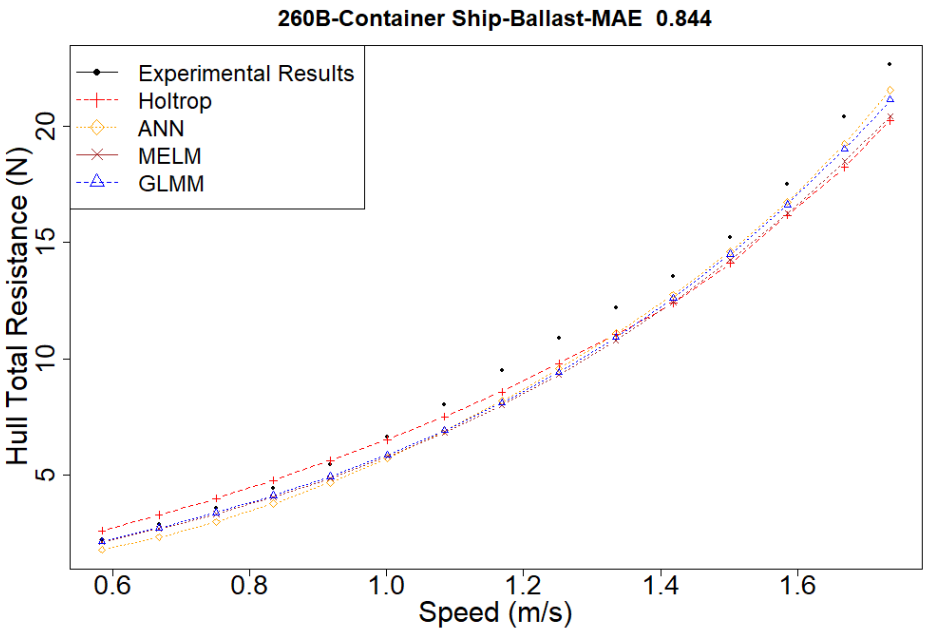

Figure D.45. Resistance Estimation of Model No: 260B, Container Ship, Loading Condition: Ballast

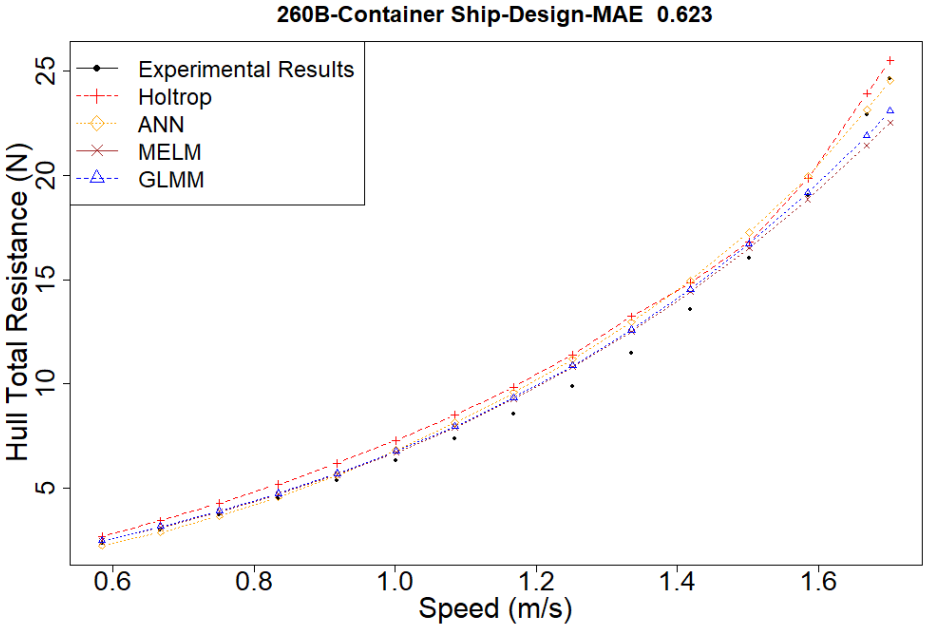

Figure D.46. Resistance Estimation of Model No: 260B, Container Ship, Loading Condition: Design

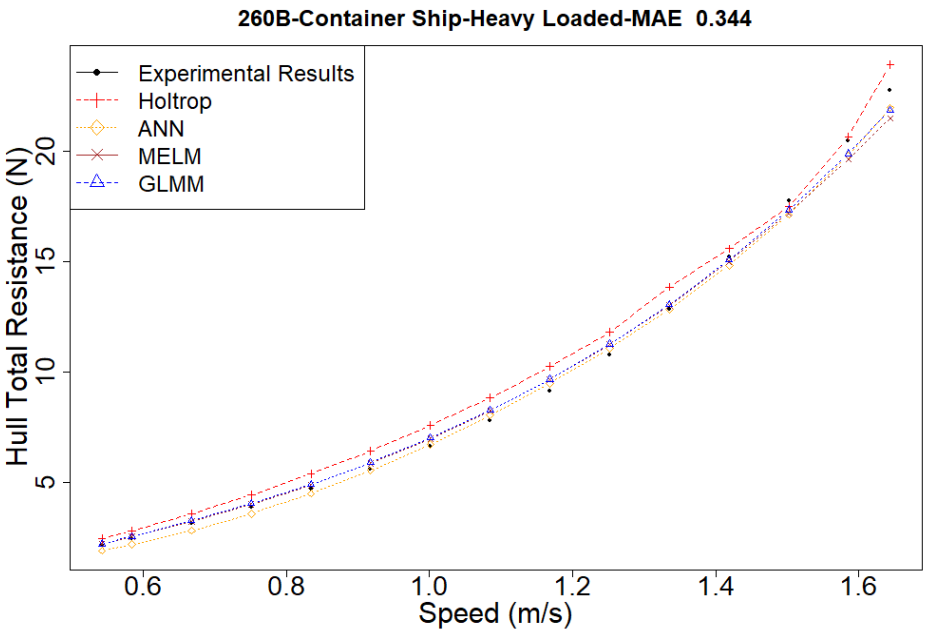

Figure D.47. Resistance Estimation of Model No: 260B, Container Ship, Loading Condition: Heavy Loaded

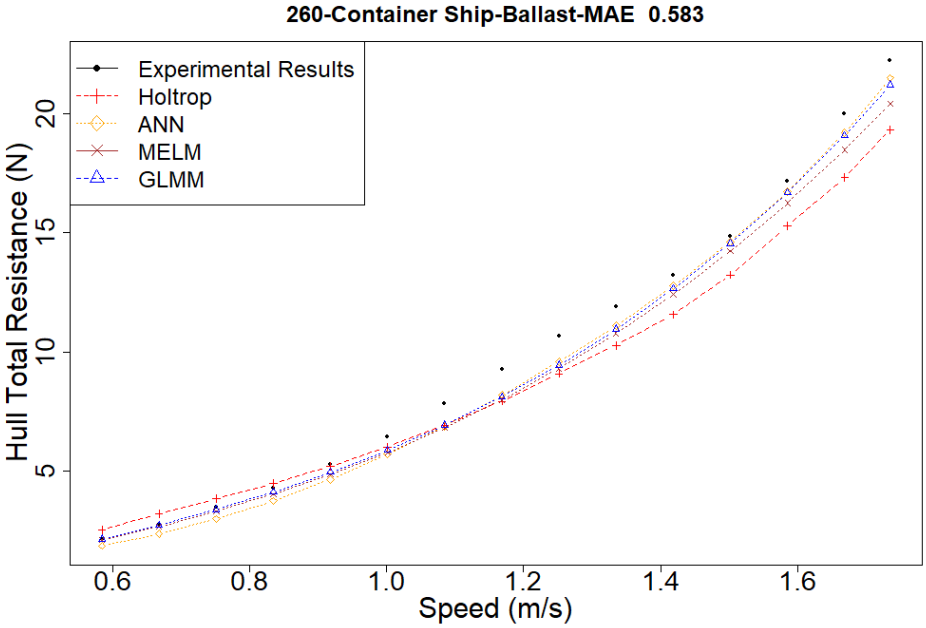

Figure D.48. Resistance Estimation of Model No: 260, Container Ship, Loading Condition: Ballast

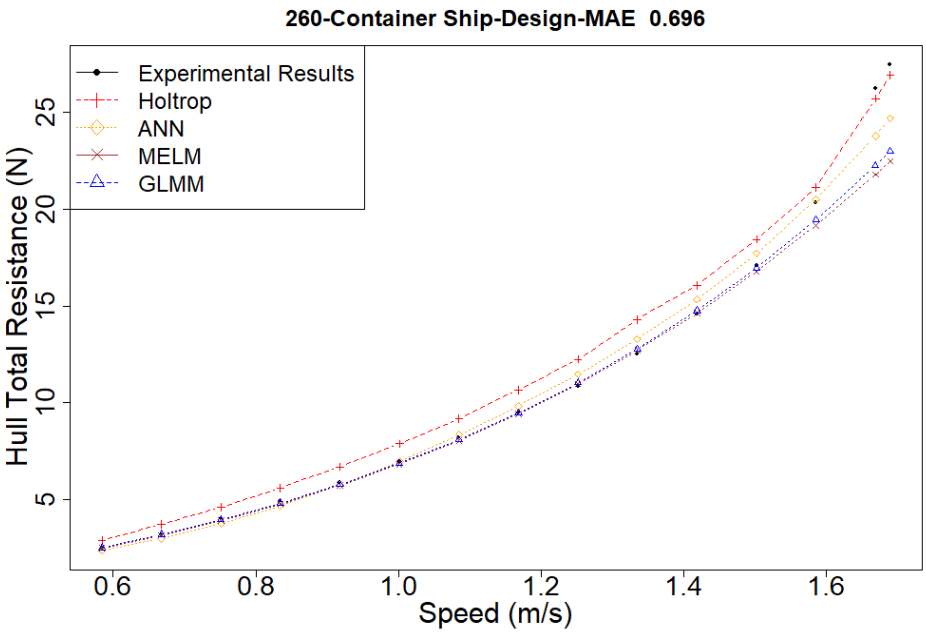

Figure D.49. Resistance Estimation of Model No: 260, Container Ship, Loading Condition: Design

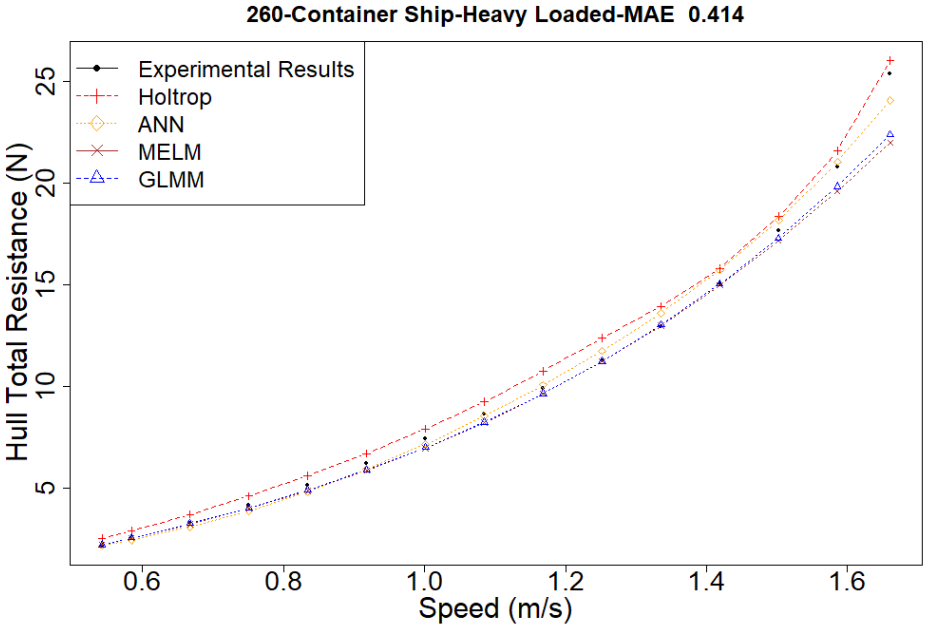

Figure D.50. Resistance Estimation of Model No: 260, Container Ship, Loading Condition: Heavy Loaded

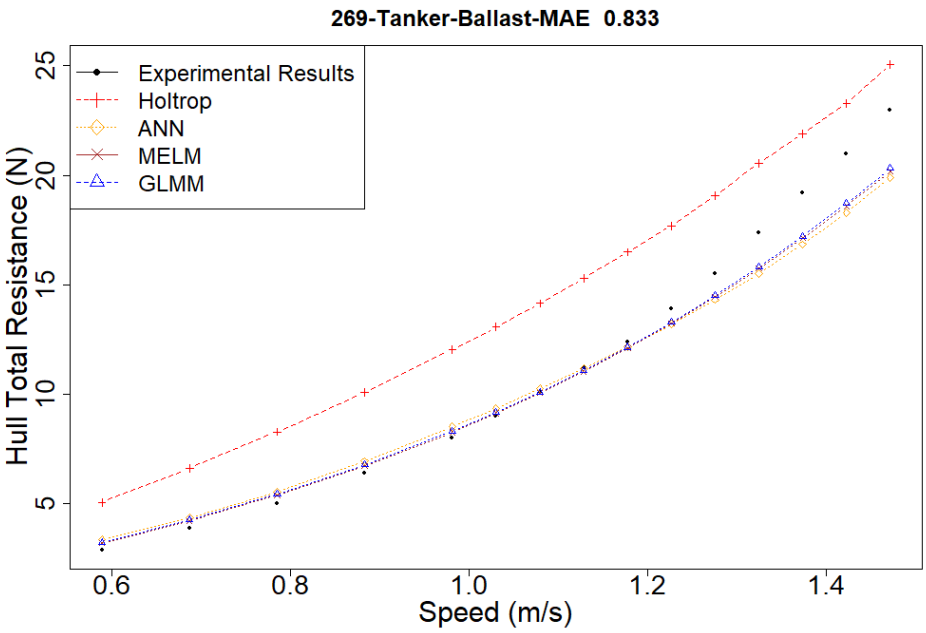

Figure D.51. Resistance Estimation of Model No: 269, Tanker, Loading Condition: Ballast

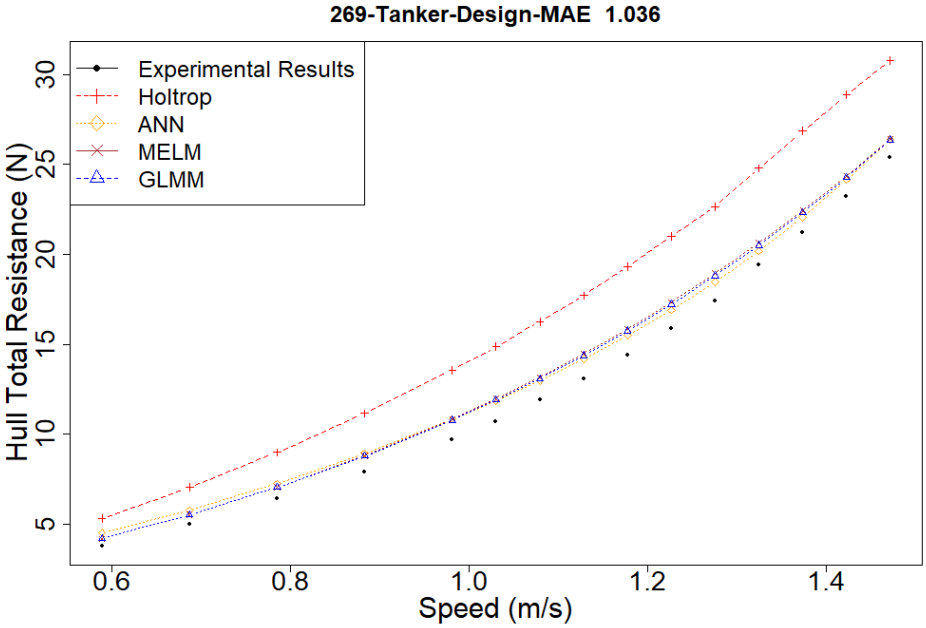

Figure D.52. Resistance Estimation of Model No: 269, Tanker, Loading Condition: Design

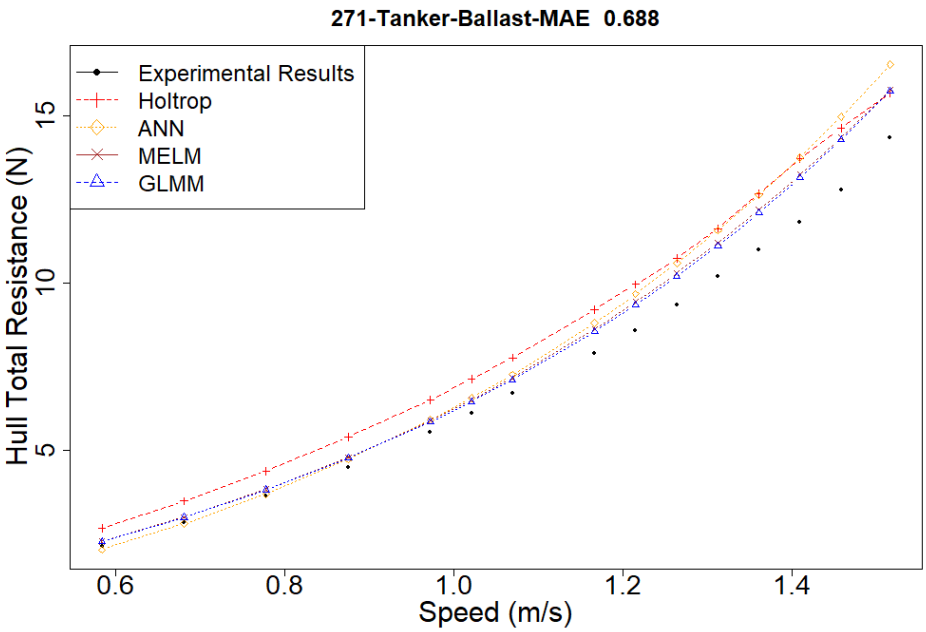

Figure D.53. Resistance Estimation of Model No: 271, Tanker, Loading Condition: Ballast

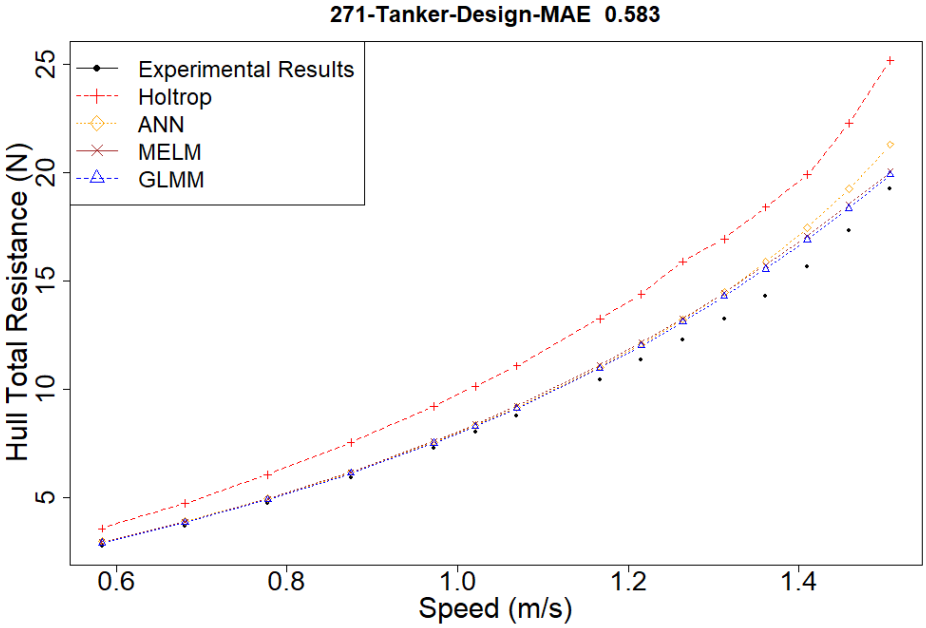

Figure D.54. Resistance Estimation of Model No: 271, Tanker, Loading Condition: Design

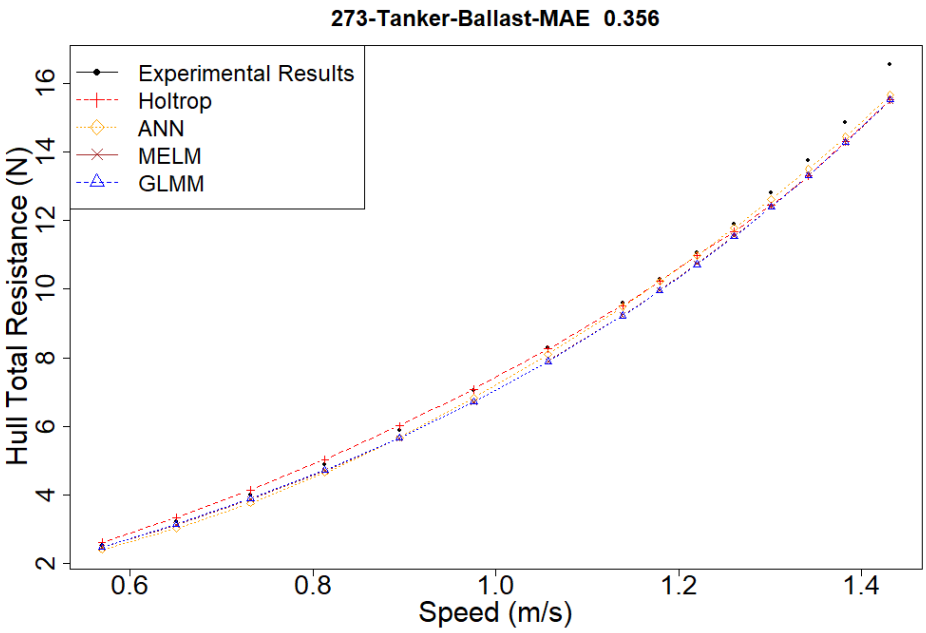

Figure D.55. Resistance Estimation of Model No: 273, Tanker, Loading Condition: Ballast

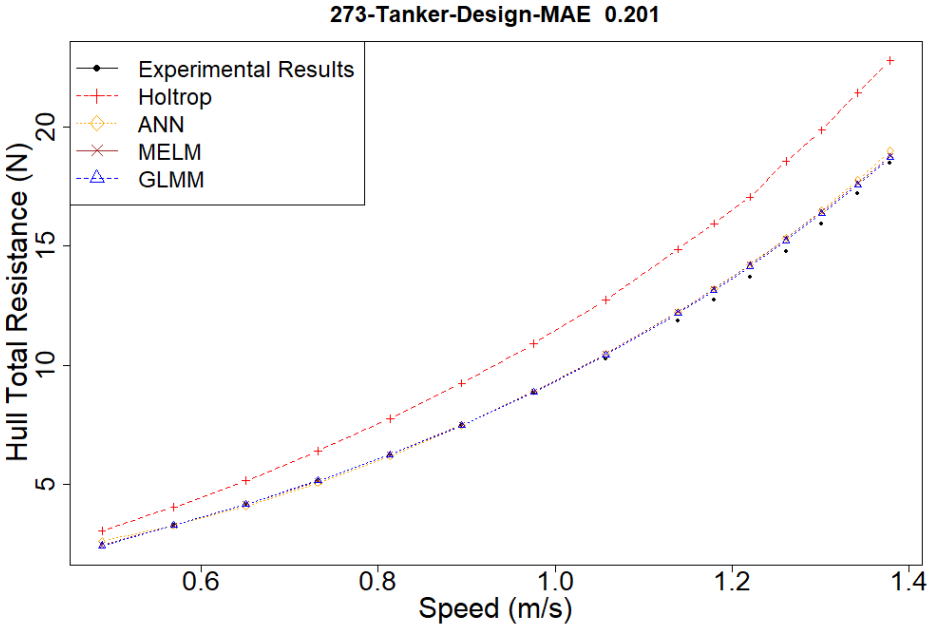

Figure D.56. Resistance Estimation of Model No: 273, Tanker, Loading Condition: Design

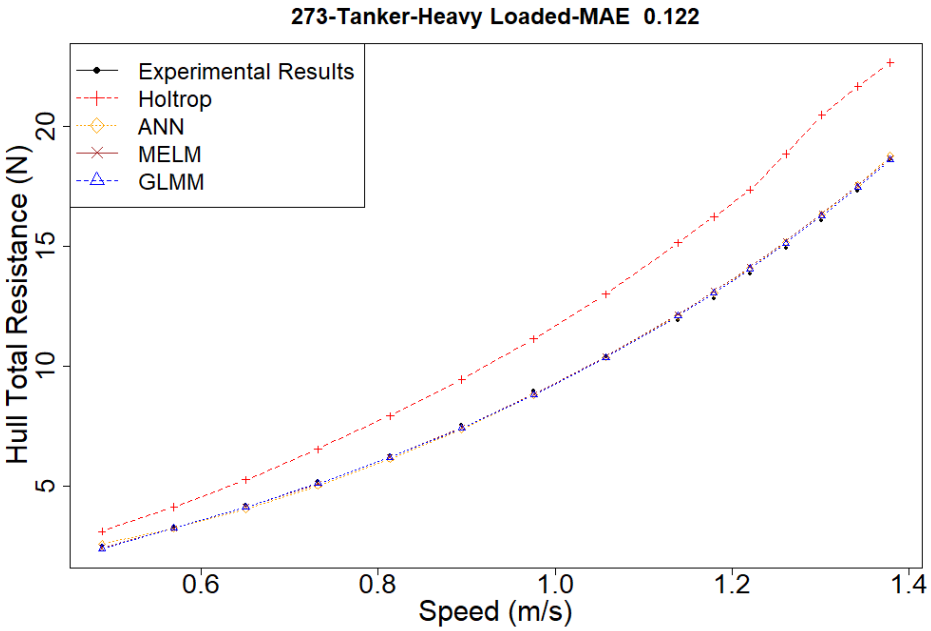

Figure D.57. Resistance Estimation of Model No: 273, Tanker, Loading Condition: Heavy Loaded

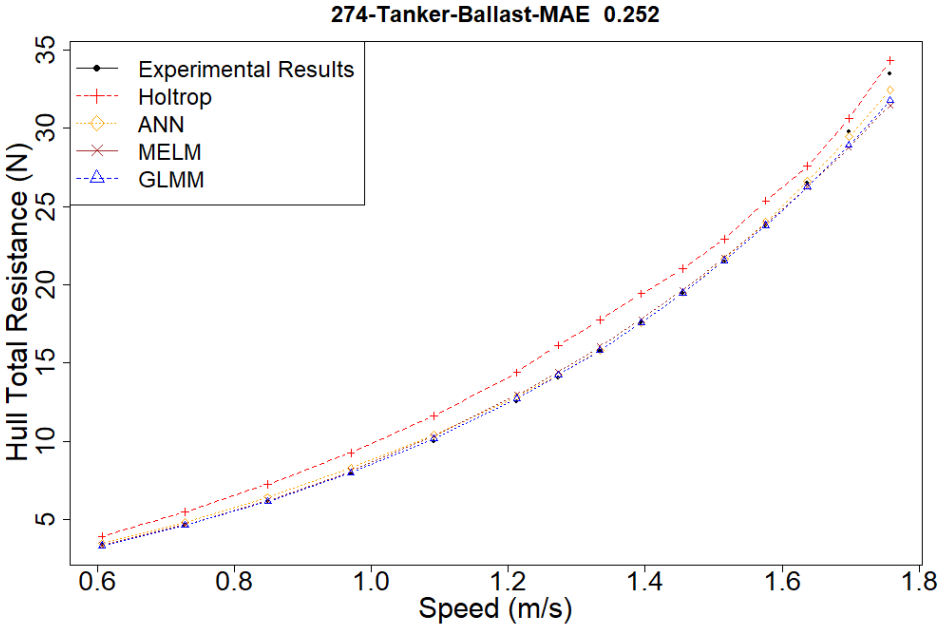

Figure D.58. Resistance Estimation of Model No: 274, Tanker, Loading Condition: Ballast

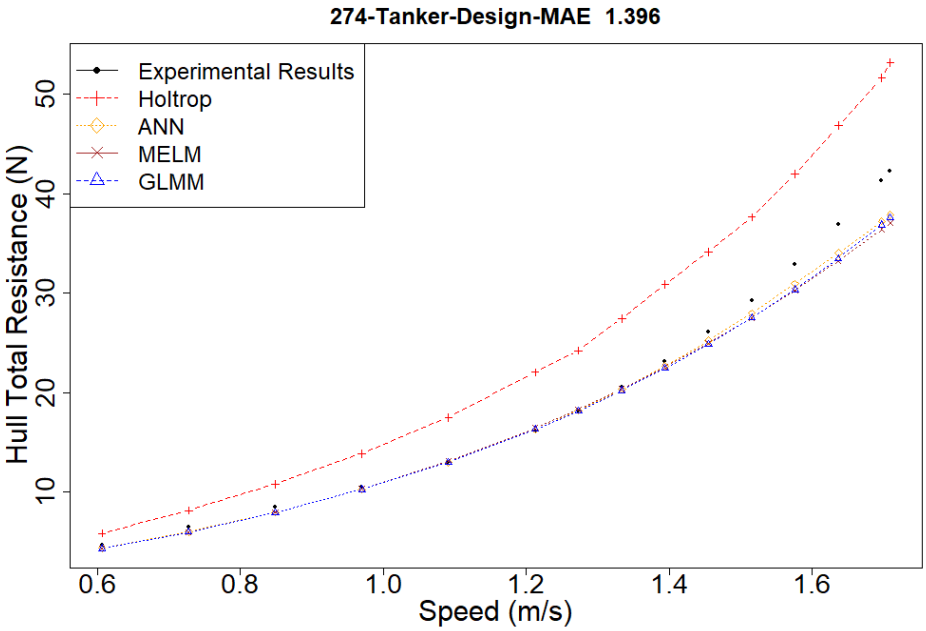

Figure D.59. Resistance Estimation of Model No: 274, Tanker, Loading Condition: Design

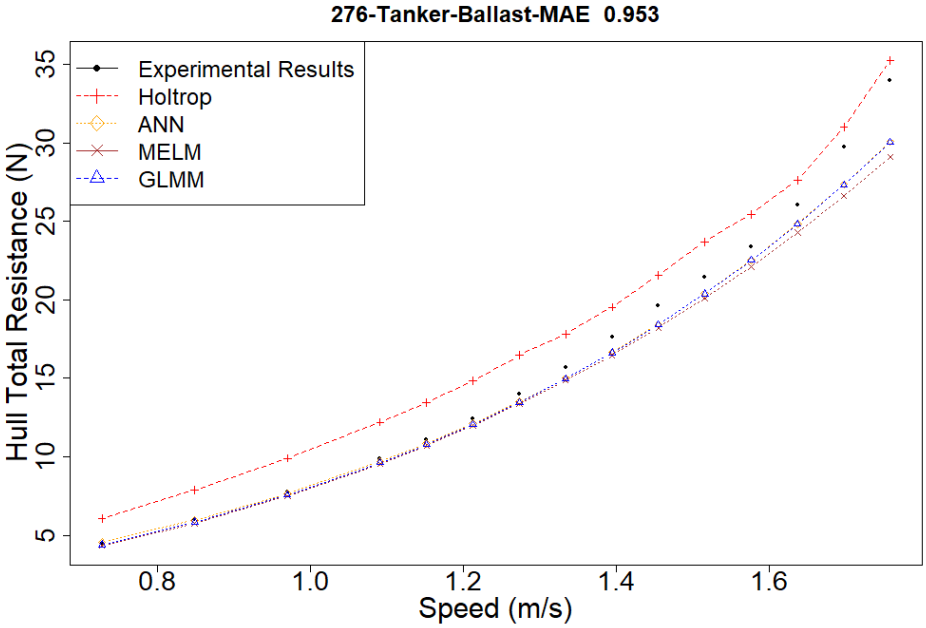

Figure D.60. Resistance Estimation of Model No: 276, Tanker, Loading Condition: Ballast

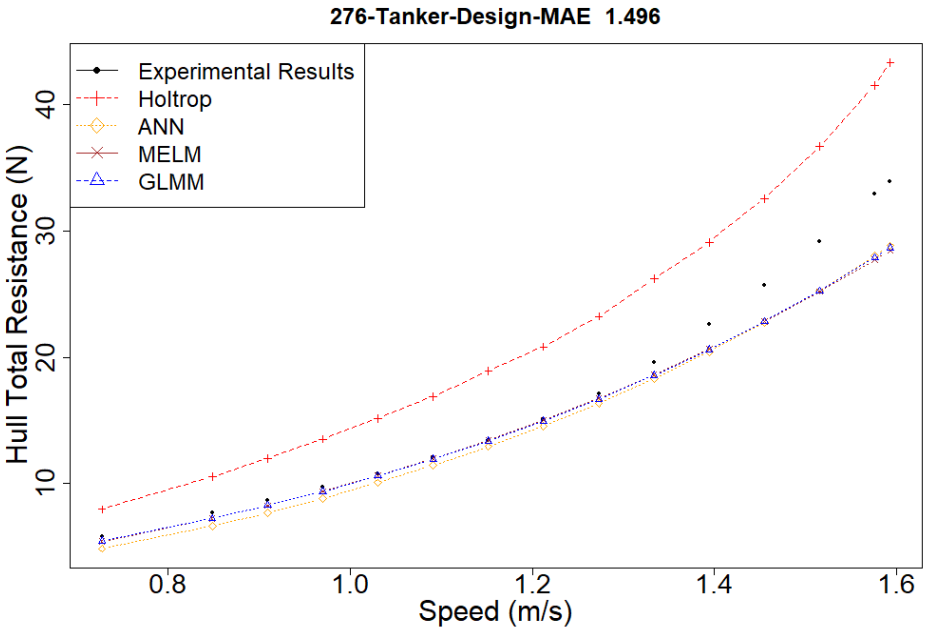

Figure D.61. Resistance Estimation of Model No: 276, Tanker, Loading Condition: Design

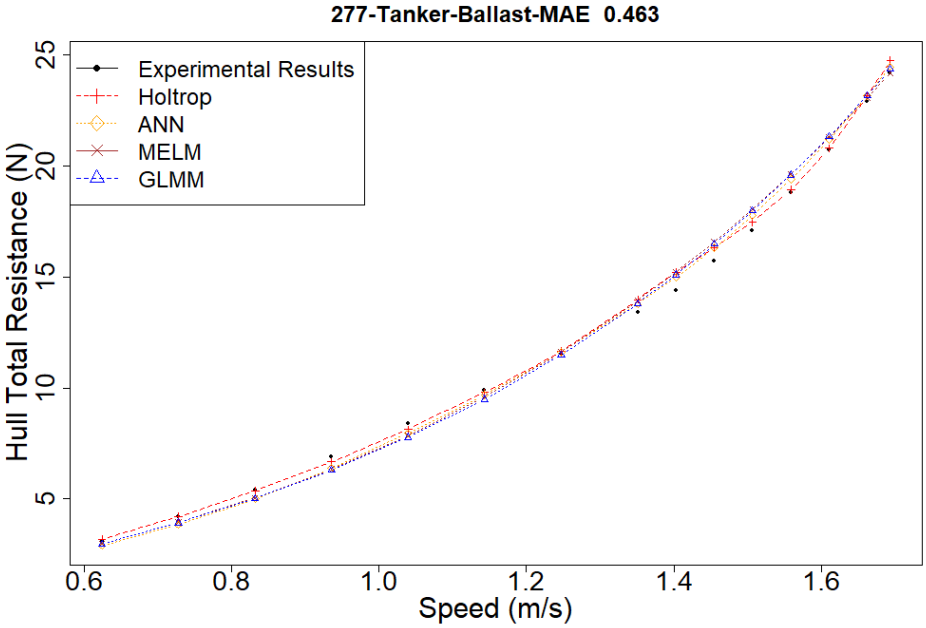

Figure D.62. Resistance Estimation of Model No: 277, Tanker, Loading Condition: Ballast

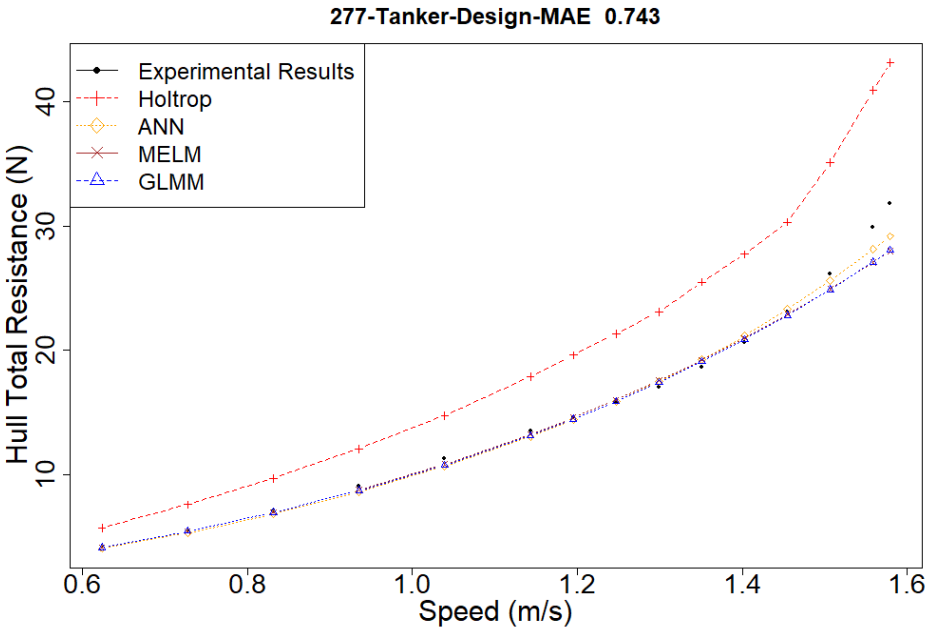

Figure D.63. Resistance Estimation of Model No: 277, Tanker, Loading Condition: Design

APPENDIX E: R CODES

```
library(openxlsx)
library(BBmisc)
library(neuralnet)
library(dplyr)
library(caret)
library(stringr)
library(ggplot2)
library(psych)
library(lme4)
library(afex)
reports_original<-
read.xlsx(``C:/Users/ahmet.pala/Desktop/Ahmet Pala
            - Personal/M.S. Thesis
            /Raporlar/reports_h.xlsx")
reports<-reports_original[,-c(3,7,33,34)]
colnames(reports)[9]<-``Ts"
plotting<-function(Hull_No,Load_Cond){
  idata<-data.frame(
    reports[reports$Model_No==Hull_No,])
  x=
    reports[reports$Model_No==Hull_No &
    reports$Loading_Condition == Load_Cond,32]
  y=
    reports[reports$Model_No==Hull_No &
    reports$Loading_Condition == Load_Cond,40]
  plot(x,y,ylab = ``Hull Resistance (N)",
       xlab = ``Speed (m/s)",
        main = paste(Hull_No,idata[1,3],Load_Cond,sep=``-"))
  #text(x,y, round(y,2), cex=1)}
```

```
# Data Summary
str(reports)
experiments_hull_based<-
  data.frame(table(reports$Model_No))
colnames(experiments_hull_based)<-
  c(``Model_No",``Total_Experiments")
experiments_hull_based<-
  merge(experiments_hull_based,unique(reports[,c(2,3)]),
        by=``Model_No",all.y = FALSE)
experiments_hull_based<-experiments_hull_based[,c(1,3,2)]
experiments_type_based<-
  data.frame(aggregate(Total_Experiments~ Ship_Type,
                       data = experiments_hull_based, sum))
d<-data.frame(table(experiments_hull_based$Ship_Type))
colnames(d)<-c(``Ship_Type",``Number_of_Hulls")
experiments_type_based<-
  merge(experiments_type_based,d,by=``Ship_Type")
experiments_type_based<-
  experiments_type_based[,c(1,3,2)]
#Total_Number_of_Ship_Types_with_Total_Number_of_Experiments
cargoships<-reports[reports$Ship_Type==``Tanker" |
                    reports$Ship_Type== ``General Cargo Ship"
                  |reports$Ship_Type==``Container Ship" |
                    reports$Ship_Type==``Bulk Carrier",]
set.seed(1) #Homogenious data splitting for train and test
cargoships$Ind<-paste(
  cargoships$Model_No,
    cargoships$Loading_Condition,sep = ``,")
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))}
cargoships.norm<-cargoships
```

```
cargoships.norm[,c(6:29,
                   32,33,41,42)] <-
  as.data.frame(lapply(
    cargoships[,c(6:29,
        32,33,41,42)], normalize))
cargoships_unique<-cargoships
cargoships_unique$merged<-
  paste(cargoships_unique$Model_No,
        cargoships_unique$Ship_Type,sep = ",")
cargoships_unique<-data.frame(
  unique(cargoships_unique$merged))
colnames(cargoships_unique)<-"merged"
cargoships_unique<-data.frame(
  str_split_fixed(cargoships_unique$merged,
                  ",", 2))
colnames(cargoships_unique)<-
  c("Model_No","
Ship_Type")
train.rows<-
  createDataPartition(
    y= cargoships_unique$Ship_Type,
          p=0.6, list = FALSE)
train_Model_Nos<-
  cargoships_unique[train.rows,]
train.data<-
  cargoships.norm[cargoships.norm$Model_No
           %in% train_Model_Nos$Model_No,]
test.data<-
  cargoships.norm[!(cargoships.norm$Model_No
      %in% train_Model_Nos$Model_No),]
lmer.cargoships<-lmer(
```

```
  log(Hull_Total_Res) ~ LWL + BWL +
   AWS + CP + CM + Bulb + Vs + I(Vs^2) +
  sqrt(Vs) + (1|Ind), data = train.data)
summary(lmer.cargoships)
AIC(lmer.cargoships)
plot(lmer.cargoships)
pairs.panels(
  cargoships[,c(7,8,14,22,23,30,32)],
              gap=0,
              pch = 21) #input relationship graph
coef(lmer.cargoships)
pairs.panels(
  cargoships[,c(7,8,10,14,17,19,22,23,24,32)],
              gap=0,
              pch = 21) #input relationship graph
pairs.panels(cargoships[,c(32,14,23,21,7,20,13)],
              gap=0,
              pch = 21) #input relationship graph
# Leave One-Out Cross Validation - GMELM Final Model
CrossValidationLeaveOneOutGMELM<-function(data){
  data_unique<-data
  data_unique$merged<-
    paste(data_unique$Model_No,
          data_unique$Ship_Type, sep = "","")
   data_unique<-data.frame(unique(data_unique$merged))
   colnames(data_unique)<-""merged"
  data_unique<-
    data.frame(str_split_fixed(
      data_unique$merged, "","", 2))
   colnames(data_unique)<-c(""Model_No"","Ship_Type")
   k=nrow(data_unique)
```

```
  result_total_ann_myv<-data.frame(
    matrix(nrow = k, ncol = 3))
  colnames(result_total_ann_myv)<-
    c(''One_Out",''MSE_Train",''MSE_Test")
  result_total_ann_myv$One_Out<-1:k
  for (i in 1:k) {
    train1.data<-data[data$Model_No!=data_unique[i,1],]
    test1.data<-data[data$Model_No==data_unique[i,1],]
    melm<-glmer(Hull_Total_Res ~ LWL + BWL +
                  AWS + CB + CWP +
                  Bulb + Vs + sqrt(Vs) + (1|Ind), family =
                  Gamma(link = ''log"),
                data = train1.data)
    pr.nn_test <- data.frame(exp(predict(melm, test1.data,
                                         allow.new.levels = TRUE)))
    #pr.nn_test<-exp(pr.nn_test)
    test.r <- test1.data$Hull_Total_Res
    pr.nn_train <- data.frame(exp(predict(melm, train1.data,
                                          allow.new.levels = TRUE)))
    #pr.nn_train<-exp(pr.nn_train)
    train.r <- train1.data$Hull_Total_Res
    MSE.nn.test <- sum((test.r - pr.nn_test)^2)/
      nrow(test1.data)
    result_total_ann_myv[i,3]<-MSE.nn.test
    MSE.nn.train <- sum((train.r - pr.nn_train)^2)/
      nrow(train1.data)
    result_total_ann_myv[i,2]<-MSE.nn.train
    result_total_ann_myv[i,4]<-data_unique[i,1]
    result_total_ann_myv[i,5]<-data_unique[i,2]
  }
  #return(mean(result_total_ann_myv$MSE_Test))
```

```
  return(result_total_ann_myv)
}
set.seed(2021)
LoutGMELM<-CrossValidationLeaveOneOutGMELM(cargoships.norm)
mean(LoutGMELM$MSE_Test)
# PRODUCING FINAL DATASET with ESTIMATIONS for FINAL GRAPHS
set.seed(2021)
RESULTSCrossValidationLeaveOneOutGLM<-function(data){
  data_unique<-data
  data_unique$merged<-
    paste(data_unique$Model_No,data_unique$Ship_Type,
          sep = ",")
  data_unique<-data.frame(unique(data_unique$merged))
  colnames(data_unique)<-"merged"
  data_unique<-
    data.frame(str_split_fixed(data_unique$merged, ",", 2))
  colnames(data_unique)<-
    c("Model_No","Ship_Type")
  k=nrow(data_unique)
  glmResult<-data.frame()
  result_total_ann_myv<-
    data.frame(matrix(nrow = k, ncol = 3))
  colnames(result_total_ann_myv)<-
    c("One_Out","MSE_Train","MSE_Test")
  result_total_ann_myv$One_Out<-1:k
  for (i in 1:k) {
    train1.data<-data[data$Model_No!=data_unique[i,1],]
    test1.data<-data[data$Model_No==data_unique[i,1],]
    glm1<-glm(Hull_Total_Res ~ LWL + BWL + TA + AWS +
                AB + AT + CP + CM + CWP +
                Vs + I(Vs^2) + I(Vs^3),
```

```
                    data=train1.data,family = Gamma(link = ''log''))
    pr.nn_test <- data.frame(exp(predict(glm1,test1.data)))
    #pr.nn_test<-exp(pr.nn_test)
    test.r <- test1.data$Hull_Total_Res
    pr.nn_train <- data.frame(
      exp(predict(glm1,train1.data)))
    #pr.nn_train<-exp(pr.nn_train)
    train.r <- train1.data$Hull_Total_Res
    MSE.nn.test <- sum((test.r -
                          pr.nn_test)^2)/nrow(test1.data)
    result_total_ann_myv[i,3]<-MSE.nn.test
    MSE.nn.train <- sum((train.r -
                          pr.nn_train)^2)/nrow(train1.data)
    result_total_ann_myv[i,2]<-MSE.nn.train
    result_total_ann_myv[i,4]<-data_unique[i,1]
    result_total_ann_myv[i,5]<-data_unique[i,2]
    glmResult<-rbind(glmResult,pr.nn_test)
  }
  #return(mean(result_total_ann_myv$MSE_Test))
  return(glmResult)
}
GLMFinalResult<-
  RESULTSCrossValidationLeaveOneOutGLM(cargoships.norm)
colnames(GLMFinalResult)<-''GLM''
MELMFinalResult<-
   RESULTSCrossValidationLeaveOneOutMELM(cargoships.norm)
colnames(MELMFinalResult)<-''MELM''
ANNFinalResult<-
  RESULTSCrossValidationLeaveOneOutANN(cargoships.norm,
                          nofNeuron = 2,thresh = 0.5)
colnames(ANNFinalResult)<-''ANN''
```

```
GMELMFinalResult<-
   RESULTSCrossValidationLeaveOneOutGMELM(cargoships.norm)
colnames(GMELMFinalResult)<-''GMELM"
#␣Tanker
setwd(''C:/Users/ahmet.pala/
␣␣␣␣␣␣Desktop/Ahmet␣Pala␣-␣Personal/
␣␣␣␣␣␣M.S.␣Thesis/Final_Plots_Last/Tanker")
data<-final_cargoships
data<-data[data$Ship_Type==''Tanker" ,]
Labels<-data.frame(unique(data$Ind))
for␣(i␣in␣1:nrow(Labels)){
␣␣aa<-data[data$Ind==Labels[i,1],]
␣␣#␣Producing␣plot
␣␣main1<-unique(aa$Ind)
␣␣main2<-unique(aa$RMSE_GLMM)
␣␣main<-paste(paste(main1,''MAE␣",sep = ''-"),main2)
␣␣png(paste(main1,''png",sep = ''."),␣width␣=␣1000,
␣␣␣␣␣␣height␣=␣684)
␣␣plot(aa$Vs,(aa$Hull_Total_Res),
␣␣␣␣␣␣␣␣ylab␣=␣''Hull␣Total␣Resistance␣(N)",
        xlab = ''Speed (m/s)",␣main␣=␣main,
␣␣␣␣␣␣␣␣ylim=range(␣c((aa$Hull_Total_Res),
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣(aa$Holtrop),(aa$MELM),
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣(aa$ANN),(aa$GMELM))␣),
␣␣␣␣␣␣␣␣type␣=␣''p", pch = 20,
        col = ''black",␣lty␣=␣1,␣lwd␣=␣1)
␣␣#␣Holtrop
␣␣lines(aa$Vs,(aa$Holtrop),␣pch␣=␣3,
␣␣␣␣␣␣␣␣␣␣col␣=␣''red", type = ''b",
␣␣␣␣␣␣␣␣␣␣lty␣=␣2,␣lwd␣=␣1)
␣␣#␣ANN
```

```
  lines(aa$Vs,(aa$ANN), pch = 5,
        col = ''orange", type = ''b",
        lty = 3, lwd = 1)
  # MELM
  lines(aa$Vs,(aa$MELM), pch = 4,
        col = ''brown", type = ''b",
        lty = 3, lwd = 1)
  # GMELM
  lines(aa$Vs,(aa$GMELM), pch = 2,
        col = ''blue", type = ''b",
        lty = 3, lwd = 1)
  # 4. Add a legend to the plot and set legend lty
  legend(''topleft",
         legend = c(''Experimental Results",
                    ''Holtrop",''ANN",''MELM",''GLMM"),
         col = c(''black",''red",''orange", ''brown",
                 ''blue"),pch = c(20,3,5,4,2),
         lty = 1:3, cex = 1.25)
  dev.off()
}

### ABLINE of GMELM Results ###
plot(final_cargoships$Hull_Total_Res,
     final_cargoships$GMELM,
     ylab = ''GLMM Results",
     xlab = ''Ata Nutku Experiment Results",
     type = ''p", pch = 20,
     col = ''black", lty = 1, lwd = 1)
abline(0,1)
###### Final Table for Thesis Production ######
FinalTable<-data.frame(matrix(ncol = 7,nrow = 12))
```

```
colnames(FinalTable)<-
  c("Ship_Type","Loading_Condition",
    "Holtrop","MELM","ANN","GLM","GMELM")
FinalTable$Loading_Condition<-
__c("Design",
    "Ballast","Heavy_Loaded")
FinalTable$Ship_Type<-
  rep(c("Bulk Carrier",
________"General_Cargo_Ship",
        "Container Ship","Tanker"),each=3)
for (i in 1:nrow(FinalTable)){
  data<-final_cargoships[final_cargoships$Ship_Type==
                           FinalTable[i,1] &
                           final_cargoships$Loading_Condition==
                           FinalTable[i,2],]
  FinalTable[i,3]<-
    mean(sqrt((data$Holtrop -
                 data$Hull_Total_Res)^2)/data$Hull_Total_Res)
  FinalTable[i,4]<-
    mean(sqrt((data$MELM -
                 data$Hull_Total_Res)^2)/data$Hull_Total_Res)
  FinalTable[i,5]<-
    mean(sqrt((data$ANN -
                 data$Hull_Total_Res)^2)/data$Hull_Total_Res)
  FinalTable[i,6]<-
    mean(sqrt((data$GLM -
                 data$Hull_Total_Res)^2)/data$Hull_Total_Res)
  FinalTable[i,7]<-
    mean(sqrt((data$GMELM -
                 data$Hull_Total_Res)^2)/data$Hull_Total_Res)
}
```

```
colMeans(FinalTable[,c(3:7)])
```