Article

Prediction of Ship Main Particulars for Harbor Tugboats Using a Bayesian Network Model and Non-Linear Regression

Ömer Emre Karaçay [1], Çağlar Karatuğ [1], Tayfun Uyanık [1], Yasin Arslanoğlu [1] and Abderezak Lashab [2,*]

1   Maritime Faculty, Istanbul Technical University, Istanbul 34940, Turkey;
    omeremrekaracay@gmail.com (Ö.E.K.); karatug@itu.edu.tr (Ç.K.); uyanikt@itu.edu.tr (T.U.);
    arslanoglu@itu.edu.tr (Y.A.)
2   Center for Research on Microgrids (CROM), Energy Department, Aalborg University, 9220 Aalborg, Denmark
*   Correspondence: abl@energy.aau.dk

Abstract: Determining the key characteristics of a ship during the concept and preliminary design phases is a critical and intricate process. In this study, we propose an alternative to traditional empirical methods by introducing a model to estimate the main particulars of diesel-powered Z-Drive harbor tugboats. This prediction is performed to determine the main particulars of tugboats: length, beam, draft, and power concerning the required service speed and bollard pull values, employing Bayesian network and non-linear regression methods. We utilized a dataset comprising 476 samples from 68 distinct diesel-powered Z-Drive harbor tugboat series to construct this model. The case study results demonstrate that the established model accurately predicts the main parameters of a tugboat with the obtained average of mean absolute percentage error values; 6.574% for the Bayesian network and 5.795%, 9.955% for non-linear regression methods. This model, therefore, proves to be a practical and valuable tool for ship designers in determining the main particulars of ships during the concept design stage by reducing revision return possibilities in further stages of ship design.

Keywords: Bayesian network; ship main particulars prediction; ship design; harbor tugboats; non-linear regression

1. Introduction

The process of ship design involves intricate phases, including concept, preliminary, and detailed design for production. This multifaceted endeavor, depicted as a spiral in Figure 1 [1], unfolds through concept design, preliminary design, contract design, and production design stages. Concept design, also known as feasibility verification, stands out as the cornerstone of ship design, given its role in translating the requirements of the mission or ship owner into precise naval architecture and engineering specifications. This phase involves conducting preliminary estimates for critical elements of the proposed vessel, including dimensions, power requirements, and alternative feature sets such as speed range, bollard pull and cargo capacity. Moreover, in the preliminary design stage, it necessitates the creation of essential technical documentation, including the ship's line plan, while further refining fundamental ship features to align with the owner's needs and economic considerations. Following concept design, the process advances to contract design, which involves the meticulous preparation of technical specifications for shipbuilding, along with the completion of necessary calculations and naval architecture drawings. Subsequently, detail design, or production design, represents the concluding phase of ship design, where detailed workshop plans for ship construction are developed. This iterative process, likened to a spiral, ensures that adjustments made to any parameter affecting ship characteristics prompt corresponding modifications throughout the design stages, ultimately leading to the realization of an optimal solution in line with economic criteria and owner requirements.
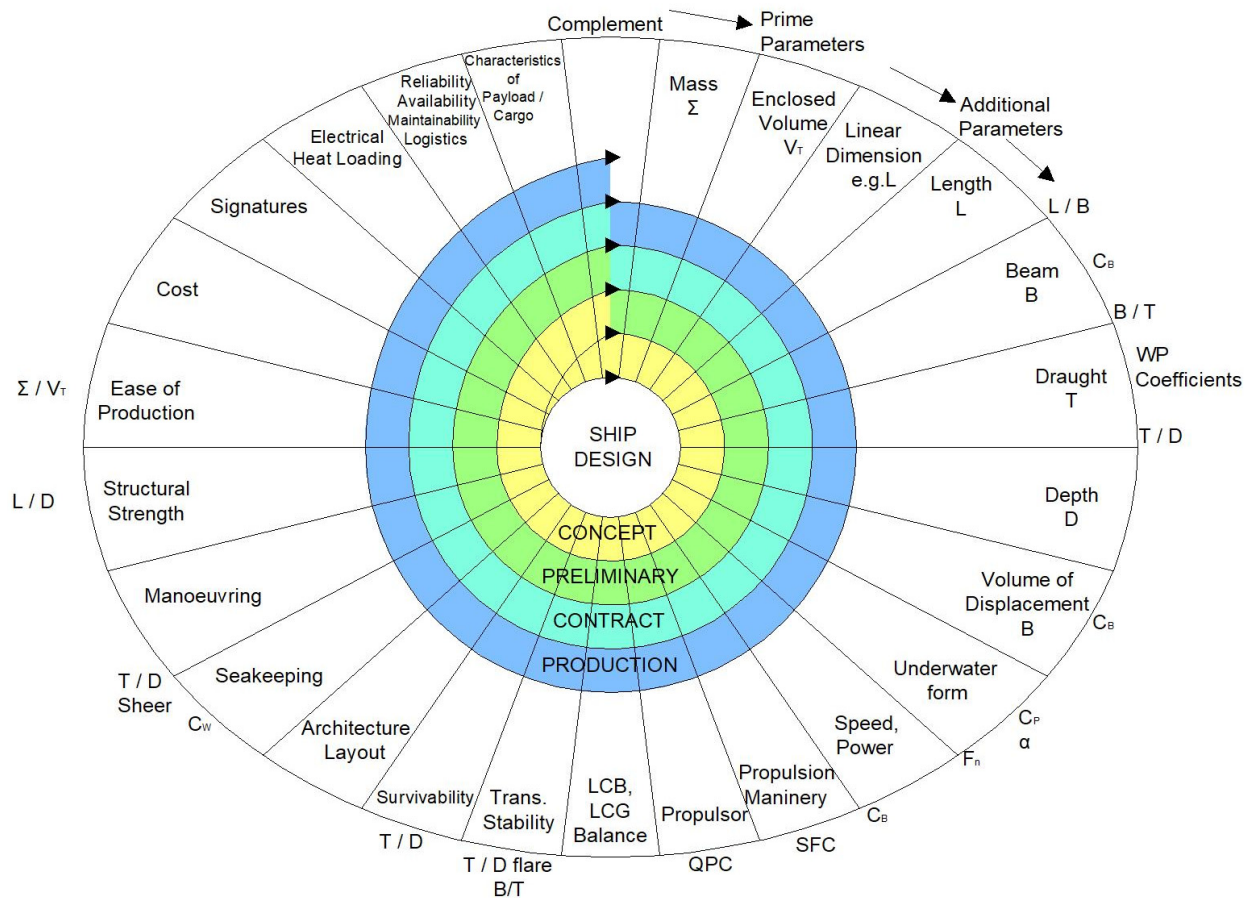
Figure 1. A ship design process model.

Throughout these design stages, the focus is on optimizing the main particulars of the tugboat to ensure it can effectively fulfill its role in servicing ships in busy ports and harbors. By considering these factors at every stage of the design process, harbor tugs must be adapted to meet the specific demands of their operating environments while ensuring safety, efficiency, and reliability.

Naval architects grapple with a plethora of ship parameters, encompassing the main dimensions, strength, hull form, displacement, resistance, powering, freeboard, machinery, endurance, capacities, trim, stability, economic considerations, efficiency, environmental impact, and cost factors [2]. Critical ship features, such as stability, power requirements, and economic efficiency, hinge significantly on the main dimensions. Thus, specifying parameters like length (L), width (B), draft (T), depth (D), freeboard (F), and block coefficient (C_B) is a pivotal aspect of ship design, where these dimensions are harmonized to meet the ship's design conditions. For example, the important considerations for harbor tugboats include a high bollard pull capacity to tow large ships at a very low speed during towing and pushing duties, a service speed that enables tugboats to reach their destinations faster, and maneuverability to navigate congested harbors.

The traditional approach involves using statistical regression equations based on data from a comparable ship with known features [3]. In the preliminary design phase, designers play a crucial role in defining a ship's major features based on implicit client requirements, including draft length, service speed range, and bollard pull capacity. Designers iteratively adjust dimensions by analyzing comparable ships or resort to empirical formulations and machine learning (ML) methods like neural networks (NN) to predict a ship's main particulars and to analyze dynamic systems, especially in early design stages [4–7]. During the early design stage of a ship, it could benefit from using an ML approach, where a large number of configurations must be tested, which could be prohibitive to achieve

using Computational Fluid Dynamics (CFD) or model experiments. It may provide fast predictions with non-linearities taken into account, overcoming the inaccuracies in linear analytical methods currently used in the early design stages [7].

Neural networks are poised to emerge as a pivotal tool in the initial phase of ship design, offering, among other capabilities, the estimation of crucial ship particulars [3]. A comprehensive examination of machine learning applications in sustainable ship design and operation is outlined in a study by [7]. Ref. [6] developed a series of artificial neural network (ANN) and regression equations to predict container ship dimensions, including the length between perpendiculars, breadth, draft, and side depth, utilizing deadweight, TEU capacity, and ship speed as input variables; ref. [8] demonstrated that augmenting the dataset with synthetic data and analyzing it using artificial neural networks (ANNs) can yield favorable outcomes concerning the main particulars of container ships. Employing artificial intelligence (AI) techniques, ref. [3] utilized multilayer perceptron and gradient-boosted trees for predicting key parameters of container ships. Ref. [2] applied non-linear regressions to ascertain the main dimensions, light ship characteristics, and dimensional relationships across various ship types.

Furthermore, neural network (NN) methods have found application in diverse maritime domains, encompassing ship resistance prediction [9–12], ship engine power [13,14] and performance forecasting [15–17], ship hydrostatics [18], ship hydrodynamics, and motion prediction [19–21], as well as condition-based maintenance of machinery systems [22] and fault diagnosis [23]. The effectiveness of regression methods in accurately estimating both point and range estimates for parameters has been well-documented through research across various disciplines of science [24,25]. However, in general, machine learning algorithms tend to replace traditional statistical methods [26].

Addressing the probabilistic dependency structure between multiple parameters necessitates a strategic approach. This strategy involves employing a learning algorithm to construct a neural network capable of answering diverse inquiries. The Bayesian method is adept at avoiding erroneous categorical decisions regarding conditional independencies, conducting model averaging for small datasets, managing missing data, and discerning between models [27]. Notably, its advantages include suitability for small and incomplete datasets, potential for structured learning, integration of diverse information sources, explicit treatment of uncertainty, and support for decision analysis and prompt responses. Leveraging these advantages, various Bayesian models have been demonstrated [28–30] for parameter identification and prediction of ship motions and maneuverability, contributing to the prediction of vessel hydrodynamics. Additionally, Bayesian networks find application in risk assessment [31–34], accident scenario analysis [35–37], reliability analysis [38,39], and fuel consumption analysis [40,41] within the maritime domain.

According to related studies, while traditional empirical methods retain value for certain aspects of ship design, the increasing complexity and evolving nature of modern ships necessitate the adoption of alternative approaches to ensure optimal performance, efficiency, and compliance with regulatory standards. Neural network (NN) methods, as one of these alternative approaches, are utilized across various disciplines within the maritime industry and are also applied in the ship design process. However, in the studies carried out to determine the main dimensions of the ship during the ship design phase, it was seen that the studies were concentrated on only a few cargo ship types such as container ships, and a specific study was required for a specific type of ship, harbor tugboats. Furthermore, this paper uniquely focuses on applying ML methods, specifically Bayesian networks and non-linear regression, to predict the main particulars of harbor tugboats during the concept design stage. While prior studies predominantly focused on cargo ship types, such as container ships, our work addresses the distinctive design parameters of harbor tugboats, providing valuable insights for decision making in ship design. The study involves developing a model, and the vessel's data are evaluated using Bayesian network and non-linear power regression approaches. The results are assessed

using error metrics like mean square error (MSE), absolute percentage error (MAPE), and determination coefficients (R) for comparison.

The primary objective of this study is to present a model with a proposed methodology, serving as a valuable and practical tool for ship designers during the concept design stage of tugboats. This approach enhances the efficiency of the design process, minimizing the need for revisions in later stages. A significant contribution of this study is its demonstration that ML methods can be applied to ship types with distinctive design parameters, such as tugboats, designed for specific purposes beyond cargo transportation.

The remainder of this paper is organized as follows: Section 2 outlines the methodology, data collection, pre-processing, and the developed prediction model. Section 3 provides findings from the case study, and finally, Section 4 concludes the study.

2. Material and Methods

2.1. Bayesian Regularization

Bayesian networks serve as statistical tools utilized to model causal connections and uncertainties among distinct events. These networks depict data relationships within a probabilistic framework, visually outlining the interrelations among a sequence of events and employing probability distributions to articulate the dependence of each event on others [42].

For example, consider the assessment of a health condition where symptoms exhibited by a patient may correlate with specific illnesses. Bayesian networks are well-suited for modeling such scenarios, as they can ascertain the likelihood that symptoms like a cough, fever, and headache indicate a particular ailment [43,44].

These networks facilitate the computation of event probabilities and provide insights into their interrelationships. In essence, Bayesian networks aid in understanding how an event unfolding within a given context may impact other scenarios. As a result, they find extensive applications across diverse domains, including decision making, prognostication, and risk assessment.

A Bayesian network is a graphical representation of an uncertain collection of quantities, also known as a directed acyclic graph (DAG), along with a set of conditional probability tables that correspond to the structure of the DAG [45]. The network consists of a sequence of probability nodes (ovals) and directed arcs connecting these nodes. Nodes represent stochastic variables, defined as a set of discrete states, with each state associated with a probability measure. The arcs between variables denote conditional probability dependencies, signifying the likelihood of a dependent variable Y (child node) being in a specific state for each combination of states of the preceding variables X (parent node). The presence of a directed arc from node X to node Y indicates that X has a direct influence on Y, specified by the conditional probability P(Y | X). Bayesian networks do not allow for directed loops. Despite its concise design, the graphical representation offers a comprehensive probabilistic depiction of the scenario. A key feature of Bayesian networks is their ability to infer from evidence observed at any node, with new knowledge traversing the network and updating all variables in the model under Bayes' rule [46].

Bayesian regularization, as a network training function, optimizes weight and bias values using Levenberg–Marquardt optimization. This method minimizes a combination of squared errors and weights, identifying the optimal combination to construct a network with robust generalization. Implemented in MATLAB R2013b as the "trainbr" trainer, the training process depends on the function's parameters. Validation vectors are employed to halt training prematurely if the network's performance on the validation vectors does not improve or remains constant for "max_fail" epochs consecutively. Bayesian regularization does not use a separate validation set but includes it in the training set [10]. Test vectors serve as an additional check for the network's generalization but do not impact the training. Enabling validation by setting "max_fail" to any strictly positive number ensures weight/bias minimization with shorter training times. Regarding the performance function, this approach shares the same constraint as "trainlm".

2.2. Non-Linear Regression

Non-linear regression, a statistical technique, is employed to model intricate relationships between variables that elude capture by linear models. Despite the challenges inherent in model selection and estimation, its versatility in capturing diverse patterns in data is invaluable for understanding complex relationships.

Non-linear regression analysis is a statistical method used to predict the value of a dependent variable based on the value of an independent variable. It assesses the association between variables by examining the coefficient values of the relevant parameters. This analysis employs a mathematical equation that characterizes the line or curve representing the best fit for the relationship between the dependent variable (Y) and the independent variable (X). The coefficient of determination, denoted as R^2, signifies the extent of variability in Y attributed to X.

As a foundation for approximating the relationship between bollard pull and the other output parameters, power functions are selected. This phrase is commonly used as power function that given in Equation (1) [47]:

$$Y = a\,X^b + E_i \,, \tag{1}$$

where *a* and *b* represent the random coefficients, and the random errors are represented by E_i.

R-squared, the coefficient of determination can be used as [48].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(t_i - o_i)^2}{\sum_{i=1}^{N}(t_i - \overline{o_i})^2} \,, \tag{2}$$

where *t* represents the goal value, *o* represents the output, $\overline{\overline{o}}$ is the average value of the samples, and *N* is the number of samples.

2.3. Prediction Error Metrics

The network performance function was determined by the mean square error (*MSE*) that is represented in Equation (3). For comparisons, the statistical approach of the mean absolute percentage error (*MAPE*) value was utilized and represented in Equation (4) [16].

$$MSE = \frac{1}{n}\sum_{i=1}^{N}(t_i - o_i)^2, \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{N}\left|\frac{t_i - o_i}{o_i}\right| * 100, \tag{4}$$

where *t* represents the goal value, *o* represents the output, and *N* is the number of samples.

2.4. Methodology

The structure of the implemented methodology for this study is depicted in Figure 2. The methodology comprises three primary steps: data collection and pre-processing, model implementation using Bayesian network and non-linear regression algorithms, and obtaining the model results. In the application of the Bayesian network algorithm, the pre-processed data are segregated into three branches: training data, test data, and validation data. The model yielding the most accurate output is determined through testing with real data reserved specifically for this purpose. The performance of the model is assessed based on the error analysis of the obtained output parameters. The closer the values of calculated error metrics are to zero, the more successful the model results are considered to be.
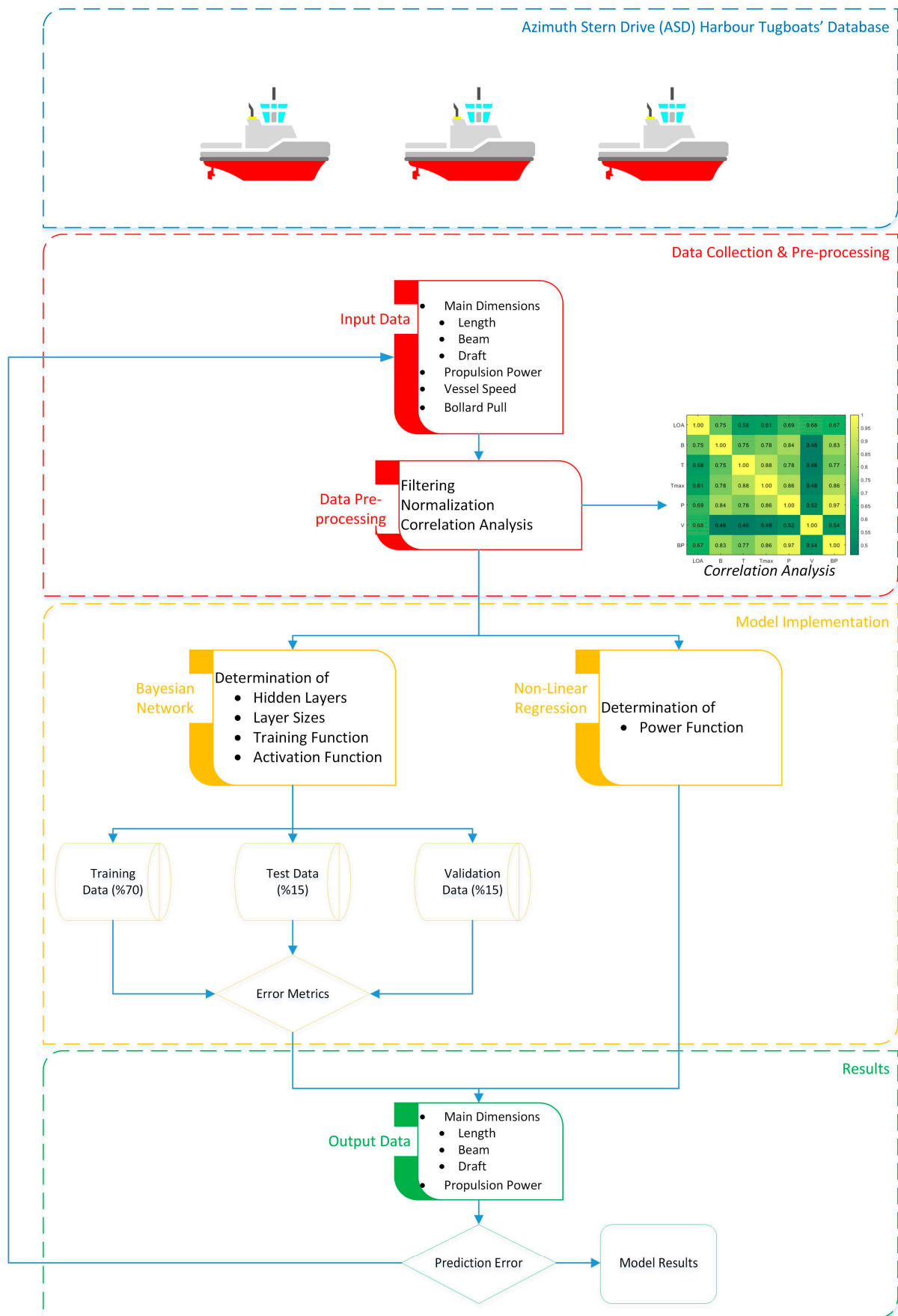
Figure 2. The methodology flow chart of case study analysis.

3. Case Study

A tugboat, though small in size, is a powerful vessel utilized for various towage operations, including search and salvage, firefighting, assisting, transporting, escorting, maneuvering, and berthing other marine vehicles when required within the tugboat's operational scope [49]. Consequently, tugboats can be designed to fulfill one or more of these functions [50].

Tugs are typically classified according to their operational context, delineated by distinctions among harbor, ocean-going, coastal, and river environments, as well as the nature of their tasks. Additionally, these vessels are further stratified based on the configuration or propulsion systems they employ [51]. Several propulsion system arrangements exist, with the four most prevalent being: conventional propulsion systems, Azimuth Stern Drive (ASD), Tractor tug with Rudder Propellers, and Voith Water Tractor, also known as Voith Schneider Propeller (VSP) tug. A profile view of the different types of tugboats is given in Figure 3 [52].
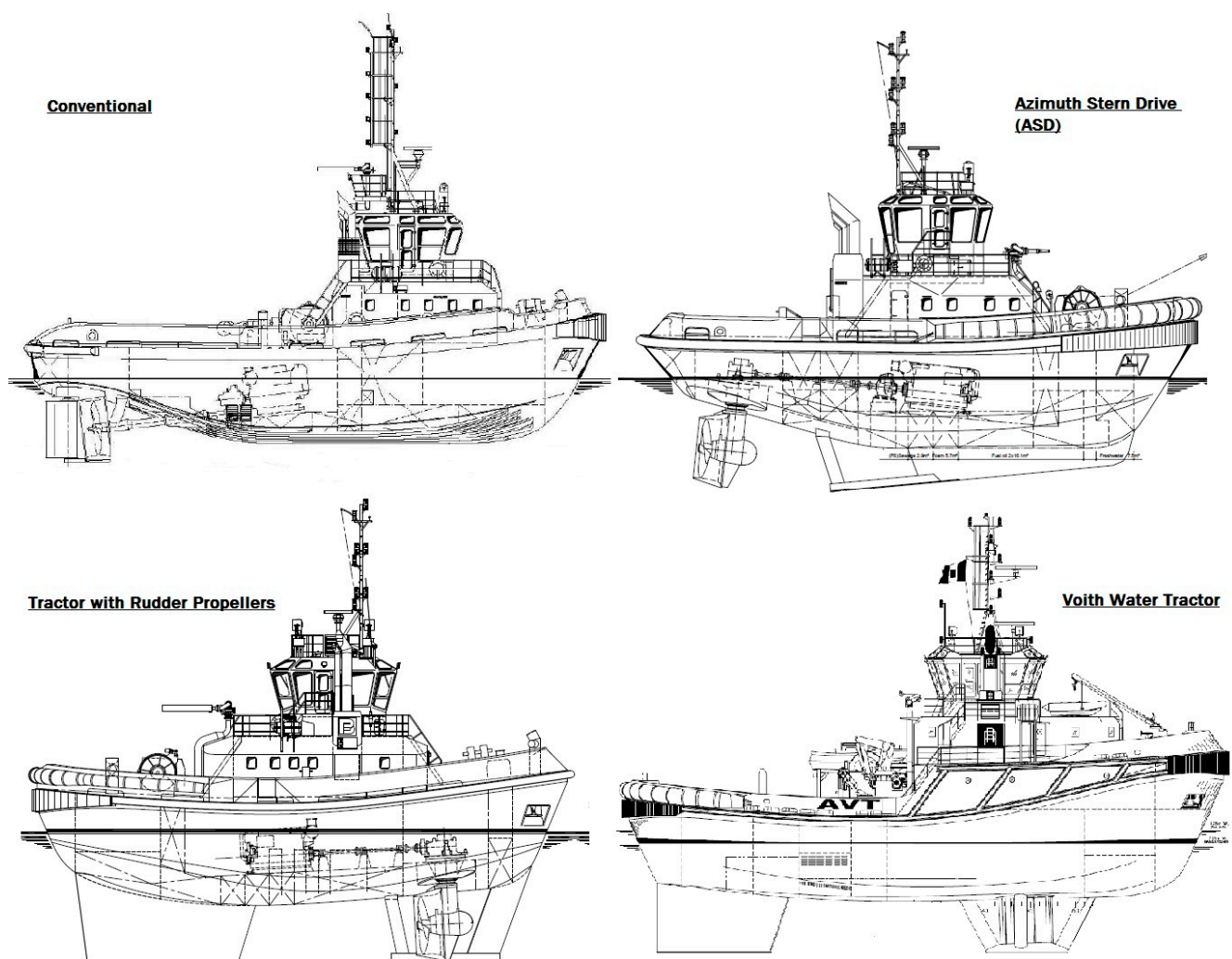
Figure 3. Profile view of different types of tugboats.

Presently, the predominant propulsion systems employed in modern ship-assist tugs consist of Z-drive which are equipped with the azimuthing propulsors or VSP configurations, with harbor tugs typically spanning from 20 to 32 m in length and featuring power outputs ranging from 2000 to 4000 kW, albeit subject to variations dictated by port size and the spectrum of ships serviced [51].

Azimuthing thrusters or azimuthing propulsors, which have been widely utilized for numerous years, are characterized by either non-ducted or ducted propeller configurations, further subdivided into pusher or tractor units, as illustrated in Figure 4.
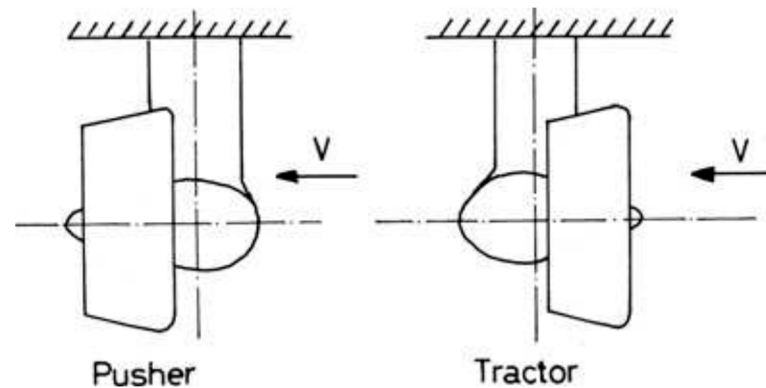
Figure 4. Pusher and tractor Azimuthing thruster units.

In general, the maximum speed and bollard pull of tugboats are critical performance measures that directly affect the effectiveness, safety, and efficiency of port tug operations. Tugboats must possess a high bollard pull to exert adequate force for safely and effectively towing or pushing large ships in various environmental conditions [53]. Tugs are constructed to exceed bollard pull forces through the optimization of their underwater hull form, leveraging hydrodynamic forces to generate higher pull forces [52]. Moreover, a higher maximum speed enables tugboats to reach their destinations more swiftly; handle larger ships efficiently, thus reducing the time and effort needed for maneuvering operations; and respond promptly to changing situations and position themselves effectively.

The paramount requirements for a tug are the bollard pull and maximum speed, determined during the concept design phase. This determination is contingent upon factors such as the size and type of the ship the tug is designed to assist, the number of tugs in the port, and environmental conditions like currents, tides, and winds prevalent in the tug's operational area. All tugs, particularly harbor tugs, must be designed to be highly maneuverable with exceptional stability.

The total engine power and hull from parameters of the tugboats with many other parameters such as propeller parameters are limitations for tugboats' maximum speed and bollard pull force [54–56]. Therefore, the dependent parameters for tugboats, namely bollard pull (BP) and ship speed (V), can be expressed as functions of the primary independent variable parameters related to a tugboat's main dimensions: length (L), width (B), draft (T), block coefficient (C_B), and main propulsion power (P). Additionally, other independent variable parameters influencing the bollard pull and ship speed include the vessel's hull form, heel and trim conditions, and the configuration of the propulsion system, encompassing the main engine(s) and power transmission equipment(s). Environmental conditions, such as currents, waves, sea state, wave dimensions, water depth, and towing rope length, also contribute to the variability of bollard pull and ship speed parameters [49].

3.1. Bayesian Network Structure

In this case study, the analysis is grounded in a compiled database encompassing the key characteristics of tugboats. The focus is on exploring the relationship between the dependent variables BP and P, and the independent variables L, B, T, and P. Each entry in the database encapsulates a joint probability distribution across the variables within the dataset. The primary aim of this section is to estimate the joint probability distribution of a set of variables utilizing a Bayesian network as a representation. Figure 5 illustrates the Bayesian network that has been trained for tugboats.
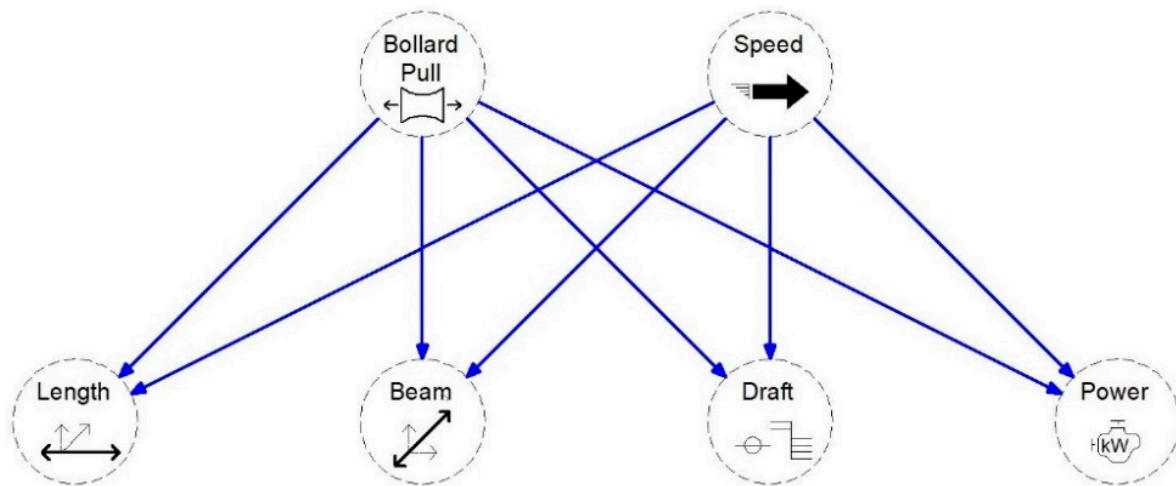
Figure 5. Bayesian network for tugboats.

3.2. Data Collection and Pre-Processing

In this study, an assessment is conducted on diesel-powered Z-Drive harbor tugboats including azimuth stern drive (ASD), Rotortug, and Z-Tractor tugboats, equipped with azimuthing propulsors, encompassing diverse main dimensions, speeds, and bollard pull values. A dataset containing the main characteristics of over 200 tugboats which are designed by well-known tugboat designers and built in leading shipyards in the tug building industry is collected and compiled for analysis. To refine the dataset, pre-processing steps are implemented, removing sister vessels that share identical designs to avoid adversely impacting analysis distribution. Table 1 illustrates the range of parameter variability within the dataset.

Table 1. Descriptive statistics of the dataset for listed tugboats.

	Maximum	Minimum	Mean	Median	Mode	Std	Var
Length [m]	42.00	18.70	29.13	29.05	24.40	4.40	19.32
Beam [m]	16.00	9.20	12.21	12.02	12.00	1.24	1.55
Draft [m]	6.50	3.40	5.12	5.10	5.30	0.67	0.45
Draft, Max. [m]	7.20	3.95	5.58	5.60	5.70	0.67	0.45
Power [kW]	7600	1939	4269.22	4190	5050	951.64	905,609.22
Speed, Max. [knot]	15.10	11.50	13.05	13.00	13.00	0.75	0.57
Bollard Pull [mT]	120.00	31.71	71.25	70.00	70.00	16.04	257.41

As depicted in Table 1, a total of 476 data samples were obtained from 68 distinct diesel-powered Z-drive harbor tugboat series equipped with azimuthing propulsors. These tugboats exhibit an average length of 29.1 m, ranging from 18.7 to 42 m, and an average bollard pull (BP) capacity of 71.25 metric tons, ranging from 31 to 120 metric tons. This investigation aims to explore the relationship between vessel length and the distribution of bollard pull and speed parameters.

Figure 6 presents the distribution of bollard pull and speed parameters, as well as the conditional probability distribution relative to vessel length for the collected dataset. When examining these conditional probability distributions, the peak probability is observed around 70 metric tons BP at approximately 30 m in length and a speed of approximately 13 knots.
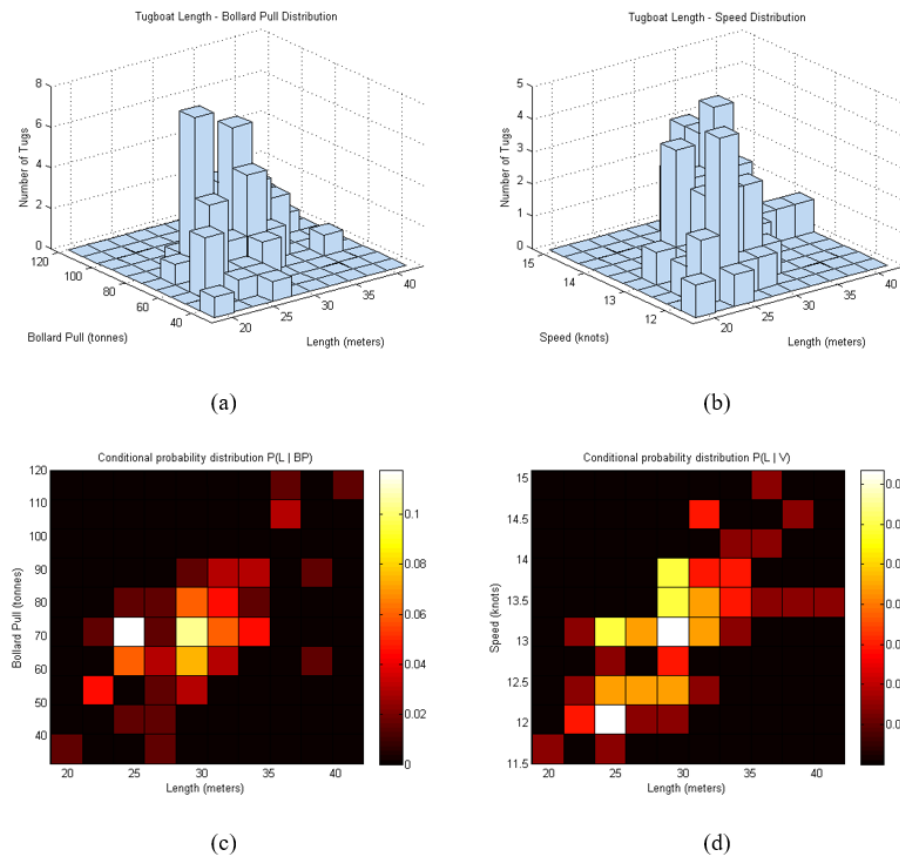
Figure 6. Bollard pull and speed parameters distribution related to vessel length for the collected dataset: (**a**) L | BP; (**b**) L | V and conditional probability distributions: (**c**) P (L | BP); (**d**) P (L | V).

Correlation Analysis

Regression analysis scrutinizes the mathematical relationship among two or more variables, while correlation analysis assesses the direction and magnitude of this relationship. Correlation analysis, as a statistical method, furnishes insights into the relationship, direction, and strength of the correlation between variables.

The correlation coefficient serves as a metric indicating the strength of the relationship between the dependent and independent variables. It gauges the linear relationship between two variables and is unit-independent, ranging between −1 and 1. A coefficient nearing 0 signifies a weak correlation, while a proximity to 1 indicates a strong correlation.

In this study, the Pearson correlation coefficient is computed from the collected data and visually represented in Figure 7. The intricate relationship between these variables is delineated in Figure 8. The figures illustrate a robust correlation between the data of length (LOA), beam (B), draft (T), draft-maximum (T_{max}), and power (P) and bollard pull (BP), whereas the correlation with speed (V) is relatively weaker.

Upon detailed examination of the correlation analysis, it becomes evident that the most robust correlations are evident between the P and BP variables, showcasing a substantial correlation coefficient of 0.97. Additionally, a notable correlation emerges between the T and Tmax variables, boasting a strong coefficient of 0.88. Subsequently, the correlations between the P-Tmax and BP-Tmax variables follow closely behind with coefficients of 0.86. Noteworthy as well is the relatively strong correlation coefficient of 0.84 observed between the P and B variables. Moreover, another salient correlation within the matrix is apparent between the B and BP variables, exhibiting a coefficient of 0.83, further accentuating the interrelatedness among the variables.
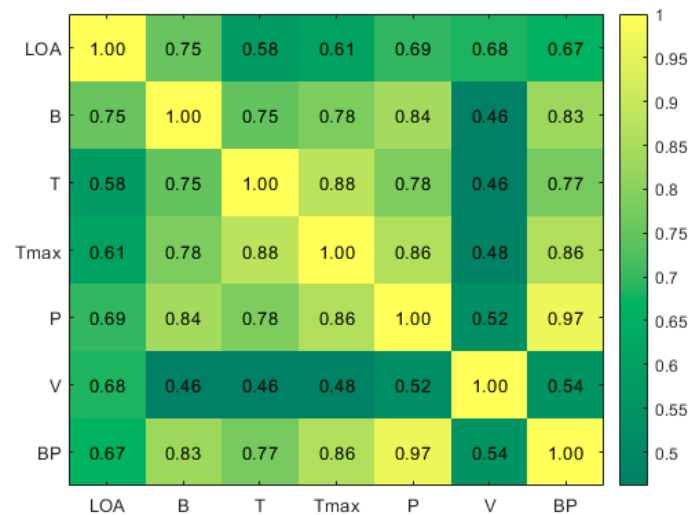
Figure 7. Pearson correlation coefficients of the collected and pre-processed data.
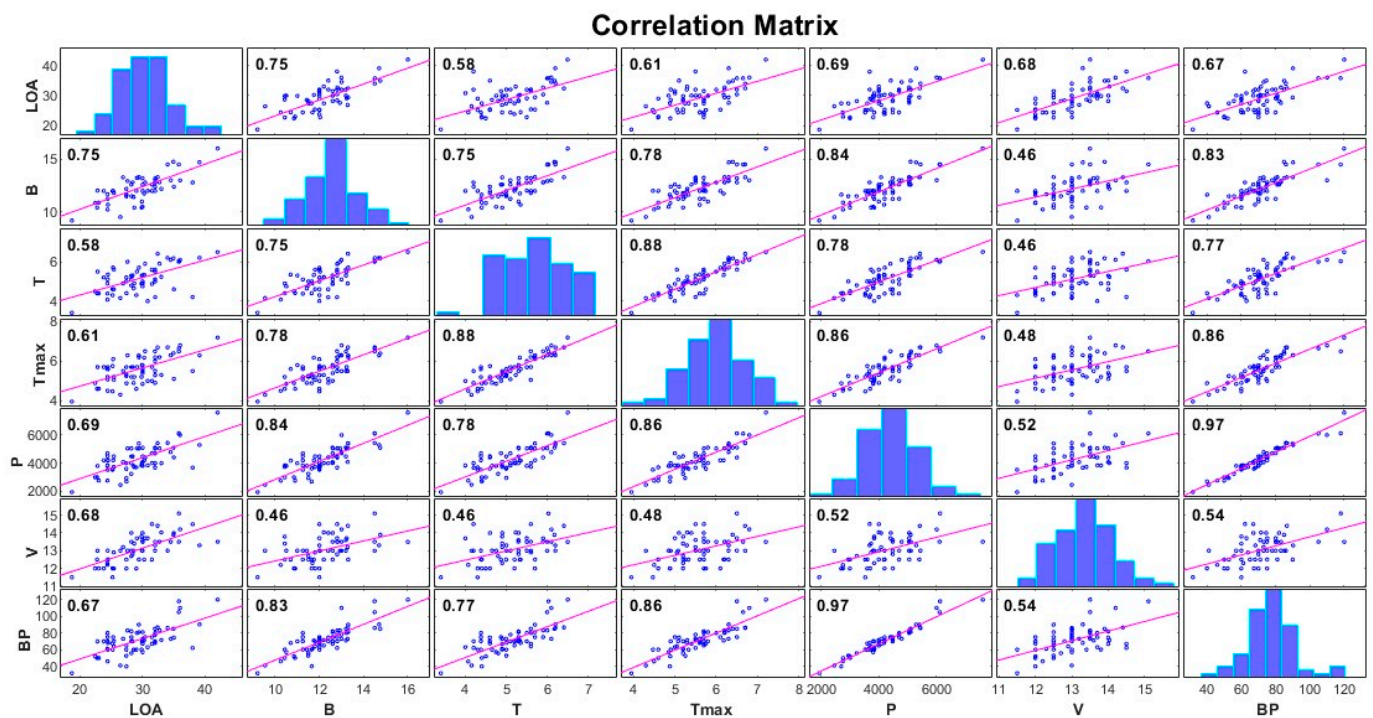

Figure 8. Correlation matrix of the collected and pre-processed data (Dots: Variables; Lines: Least-squares references; Histogram; Distribution of a variables).

3.3. Implementation of Model

In the present analysis, a model utilizing Bayesian network and non-linear regression methods was trained and assessed using the MATLAB program. As mentioned in the preceding section, the validation dataset was employed to address the overfitting issue, with a maximum validation failure set at 100. Back-propagation learning was conducted on a single hidden layer with 10 hidden nodes. A total of 476 data samples from 68 distinct diesel-powered Z-drive harbor tugboat series were collected for this investigation [6]. The dataset was partitioned such that 70% of the samples were allocated to the training set, 15% to the validation set, and another 15% to the test set. The training set exclusively served for training the network, while the test set was used to evaluate network performance. The activation functions "logsig" and "purelin" were applied to the hidden layer and output layer, respectively [5]. The distribution graphs of output parameters, specifically those

related to vessel length (L) with input parameters bollard pull (BP) and ship speed (V), are presented in Figure 6 along with their conditional probability distributions.

The structure of a neural network (NN) comprises multiple processing units capable of bidirectional communication through connections with varying weighting factors. Generally, not all neural networks possess a structure that includes the following components: input layer, signals, hidden layer(s), and output layer. The number of neurons in the input and output layers depends on the nature of the problem being addressed, considering the number of variables and outcomes.

Input values (x_i) from previous layers are processed in a single artificial neuron using bias (b) and weights (w_i), as seen in the below Equation (5) [17]:

$$S = b + \sum_{i=1}^{n} w_i * x_i, \tag{5}$$

The structure of the Bayesian network model for this investigation is depicted in Figure 9. This model has two inputs (bollard pull and speed) and ten hidden layers with five outputs (length, beam, draft, draft-maximum, and power).

Figure 9. Bayesian network model framework diagram of the implemented model.

3.4. Performance of Model

The established model is evaluated with the parameters of the collected and pre-processed dataset in the MATLAB program. After the 754th attempt, both the validation and test sets had an upward trend for the Bayesian network. The best validation performance was thus reached at epoch 754. Figure 10 depicts the performance graph of the constructed Bayesian network.
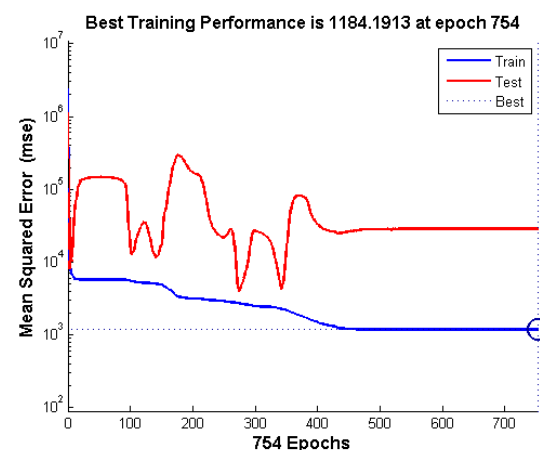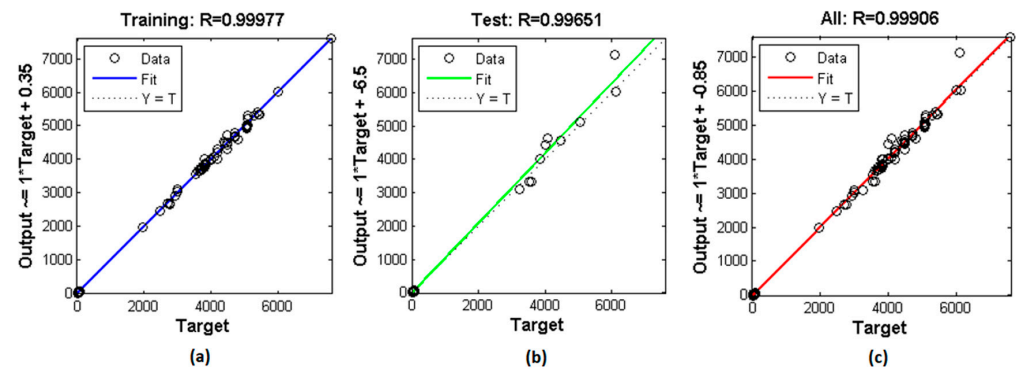

Figure 10. Performance plot of the developed network.

Figure 11 depicts the regression graph between the estimated Bayesian network model values and actual ship data. The observed determination of coefficients (R) for training, test, and the total process were 0.99977, 0.99651, and 0.99906, respectively.

Figure 11. Bayesian regression graphics for (**a**) training, (**b**) test and (**c**) total process, of the developed network.

Non-linear regressions were presented in two distinct groups based on the input variables, bollard pull (BP), and vessel speed (V) in this case study. In the first group, power functions were determined with respect to bollard pull (BP) values, whereas in the second group, power functions were determined with respect to vessel speed (V) values.

The power function coefficients are predicted with a 95% confidence level for the bounds. The equations of predicted non-linear regression functions with their coefficients of determination values for this case study are presented in Table 2.

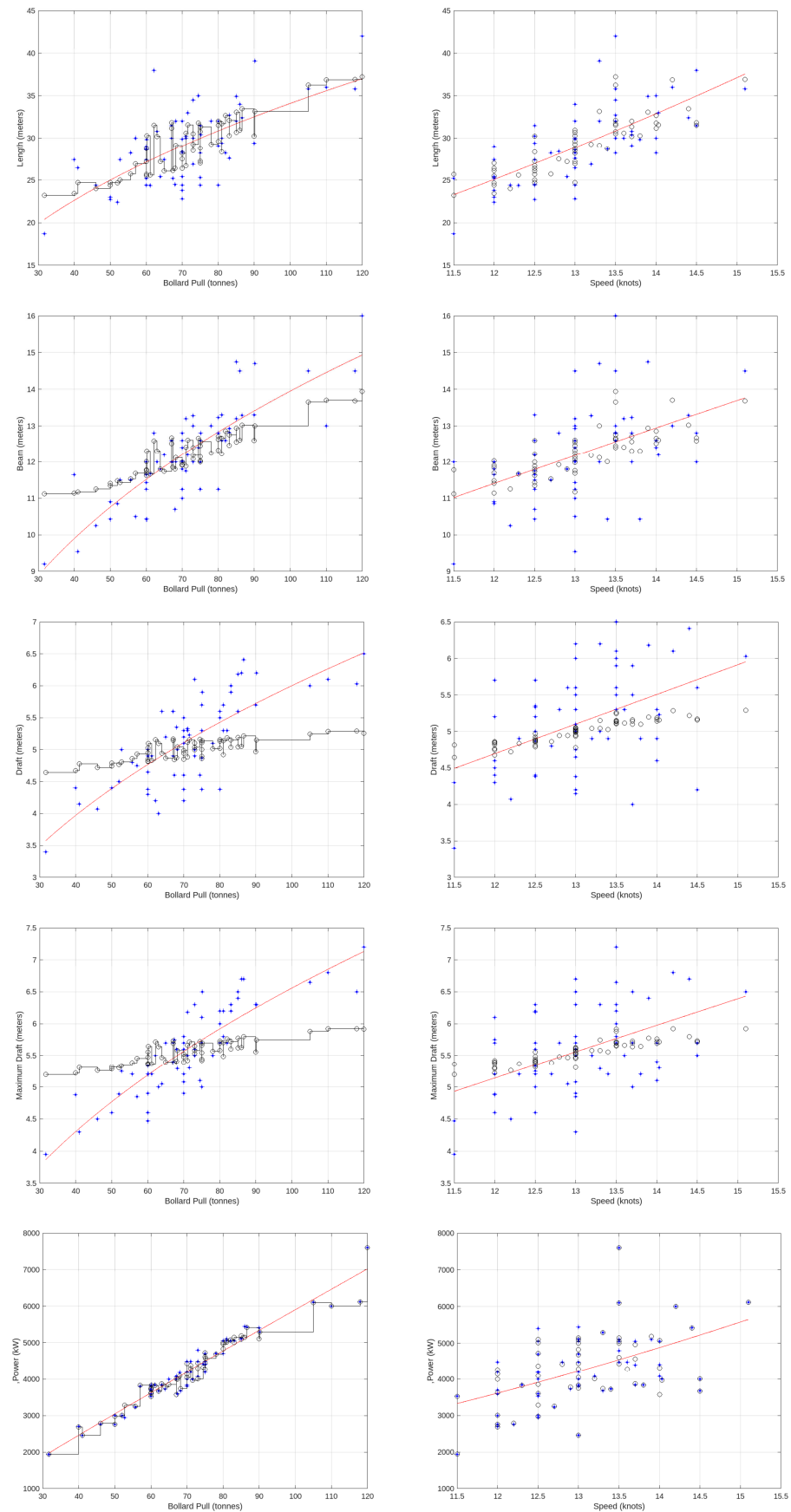**Table 2.** Non-linear regressions' power functions and coefficients of determination of case study.

	Non-Linear Regression Group 1 Functions	Coefficient of Determination (R^2)	Non-Linear Regression Group 2 Functions	Coefficient of Determination (R^2)
Length [m]	$BP = 4.338\,L^{0.4477}$	0.4383	$V = 0.319\,L^{1.757}$	0.4626
Beam [m]	$BP = 2.484\,B^{0.3747}$	0.6891	$V = 1.497\,B^{0.817}$	0.2159
Draft [m]	$BP = 0.7528\,T^{0.4508}$	0.6064	$V = 0.360\,T^{1.034}$	0.2129
Draft, Max. [m]	$BP = 0.79\,T_{max}{}^{0.4596}$	0.7527	$V = 0.455\,T_{max}{}^{0.976}$	0.2266
Power [kW]	$BP = 73.15\,P^{0.9535}$	0.9368	$V = 29.020\,P^{1.942}$	0.2656

3.5. Model Outputs

This study aimed to develop a Bayesian network model and simple power regression with a power function equation to predict the key characteristics of a tugboat during the early design phase. Bollard pull (BP) and vessel speed (V) were selected as the input variables, with length, beam, draft, maximum draft, and power serving as the output variables.

Moreover, Figure 12 illustrates the actual and predicted output values using a Bayesian network and power regression for length (L), beam (B), draft (T), and power (P). These are presented in plots depicting the conditional mean as a function of bollard pull (BP) and vessel speed (V).

The results depicted in Figure 12 indicate that both the Bayesian network and power regressions yield moderate results compared to the actual data, confirming the coherence of the predictions. However, while the results for the bollard pull values between 60 and 80 metric tons are more accurate, deviations are observed for the maximum and minimum pull values.

Figure 12. Comparison of the design parameters: actual values ('+'); Bayesian network predictions ('o', step curve); power regressions (power function curve).

3.6. Model Error

The mean squared error (MSE) histogram of the developed model is summarized in Table 3. Upon scrutinizing the error rates resulting from the implementation of this collected statistical data with the developed model, it is observed that the Bayesian network method exhibits lower mean absolute percentage error rates in predicting the length and main engine power output parameters. Conversely, the power regression method demonstrates superior error rates in determining the other output parameters.

Table 3. Prediction errors of the developed model—MSEs.

Output Parameter	Mean Squared Errors (MSEs)		
	Bayesian Network	**Power Regression Group 1**	**Power Regression Group 2**
Length	7.15	10.69	10.23
Beam	0.68	0.47	1.20
Draft	0.35	0.18	0.35
Draft, Max.	0.30	0.11	0.34
Power	20,181.19	56,410.33	655,323.72
Length	7.15	10.69	10.23

To assess the efficacy of the developed model, error metrics were computed from the prediction results for both the Bayesian network and power regression. The calculated mean absolute percentage errors (MAPEs) for each output parameter—length, beam, draft, maximum draft, and power—are presented in Figure 13, while the averages of the MAPEs are depicted in Figure 14.
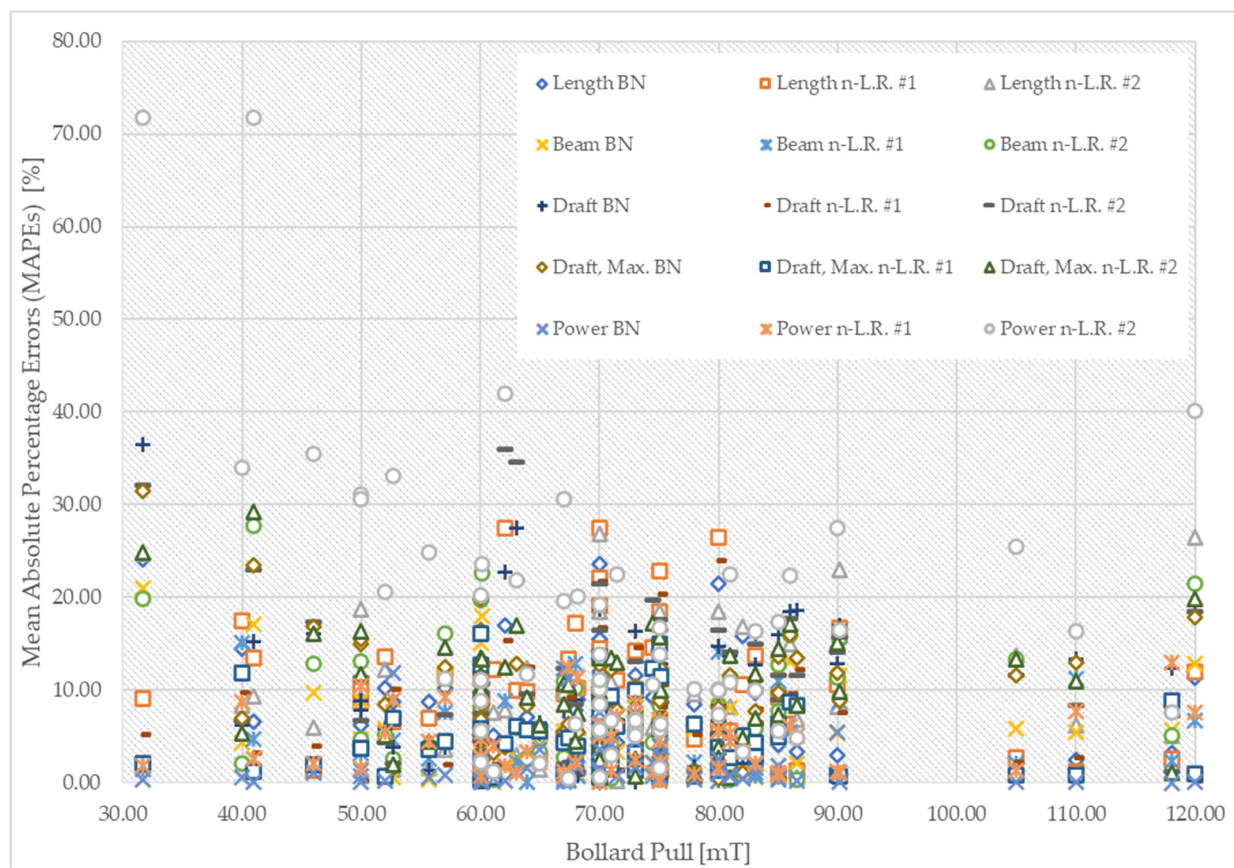
Figure 13. Error histogram of the developed model—MAPEs.
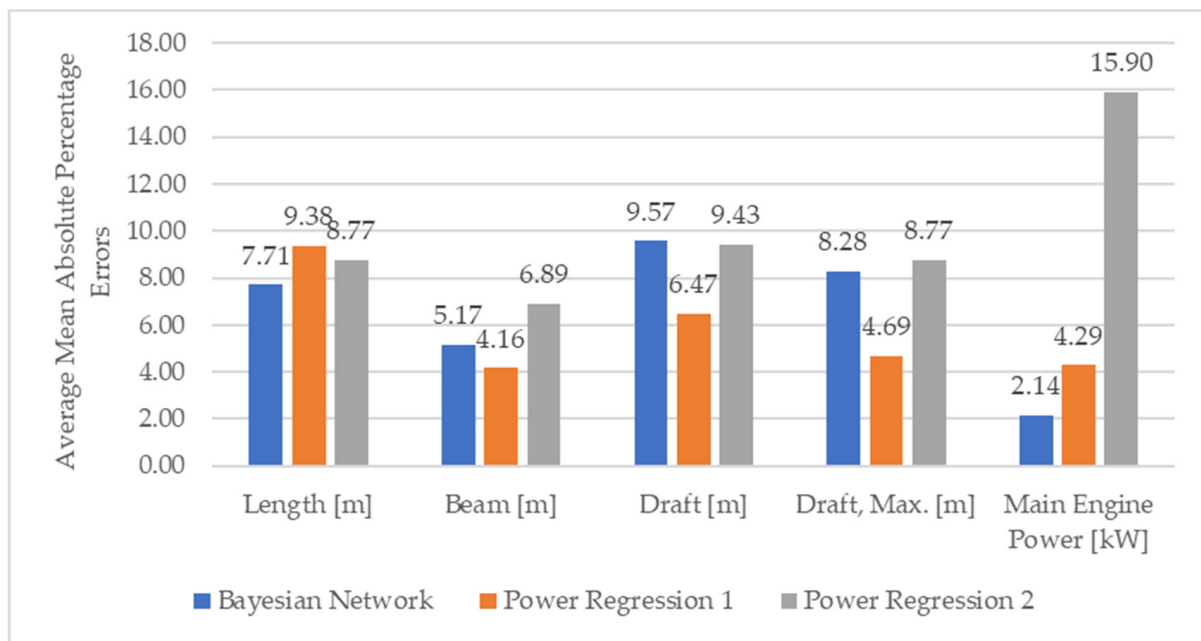
Figure 14. Error histogram of the developed model—average of MAPEs.

4. Results and Discussion

The aim of this study was to develop a Bayesian network model to predict the main particulars of a Z-Drive harbor tugboat equipped with azimuthing propulsors at the concept and preliminary design phases; the bollard pull (BP) and ship speed (V) were used as the input layer, and the length (LOA), beam (B), draft (T), draft-maximum (Tmax), and power (P) were used as the output layer with ten hidden layers.

When analyzing the performance graph of the constructed Bayesian network, as visualized in Figure 10 and depicted in Figure 11, it becomes apparent that the optimal validation performance was achieved at epoch 754, with a mean squared error (MSE) calculated as 1184.1913. Moreover, the determination coefficients (R^2) for the training, testing, and overall process of the Bayesian network were found to be 0.99977, 0.99651, and 0.99906, respectively, indicating a high level of consistency between the actual and expected values. These findings underscore the reliability of the model. In the case of the implemented power regressions group 1 and 2, the coefficients of determination were 0.4383 and 0.4626, 0.6891 and 0.2159, 0.6064 and 0.2129, 0.7527 and 0.2266, and 0.9368 and 0.2656 for LOA, B, T, Tmax, and P, respectively. Notably, although the coefficients of determination for the speed (V) parameter are relatively low, this can be attributed to the weaker correlation with speed (V) in the dataset used.

When examining the actual and predicted values of the output parameters depicted in Figure 12, it becomes apparent that the actual values closely align with the predicted values. Furthermore, upon evaluating the mean absolute percentage error (MAPE) values presented in Figure 13, it is evident that the model yields results with a lower error rate, particularly within the range of 60 to 80 metric tons.

Figure 14 illustrates the average mean absolute percentage errors (MAPEs) for the developed model. It can be observed that the MAPE values for all datasets were determined as 7.71%, 5.17%, 9.57%, 8.28%, and 2.14% for LOA, B, T, Tmax, and P in the Bayesian network, respectively. Meanwhile, for the power regressions group 1 and 2, the corresponding values were 9.38% and 8.77%, 4.16% and 6.89%, 6.47% and 9.43%, 4.69% and 8.77%, and 4.29% and 15.90% for LOA, B, T, Tmax, and P, respectively. The average of these values was calculated as 6.574% for the Bayesian network and 5.795% and 9.955% for power regressions groups 1 and 2, respectively. Notably, among the output parameters, the maximum mean error occurred at 71.81%, while the mean error for all datasets was

relatively acceptable. However, it is worth mentioning that power regression 2 exhibited a relatively high MAPE value, despite the overall error being within an acceptable range.

In the context of power regression, the coefficient of determination (R-squared) is commonly computed by software to assess the goodness of fit of the model, similar to its application in linear regression. However, given that power regression entails a non-linear relationship between variables, the calculation of R-squared entails comparing the variation accounted for by the model to the total variation present in the dataset. Meanwhile, the MSE and MAPE results demonstrate that the developed Bayesian network model, along with the implemented power regression models, presents a promising approach for predicting the main characteristics of the azimuthing-propulsor-equipped Z-Drive harbor tugs during the initial stages of ship design.

5. Conclusions

The determination of a ship's main particulars during the design phases constitutes a significant and intricate process, necessitating the evaluation of relationships among numerous parameters. Traditionally, these relationships are established through statistical methods or empirical formulas based on data from existing ships. However, conventional statistical methods and empirical formulas often depict relationships between two factors in isolation, disregarding others. Consequently, novel approaches have gained prominence in the determination of a ship's main particulars.

In this study, a model employing a Bayesian network and non-linear regression was developed to predict the primary characteristics of a Z-Drive harbor tugboat equipped with azimuthing propulsors capable of meeting the specified bollard pull and speed requirements. The case study utilized a dataset comprising main particulars from 68 distinct azimuthing-propulsor-equipped Z-Drive harbor tugboat series, derived from a collection of over 200 existing tugboats, with sister ships eliminated. The dataset was analyzed using the developed model, and the results were compared.

Upon examining the distribution of the bollard pull input relative to the other output parameters in the calculated results, it was observed that the accuracy of the results was higher within the 60 to 80 metric tons bollard pull range, while deviations were noted for the maximum and minimum bollard pull values. In conclusion, the developed model proves most effective for the early design of tugs with bollard pull capacities ranging from 60 to 80 metric tons. The results of the case study, the presented descriptive methodology, and the developed model collectively emerge as crucial tools for decision makers and ship designers in the conceptual design process for determining a ship's main dimensions.

Furthermore, this study demonstrates that machine learning methods can be applied to the design of ship types with unique parameters, such as tugboats designed for specific purposes rather than cargo-carrying ships. Neural network models, exemplified by the Bayesian network in this study, present themselves as viable alternatives to traditional statistical and empirical methods. This model proves to be a practical and valuable tool for ship designers in determining the main particulars of ships during the concept design stage by reducing revision returns possibilities in further stages of ship design like expensive Computational Fluid Dynamics (CFD) and conventional ship model tests. Therefore, the neural networks streamline this process by offering faster and more cost-effective predictions compared to traditional methods, which entail time-consuming and expensive extensive physical testing and computational simulations, thus diminishing the necessity for expensive prototyping and testing. This approach not only reduces costs at the preliminary design stage but also enhances time efficiency and mitigates risks which are associated with proposed designs.

The utilization of machine learning methodologies, particularly neural network models such as the Bayesian network, in the design of ships with unique parameters like tugboats is poised to yield numerous advantages, including enhanced accuracy, expedited design iterations, adaptability, and improved time and cost efficiency. These advancements herald an innovative shift in ship design, enabling the utilization of data-driven insights

to develop vessels that are safer, more efficient, and technologically advanced, tailored to meet contemporary design requirements.

The models proposed within this research are dynamic in nature and specifically tailored for tugboats. However, in forthcoming studies, these models have the potential to be modified for various other vessel types. Achieving this adaptation necessitates updating the inputs in accordance with the specific characteristics of the target ship type. Furthermore, future research endeavors hold the promise of enhancing model robustness through parameter optimization techniques. Future studies may delve into incorporating additional ship parameters such as displacement, tonnage, vessel hull form coefficients, towing speed, the number of propellers, propeller diameters, which are not considered here, constructing models with greater accuracy through larger datasets, or exploring alternative neural network modeling methods for different specific ship types. Furthermore, neural network models can be integrated into existing design software utilized by ship designers and naval architects, enabling predictive modeling capabilities to be seamlessly incorporated into the ship design workflow. This enables ship designers to leverage advanced analytical tools without requiring specialized expertise in machine learning.

References

1. Ma, X.; Ping, W. Design Spiral. In *Encyclopedia of Ocean Engineering*; Springer: Singapore, 2019; pp. 1–11. [CrossRef]
2. Papanikolaou, A. *Ship Design: Methodologies of Preliminary Design*; Springer: Berlin/Heidelberg, Germany, 2014; ISBN 9789401787512.
3. Majnarić, D.; Baressi Šegota, S.; Lorencin, I.; Car, Z. Prediction of Main Particulars of Container Ships Using Artificial Intelligence Algorithms. *Ocean Eng.* **2022**, *265*, 112571. [CrossRef]
4. Uyanık, T.; Yalman, Y.; Kalenderli, Ö.; Arslanoğlu, Y.; Terriche, Y.; Su, C.-L.; Guerrero, J.M. Data-Driven Approach for Estimating Power and Fuel Consumption of Ship: A Case of Container Vessel. *Mathematics* **2022**, *10*, 4167. [CrossRef]
5. Gurgen, S.; Altin, I.; Ozkok, M. Prediction of Main Particulars of a Chemical Tanker at Preliminary Ship Design Using Artificial Neural Network. *Ships Offshore Struct.* **2018**, *13*, 459–465. [CrossRef]
6. Cepowski, T.; Chorab, P. Determination of Design Formulas for Container Ships at the Preliminary Design Stage Using Artificial Neural Network and Multiple Nonlinear Regression. *Ocean Eng.* **2021**, *238*, 109727. [CrossRef]
7. Huang, L.; Pena, B.; Liu, Y.; Anderlini, E. Machine Learning in Sustainable Ship Design and Operation: A Review. *Ocean Eng.* **2022**, *266*, 112907. [CrossRef]
8. Majnari, D.; Šegota, S.B.; An, N. Improvement of Machine Learning-Based Modelling of Container Ship's Main Particulars with Synthetic Data. *J. Mar. Sci. Eng.* **2024**, *12*, 273. [CrossRef]
9. Mittendorf, M.; Nielsen, U.D.; Bingham, H.B. Data-Driven Prediction of Added-Wave Resistance on Ships in Oblique Waves—A Comparison between Tree-Based Ensemble Methods and Artificial Neural Networks. *Appl. Ocean Res.* **2022**, *118*, 102964. [CrossRef]
10. Yang, K.; Duan, W.; Huang, L.; Zhang, P.; Ma, S. A Prediction Method for Ship Added Resistance Based on Symbiosis of Data-Driven and Physics-Based Models. *Ocean Eng.* **2022**, *260*, 112012. [CrossRef]
11. Ao, Y.; Li, Y.; Gong, J.; Li, S. An Artificial Intelligence-Aided Design (AIAD) of Ship Hull Structures. *J. Ocean Eng. Sci.* **2023**, *8*, 15–32. [CrossRef]
12. Nazemian, A.; Boulougouris, E. Utilizing Machine Learning Tools for Calm Water Resistance Prediction and Design Optimization of a Fast Catamaran Ferry. *J. Mar. Sci. Eng.* **2024**, *12*, 216. [CrossRef]

13. La Ferlita, A.; Qi, Y.; Di Nardo, E.; Moenster, K.; Schellin, T.E.; EL Moctar, O.; Rasewsky, C.; Ciaramella, A. Power Prediction of a 15,000 TEU Containership: Deep-Learning Algorithm Compared to a Physical Model. *J. Mar. Sci. Eng.* **2023**, *11*, 1854. [CrossRef]
14. Kim, Y.C.; Kim, K.S.; Yeon, S.; Lee, Y.Y.; Kim, G.D.; Kim, M. Power Prediction Method for Ships Using Data Regression Models. *J. Mar. Sci. Eng.* **2023**, *11*, 1961. [CrossRef]
15. Lazakis, I.; Raptodimos, Y.; Varelas, T. Predicting Ship Machinery System Condition through Analytical Reliability Tools and Artificial Neural Networks. *Ocean Eng.* **2018**, *152*, 404–415. [CrossRef]
16. Uyanık, T.; Karatuğ, Ç.; Arslanoğlu, Y. Machine Learning Approach to Ship Fuel Consumption: A Case of Container Vessel. *Transp. Res. Part D Transp. Environ.* **2020**, *84*, 102389. [CrossRef]
17. Cepowski, T.; Chorab, P. The Use of Artificial Neural Networks to Determine the Engine Power and Fuel Consumption of Modern Bulk Carriers, Tankers and Container Ships. *Energies* **2021**, *14*, 4827. [CrossRef]
18. Lee, D.; Lim, C.; Oh, S.J.; Kim, M.; Park, J.S.; Shin, S.C. Predictive Model for Hydrostatic Curves of Chine-Type Small Ships Based on Deep Learning. *J. Mar. Sci. Eng.* **2024**, *12*, 180. [CrossRef]
19. Bakhtiari, M.; Ghassemi, H. CFD Data Based Neural Network Functions for Predicting Hydrodynamic Performance of a Low-Pitch Marine Cycloidal Propeller. *Appl. Ocean Res.* **2020**, *94*, 101981. [CrossRef]
20. Silva, K.M.; Maki, K.J. Data-Driven System Identification of 6-DoF Ship Motion in Waves with Neural Networks. *Appl. Ocean Res.* **2022**, *125*, 103222. [CrossRef]
21. Guo, X.; Zhang, X.; Tian, X.; Li, X.; Lu, W. Predicting Heave and Surge Motions of a Semi-Submersible with Neural Networks. *Appl. Ocean Res.* **2021**, *112*, 102708. [CrossRef]
22. Karatuğ, Ç.; Arslanoğlu, Y.; Guedes Soares, C. Design of a Decision Support System to Achieve Condition-Based Maintenance in Ship Machinery Systems. *Ocean Eng.* **2023**, *281*, 114611. [CrossRef]
23. Karatuğ, Ç.; Arslanoğlu, Y. Importance of Early Fault Diagnosis for Marine Diesel Engines: A Case Study on Efficiency Management and Environment. *Ships Offshore Struct.* **2022**, *17*, 472–480. [CrossRef]
24. Wang, W.; Cui, Z.; Chen, R.; Wang, Y.; Zhao, X. Regression Analysis of Clustered Panel Count Data with Additive Mean Models. *Stat. Pap.* **2023**, 1–22. [CrossRef]
25. Zhu, L.; Tong, X.; Zhao, H.; Sun, J.; Srivastava, D.K.; Leisenring, W.; Robison, L.L. Statistical Analysis of Mixed Recurrent Event Data with Application to Cancer Survivor Study. *Stat. Med.* **2013**, *32*, 1954–1963. [CrossRef] [PubMed]
26. Wang, Z.; Wu, X.; Wu, Y. A Spatiotemporal XGBoost Model for PM2.5 Concentration Prediction and Its Application in Shanghai. *Heliyon* **2023**, *9*, e22569. [CrossRef] [PubMed]
27. Neapolitan, R.E. *Learning Bayesian Networks*; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2007; p. 1. [CrossRef]
28. Han, X.; Leira, B.J.; Sævik, S. Vessel Hydrodynamic Model Tuning by Discrete Bayesian Updating Using Simulated Onboard Sensor Data. *Ocean Eng.* **2021**, *220*, 108407. [CrossRef]
29. Xue, Y.; Liu, Y.; Ji, C.; Xue, G. Hydrodynamic Parameter Identification for Ship Manoeuvring Mathematical Models Using a Bayesian Approach. *Ocean Eng.* **2020**, *195*, 106612. [CrossRef]
30. Moreira, L.; Guedes Soares, C. Simulating Ship Manoeuvrability with Artificial Neural Networks Trained by a Short Noisy Data Set. *J. Mar. Sci. Eng.* **2022**, *11*, 15. [CrossRef]
31. Guo, C.; Wu, W. Fuzzy Dynamic Bayesian Network Based on a Discrete Aggregation Method for Risk Assessment of Marine Nuclear Power Platform Hinge Joints Accidents. *Appl. Ocean Res.* **2023**, *138*, 103674. [CrossRef]
32. Sokukcu, M.; Sakar, C. Risk Analysis of Collision Accidents during Underway STS Berthing Maneuver through Integrating Fault Tree Analysis (FTA) into Bayesian Network (BN). *Appl. Ocean Res.* **2022**, *126*, 103290. [CrossRef]
33. Aydin, M.; Akyuz, E.; Turan, O.; Arslan, O. Validation of Risk Analysis for Ship Collision in Narrow Waters by Using Fuzzy Bayesian Networks Approach. *Ocean Eng.* **2021**, *231*, 108973. [CrossRef]
34. Elidolu, G.; Sezer, S.I.; Akyuz, E.; Arslan, O.; Arslanoglu, Y. Operational Risk Assessment of Ballasting and De-Ballasting on-Board Tanker Ship under FMECA Extended Evidential Reasoning (ER) and Rule-Based Bayesian Network (RBN) Approach. *Reliab. Eng. Syst. Saf.* **2023**, *231*, 108975. [CrossRef]
35. Cao, Y.; Wang, X.; Wang, Y.; Fan, S.; Wang, H.; Yang, Z.; Liu, Z.; Wang, J.; Shi, R. Analysis of Factors Affecting the Severity of Marine Accidents Using a Data-Driven Bayesian Network. *Ocean Eng.* **2023**, *269*, 113563. [CrossRef]
36. Li, H.; Ren, X.; Yang, Z. Data-Driven Bayesian Network for Risk Analysis of Global Maritime Accidents. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108938. [CrossRef]
37. Sakar, C.; Toz, A.C.; Buber, M.; Koseoglu, B. Risk Analysis of Grounding Accidents by Mapping a Fault Tree into a Bayesian Network. *Appl. Ocean Res.* **2021**, *113*, 102764. [CrossRef]
38. Abaei, M.M.; Abbassi, R.; Garaniya, V.; Chai, S.; Khan, F. Reliability Assessment of Marine Floating Structures Using Bayesian Network. *Appl. Ocean Res.* **2018**, *76*, 51–60. [CrossRef]
39. Gao, C.; Guo, Y.; Zhong, M.; Liang, X.; Wang, H.; Yi, H. Reliability Analysis Based on Dynamic Bayesian Networks: A Case Study of an Unmanned Surface Vessel. *Ocean Eng.* **2021**, *240*, 109970. [CrossRef]
40. Zhu, J.; Zhou, D.; Yang, W.; Qian, Y.; Mao, Y.; Lu, X. Investigation on the Potential of Using Carbon-Free Ammonia in Large Two-Stroke Marine Engines by Dual-Fuel Combustion Strategy. *Energy* **2023**, *263*, 125748. [CrossRef]
41. Gkerekos, C.; Lazakis, I.; Theotokatos, G. Machine Learning Models for Predicting Ship Main Engine Fuel Oil Consumption: A Comparative Study. *Ocean Eng.* **2019**, *188*, 106282. [CrossRef]

42. Wang, J.; Chen, Z.; Song, Y.; Liu, Y.; He, J.; Ma, S. Data-Driven Dynamic Bayesian Network Model for Safety Resilience Evaluation of Prefabricated Building Construction. *Buildings* **2024**, *14*, 570. [CrossRef]

43. McLachlan, S.; Dube, K.; Hitman, G.A.; Fenton, N.E.; Kyrimi, E. Bayesian Networks in Healthcare: Distribution by Medical Condition. *Artif. Intell. Med.* **2020**, *107*, 101912. [CrossRef]

44. Lucas, P.J.F.; Van Der Gaag, L.C.; Abu-Hanna, A. Bayesian Networks in Biomedicine and Health-Care. *Artif. Intell. Med.* **2004**, *30*, 201–214. [CrossRef] [PubMed]

45. Butz, C.J.; Oliveira, J.S.; dos Santos, A.E.; Madsen, A.L. An Empirical Study of Bayesian Network Inference with Simple Propagation. *Int. J. Approx. Reason.* **2018**, *92*, 198–211. [CrossRef]

46. Clausen, H.B.; Lützen, M.; Friis-Hansen, A.; Bjørneboe, N. Bayesian and Neural Networks for Preliminary Ship Design. *Mar. Technol. SNAME News* **2001**, *38*, 268–277. [CrossRef]

47. Huang, H.H.; He, Q. Nonlinear Regression Analysis. In *International Encyclopedia of Education*, 4th ed.; Elsevier Ltd.: Amsterdam, The Netherlands, 2022; pp. 558–567, ISBN 9780128186299.

48. Xie, X.; Sun, B.; Li, X.; Olsson, T.; Maleki, N.; Ahlgren, F. Fuel Consumption Prediction Models Based on Machine Learning and Mathematical Methods. *J. Mar. Sci. Eng.* **2023**, *11*, 738. [CrossRef]

49. Karaçay, Ö.E.; Özsoysal, O.A. Techno-Economic Investigation of Alternative Propulsion Systems for Tugboats. *Energy Convers. Manag. X* **2021**, *12*, 100140. [CrossRef]

50. Çakıroğlu, G.; Şener, B.; Balın, A. Applying a Fuzzy-Ahp for the Selection of a Suitable Tugboat Based on Propulsion System Type. *Brodogradnja* **2018**, *69*, 1–13. [CrossRef]

51. McIvor, J. Tug and Workboat Design Guide. *BC Shipp. News* **2012**, *2*, 26–33.

52. Association, E.T. *Guidelines for Safe Harbour Towage Operations*, 1st ed.; European Tugowners Association: Bruxelles, Belgium, 2015.

53. Nitonye, S.; Adumene, S.; Howells, U.U. Numerical Design and Performance Analysis of a Tug Boat Propulsion System. *J. Power Energy Eng.* **2017**, *5*, 80–98. [CrossRef]

54. Koznowski, W.; Łebkowski, A. Analysis of Hull Shape Impact on Energy Consumption in an Electric Port Tugboat. *Energies* **2022**, *15*, 339. [CrossRef]

55. Carlton, J.S. *Marine Propellers and Propulsion*, 2nd ed.; Buttersworth-Heineman Ltd., Ed.; Elsevier Ltd.: Oxford, UK, 2007; ISBN 978-07506-8150-6.

56. Vrijdag, A.; Martelli, M. Parameter Identification of a Model Scale Ship Drive Train. *J. Mar. Sci. Eng.* **2021**, *9*, 268. [CrossRef]