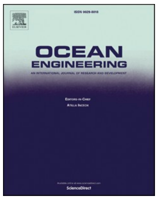
Ocean Engineering 292 (2024) 116499

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Research paper

# Regression analysis for container ships in the early design stage

B. Rinauro [a,*], E. Begovic [a], F. Mauro [b,c], G. Rosano [a]

[a] University of Naples Federico II, Department of Industrial Engineering, Via Claudio 21, 80125, Naples, Italy
[b] Department of Maritime and Transport Technology, Faculty of Mechanical. Maritime and Materials Engineering, Delft University of Technology, Leegwaterstraat 17, 2628, CA, Delft, the Netherlands
[c] Sharjah Maritime Academy, 180018, Khorfakkan, Sharjah, United Arab Emirates

ARTICLE INFO

*Keywords:*
Container ships
Database statistics
Simple regressions
Multivariable regressions
Forest trees
Ship design

ABSTRACT

The seaway trade market has expanded in the last years and container ship dimensions are constantly increasing for higher cargo capacity. In the early design stage, main dimensions are usually determined based on an existing ship database from which regression formulas are derived. In the present paper, a database of 260 non-sister container ships built from 1979 to 2022, representing 20% of the world fleet, has been considered to derive and compare different types of regressions. Simple regressions have been developed and compared with equivalent formulations available in literature, proving better approximations of the trends. The study has been further extended by multivariable regressions and forest tree algorithms, which allow the use of more than one independent variable and provide a better fitting compared to simple regressions. Forest tree regressions return the highest values of fitting coefficients, but the technique is not of easy application due to the absence of mathematical expressions. The main contribution is the updated set of simple and multivariable regression formulas which have a higher goodness of fit than previous works and can be easily employed by designers in the early design stage and in multi-attribute design procedures.

## 1. Introduction

Within today's shipping, container ships are the trendsetters which have revolutionized the transport of goods. Even though the first "container ship" dates back to 1956, the massive "containerization" started only in January 1968 with the introduction of ISO 668 (1968), which defined terminology, dimensions and ratings for freight containers. In the last two decades, the shipping market witnessed the trend of a continuous increase of goods transported by containers, leading to an enlargement of the container ship fleet and the entrance of the Very-Large Container Ships, VLCS, and the Ultra-Large Container Ships, ULCS, with up to 23000 TEU. In parallel, two aspects related to container ships are drawing the attention of experts in the transport and logistic chain and of naval architects: ultimate limit ship dimensions and the yearly number of containers lost at sea.

Economics and logistics experts (Malchow (2017), Saxon and Stone (2017); Garrido et al. (2020)) analyzed the trends of container ships growth based on economies of scale, port infrastructure, demand, and environmental tendencies, to predict the ship size limits. According to Malchow (2017), a 30000 TEU container ship with approximately 20 m draught, should be the ultimate limit because of the depth constraints in the Malacca Strait and the Suez Canal. Saxon and Stone (2017) envisaged even 50000 TEU container ships in the next 50 years. Garrido et al. (2020) analyzed the design restrictions and stability regulations, the geographic and port restrictions, the economies of scale of container shipping lines, the $CO_2$ emissions of vessels, and the world economy, demand, and global trends. The authors concluded that all trends indicate 30000 TEU container ships on the market by 2030 and reported that according to the capital cost, the optimal ship size has 418 m length, 69 m breadth, 35 m moulded depth, and 17 m draught. They calculated the metacentric height and stated that its reasonable value should be between 4% and 5% of the breadth to avoid stability problems. Finally, the authors compared the "ideal" dimensions predicted by different authors, namely Park and Suh (2019), Kristensen (2012) and Korea Maritime Institute (2012). Under the assumption of a 30000 TEU ship, the predicted ship length, breadth and draught are then equal to 453 m–72 m – 17.3 m (Park and Suh, 2019); 483 m–71.5 m – 18.7 m (Kristensen, 2012); and 517 m–65 m – 19.4 m (Korea Maritime Institute, 2012). As reported, the differences in draught and breadth are relatively small as they have upper limits dictated by water depth in ports and canals and by the arm-length of port cranes. The decreased "optimal" length clearly indicates the change in design trends of the last decade.

[*] Corresponding author.
*E-mail address:* barbara.rinauro@unina.it (B. Rinauro).

https://doi.org/10.1016/j.oceaneng.2023.116499
Received 25 August 2023; Received in revised form 1 December 2023; Accepted 1 December 2023
Available online 11 December 2023

Another relevant issue is the number of containers lost at sea each year. In 2011 the World Shipping Council (WSC) started a survey among its members (covering more than 90% of the global container ship capacity) to accurately estimate the number of containers lost at sea each year. Reviewing the results of the total surveyed period 2008–2022 (WSC, 2023), the WSC estimated that there was, on average, a total of 1566 containers lost at sea each year. Average losses for the last three years were 2301 containers per year (2020–2022). Up-to-date results indicated the parametric roll as one of the main reasons for container losses. Actions to preclude further accidents for the existing ships are the training of mariners to recognize and prevent the parametric roll (Galeazzi et al., 2013) and the application of operational guidance (Begovic et al., 2023) which clearly identifies speeds and headings where the ship may be vulnerable. For the new vessels, operators should consider from the design stage the vulnerability to stability failure modes, such as parametric roll (France et al., 2003), excessive acceleration (IMO SLF 54/INF.6, 2011) or pure loss of stability.

It is evident that, for such a competitive and demanding market, ship design must assure maximum performance in all these aspects using up-to-date knowledge, software and technology. A reliable preliminary design assessing the main ship's characteristics is essential. As underlined by Papanikolaou (2014), the values of these characteristics mainly depend on the four main basic demands: cargo capacity, top speed, range/autonomy and class, but dominated by constraints such as the maximum value of breadth or draught to pass through the canals or enter in ports. The "traditional" approach to determe the main ship characteristics employs regression formulas obtained from a database of similar ships. When following the design spiral, few iterations are needed to retrieve the "optimal" main particulars.

Kristensen (2012) published a set of linear and nonlinear regressions to estimate main dimensions, deadweight and various design parameters as a function of the number of TEU. For the development of these regressions, container ships built between 1990 and 2010 were classified into three groups (Small, Panamax and Post Panamax).

In the multi-attribute and multi-objective ship design (Zanic et al. (1992), Trincas et al. (1994), Grubisic and Begovic (2001, 2011), Mauro et al. (2019), Ljulj et al. (2020), regression formulas or "low-fidelity" codes are the basis of different modules which calculate ship attributes (power, deadweight, structural weight, total cost, etc.) based only on ship main parameters. It is evident that the accuracy of the obtained results will be strongly affected by the accuracy of the input regression formula.

In the last two decades, the application of artificial intelligence (AI) techniques, such as genetic algorithms, neural networks, and machine learning, is increasing in all phases of ship design. AI techniques are in continuous development, but the main idea is to find the optimal design starting from a few macroscopic parameters provided by the owner (such as cargo capacity, maximum speed, and range/autonomy). All other parameters are then estimated through a regression analysis of an existing ships database and/or applying some AI techniques to perform the optimization. One example is the automatic hull form generation, as shown by Islam et al. (2001). The authors performed a three-steps procedure starting from 104 ships' half-breadths, then used neural networks to adjust parameter values and finally used the genetic algorithm to design the body plan. Their work is one of the first examples where the advantages of both techniques have been combined: neural networks are used to identify the data pattern and to predict the required parameters and genetic algorithms are used for search-based optimization problems (i.e. maximization or minimization of the objective function), which are difficult and time-intensive to solve by other general algorithms. Clausen et al. (2001) acquired a database of 87000 ships and applied regression analysis, Bayesian and neural networks to encode the relations between ship main characteristics and loading capacity for different ship types. They concluded that neural networks are easier to implement and yield smaller estimate errors than Bayesian networks. This work remains, up to now, the one with the most extensive database used for the neural network.

The applications of AI as a predictive tool for seakeeping (Romero-Tello et al., 2022), fuel consumption (Uyanık et al., 2020), and corrosion damage detection (Yao et al., 2019) are numerous, showing their great potential when the model is developed from a big data sample, as for example from the data monitored on board during voyages.

Recent works on the application of AI in preliminary ship design, such as Ekinci et al. (2011), Gurgen et al. (2018), Cepowski and Chorab (2021), and Majnaric et al. (2022), used artificial neural networks, machine learning and multiple regression analysis to develop design formulas considering databases of a few hundred ships. Concerning the size of the database, the efficiency of AI techniques can be discussed because, as the "rule of thumb", the minimum sample should be at least 30 times the number of weights (Alwosheel et al., 2018).

The present work reports the statistical and regression analyses of the database obtained by the Hyundai catalogue (HHI shipbuilding group performance record, 2022) of container ships built from 1979 to 2022. The objective of this work is to present an overview of the possible methodologies implementable for the determination of main design parameters for container ships. The application of AI technique (forest tree) has been explored to obtain the estimations with the highest coefficient of determination $R^2$. Even though it is a black-box method and does not return any mathematical expressions, it provides better results than simple and multivariable regression methodologies. Therefore, the order of multivariable regressions models has been further varied to improve $R^2$ values.

The different regression methodologies are described in Section 2. In Section 3, a complete database of about 1000 ships is presented, but for all regression analyses a reduced database of 260 vessels without sisterships has been considered. In Section 4, simple regression formulas, of the same form as published in Papanikolaou (2014), have been developed to investigate the trend of new container ships. In Section 5, the simple regression formulas obtained from the present database have been compared with the ones developed by Cepowski and Chorab (2021) and the ones presented in Papanikolaou (2014). In Section 6, a new nonlinear multiple variable regression is performed, and the accuracy of the predictive models is discussed according to their mathematical formulation and different fit coefficients. In Section 7, a forest tree algorithm has been used to identify non-correlated parameters for the multiple regression analysis. In Section 8, an example of the design parameters prediction with all the regressions methodologies is provided for a container ship of 20000 TEU at a speed of 23 knots. Discussions and Conclusion are presented in Section 9.

## 2. Regression analysis methodologies

The regression formulas can be used to predict the value of one parameter based on the value of one or more variables and in this section the applied methodologies are described.

At first, simple regressions, have been developed in the forms of Equation (1) (polynomial, power or logarithmic) where one parameter $y$ is function of only one variable $x$ with the coefficients $a$ and $b$:

$$y = ax + b \text{ or } y = ax^b \text{ or } y = a\,ln\,(x) + b \tag{1}$$

To increase the accuracy of the predicted values, multivariable regressions have been developed, where one parameter can be estimated based on two or more variables. The general model for multiple linear regression is given by the following matrix formulation:

$$\boldsymbol{Y} = \boldsymbol{c}\boldsymbol{X} + \boldsymbol{d} \tag{2}$$

where $\boldsymbol{Y}$ is the matrix of measured values, $\boldsymbol{X}$ is the matrix of independent variables, $\boldsymbol{c}$ is the matrix of coefficient and $\boldsymbol{d}$ is the matrix of errors.

Using the matrix formulation, the unknown of the problem is the matrix $\boldsymbol{c}$, obtained as follows:

$$\boldsymbol{c} = (\boldsymbol{X}^{'}\boldsymbol{X})^{-1} * \boldsymbol{X}^{'}\boldsymbol{Y} \tag{3}$$

where $\mathbf{X}^{'}$ is the transpose of matrix $\mathbf{X}$ and $(\mathbf{X}^{'}\mathbf{X})^{-1}$ is the inverse of $(\mathbf{X}^{'}\mathbf{X})$.

Finally, forest tree algorithms have been developed with different set of input variables to explore another technique and compare the various regression methods.

Forest tree is one of the most popular and commonly used algorithms by data scientists. Forest tree is a Supervised Machine Learning Algorithm widely used in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. Random forest is a versatile machine learning algorithm that leverages an ensemble of multiple decision trees to generate predictions or classifications. The random forest algorithm delivers a consolidated and more accurate result by combining the outputs of these trees. Its widespread popularity stems from its user-friendly nature and adaptability, which enables the effective tackling of classification and regression problems. The algorithm's strength lies in its ability to handle complex datasets and mitigate overfitting, making it a valuable tool for various predictive tasks in machine learning. One of the most relevant features of the Random Forest Algorithm is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It performs better for classification and regression tasks.

For all the different regression methodologies the following fit coefficients have been determined and compared to assess the accuracy of the regressions: coefficient of determination ($R^2$), Pearson coefficient, MAPE (Mean Absolute Percentage Error), RMSR (Relative Root Mean Square Error), and RRMSE (Relative Root Mean Square Error). In particular, for multivariable regressions the additional values of $R^2_{adj}$, SE (Standard Error), t-stud, and p-value have been evaluated.

The formulations of the fit coefficients are defined as follows:

$$R^2 = 1 - \frac{SS_E}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{4}$$

$$R^2_{adj} = 1 - \left(1 - R^2\right)\frac{n-1}{n - n_p - 1} \tag{5}$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|y_i - y_i^*\right|}{n} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{n}} \tag{7}$$

$$RRMSE = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{\sum_{i=1}^{n}\left(y_i^*\right)^2}} \tag{8}$$

$$Pearson = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(y_i - \overline{y^*})}{\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)^2}} \tag{9}$$

where $y_i$ are the n observations, $y_i$* the predicted values, $\overline{y}$ is the mean value of the observations, $\overline{y^*}$ is the mean value of the predicted variable and $n_p$ is the number of parameters used in the regression model.

The use of one or the other regression, depends on the available input parameters: in case only the number of TEU is available as an input, simple regression or forest tree have to be used; if also speed is known, multiple regressions or forest tree function of speed and TEU have to be adopted. Moreover, if the main dimensions are also available the corresponding multiple regressions or forest tree can be employed. In any case, being the forest tree algorithm a black-box methodology without giving the equation coefficients, the reproducibility of the results may be achieved only through trials of simple and multiple regressions.

## 3. Ship database and statistics

The database used for this study has been generated from the "Hyundai Heavy Industries Shipbuilding Group" catalogue (HHI shipbuilding group performance record, 2022), which includes 971 ships, 260 of which are not sisterships, and represents about 20% of container ships of the world fleet. The ships, built from 1979 to 2022, have been classified by main characteristics: length, breadth, moulded depth, draught, maximum speed, TEU, engine power, and delivery date, as summarized in Table 1. Only 13 ships are twin screws, all the others are single screw.

This database has been chosen since Hyundai can be considered the world's largest builder of container ships and no other trustful data has been available for this study. The range of data is very wide ensuring a considerable variability as shown in Table 1. Furthermore, the database ensures the use of the same definitions for all dimensions and variables among all the vessels.

The ships of both complete and non-sistership database have been divided into classes as shown in Fig. 1: Small Feeder (up to 1000 TEU); Feeder (1001–2000 TEU); Feedermax (2001–3000 TEU); Panamax (3001–5100 TEU); Post-Panamax (5101–10000 TEU); New Panamax (10000-14500 TEU); Ultra Large Container Vessel (14501 TEU and higher). This range division follows Kristensen (2012) but each grouped database would have been too small to perform the regression analysis. Most of the ships fall in the classes of Panamax and Post-Panamax and only about 10% of ships (102 out of the complete database) have less than 2000 TEU. Fig. 1 also presents the number of ships delivered each year, and, although the first ship was built in 1978, only 33 non-sisterships (and 111 out of the complete database) were built before 2000, 227 non-sisterships were built from 2000 to 2022 and 55 ships of them are from 2015 to 2022.

The ships have been grouped based on the main dimensions. Based on ship length the samples have been sorted in intervals of 10 m, for both the total number of ships and for the non-sisterships, as shown in Fig. 2. It can be noticed that almost one-third of the complete database is composed of ships having length between 275 and 295 m (223 ships) and 345–355 m (95 ships).

From now on, the statistics and all data analysis will consider only the 260 non-sisterships. The container ships' main dimensions are not continuous variables due to the container dimensions, therefore the interval for the ship breadth and depth has been chosen as 2.6 m (2.54 m for container breadth/height and 0.06 m for the spacing). As shown in Fig. 3, most of the ships have breadth between 29.7 and 32.3 m, and

**Table 1**
Container ship database characteristics.

| Main characteristics | | SI Unit | Minimum | Maximum |
|---|---|---|---|---|
| **Length** | **L** | m | 135 | 400.0 |
| **Breadth** | **B** | m | 22.0 | 61.5 |
| **Moulded depth** | **D** | m | 11.2 | 33.5 |
| **Draught** | **T** | m | 7.4 | 17.0 |
| **Maximum speed** | **V** | kn | 16.0 | 27.3 |
| **Froude number** | **Fn** | – | 0.172 | 0.269 |
| **Length over breadth** | **L/B** | – | 5.48 | 9.13 |
| **Breadth over draught** | **B/T** | – | 2.39 | 4.04 |
| **Breadth over depth** | **B/D** | – | 1.38 | 2.40 |
| **Draught over depth** | **T/D** | – | 0.42 | 0.73 |
| **LBD** | **LBD** | $m^3$ | 34322 | 821215 |
| **Twenty-foot Equivalent Unit** | **TEU** | – | 1000 | 23992 |
| **Engine power** | $\mathbf{P_{ENG}}$ | kW | 6770 | 83991 |
| **Delivery date** | **Year** | | 1979 | 2022 |

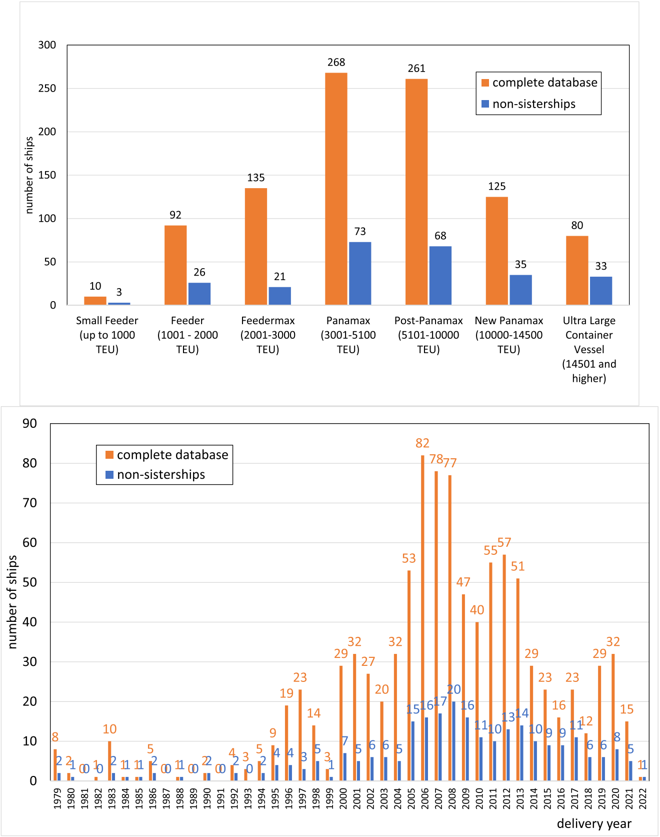


**Fig. 1.** Database statistics by class dimensions and delivery date.

draught between 11 and 13 m, which are the maximum allowable dimensions to cross the Panama Canal. The statistics on nondimensional values reported in Figs. 4 and 5 highlight that: a considerable part of ships (138 out of 260) has L/B in a range of 6.5–7.5; 73 ships have B/T between 3.2 and 3.4; and 183 have T/D between 0.48 and 0.58. The most frequent depth values go from 21 to 25 m as shown in Fig. 5.

The ship length and number of TEU have been reported as a function of the delivery date and are shown in Fig. 6, while the relation between ship maximum speed, number of TEU and years is shown in Fig. 7. The upper limit of the length and the number of TEU can be easily described by linear and exponential trendlines, respectively. On the other hand, the speed is not correlated with the number of TEU, and it is not relatable to a known simple distribution. Before the 21st century, ships reached a maximum length value of 300 m and a capacity of 7500 TEU. Ships of 350 m and 10000 TEU started to be built around 2008 and ships of 400 m and 20000 TEU appeared in 2015. The fluctuating tendency of ship speed and its decrease during the last years, despite the increase in the number of TEU and dimensions, are probably related to the limit of marine engines of 80 MW and the optimization of the route for the greenhouse gas emissions and for scheduled travelling days. Nonetheless the number of TEU, the speed is in a range between 21 and 26 kn.

## 4. Simple regression analysis

In this section, simple regressions have been developed to estimate the different parameters. In the early design stage of a container vessel, the main requested parameters are the number of TEU and/or DWT and the speed. Therefore, these were chosen as independent variables for the regression analysis. Different types of simple regressions (power, logarithmic and polynomial), as reported in Appendix A, have been performed and their coefficient of determination ($R^2$) have been compared in Appendix C, Table C1. Moreover, for each ship dimension as a function of TEU the fit coefficients MAPE, RMSE, RRMSE and Pearson have been determined to identify the best type of regression and reported in Table C2.

The whole set of best regressions for ship main characteristics and the related formulas are reported in Equations (10)–(31) and have been presented in Figs. 8–18. Particular attention should be given to the regressions of the engine power, $P_{ENG}$, when 40 MW are exceeded: as can be seen in Fig. 16, there is a huge scattering of data after this value and Equations (20) and (21) are valid only up to this power limit (a more detailed explanation is given in the following).

$$B = 0.334 \bullet L^{0.845} \; R^2 = 0.803 \tag{10}$$

$$\frac{B}{T} = -0.19495 \bullet \frac{L}{B} + 4.565 \; R^2 = 0.202 \tag{11}$$

$$D = 0.1653 \bullet L^{0.8747} \; R^2 = 0.8762 \tag{12}$$

$$\frac{B}{D} = -0.138 \bullet \frac{L}{B} + 2.71 \; R^2 = 0.4457 \tag{13}$$

$$LBT = 111.16 \; \bullet TEU^{0.8139} \;\; R^2 = 0.9826 \tag{14}$$

$$LBD = 90.303 \; \bullet TEU^{0.9074} \; R^2 = 0.9699 \tag{15}$$

$$L = 84.5 \; \bullet ln(TEU) - 450 \; R^2 = 0.948 \tag{16}$$

$$B = 2.81 \; \bullet TEU^{0.301} \; R^2 = 0.939 \tag{17}$$

$$D = 1.82 \; \bullet TEU^{0.2897} \; R^2 = 0.8893 \tag{18}$$

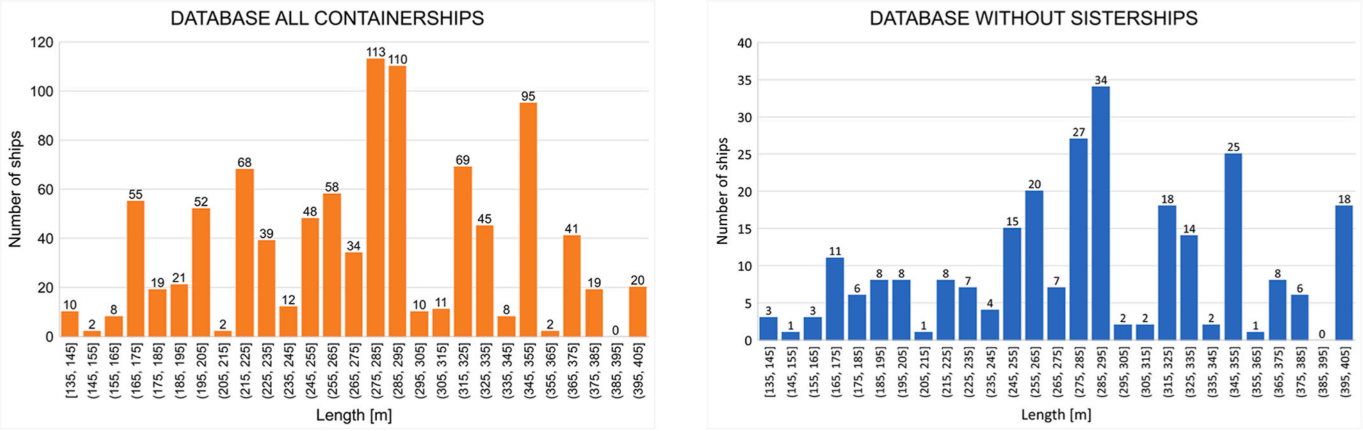


**Fig. 2.** Length of container ships in database.

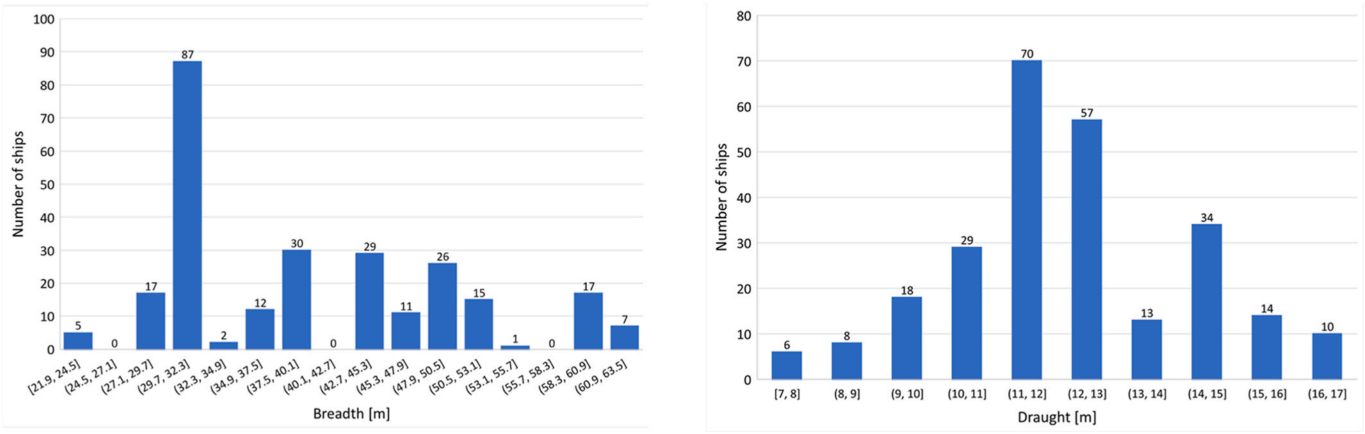


**Fig. 3.** Breadth and draught of container ships in database.

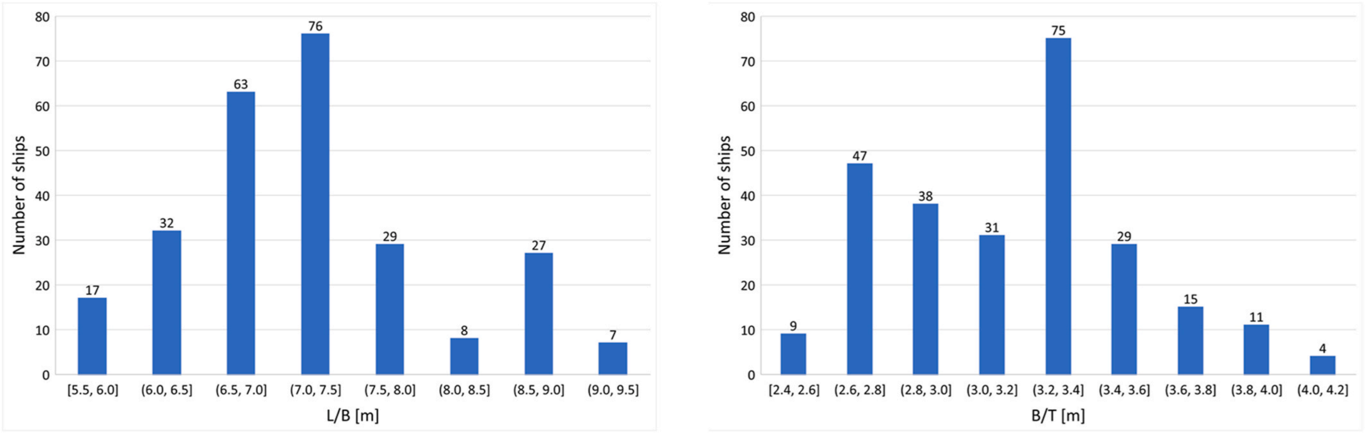


**Fig. 4.** Nondimensional ratios of container ships in database.

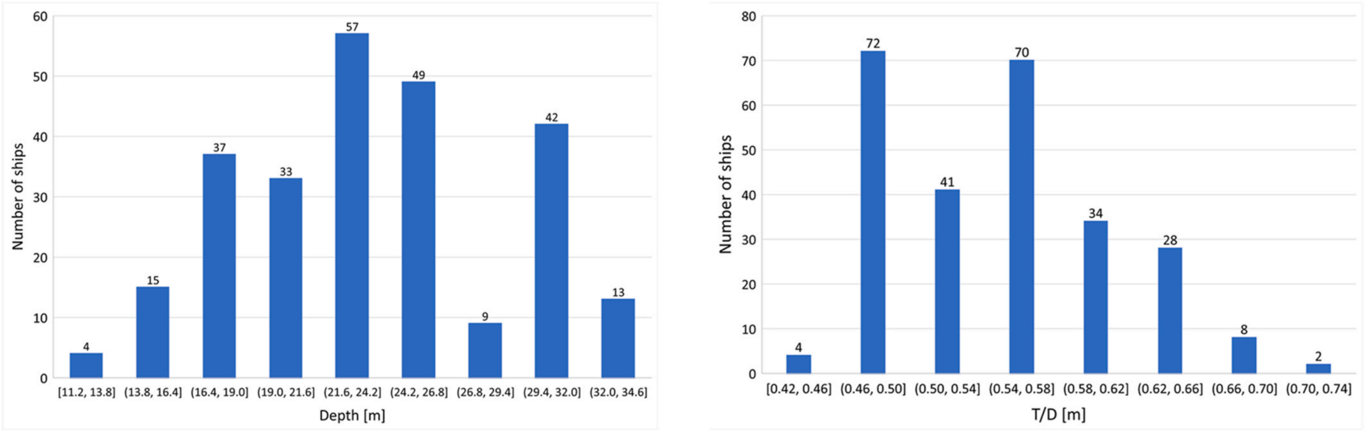


**Fig. 5.** Depth and T/D ratio of container ships in database.

$$T = 2.2404 \bullet TEU^{0.1961} \; R^2 = 0.8477 \tag{19}$$

$$LBT = 1.6354 \bullet DWT + 10387 \; R^2 = 0.981 \tag{20}$$

$$LBD = 3.355 \bullet DWT - 1703 \; R^2 = 0.972 \tag{21}$$

$$L = 92.7 \bullet ln(DWT) - 754 \; R^2 = 0.9460 \tag{22}$$

$$D = 0.6353 \bullet DWT^{0.319} \; R^2 = 0.8941 \tag{23}$$

$$B = 0.9732 \bullet DWT^{0.3288} \; R^2 = 0.9276 \tag{24}$$

$$T = 1.1025 \bullet DWT^{0.2157} \; R^2 = 0.8460 \tag{25}$$

$$P_{ENG} = 0.0006 \bullet L^{1.9793} \; R^2 = 0.8139 \tag{26}$$

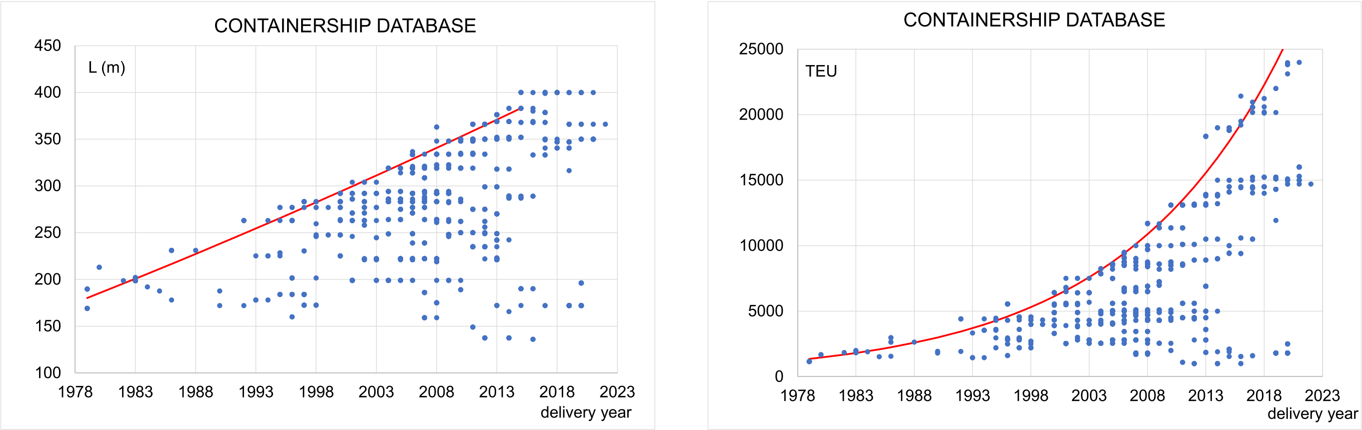


**Fig. 6.** Container ships statistics of length and TEU.

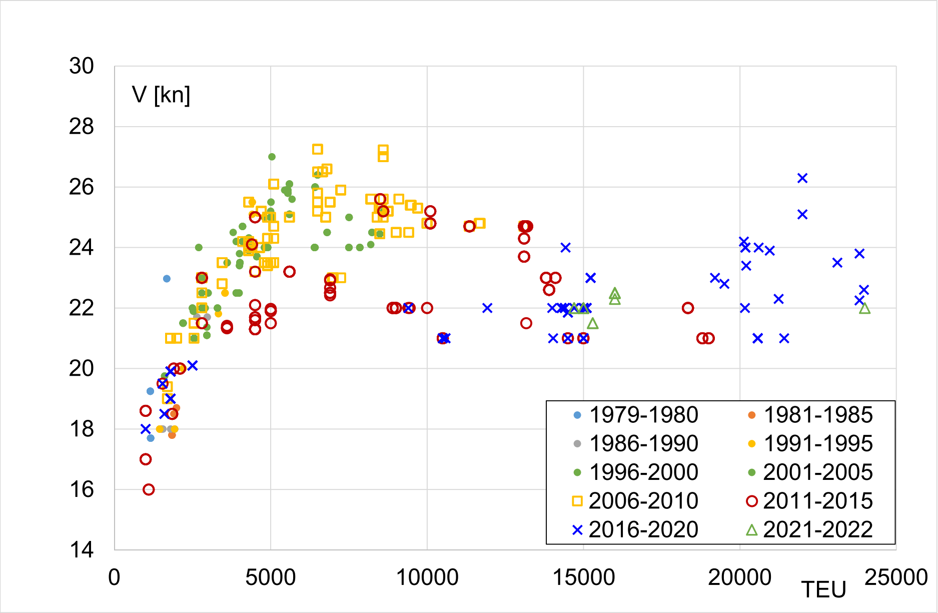


**Fig. 7.** Container ships statistics of ship speed as a function of number of TEU over years.

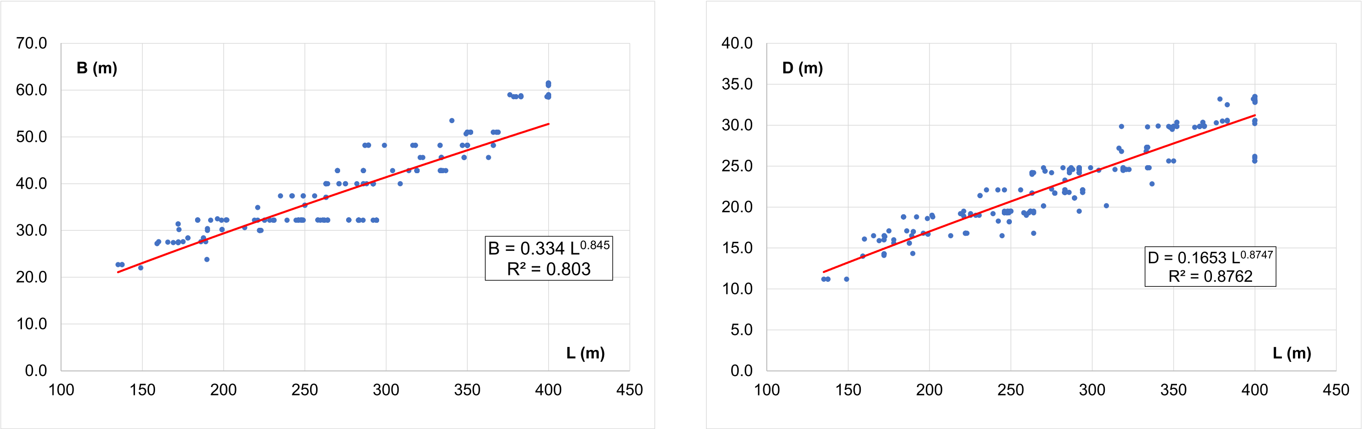


**Fig. 8.** Regression analysis of main ship dimensions.

$$P_{ENG} = 0.001 \bullet V^{3.2961}\ R^2 = 0.7456 \tag{27}$$

$$P_{ENG} = 0.3084 \bullet TEU^{0.5636}\ R^2 = 0.5448 \tag{28}$$

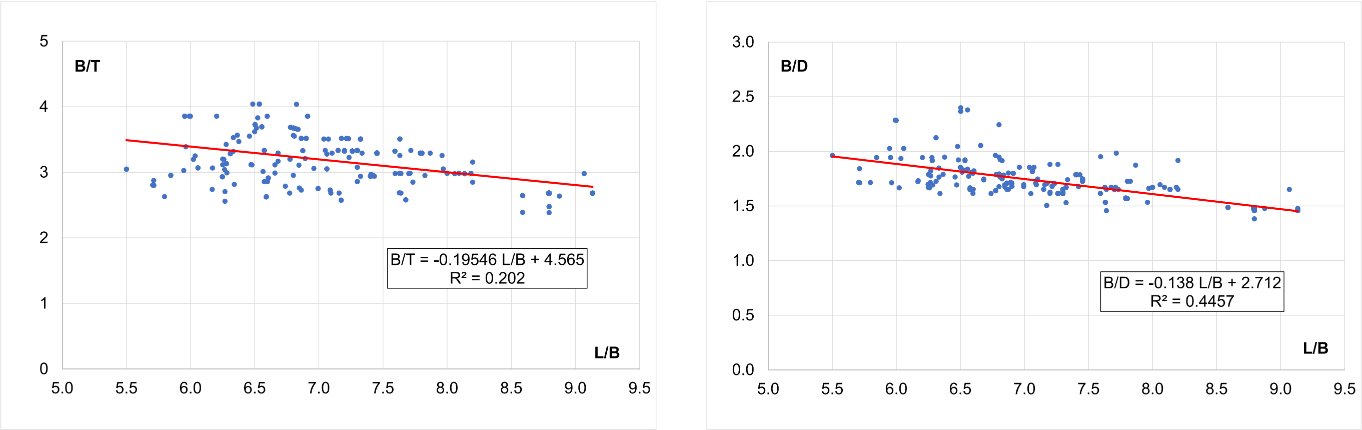


**Fig. 9.** Regression analysis of ship nondimensional ratios.

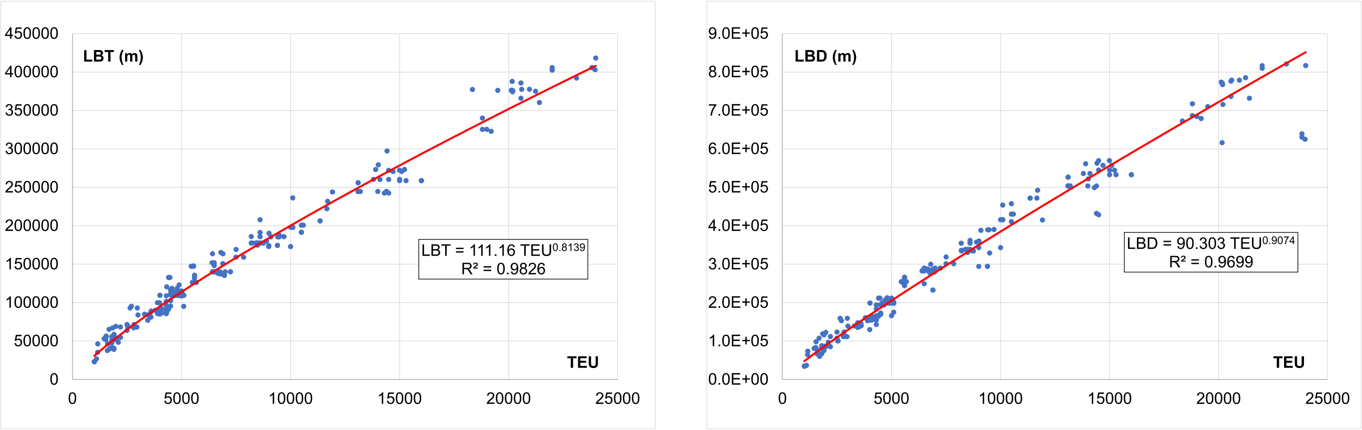


**Fig. 10.** Regression analysis of ship dimensions on TEU.

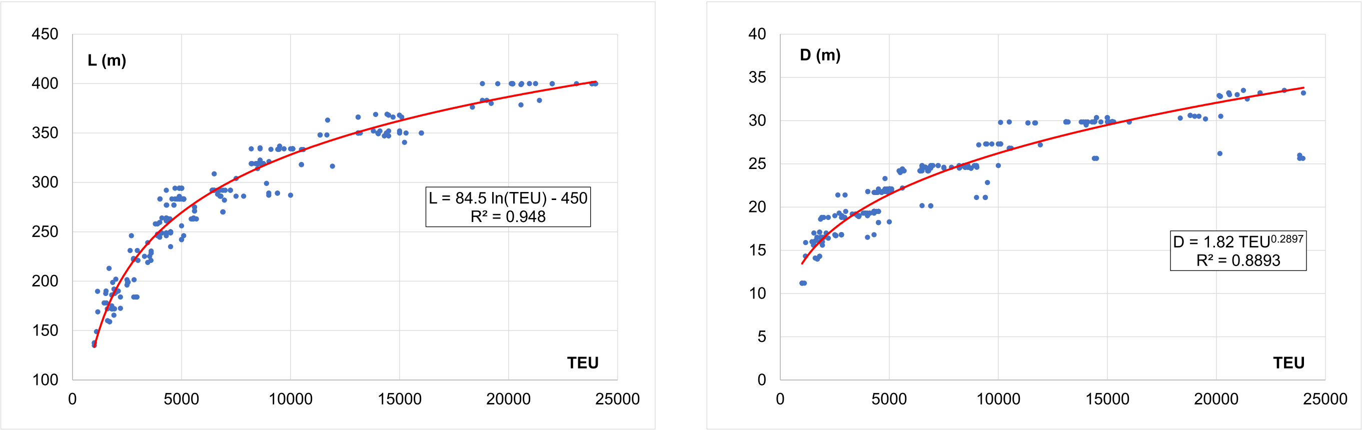


**Fig. 11.** Regression analysis of ship dimensions on TEU.

$$P_{ENG} = 0.0356 \bullet DWT^{0.6306} \quad R^2 = 0.5683 \tag{29}$$

$$V = 13.887 \bullet TEU^{0.0582} \quad R^2 = 0.1592 \tag{30}$$

$$V = 10.761 \bullet DWT^{0.068} \quad R^2 = 0.181 \tag{31}$$

The relations between main dimensions are well approximated by a power function and the data points are close to the regression curve, as shown in Fig. 8. It can be noted that in many cases the B value is constant for an increasing length and the increase in B has a discrete step strictly due to the container dimensions. For the nondimensional ratios B/T and B/D as function of L/B shown in Fig. 9, the spreading of data indicates no correlation between the variables, and it is confirmed also by the low $R^2$ of the obtained linear trendline.

Figs. 10–12 show the variation of each dimension as a function of the number of TEU. All dimensions are approximated with a power function except the length, which is better estimated by a logarithmic function. All data are well gathered around the tendency line, with values of $R^2$

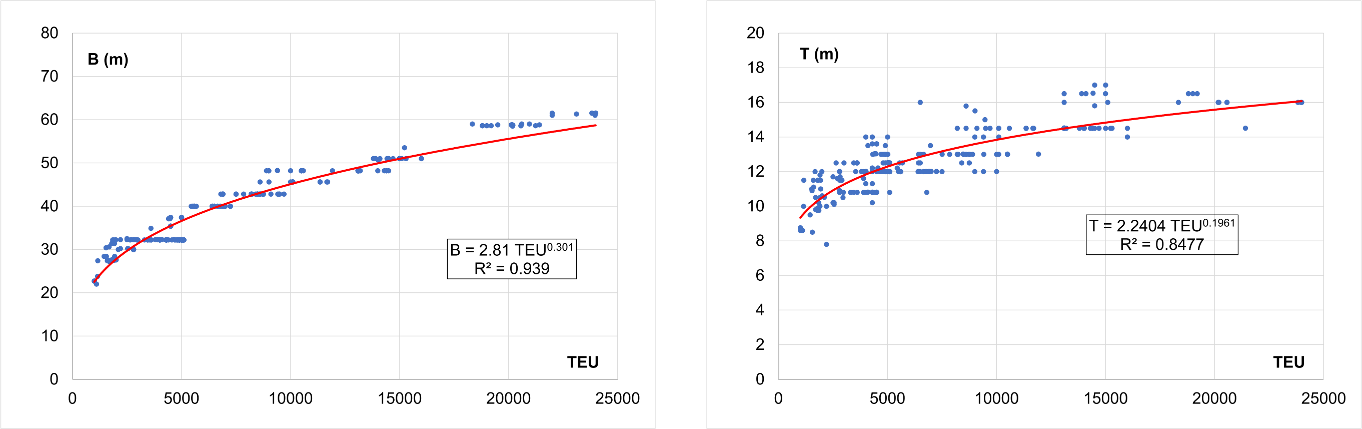


**Fig. 12.** Regression analysis of ship dimensions on TEU.

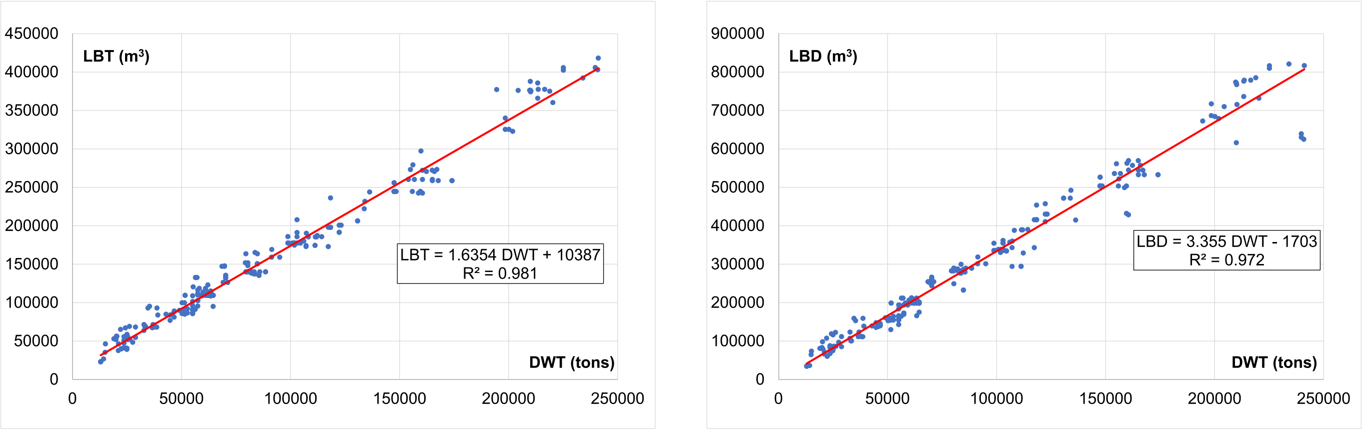


**Fig. 13.** Regression analysis of ship dimensions on DWT.

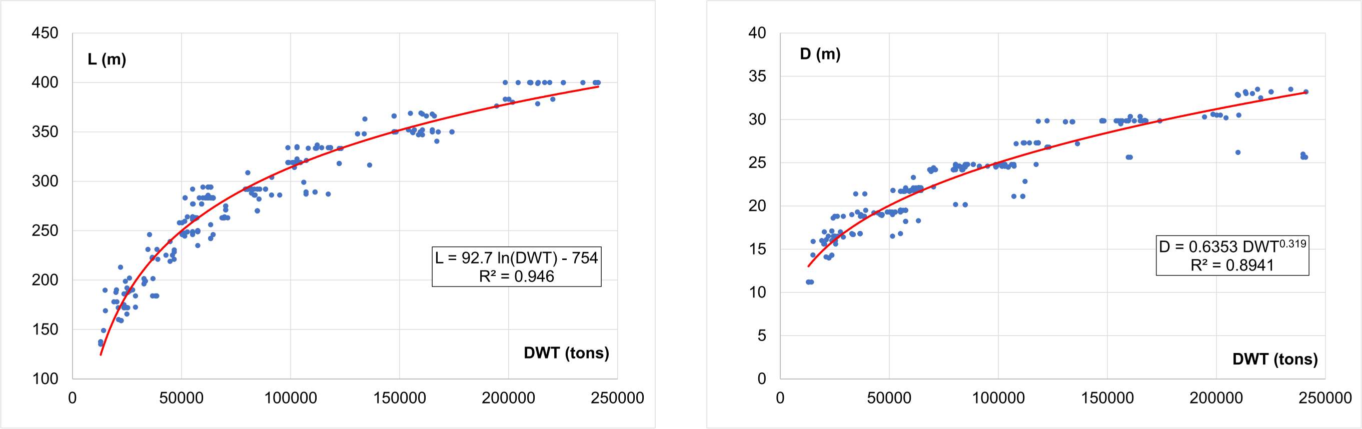


**Fig. 14.** Regression analysis of ship dimensions on DWT.

around 0.95–0.98, except for the draught and depth which are more scattered and the $R^2$ value is about 0.85 and 0.89, respectively.

In the present database, the value of the deadweight (DWT) was not available, therefore the correlation between TEU and DWT, reported in Equation (32), from Abramowski et al. (2018), has been adopted to generate the necessary data to obtain the regressions as a function of DWT, as presented in Figs. 13–15.

$$DWT = 1317.745 + 2.24 \bullet 10^{-3} \bullet (ln(TEU))^{8} \tag{32}$$

DWT regressions were best fitted by linear equations for the product of the main dimensions, as in Fig. 13, and by power formulas for the single dimensions (Figs. 14 and 15), except for the length, fitted by a logarithmic function. In all cases, the data samples are close to the regression curves; few points are more scattered when analyzing the depth values, especially for D = f(DWT) where the $R^2$ value decreases to 0.846. This may be due to the change of D value when keeping constant the B and T values.

It can be noted that the increasing step of B and T variables is discrete

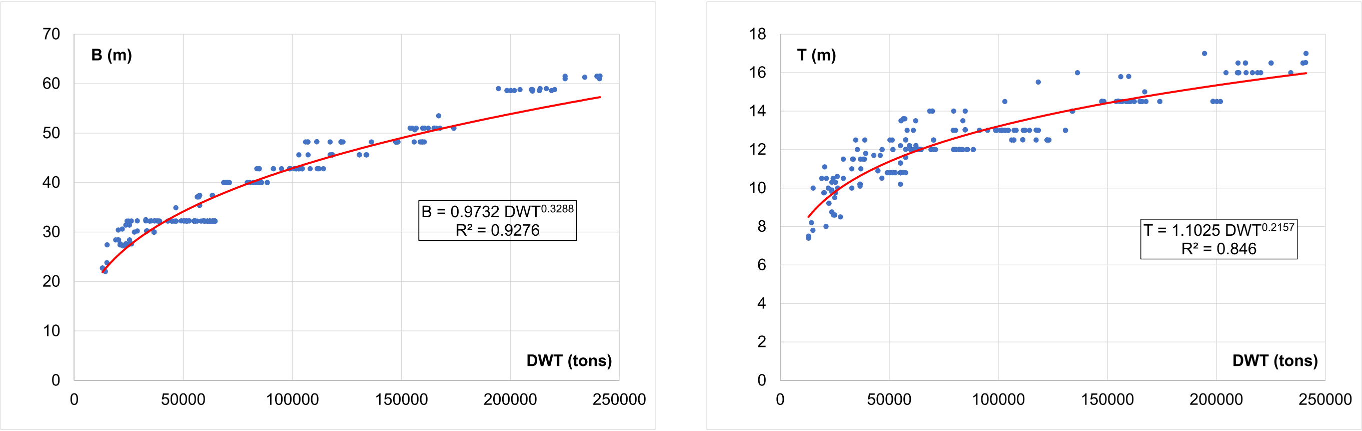


**Fig. 15.** Regression analysis of ship dimensions on DWT.

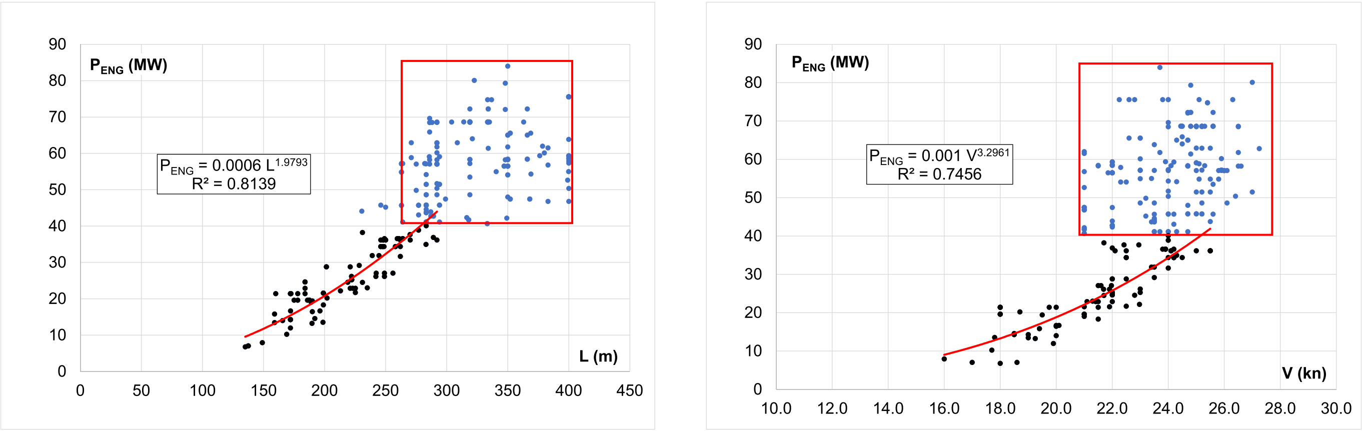


**Fig. 16.** Regression analysis of engine power on ship length and speed.

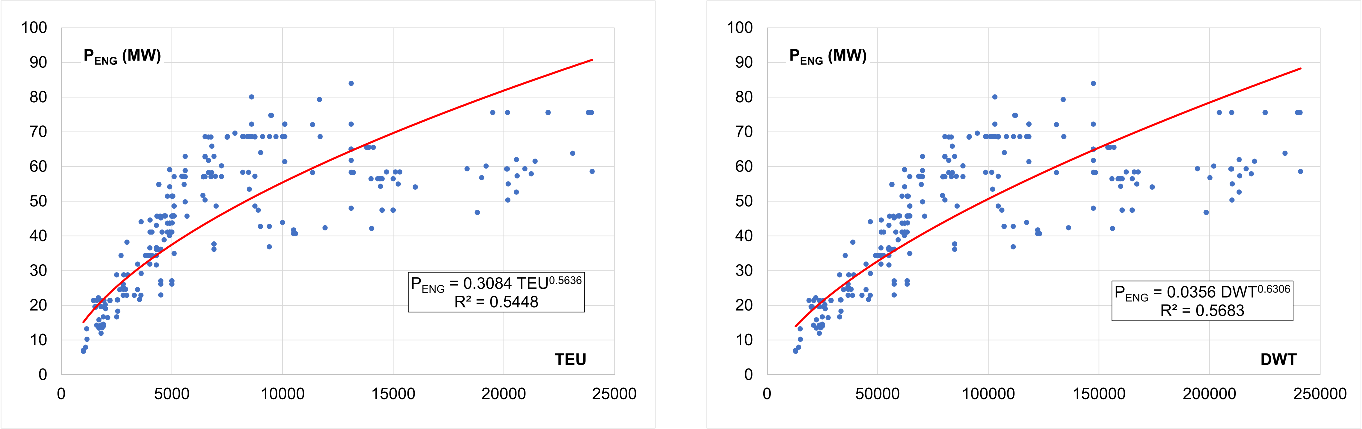


**Fig. 17.** Regression analysis of engine power on TEU and DWT.

for both TEU and DWT regressions, due to container sizes and limiting channels dimensions.

In the analysis of the installed power as a function of ship length or speed, two regions have been identified in Fig. 16: a regression curve and a rectangular area. When ships have an engine power lower than 40 MW (data represented by black dots) the regression curve has been approximated with a power function, reporting the formula in the graphs. For ships with higher engine power, the dependency of engine power from ship length or speed can no longer be represented by a regression curve. The data are very scattered and a rectangular region can be identified. It is clearly visible that for the VLC and ULCS the ship speed varies from 21 to 26 knots, probably set as the design requirement to serve some specific route in a certain number of days.

The engine power and the ship speed as function of TEU or DWT are presented in Figs. 17 and 18, respectively. The data is scattered and uniformly spread around the graph. Although the best regression is represented by a power function, the $R^2$ values are lower than 0.6 for engine power and lower than 0.2 for speed, highlighting a low direct

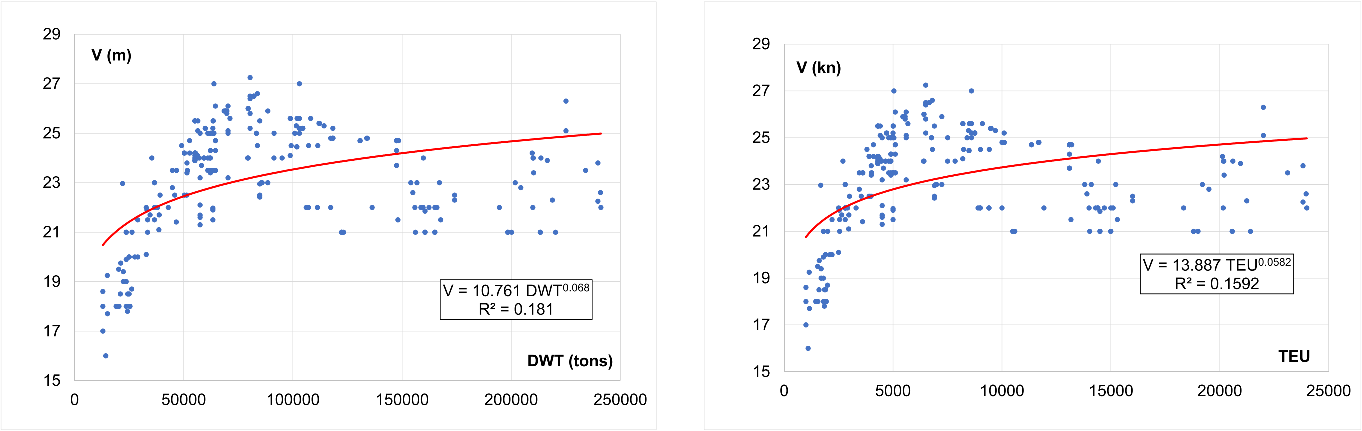


**Fig. 18.** Regression analysis of ship speed on TEU and DWT.

dependency of these two parameters from cargo capacity.

## 5. Comparison with previous studies

The present regression equations have been compared with the ones reported by Cepowski and Chorab (2021) and in Papanikolaou (2014) who recalled formulations and regressions developed by Kolakarinos et al. (2000–2005).

Since in Papanikolaou (2014) all ship dimensions are estimated only as a function of the DWT, the correlation between DWT and TEU, obtained by the regression analysis in Papanikolaou (2014), is reported in Equation (33) and has been used for the comparison in terms of TEU.

$$Payload = 0.75\ DWT = 10 \bullet TEU$$
$$\rightarrow \boldsymbol{DWT} = 13.33\ \boldsymbol{TEU} \tag{33}$$

As already mentioned in Section 2, the value of DWT for the present database has been obtained using Equation (32) (Abramowski et al., 2018).

Table 2 and Figs. 19–23, illustrate the comparison between the formulas of the regression analysis for the present database and the ones reported in Papanikolaou (2014) and Cepowski and Chorab (2021).

Fig. 19 compares the results of the present analysis and the ones from Papanikolaou (2014) in terms of ship dimensions, velocity, and number of TEU. While the dependency of breadth on length has a similar tendency for both regressions in all the dimensions range, the estimations of the product of LBD follow two different tendencies for values greater than 5000 TEU. The database used in Papanikolaou (2014) included ships built up to 300 m and 80000 DTW (about 6000 TEU), and the estimation of bigger ships with the extension of the original regression formula is overpredicted. The present regression gives a better overall approximation of the database for velocity dependency on TEU, even though the higher speed values are not caught with a power regression.

Figs. 20 and 21 show the comparison between the three regressions for the estimation of ship dimensions as functions of TEU. While breadth and draught approximations are in good agreement in all three cases, for length and depth it is clear that: when extrapolating the formulas of regression in Papanikolaou outside the limits of the referenced database (about 1000 TEU), the values are overestimated and do not follow the tendencies of new built ships; and that the formulas in Cepowski and Chorab (2021) have been reported with some typo-errors.

The obtained regression formulas for the main ship dimensions have been used for an hypothetical 30000 TEU ship and compared with Garrido et al. (2020). The estimated length for 30000 TEU results equal to 421 m for the present database regression, 433 m calculated with Ceposwki's formulation, and 531 m following Papanikolaou regression formula. Since the length evaluated by Garrido et al. (2020) is about 420 m, the best fitting curve that estimates the closest value is the one obtained by the present database.

Ship dimensions as functions of the DWT are reported in Figs. 22 and 23; in this case, the estimations of Papanikolaou are in good agreement with the other two analyses. This difference with the tendency found for TEU dependencies may be attributed to the different correlation formulas adopted for DWT and TEU.

In the works of Papanikolaou (2014) and Cepowski and Chorab (2021), $R^2$ values were not available for all formulas, therefore, to evaluate the goodness of the formulas with the present database, the $R^2$ value have been calculated using Equation (4) considering the present database as the observed value and the estimated values have been predicted using the formulas as reported in Table 2. The calculated $R^2$ values for the ship dimensions as function the number of TEU are presented in Table 3. For Cepowski and Chorab (2021) the $R^2$ values have been calculated only considering the DWT formulation due to the typos in the formulas as function of TEU. It can be seen that except for the B value, all others are better predicted by the formulas of Cepowski and Chorab (2021). The comparison for ships speed has not been reported since the correlation with TEU is very low.

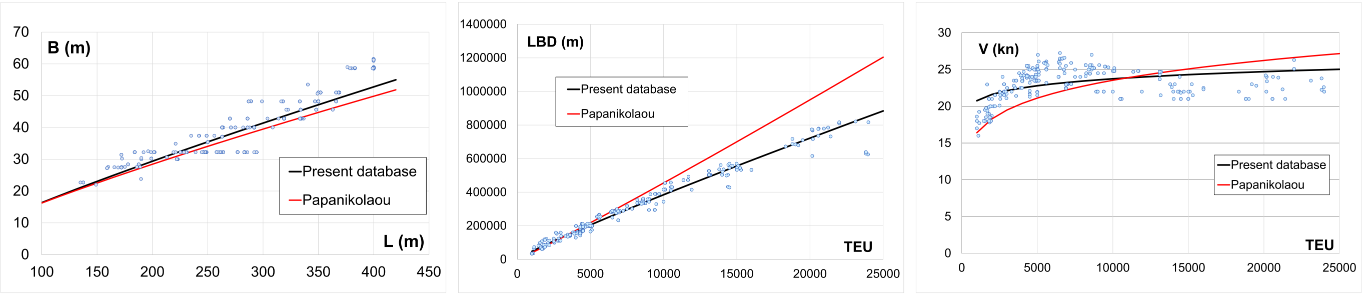


**Fig. 19.** Ship dimensions and TEU regressions compared to Papanikolaou (2014).

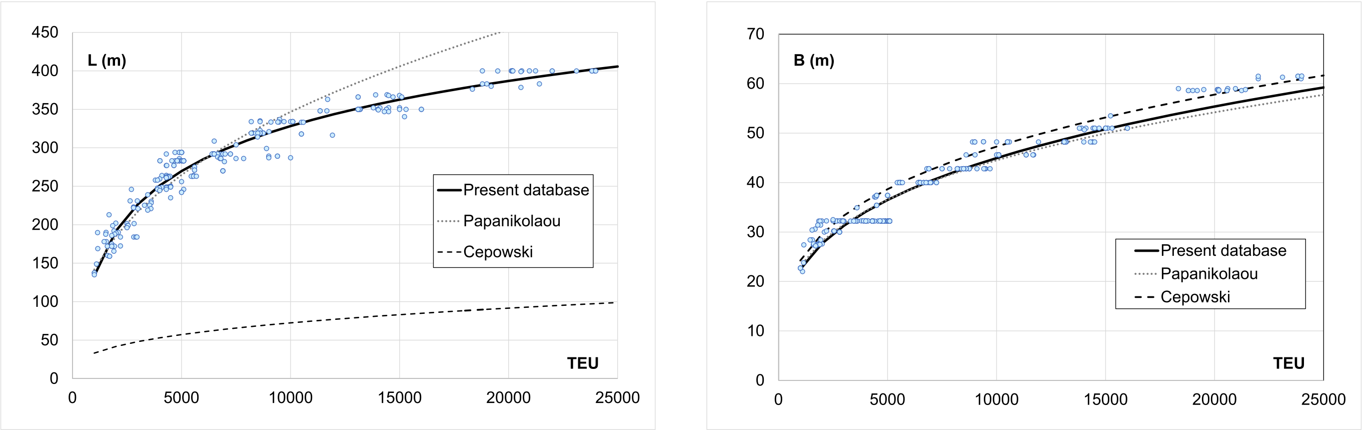


**Fig. 20.** Ship dimensions and TEU regressions compared to Papanikolaou (2014) and Cepowski and Chorab (2021).

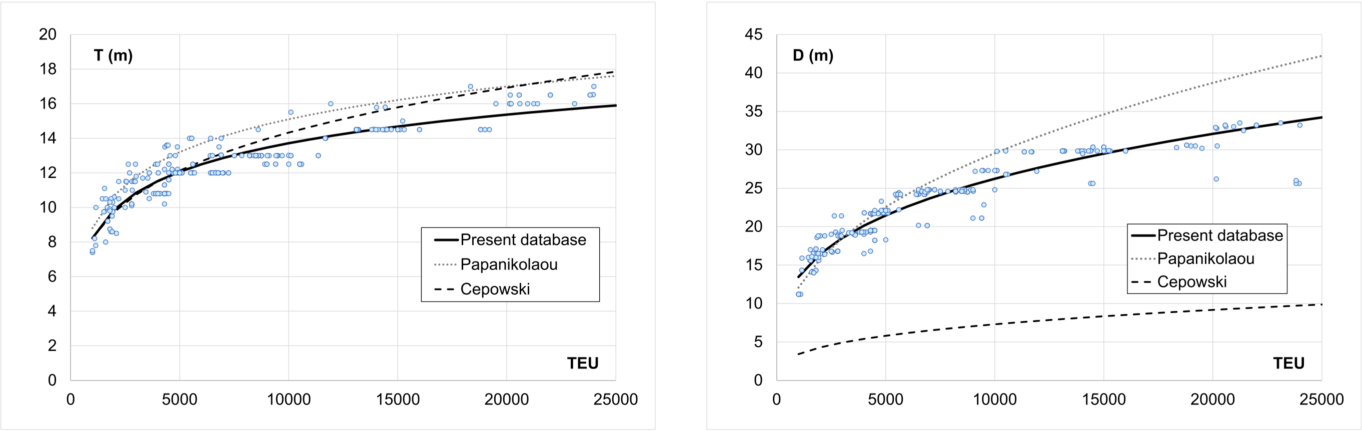


**Fig. 21.** Ship dimensions and TEU regressions compared to Papanikolaou (2014) and Cepowski and Chorab (2021).

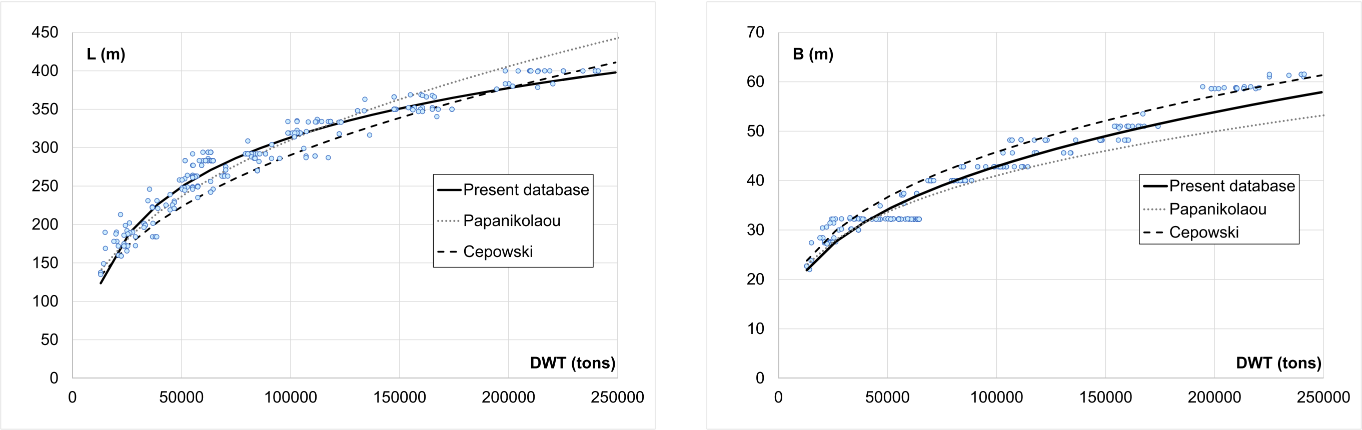


**Fig. 22.** Ship dimensions and DWT regressions compared to Papanikolaou (2014) and Cepowski and Chorab (2021).

## 6. Multivariable regressions

In this section, different multivariable regressions (MR) are presented in the following forms.

- MR1 for ship dimension and engine power function of V and TEU;
- MR2 for ship speed and engine power function of L, D and TEU;
- MR3 for ship speed and engine power function of L, T and TEU;
- MR4 for ship speed and engine power function of L, B, D and TEU.

For each regression the values of estimates, SE, t-stud, and p-value have been evaluated and reported in Appendix B. The whole set of fit coefficients, $R^2$, $R^2_{adj}$ MAPE, RMSE, RRMSE and Pearson, has been reported in Table C3.

An extended analysis of the multiple linear regression has been conducted. The multicollinearity has been checked according to the VIF (Variance Inflation Factor), highlighting no collinearities for all the regressions function of V and TEU, as shown in Table C5 in Appendix C. There is multicollinearity in the case of the regressions provided as a

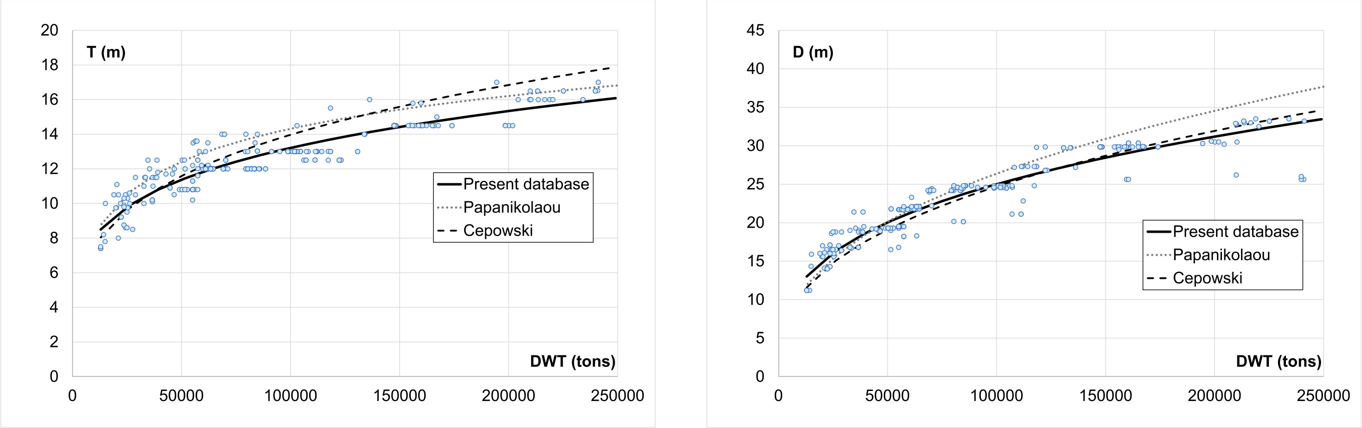


**Fig. 23.** Ship dimensions and DWT regressions compared to Papanikolaou (2014) and Cepowski and Chorab (2021).

**Table 2**
Comparison of the obtained regression formulas from the literature.

| Present Database | | Papanikolaou (2014) | Cepowski and Chorab (2021) | |
|---|---|---|---|---|
| TEU | DWT | | TEU | DWT |
| B = 0.3339 $L^{0.845}$ | | B = 0.3899 $L_{OA}^{0.8096}$ | | |
| L = 84.5 ln (TEU) - 450 | L = 92.7 ln (DWT) - 754 | $L_{BP}$ = 3.54132 $DWT^{0.388442}$ | $L_{BP}$ = 3.16 $TEU^{0.34}$ | $L_{BP}$ = 3.656 $DWT^{0.38}$ |
| B = 2.81 $TEU^{0.301}$ | B = 0.9732 $DWT^{0.3288}$ | B = 1.55219 $DWT^{0.284381}$ | B = 3.27 $TEU^{0.29}$ | B = 1.15 $DWT^{0.32}$ |
| T = 2.3834 ln (TEU)- 8.2372 | T = 1.1025 $DWT^{0.2157}$ | T = −17.1581 + 2.72338 ln(DWT) | T = 1.571 $TEU^{0.24}$ | T = 0.624 $DWT^{0.27}$ |
| D = 1.82 $TEU^{0.2897}$ | D = 0.6353 $DWT^{0.319}$ | D = 0.299394 $DWT^{0.38902}$ | D = 0.349 $TEU^{0.33}$ | D = 0.349 $DWT^{0.37}$ |
| LBD = 90.303 $TEU^{0.9074}$ | LBD = 3.4256 $DWT^{0.9971}$ | LBD = 1.64898 $DWT^{1.06169}$ | | |
| V = 13.887 $TEU^{0.0582}$ | V = 10.761 $DWT^{0.068}$ | V = 3.7087 $DWT^{0.1566}$ | | |

**Table 3**
Comparison of coefficients of determination with Papanikolaou (2014) and Cepowski and Chorab (2021).

| $R^2$ | L = f(TEU) | B = f(TEU) | D = f(TEU) | T = f(TEU) |
|---|---|---|---|---|
| **Present database** | 0.948 | 0.939 | 0.889 | 0.848 |
| **Papanikolaou** | 0.779 | 0.925 | 0.500 | 0.436 |
| **Cepowski (DWT)** | 0.854 | 0.851 | 0.863 | 0.743 |

function of L, B, D and TEU. Power and speed regressions function of L, D, TEU and L, T, TEU, highlight a moderate collinearity compared to the previous regressions. For all the regressions the normality of the residuals has been checked with the Kolmogorov-Smirnov test, giving positive results for all tested cases. Furthermore, heteroscedasticity has been evaluated according to the Breush-Pagan test, as shown in Table C6 in Appendix C, detecting homoscedasticity only for the regression of B as a function of V and TEU and for V as a function of L, T, TEU. All other cases are affected by heteroscedasticity of data, decreasing the reliability of the final regression. In the MR4 regressions, with 4 dependent variables, heteroscedasticity is associated with the presence of multicollinearity, in other cases deals only with the nature of data. In any case, being the objective of the regressions the estimation of the independent variable and not the influence of each parameter on the final regression, the detection of non-constant variance does not require the manipulation of input data to eliminate the problem. The normality of multiple linear regression residuals can be found in Appendix D.

Each ship characteristic estimated by the different multivariable regressions is compared with original corresponding value of the database, as shown in Figs. 24–29. In each figure the spreading of data around the bisector indicates the goodness of fit of each regression; the points are gathered around the bisector line and the closer they are to the bisector line the better they are estimated by the regression formulas.

In Figs. 24 and 25, the estimated ship dimensions (L, B, D and T) are functions of V and TEU, and while for length and breadth the values are well gathered around the bisector line, highlighting the goodness of the regression formulas, for the depth and the draught the values are more spread. The fit coefficients in Table C3 confirm these tendencies, with $R^2$, $R^2_{adj}$, RMSE and Pearson coefficients higher and MAPE and RRMSE coefficients lower for L and B dimensions. In the graph reporting the breadth, the discrete step linked to container ship sizes is recurring.

Figs. 26–29 present the comparison of different regression methods for engine power and ship speed. As confirmed also by the fit coefficients in Table C3 the MR4 regressions are better estimated, even though the small improvement may not be worth the increment of input variables. In particular $R^2$, $R^2_{adj}$ and Pearson coefficients are higher for MR4 and MAPE, RMSE and RRMSE coefficients are lower. It is worth noting the high scattering data around the ship speed of 21 and 22 knots where for a constant value of the present database the estimated values vary in a range of about 2 knots. A similar tendency can be seen for the engine power around 68000 kW. This may be due to the limitations in actual engine performances and fixed ship speed in trip voyages for different ships.

## 7. Forest trees

Besides multiple linear regressions, forest tree regressions are a suitable advanced technique to investigate the dependencies of the main dimensions of the container ships from one or more parameters. The forest tree algorithm allows the classification of the output through the averaged prediction of more individual trees (Ho, 1998), thus reducing the overfitting problem of individual trees. Here, the MATLAB application for the determination of forest tree is applied to the database, providing regression for the quantities of interest. The forest trees for simple regressions (parameters as a function of TEU) and multivariable regressions (MR1, MR2, MR3 and MR4) have been performed to estimate ship main dimensions, ship speed and engine power. The values estimated by the forest tree algorithm have been compared with the original ones of the present database, as described for the multivariable regression and are reported in Figs. 30–38 with the corresponding coefficient of determination. In the graphs the subscript SR defines the simple regression and $MR_i$ the multivariable regression approximations.

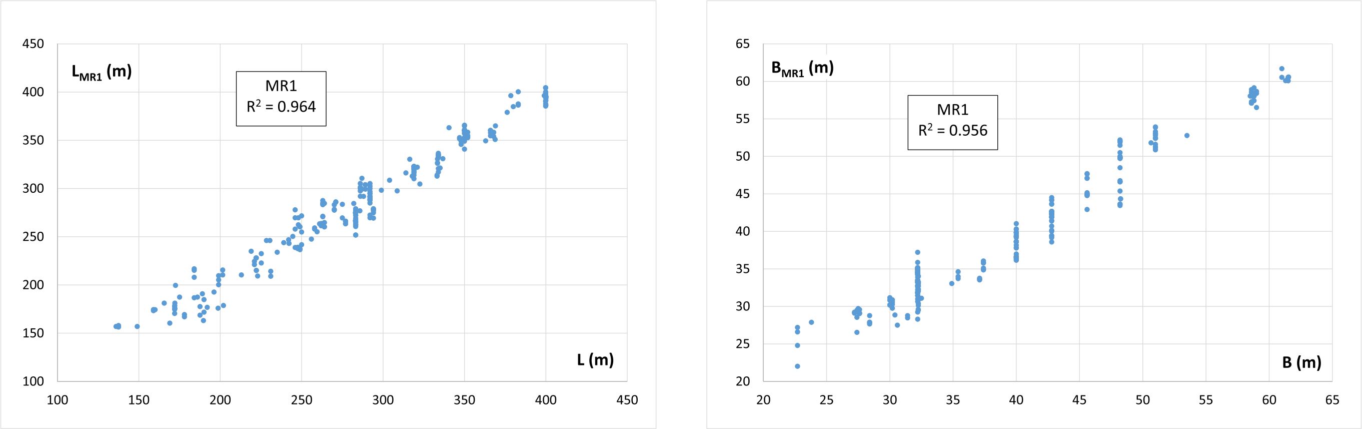


**Fig. 24.** Multivariable regression of ship dimensions function of V and TEU.

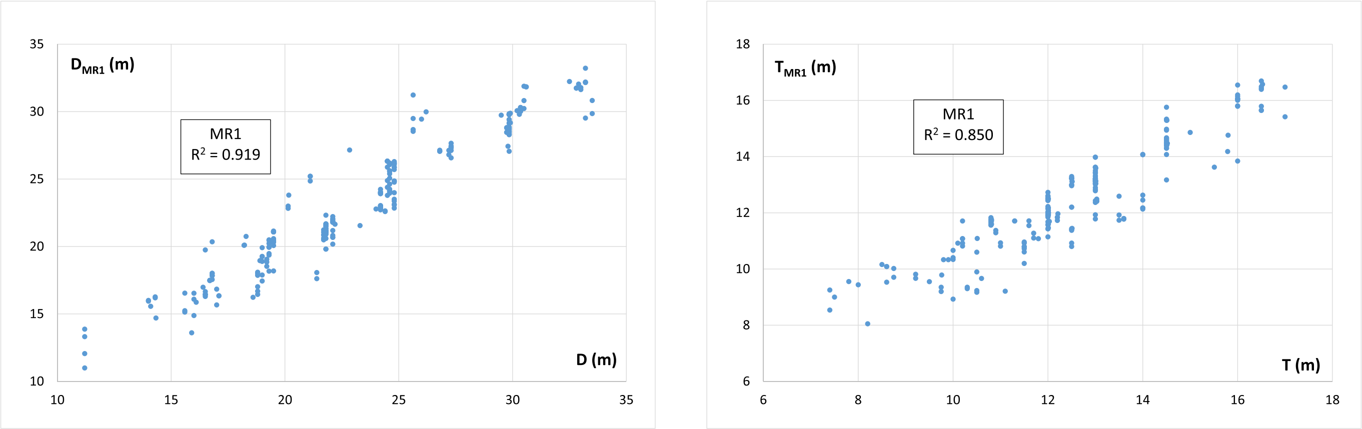


**Fig. 25.** Multivariable regression of ship dimensions function of V and TEU.

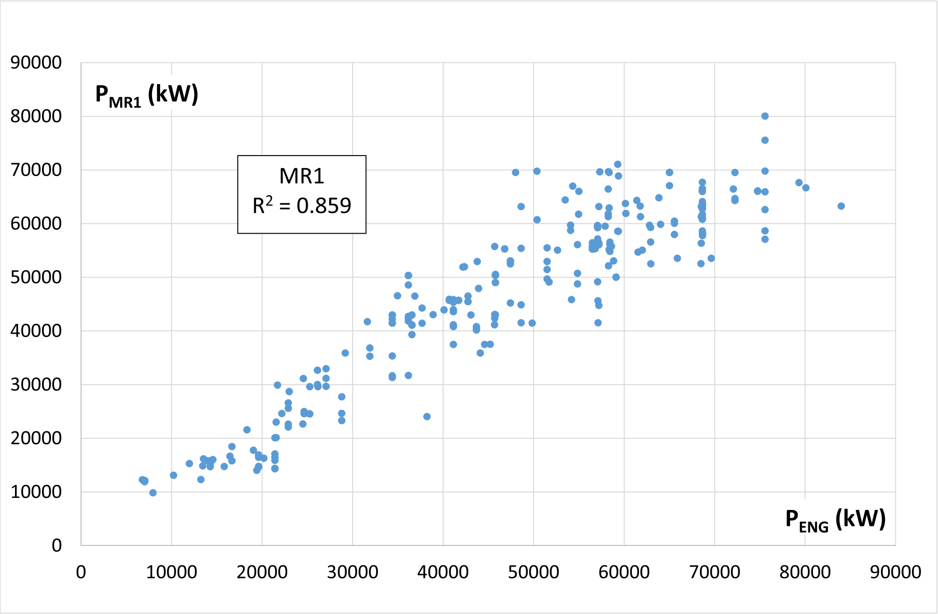


**Fig. 26.** Multivariable regression of engine power function of V and TEU.

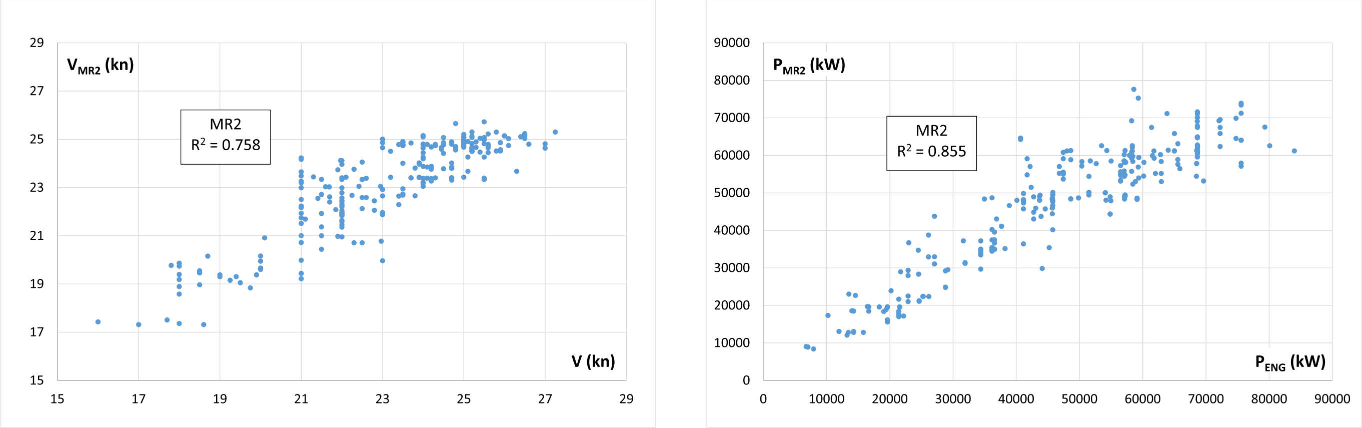


**Fig. 27.** Multivariable regression of ship speed and engine power function of L, D and TEU.

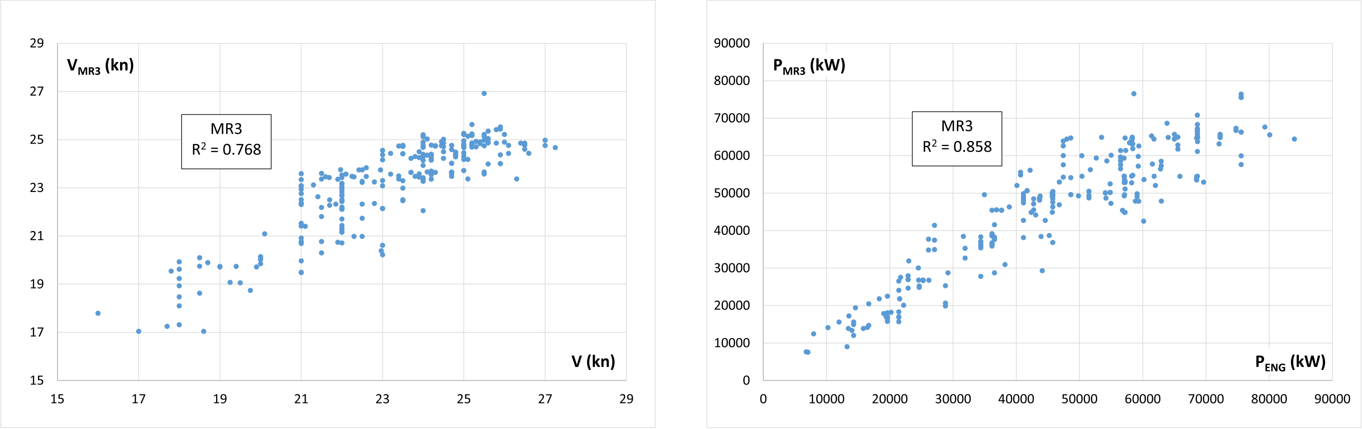


**Fig. 28.** Multivariable regression of ship speed and engine power function of L, T and TEU.

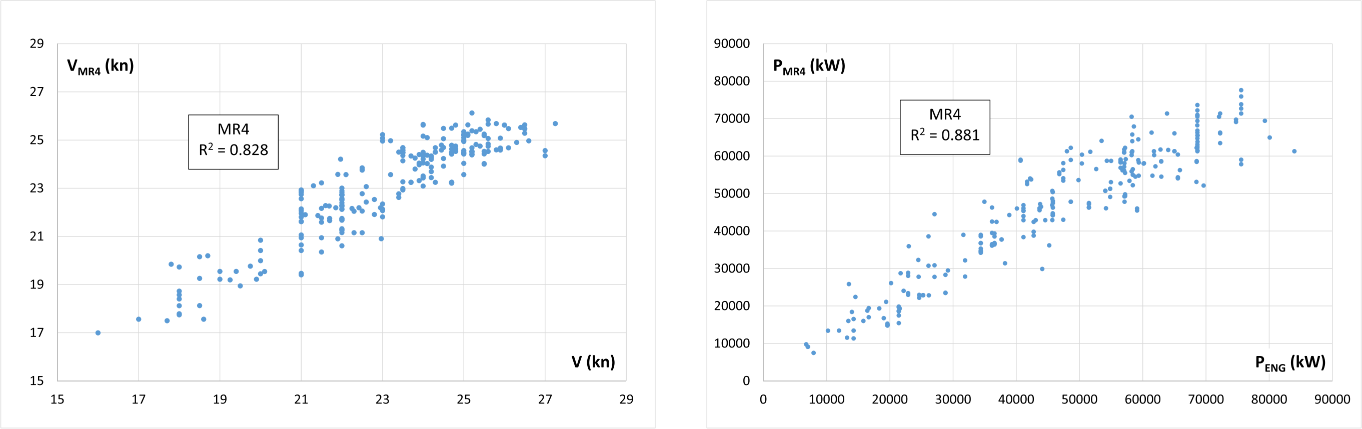


**Fig. 29.** Multivariable regression of ship speed and engine power function of L, B, D and TEU.

As expected, for all parameters, except for the depth, data is less scattered for multivariable regressions than for simple ones and it is confirmed by the high value of $R^2$. In particular, Figs. 30 and 31, presenting L and B values, have well-gathered data around the bisector, while T and D values, shown in Figs. 32 and 33, are more scattered, as already observed for multivariable regressions in Section 6. Figs. 34 and 35 present the different regressions for the ship speed and Figs. 36–38 the regressions for engine power. The best fitting, with higher $R^2$, is estimated by MR2, differently than in the case of multivariable regressions of Section 6, where the best fitting was found for MR4. Overall, the forest tree algorithm better fits the database compared to the linear multivariable regression in Section 6, as can be seen from the graphs and the values of all fitting coefficients in Table C4 of Appendix C.

The disadvantage of a forest tree is the absence of a simple regression formula for determining the desired variables. However, having at disposal a database to determine the forest tree allows for applying this method also in the early design stage. In particular knowing the

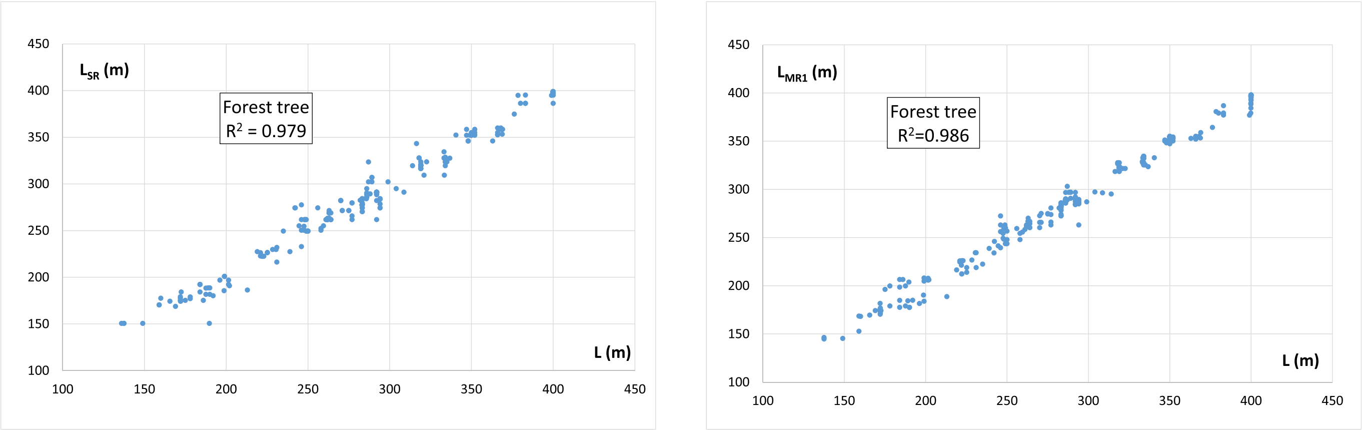


**Fig. 30.** Length estimation by forest tree simple (left) and multivariable MR1 (right) regressions.

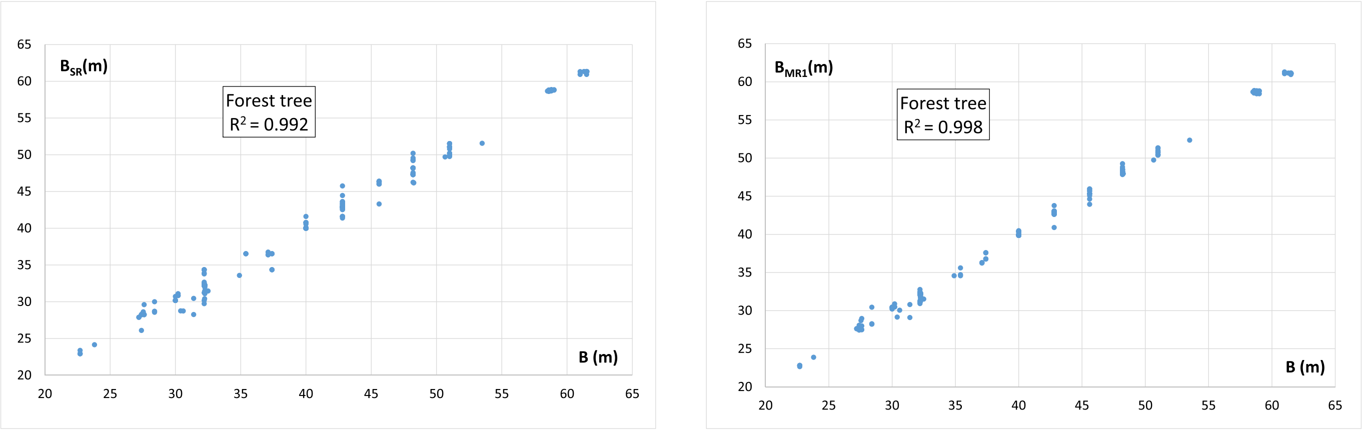


**Fig. 31.** Breadth estimation by forest tree simple (left) and multivariable MR1 (right) regressions.

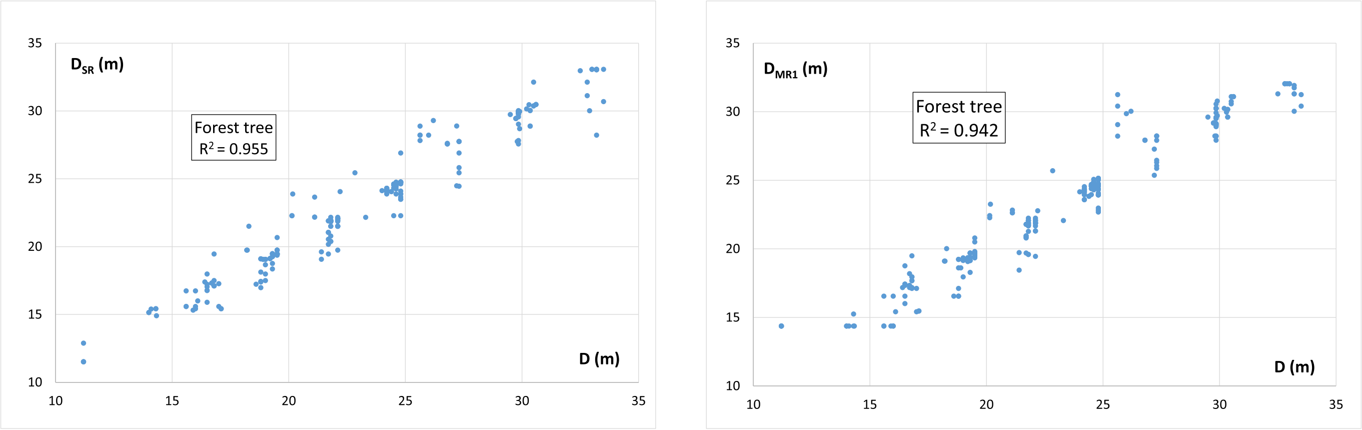


**Fig. 32.** Moulded depth estimation by forest tree simple (left) and multivariable MR1 (right) regressions.

maximum coefficient of determination $R^2$ from forest tree, it is possible to investigate which of simple or multiple, linear or nonlinear regression formulas reaches this value.

## 8. Application example

To have an insight into the possible results when estimating ship parameters in the early design stage, a container ship of 20000 TEU designed to sail at a speed of 23 kn has been considered as a test case.

The different regression methods have been used to estimate ship dimensions, velocity, and engine power and the results are shown in Table 4. For the single regressions the parameters are function only of the TEU variable and Equations (16)–(19) and (28) and (30) have been applied. For the multivariable regressions ship dimensions and engine power have been estimated from the equations in Appendix B in the form of Y = f (TEU, V).

The generic equation for a specific parameter can be written as:

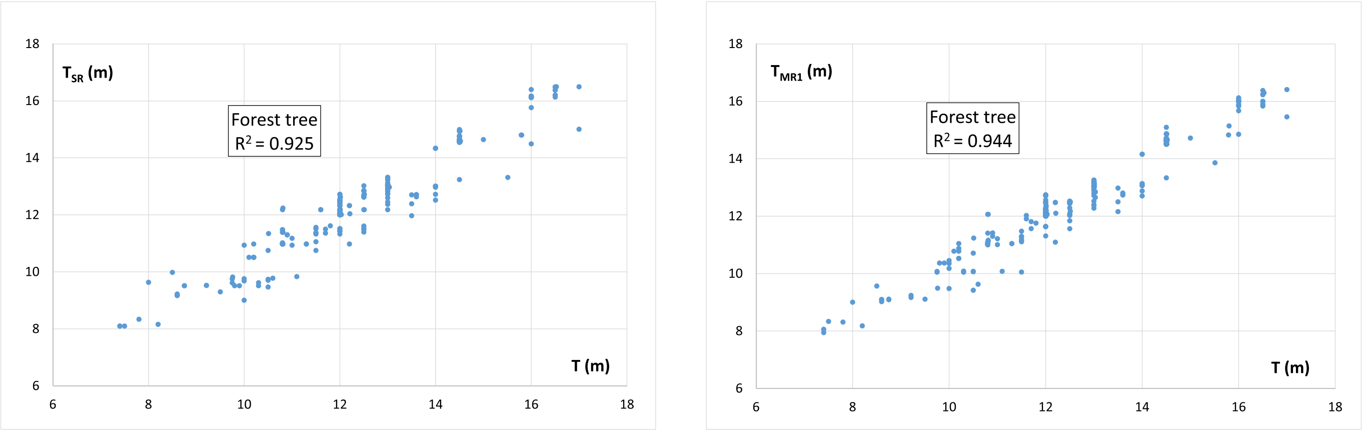


**Fig. 33.** Draught estimation by forest tree simple (left) and multivariable MR1 (right) regressions.

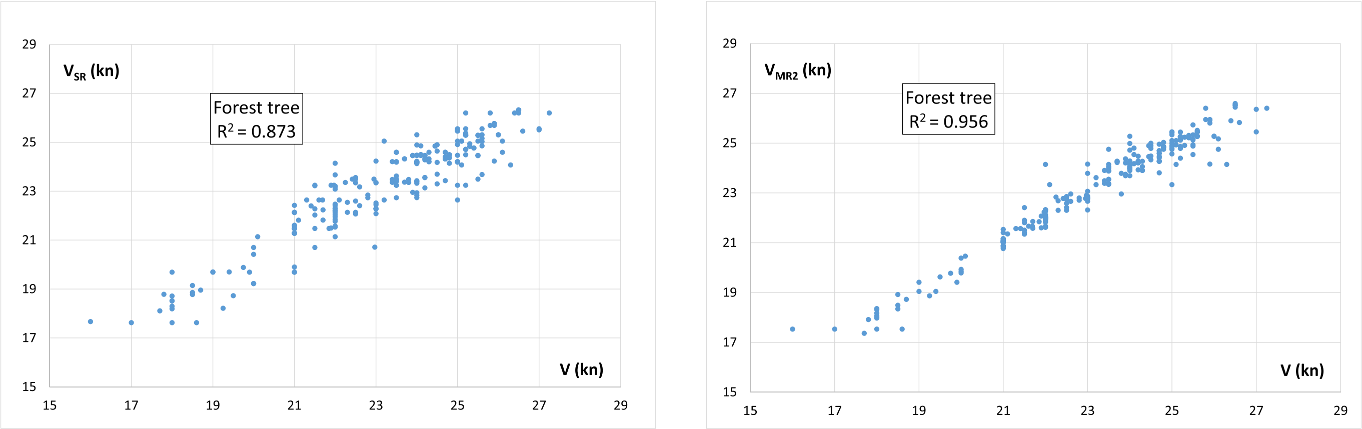


**Fig. 34.** Ships speed estimation by forest tree simple (left) and multivariable MR2 (right) regressions.

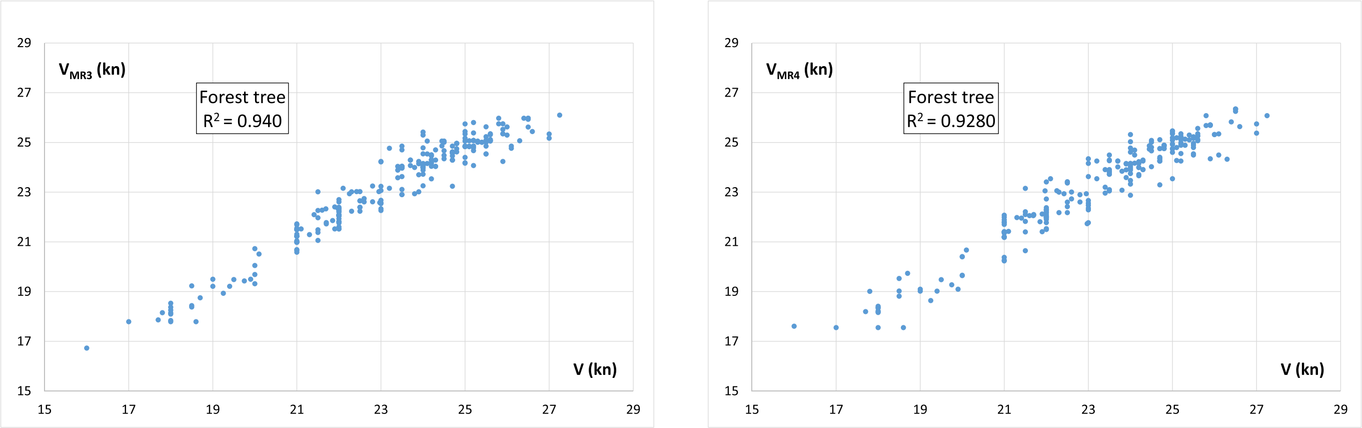


**Fig. 35.** Ships speed estimation by multivariable MR3 and MR4 forest tree regressions.

$$Y=\sum_i\sum_k\sum_t c_i V^k TEU^t + d \tag{34}$$

where $c_i$ are the estimates of the variables $V^k$ and $TEU^t$ and $d$ is the intercept.

Forest tree regressions have been also applied.

Table 5 presents the percentage difference between the values estimated using forest trees and multivariable regressions and the values obtained by the single regressions. It can be noticed that a great difference appears for the estimation of engine power, highlighting the better regressions obtained by multivariable formulations and forest tree algorithms.

## 9. Discussion and conclusions

In the present work, a database of container ships representing the 20% of the world fleet, with vessels built up to 2022, has been analyzed. The database includes about 1000 ships, 260 of which are non-sisterships. A complete regression analysis, using methods of different

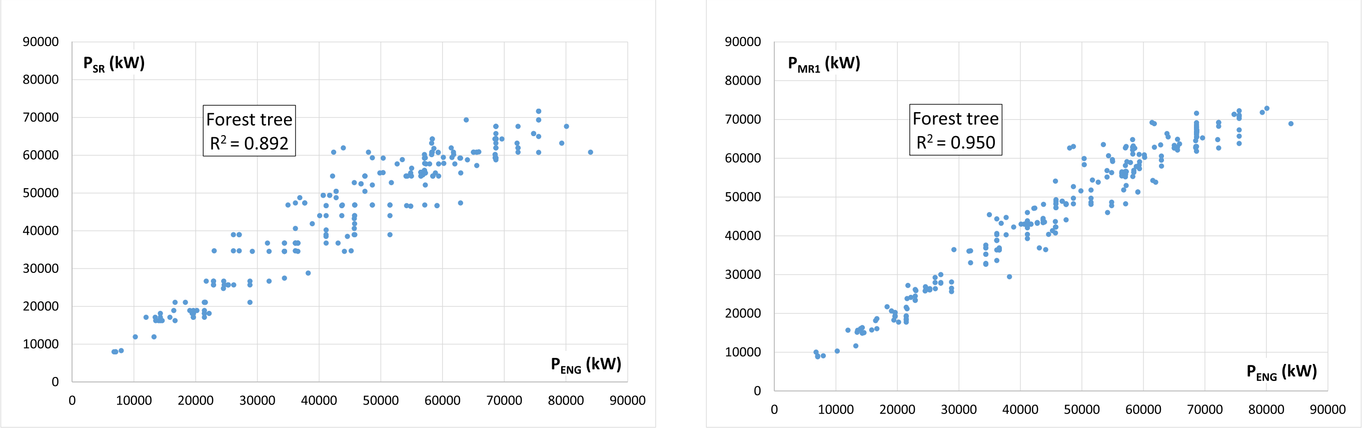


**Fig. 36.** Engine power estimation by forest tree simple (left) and multivariable MR1 (right) regressions.

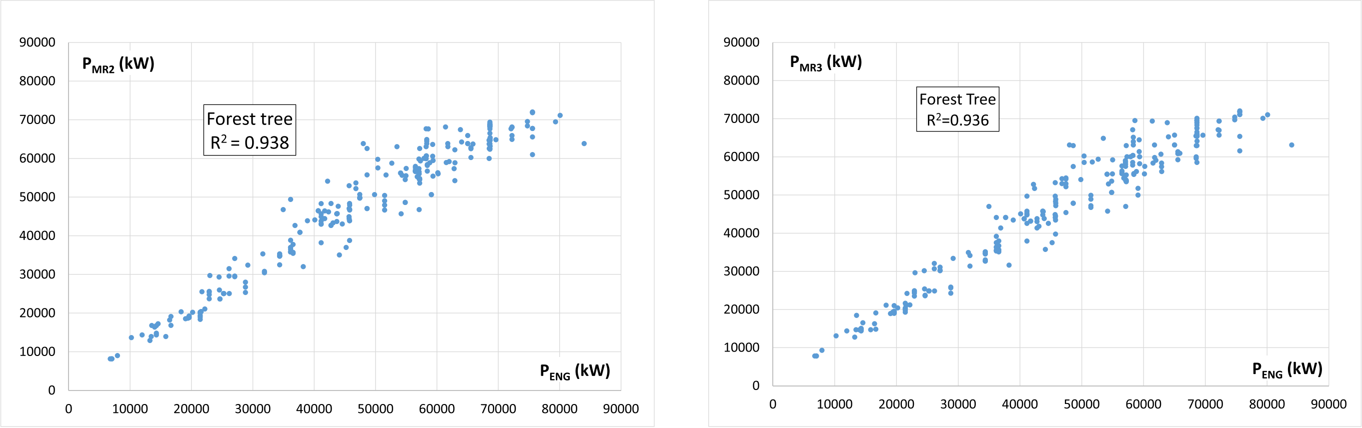


**Fig. 37.** Engine power estimation by forest tree multivariable MR2 and MR3 regressions.

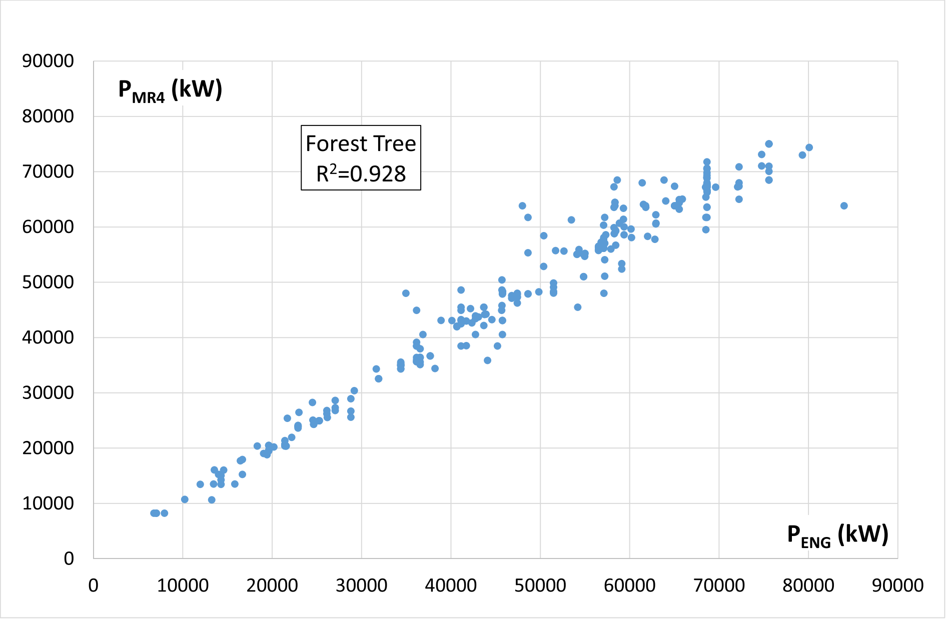


**Fig. 38.** Engine power estimation by forest tree multivariable MR4 regression and all regressions together.

**Table 4**
Estimated values of ship parameters for a container ship of 20000 TEU at 23 kn.

| Application case for container ship of 20000 TEU at 23 kn | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Multiple linear regressions** | | **Simple regressions** | | **Trees simple** | | **Trees multiple** | |
| **L** | 388.13m | **L** | 386.84m | **L** | 398.37m | **L** | 397.98m |
| **B** | 57.85m | **B** | 55.38m | **B** | 58.60m | **B** | 58.88m |
| **T** | 15.92m | **T** | 15.62m | **T** | 16.13m | **T** | 15.91m |
| **D** | 30.29m | **D** | 32.07m | **D** | 30.63m | **D** | 30.70m |
| | | **V** | 24.71kn | **V** | 23.21kn | | |
| **P** | 63681.1625 kW | **P** | 81879.44 kW | **P** | 63526.44 kW | **P** | 61757.17 kW |

**Table 5**
Percentage difference of forest tree and multivariable regressions compared to single regressions.

| Percentage difference compared to simple regressions | | | | | |
|---|---|---|---|---|---|
| **Trees simple** | | **Trees multiple** | | **Multivariable regressions** | |
| **B** | 5.8% | **B** | 6.3% | **B** | 4.5% |
| **T** | 3.2% | **T** | 1.9% | **T** | 1.9% |
| **D** | 4.5% | **D** | 4.3% | **D** | 5.5% |
| **V** | 6.1% | **P** | 24.6% | **P** | 22.2% |
| **P** | 22.4% | | | | |

complexity, has been performed to estimate ship main dimensions, speed, and engine power in a preliminary stage of ship design. Simple regression formulas obtained from the present database have been compared with the results reported in Papanikolaou (2014) and Cepowski and Chorab (2021). The most evident difference in the design trends is shown when comparing the present results to the ones in Papanikolaou (2014), since the ship length and DWT from the database mentioned in his work had lower limits and the extrapolation of the regression formulas outside the limits overestimated the values of new built ships. Overall, the present regression is in line with the results in Cepowski and Chorab (2021) which used a similar database, except for two cases where probably some typo errors occurred in their paper.

To increase the goodness of the developed models, the regression analysis is further studied by multivariable regressions and forest trees algorithms, providing regression solutions as a function of more than one design variable. The results of the more complex regression techniques show an improvement compared to simple regression models, especially by employing forest trees. It can be appreciated that, for the engine power $P_{ENG}$ prediction, the $R^2$ has been significantly improved from 0.545 when using simple regression up to 0.88 for the multiple regression $P_{ENG} = f$ (L, B, D, TEU). The ANN approach in Majnaric et al. (2022) proposes the form with the final $R^2$ 0.66.

A comparison of the different fitting coefficients, $R^2$, MAPE, RMSE, RRMSE and Pearson, as indicators of prediction goodness, are provided for each of the regression methods, all indicating the improvement obtained by using more sophisticated models. In accordance with the literature (Padhma, 2023) estimations can be classified into four categories based on the RRMSE criterion: excellent (less than 10%), good (between 10% and 20%), acceptable (between 20% and 30%), and unacceptable (greater than 30%). The values of Relative Root Mean Square Error (RRMSE) reported in Appendix C are all between 0 and 2% and therefore the estimations can all be classified as excellent.

For all the regression formulas provided in this paper, the coefficient of determination has been increased maintaining still very simple formulations that are function of a few input values. As such, they are easily applicable to assess the main dimensions and parameters of a container vessel in the early design stage. The provided models are a valid support to designers in finding initial solutions for the design of a modern container vessel.

Based on the obtained results, the forest tree algorithm is the most accurate regression but, being a black-box method, it has the drawback that does not provide any regression formula. Therefore, the multivariable regressions, with higher $R^2$ values and depending on available input parameters, are considered the most recommended. Future works could include other hull parameters and the development of advanced regression models for hydrodynamic performances.

## CRediT authorship contribution statement

**B. Rinauro:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **E. Begovic:** Conceptualization, Data curation, Methodology, Supervision, Writing – original draft, Writing – review & editing. **F. Mauro:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Writing – review & editing. **G. Rosano:** Conceptualization, Data curation, Supervision, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Appendix A

*Simple regressions*

**L-TEU.**
*Power regression:* $L = 17.48756101\ TEU^{0.3168693342}$
*Logarithmic regression:* $L = 84.50646382 \ln(TEU)$- 450.2988469.
*Polynomial regression:* $L = 102.9043417 + 0.04282532441TEU + 2.463170177^{E}\text{-}006\ TEU^{2}\text{-}1.704594192^{E}\text{-}009TEU^{3}$
**B-TEU.**
*Power regression*: $B = 2.821629922\ TEU^{0.3008998037}$
*Logarithmic regression*: $B = 12.03526291 \ln(TEU)$- 64.94743548.
*Polynomial regression:* $B = 13.16232382 + 0.01679801069TEU - 6.668772272E\text{-}006TEU^{2} + 1.425392983E\text{-}009TEU^{3}$
**D-TEU.**
*Power regression*: $D = 1.820019294TEU^{0.2896828926}$
*Logarithmic regression*: $D = 6.442566143 \ln(TEU) – 32.84293138$.
*Polynomial regression:* $D = 10.39593358 + 0.002777397645TEU + 3.841705464E\text{-}007TEU^{2}\ \text{-}2.125296432E\text{-}010TEU^{3}$
**T-TEU.**

*Power regression*: $T = 2.240354683TEU^{0.1961444783}$
*Logarithmic regression*: $T = 2.383434526\ \ln(TEU)$- 8.237213727.
*Polynomial regression:* $T = 5.261912689 + 0.003533447633TEU - 7.092875169\text{E-}007TEU^{2} + 6.726374147\text{E-}011TEU^{3}$
**V-TEU.**
*Power regression:* $V = 13.88703861TEU^{0.05818503954}$
*Logarithmic regression:* $V = 1.221004058\ \ln(TEU)+ 12.526487$.
*Polynomial regression:* $V = 15.31103547 + 0.001681375982TEU + 7.342533865^{E}\text{-}007TEU^{2}\ \text{-}2.619085271^{E}\text{-}010TEU^{3}$
**TEU,P-TEU.**
*Power regression:* $P = 308.359455TEU^{0.5636479558}$
*Logarithmic regression:* $P = 19800.78429\ \ln(TEU)$- 126083.3925.
*Polynomial regression:* $P = 18500.70682–13.07205522TEU + 0.009094797463TEU^{2}\ \text{-}1.663366635^{E}\text{-}006TEU^{3}$

## Appendix B

### *Multivariable regressions*

Four types of multivariable regressions have been determined with the following general formulation.

•MR1: Y = f (V, TEU) for L, B, D, T, $P_{ENG}$;
•MR2: Y = f (L, D, TEU) for V and $P_{ENG}$;
•MR3: Y = f (L, T, TEU) for V and $P_{ENG}$;
•MR4: Y = f (L, B, D, TEU) for V and $P_{ENG}$.

An example of application is given in Section 8.

Multivariable regressions

| MR1 – L = f (V, TEU) | | | | |
|---|---|---|---|---|
| Variables | Estimate | SE | t-stud | p-value |
| intercept | −7020.6859 | 3253.714 | −2.15775 | 0.0319 |
| V | 1476.2432 | 604.2363 | 2.443155 | 0.015252 |
| TEU | 0.1056 | 0.030347 | 3.48038 | 0.000591 |
| $V^2$ | −113.4121 | 41.80004 | −2.7132 | 0.007127 |
| TEU V | −0.0072 | 0.002646 | −2.71232 | 0.007145 |
| $TEU^2$ | −3.26E-07 | 2.68E-08 | −12.1813 | 4.16E-27 |
| $V^3$ | 3.8307 | 1.2771 | 2.999526 | 0.002977 |
| $V^2$ TEU | 0.0001 | 5.73E-05 | 2.511556 | 0.012651 |
| $V^4$ | −0.0477 | 0.014546 | −3.28268 | 0.001175 |
| **MR1 – B = f (V, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| intercept | −380.6165 | 82.09207 | −4.63646 | 5.69E-06 |
| V | 56.2597 | 11.28242 | 4.986494 | 1.15E-06 |
| TEU | 0.0020 | 0.000357 | 5.478681 | 1.04E-07 |
| $V^2$ | −2.5729 | 0.511914 | −5.02604 | 9.50E-07 |
| TEU V | 2.86E-05 | 1.59E-05 | 1.798981 | 0.073218 |
| $TEU^2$ | −4.40E-08 | 4.31E-09 | −10.2081 | 1.06E-20 |
| $V^3$ | 0.0387 | 0.007675 | 5.043116 | 8.76E-07 |
| **MR1 – D = f (V, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| intercept | −316.7987 | 95.69022 | −3.31067 | 0.001068 |
| V | 42.6892 | 13.22484 | 3.227956 | 0.001414 |
| TEU | 0.1637 | 0.032708 | 5.00569 | 1.05E-06 |
| $V^2$ | −1.8414 | 0.606432 | −3.03646 | 0.002647 |
| TEU V | −0.0206 | 0.004221 | −4.87796 | 1.91E-06 |
| $TEU^2$ | −3.86E-08 | 3.20E-09 | −12.0874 | 8.56E-27 |
| $V^3$ | 0.0265 | 0.009224 | 2.877346 | 0.004357 |
| $V^2$ TEU | 0.0009 | 0.000181 | 4.791969 | 2.83E-06 |
| $V^3$ TEU | −1.22E-05 | 2.59E-06 | −4.69684 | 4.36E-06 |
| **MR1 – T = f (V, TEU)** | | | | |
| **Variables** | **coefficients** | **SE** | **t-stud** | **p-value** |
| intercept | −9.0068 | 4.122446 | −2.18482 | 0.029817 |
| V | 1.5253 | 0.371184 | 4.107214 | 5.40E-05 |
| TEU | 0.0004 | 3.72E-05 | 10.52903 | 9.50E-22 |
| V2 | −0.0303 | 0.008222 | −3.68667 | 0.000278 |
| TEU2 | −4.8235E-09 | 1.59E-09 | −3.04312 | 0.002587 |
| **MR1 – $P_{ENG}$ = f (V, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| intercept | −3826801.116 | 1765367 | −2.16771 | 0.031117 |
| V | 754690.6349 | 327195.8 | 2.306541 | 0.021892 |
| TEU | 4.8726 | 0.3386 | 14.39042 | 1.12E-34 |
| $V^2$ | −55037.79937 | 22588.41 | −2.43655 | 0.015521 |
| $TEU^2$ | −0.0001 | 1.44E-05 | −8.80295 | 2.18E-16 |

(*continued on next page*)

(*continued*)

| MR1 – L = f (V, TEU) | | | | |
|---|---|---|---|---|
| Variables | Estimate | SE | t-stud | p-value |
| $V^3$ | 1758.773976 | 688.6039 | 2.554116 | 0.011235 |
| $V^4$ | −20.7193 | 7.823848 | −2.64823 | 0.008602 |
| **MR2 – V = f (L, D, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| Intercept | 24.56265605 | 13.71708 | 1.790662 | 0.074603 |
| L | −0.540577487 | 0.242591 | −2.22835 | 0.02678 |
| D | 4.111193311 | 2.598558 | 1.582106 | 0.114937 |
| TEU | 0.010813276 | 0.006204 | 1.743074 | 0.082596 |
| Lˆ2 | 0.003648354 | 0.001006 | 3.626227 | 0.000351 |
| LD | −0.005247444 | 0.010848 | −0.48372 | 0.629023 |
| Dˆ2 | −0.312480065 | 0.134968 | −2.31521 | 0.021442 |
| LTEU | −0.000275754 | 5.90E-05 | −4.67148 | 4.97E-06 |
| DTEU | 0.002420124 | 0.00073 | 3.316304 | 0.001053 |
| TEUˆ2 | −2.97E-07 | 1.05E-07 | −2.81271 | 0.005317 |
| Lˆ2D | −0.000194496 | 5.22E-05 | −3.72391 | 0.000244 |
| LDˆ2 | 0.002162734 | 0.0006 | 3.603479 | 0.000381 |
| Lˆ2TEU | 5.72E-07 | 1.28E-07 | 4.473356 | 1.19E-05 |
| LDTEU | 4.50E-06 | 1.71E-06 | 2.627049 | 0.009165 |
| Dˆ2TEU | −0.000109895 | 2.63E-05 | −4.18328 | 4.03E-05 |
| DTEUˆ2 | 9.85E-09 | 3.53E-09 | 2.793309 | 0.005636 |
| Lˆ2DTEU | −2.05E-08 | 5.12E-09 | −3.99291 | 8.67E-05 |
| LDˆ2TEU | 2.19E-07 | 5.95E-08 | 3.679987 | 0.000288 |
| **MR2 – $P_{ENG}$ = f (L, D, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| Intercept | −51214.97917 | 82331.06 | −0.62206 | 0.534485 |
| L | −174.4098865 | 1345.627 | −0.12961 | 0.89698 |
| D | 18961.94925 | 11059.41 | 1.714553 | 0.087702 |
| TEU | −53.94155239 | 16.32385 | −3.30446 | 0.001095 |
| Lˆ2 | 15.30319721 | 7.224104 | 2.118352 | 0.035161 |
| LD | −384.3496449 | 1.35E+02 | −2.85021 | 4.74E-03 |
| Dˆ2 | 1008.661712 | 880.2562 | 1.145873 | 0.252975 |
| LTEU | 2.61E-01 | 9.26E-02 | 2.817848 | 0.005232 |
| DTEU | 3.498457681 | 1.25E+00 | 2.807537 | 0.005397 |
| TEUˆ2 | −0.003742477 | 0.000645 | −5.80406 | 2.01E-08 |
| Lˆ2D | −9.83E-01 | 3.96E-01 | −2.48016 | 1.38E-02 |
| LDˆ2 | 2.43E+01 | 7.46E+00 | 3.264731 | 0.001253 |
| Dˆ3 | −114.4079426 | 3.86E+01 | −2.96259 | 3.35E-03 |
| Lˆ2TEU | 5.46E-04 | 2.15E-04 | 2.536243 | 0.011832 |
| LDTEU | −2.11E-02 | 4.61E-03 | −4.57167 | 7.71E-06 |
| DTEUˆ2 | 1.34E-04 | 2.13E-05 | 6.308056 | 1.33E-09 |
| **MR3 – V = f (L, T, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| Intercept | −58.8249432 | 31.52244 | −1.86613 | 0.063238 |
| L | −0.959180192 | 0.260014 | −3.68896 | 0.000278 |
| T | 41.75581086 | 11.32448 | 3.687216 | 0.00028 |
| TEU | 0.00166132 | 0.012486 | 0.133056 | 0.89426 |
| Lˆ2 | 0.004048978 | 0.00085 | 4.765167 | 3.26E-06 |
| LT | 0.071119467 | 0.023257 | 3.057944 | 0.00248 |
| Tˆ2 | −5.38304808 | 1.291151 | −4.16918 | 4.27E-05 |
| LTEU | −0.000354577 | 7.94E-05 | −4.46427 | 1.23E-05 |
| TTEU | 0.007613745 | 0.003447 | 2.208876 | 0.028125 |
| TEUˆ2 | −6.67E-07 | 1.34E-07 | −4.98963 | 1.16E-06 |
| Lˆ2T | −0.000315294 | 7.03E-05 | −4.48815 | 1.11E-05 |
| Tˆ3 | 0.208585136 | 0.04628 | 4.507024 | 1.03E-05 |
| Lˆ2TEU | 3.05E-07 | 8.37E-08 | 3.644103 | 0.000329 |
| LT TEU | 3.02E-05 | 6.86E-06 | 4.408142 | 1.57E-05 |
| Tˆ2TEU | −0.000757271 | 0.000281 | −2.69655 | 0.0075 |
| T TEUˆ2 | 3.82E-08 | 8.11E-09 | 4.709469 | 4.20E-06 |
| Lˆ2TTEU | −2.67E-08 | 6.87E-09 | −3.88313 | 0.000133 |
| Tˆ3TEU | 9.94E-06 | 5.35E-06 | 1.857482 | 0.064462 |
| **MR3 – $P_{ENG}$ = f (L, T, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| Intercept | −300595.5842 | 355187.2 | −0.8463 | 0.398235 |
| L | 3818.39297 | 3310.125 | 1.15355 | 0.249842 |
| T | 11491.62815 | 120369 | 0.09547 | 0.924022 |
| TEU | −150.2089654 | 56.79532 | −2.64474 | 0.008719 |
| Lˆ2 | −67.06777786 | 29.72797 | −2.25605 | 0.024976 |
| LT | 1476.089831 | 1197.314 | 1.232834 | 0.218854 |
| Tˆ2 | −7451.858834 | 10789.45 | −0.69066 | 0.490451 |
| LTEU | −0.262639218 | 0.240779 | −1.09079 | 0.276468 |
| TTEU | 34.66496327 | 9.786817 | 3.542006 | 0.000478 |
| TEUˆ2 | −0.006746098 | 0.000995 | −6.77854 | 9.48E-11 |
| Lˆ3 | −0.098690935 | 0.088331 | −1.11728 | 0.265 |
| Lˆ2T | 17.96549814 | 5.849477 | 3.0713 | 0.002379 |

(*continued on next page*)

(*continued*)

| MR1 – L = f (V, TEU) | | | | |
|---|---|---|---|---|
| Variables | Estimate | SE | t-stud | p-value |
| LTˆ2 | −464.1148458 | 163.5258 | −2.83817 | 0.004929 |
| Tˆ3 | 2490.556459 | 681.3501 | 3.655326 | 0.000316 |
| Lˆ2TEU | −0.000935618 | 0.000567 | −1.65142 | 0.099972 |
| LTTEU | 0.074302192 | 0.037586 | 1.976834 | 0.049215 |
| Tˆ2TEU | −2.402734551 | 0.655566 | −3.66513 | 0.000305 |
| T TEUˆ2 | 0.000411303 | 6.74E-05 | 6.102632 | 4.20E-09 |
| Lˆ4 | 0.00059148 | 0.000249 | 2.379707 | 0.018115 |
| Lˆ3Tˆ4 | −0.044457043 | 0.016557 | −2.68511 | 0.007761 |
| Lˆ2Tˆ2 | 0.725949793 | 0.29578 | 2.454357 | 0.014831 |
| **MR4 – V = f (L, B, D, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| intercept | −123.6176 | 70.49079 | −1.75367 | 0.080837 |
| L | 0.4986 | 0.490953 | 1.015516 | 0.310941 |
| B | 17.2011 | 7.704167 | 2.232698 | 0.026545 |
| D | 3.8256 | 3.907033 | 0.97916 | 0.328543 |
| TEU | 0.0383 | 0.009429 | 4.056885 | 6.84E-05 |
| $L^2$ | 0.0066 | 0.001874 | 3.533402 | 0.000497 |
| L B | −0.1919 | 0.052607 | −3.64778 | 0.000328 |
| $B^2$ | −0.6788 | 0.212604 | −3.19288 | 0.001608 |
| L D | 0.0226 | 0.034194 | 0.660233 | 0.509774 |
| B D | 0.9622 | 0.300336 | 3.203726 | 0.001551 |
| $D^2$ | −1.1840 | 0.341681 | −3.46517 | 0.000633 |
| L TEU | −9.84E-05 | 7.61E-05 | −1.29271 | 0.197425 |
| B TEU | −0.0008 | 0.000626 | −1.32782 | 0.185572 |
| D TEU | 0.0001 | 0.00113 | 0.099364 | 0.920937 |
| $TEU^2$ | 2.42E-06 | 4.38E-07 | 5.526168 | 8.95E-08 |
| $L^3$ | −1.01E-05 | 2.87E-06 | −3.53184 | 0.0005 |
| L $B^2$ | 0.0057 | 0.001347 | 4.265583 | 2.93E-05 |
| $B^3$ | 0.0039 | 0.001887 | 2.077272 | 0.038901 |
| L $D^2$ | −0.0003 | 0.000872 | −0.31882 | 0.750156 |
| B $D^2$ | −0.0168 | 0.006593 | −2.5482 | 0.011489 |
| $D^3$ | 0.0253 | 0.009126 | 2.770462 | 0.006061 |
| $L^2$ TEU | 3.82E-07 | 8.94E-08 | 4.273276 | 2.84E-05 |
| $B^2$ TEU | 2.82E-05 | 1.52E-05 | 1.860976 | 0.06404 |
| L D TEU | −1.15E-05 | 5.99E-06 | −1.92552 | 0.055413 |
| B D TEU | −0.0001 | 3.89E-05 | −3.72129 | 0.00025 |
| $D^2$ TEU | 0.0002 | 6.74E-05 | 3.003014 | 0.002973 |
| B $TEU^2$ | −5.04E-08 | 8.75E-09 | −5.75894 | 2.73E-08 |
| L $B^3$ | −5.28E-05 | 1.14E-05 | −4.6365 | 5.98E-06 |
| $B^3$ TEU | 2.15E-07 | 1.20E-07 | 1.795408 | 0.073919 |
| L $D^2$ TEU | 2.00E-07 | 9.53E-08 | 2.101558 | 0.036695 |
| B $D^2$ TEU | 2.56E-06 | 6.64E-07 | 3.852357 | 0.000152 |
| $D^3$ TEU | −4.80E-06 | 1.19E-06 | −4.03768 | 7.38E-05 |
| **MR4 – $P_{ENG}$ = f (L, B, D, TEU)** | | | | |
| **Variables** | **Estimate** | **SE** | **t-stud** | **p-value** |
| intercept | −1128750.7849 | 533838.7 | −2.1144 | 0.035572 |
| L | 9643.1847 | 3709.667 | 2.599474 | 0.009949 |
| B | 116584.3452 | 57579.52 | 2.024754 | 0.044064 |
| D | −19552.1683 | 27011.72 | −0.72384 | 0.469909 |
| TEU | 66.9804 | 72.46794 | 0.924276 | 0.356324 |
| $L^2$ | 29.1442 | 17.34413 | 1.680347 | 0.094265 |
| L B | −1309.8360 | 383.5003 | −3.41548 | 0.000754 |
| $B^2$ | −3268.4354 | 1830.299 | −1.78574 | 0.075476 |
| L D | −195.6562 | 201.0299 | −0.97327 | 0.331455 |
| B D | 3652.5626 | 1797.192 | 2.032372 | 0.04328 |
| $D^2$ | −867.6640 | 2068.029 | −0.41956 | 0.675203 |
| L TEU | 0.8854 | 0.420075 | 2.107792 | 0.036146 |
| B TEU | −4.7628 | 4.488496 | −1.06112 | 0.289763 |
| D TEU | −0.6786 | 7.904148 | −0.08586 | 0.931656 |
| $TEU^2$ | 0.0136 | 0.003545 | 3.827717 | 0.000167 |
| $L^3$ | −0.0520 | 0.030077 | −1.72831 | 0.085293 |
| L $B^2$ | 33.4326 | 10.44596 | 3.200535 | 0.001568 |
| $B^3$ | 20.3311 | 23.09544 | 0.880307 | 0.379624 |
| L B D | 12.8520 | 7.248143 | 1.77314 | 0.077546 |
| B $D^2$ | −147.0299 | 61.44205 | −2.39299 | 0.017526 |
| $D^3$ | 60.3298 | 42.25664 | 1.4277 | 0.154753 |
| $L^2$ TEU | 0.0026 | 0.001273 | 2.033401 | 0.043176 |
| L B TEU | −0.0357 | 0.0178 | −2.00513 | 0.046136 |
| $B^2$ TEU | 0.1593 | 0.107118 | 1.486729 | 0.138474 |
| L D TEU | −0.0967 | 0.026393 | −3.66399 | 0.000309 |
| B D TEU | −0.4317 | 0.130252 | −3.31411 | 0.00107 |
| $D^2$ TEU | 0.9906 | 0.450182 | 2.200445 | 0.028783 |
| B $TEU^2$ | −0.0003 | 7.30E-05 | −3.76158 | 0.000215 |
| L $B^3$ | −0.3174 | 0.095704 | −3.31681 | 0.00106 |

(*continued on next page*)

(*continued*)

| MR1 – L = f (V, TEU) | | | | |
|---|---|---|---|---|
| Variables | Estimate | SE | t-stud | p-value |
| B$^3$ TEU | 0.0014 | 0.000722 | 1.881053 | 0.061245 |
| L B D TEU | 0.0012 | 0.000322 | 3.721751 | 0.00025 |
| D$^3$ TEU | −0.0087 | 0.004327 | −2.00619 | 0.046023 |

## Appendix C

In this Appendix the coefficients of determination for simple and multivariable regressions are compared.

**Table C1**
Coefficients of determination R$^2$ for simple regressions

| **Simple regressions** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Dependent variable** | ***L = f(TEU)*** | ***B = f(TEU)*** | ***D = f(TEU)*** | ***T = f(TEU)*** | ***V = f(TEU)*** | ***P = f(TEU)*** |
| **Type of regression** | **forest tree** | 0.97940 | 0.99180 | 0.95520 | 0.92520 | 0.87330 | 0.89210 |
| | **power** | 0.90740 | 0.93850 | 0.88930 | 0.84770 | 0.15920 | 0.54480 |
| | **logarithm** | 0.94826 | 0.89373 | 0.89285 | 0.84289 | 0.17299 | 0.66658 |
| | **polynomial 1st order** | 0.81945 | 0.92987 | 0.78482 | 0.78220 | 0.02250 | 0.43856 |
| | **polynomial 2nd order** | 0.91662 | 0.95113 | 0.89805 | 0.81818 | 0.27027 | 0.65632 |
| | **polynomial 3$^{rd}$ order** | 0.94203 | 0.95250 | 0.89901 | 0.83870 | 0.61717 | 0.77400 |

**Table C2**
Comparison of goodness of fit coefficients for the simple regressions

| **Simple** | **L** | **B** | **D** | **T** | **V** | **P** |
|---|---|---|---|---|---|---|
| **R$^2$** | 0.948 | 0.939 | 0.889 | 0.848 | 0.159 | 0.545 |
| **MAPE** | 4.49% | 5.12% | 5.63% | 4.89% | 7.66% | 24.51% |
| **RMSE** | 14.75 | 2.42 | 1.70 | 0.76 | 2.02 | 13338.88 |
| **RRMSE** | 3.13E-03 | 3.71E-03 | 4.46E-03 | 3.73E-03 | 5.43E-03 | 1.69E-02 |
| **Pearson** | 0.974 | 0.969 | 0.943 | 0.921 | 0.399 | 0.738 |

**Table C3**
Comparison of goodness of fit coefficients for multivariable regressions

| **Multivariable** | **L MR1** | **B MR1** | **D MR1** | **T MR1** | **V** | | | **P** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **MR2** | **MR3** | **MR4** | **MR1** | **MR2** | **MR3** | **MR4** |
| **R$^2$** | 0.964 | 0.956 | 0.919 | 0.850 | 0.758 | 0.768 | 0.828 | 0.859 | 0.855 | 0.858 | 0.881 |
| **R adj** | 0.963 | 0.955 | 0.92 | 0.85 | 0.74 | 0.75 | 0.80 | 0.86 | 0.85 | 0.85 | 0.86 |
| **MAPE** | 3.78% | 4.40% | 5.00% | 4.77% | 3.78% | 3.73% | 3.17% | 13.14% | 12.62% | 12.61% | 11.73% |
| **RMSE** | 12.23 | 2.00 | 1.45 | 0.75 | 1.08 | 1.06 | 0.91 | 6803.67 | 6912.56 | 6840.10 | 6260.56 |
| **RRMSE** | 2.60E-03 | 3.04E-03 | 3.80E-03 | 3.70E-03 | 2.89E-03 | 2.83E-03 | 2.43E-03 | 8.59E-03 | 8.73E-03 | 8.63E-03 | 7.89E-03 |
| **Pearson** | 0.982 | 0.978 | 0.958 | 0.922 | 0.871 | 0.876 | 0.910 | 0.927 | 0.924 | 0.926 | 0.938 |

**Table C4**
Comparison of goodness of fit coefficients for forest tree

| **Forest Tree** | **L** | | **B** | | **D** | | **T** | |
|---|---|---|---|---|---|---|---|---|
| | **SR** | **MR1** | **SR** | **MR1** | **SR** | **MR1** | **SR** | **MR1** |
| **R$^2$** | 0.979 | 0.986 | 0.992 | 0.998 | 0.955 | 0.942 | 0.925 | 0.944 |
| **MAPE** | 2.41% | 2.23% | 1.42% | 0.74% | 3.02% | 3.81% | 3.06% | 2.70% |
| **RMSE** | 9.32 | 7.67 | 0.86 | 0.44 | 1.08 | 1.22 | 0.53 | 0.46 |
| **RRMSE** | 1.98E-03 | 1.63E-03 | 1.31E-03 | 6.70E-04 | 2.83E-03 | 3.21E-03 | 2.61E-03 | 2.26E-03 |
| **Pearson** | 0.990 | 0.993 | 0.996 | 0.999 | 0.977 | 0.971 | 0.962 | 0.973 |

| Forest Tree | V | | | | P | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | SR | MR2 | MR3 | MR2 | SR | MR1 | MR2 | MR3 | MR4 |
| **$R^2$** | 0.873 | 0.956 | 0.940 | 0.928 | 0.892 | 0.950 | 0.938 | 0.936 | 0.928 |
| **MAPE** | 2.68% | 1.28% | 1.74% | 1.97% | 10.60% | 7.17% | 7.47% | 7.59% | 5.31% |
| **RMSE** | 0.78 | 0.46 | 0.54 | 0.59 | 5954.69 | 4049.19 | 4498.59 | 4586.14 | 3611.75 |
| **RRMSE** | 2.09E-03 | 1.24E-03 | 1.44E-03 | 1.57E-03 | 7.63E-03 | 5.10E-03 | 5.67E-03 | 5.77E-03 | 4.53E-03 |
| **Pearson** | 0.935 | 0.978 | 0.970 | 0.964 | 0.946 | 0.975 | 0.969 | 0.968 | 0.980 |

**Table C5**
VIF values for multiple regressions

| MR1 | | MR2 | | MR3 | | MR4 | |
|---|---|---|---|---|---|---|---|
| | VIF | | VIF | | VIF | | VIF |
| **V** | 1.023 | **L** | 10.4526 | **L** | 7.5409 | **L** | 10.5015 |
| **TEU** | 1.023 | **D** | 8.7705 | **T** | 6.2513 | **B** | 17.6318 |
| | | **TEU** | 5.9191 | **TEU** | 6.4042 | **D** | 10.1556 |
| | | | | | | **TEU** | 16.2022 |
| No multicollinearity | | Moderate multicollinearity for L | | Almost no collinearity | | Presence of strong multicollinearity for B and TEU | |

**Table C6**
Breush-Pagan test results for multiple regressions

| Regression | T | df | p-value | Heteroscedasticity |
|---|---|---|---|---|
| **Length MR1** | 14.432 | 2 | 0.00073474 | YES |
| **Breadth MR1** | 3.4136 | 2 | 0.1814463 | NO |
| **Depth MR1** | 29.2087 | 2 | 4.54E-07 | YES |
| **Draught MR1** | 25.6924 | 2 | 2.63E-06 | YES |
| **Power MR1** | 26.9584 | 2 | 1.40E-06 | YES |
| **Velocity MR2** | 14.1158 | 4 | 6.90E-03 | YES |
| **Power MR2** | 36.0808 | 4 | 2.79E-07 | YES |
| **Velocity MR3** | 13.1311 | 3 | 4.36E-03 | YES |
| **Power MR3** | 31.6397 | 3 | 6.23E-07 | YES |
| **Velocity MR4** | 3.7615 | 3 | 0.28839939 | NO |
| **Power MR4** | 23.8087 | 3 | 2.74E-05 | YES |

## Appendix D

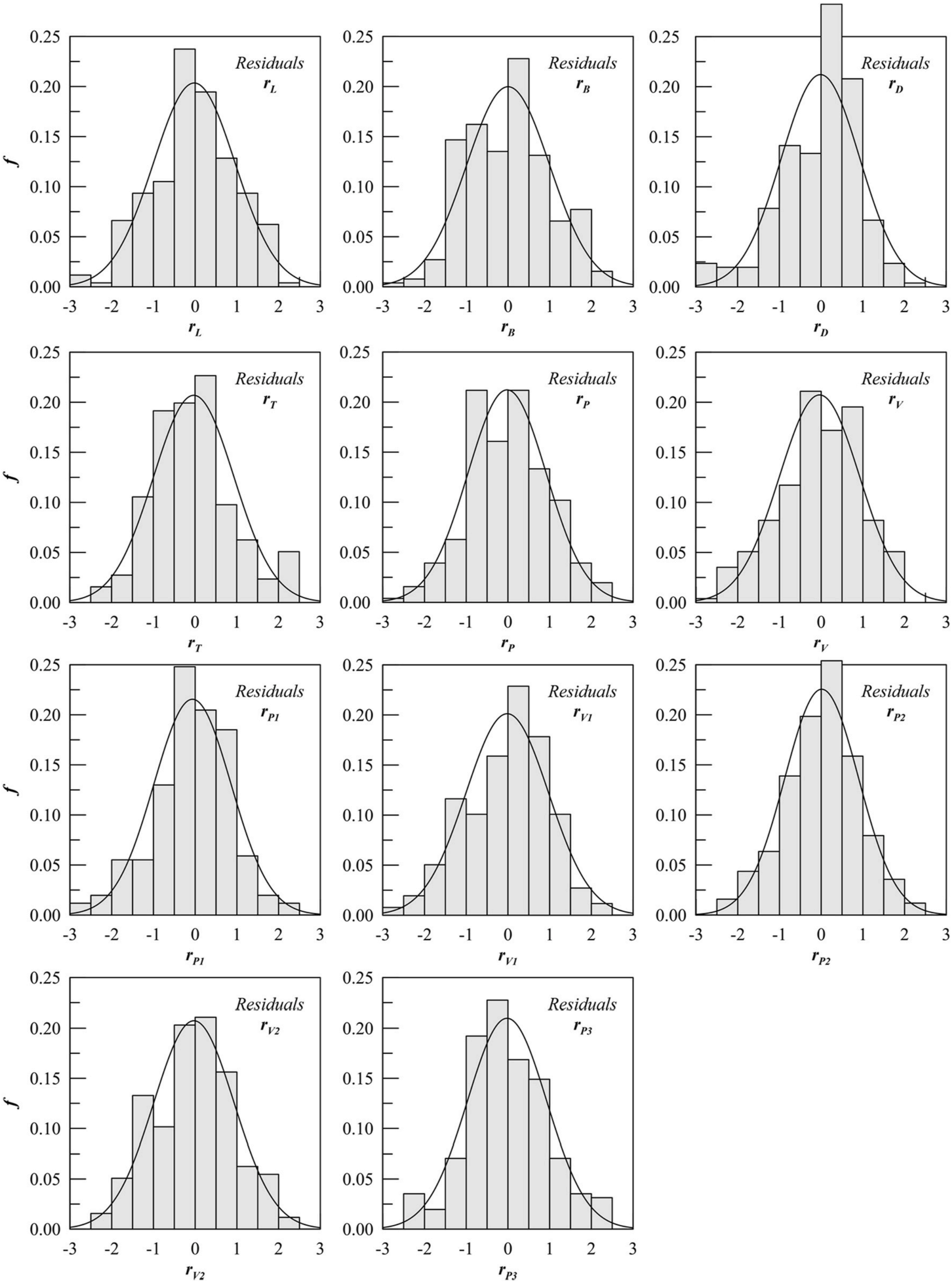


**Fig. D1.** Normality of multiple linear regression residuals.

## References

Abramowski, T., Cepowski, T., Zvolensky, P., 2018. Determination of regression formulas for key design characteristics of container ships at preliminary design stage. New trends in production engineering 1, 247–257. https://doi.org/10.2478/ntpe-2018-0031.

Alwosheel, A., van Cranenburgh, S., Chorus, C.G., 2018. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. J. Choice Model. 28, 167–182. https://doi.org/10.1016/j.jocm.2018.07.002.

Begovic, E., Bertorello, C., Rinauro, B., Rosano, G., 2023. Simplified operational guidance for second generation intact stability criteria. Ocean Eng. 270, 113583 https://doi.org/10.1016/J.OCEANENG.2022.113583.

Cepowski, T., Chorab, P., 2021. Determination of design formulas for container ships at the preliminary design stage using artificial neural network and multiple nonlinear regression. Ocean Eng. 238, 109727 https://doi.org/10.1016/j.oceaneng.2021.109727.

Clausen, H.B., Lutzen, M., Friis-Hansen, A., Bjørneboe, N., 2001. Bayesian and neural networks for preliminary ship design. Mar. Technol. 38 (4), 268–277. https://doi.org/10.5957/mt1.2001.38.4.268. SNAME N.

Ekinci, S., Celebi, U.B., Bal, M., Amasyali, M.F., Boyaci, U.K., 2011. Predictions of oil/chemical tanker main design parameters using computational intelligence techniques. Appl. Soft Comput. 11, 2356–2366.

France, W.N., Levadou, M., Treakle, T.W., Paulling, J.R., Michel, R.K., Moor, C., 2003. An investigation of head-sea parametric rolling and its influence on container lashing systems. SNAME Annual Meeting 2001 Presentation. https://doi.org/10.5957/mt1.2003.40.1.1.

Galeazzi, R., Blanke, M., Poulsen, N.K., 2013. Early detection of parametric roll resonance on container ships. IEEE Trans. Control Syst. Technol. 21, 489–503. https://doi.org/10.1109/TCST.2012.2189399.

Garrido, J., Sauri, S., Marrero, A., Gul, U., Rua, C., 2020. Predicting the future capacity and dimensions of container ships. Transport. Res. Rec. 2674 (9), 177–190. https://doi.org/10.1177/0361198120927395.

Grubisic, I., Begovic, E., 2001. Multi-attribute concept design model of the Adriatic type of fishing vessel. Brodogradnja 49 (1), 39–54.

Grubisic, I., Begovic, E., 2011. Reliability of attribute prediction in small craft concept design. In: Sustainable Maritime Transportation and Exploitation of Sea Resources – Proceedings of the 14th International Congress of the International Maritime Association of the Mediterranean. IMAM.

Gurgen, S., Altin, I., Ozkok, M., 2018. *Prediction Of Main Particulars of a Chemical Tanker at Preliminary Ship Design Using Artificial Neural Network*, Ships and Offshore Structures.

HHI, 2022. shipbuilding group performance record by Hyundai Industry working group web site: https://english.hhi.co.kr/img/filedown/HHI_PerformanceRecord_220511.pdf.

Ho, K.T., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20 (8), 832–844.

IMO SLF 54/INF.6, 2011. Theoretical Investigation into the Loss of Containers of the Pacific Adventurer off Cape Morton Queensland.

Islam, M.M., Reaz, M., Khondoker, H., Rahman, C.M., 2001. Application of artificial intelligence techniques in automatic hull form generation. Ocean Eng. 28, 1531–1544.

ISO 668, 1968. Series 1 Freight Containers; Classification, Dimensions and Ratings.

Kalokairinos, E., Mavroeidis, T., Radou, G., Zachariou, Z., 2000-2005. *Regression Analysis of Basic Ship Design Values for Merchant Ships*, Diploma Theses. National Technical University of Athens.

Korea Maritime Institute, 2012. Report on Technology, Development of Smart Green Container Terminal. Ministry Land and Ocean, Busan.

Kristensen, H.O., 2012. *Statistical Analysis And Determination Of Regression Formulas For Main Dimensions Of Container Ships Based On IHS Fairplay Data. Project No. 2010-56*, Emissionsbeslutningsstøttesystem, Work Package 2, Report No. 03. Technical University of Denmark, Odense.

Ljulj, A., Slapnicar, V., Grubisic, I., 2020. Multi-attribute concept design procedure of a generic naval vessel. Alex. Eng. J. 59 (3), 1725–1734. https://doi.org/10.1016/j.aej.2020.04.038.

Majnaric, D., Segota, S.B., Lorencin, I., Car, Z., 2022. Prediction of main particulars of container ships using artificial intelligence algorithms. Ocean Eng. 265, 112571 https://doi.org/10.1016/j.oceaneng.2022.112571.

Malchow, U., 2017. Growth in Container Ship Sizes to Be Stopped. ISSN: 2397-3757. Maritime Business Review.

Mauro, F., Braidotti, L., Trincas, G., 2019. Determination of an optimal fleet for CNG transportation scenario in the Mediterranean Sea. Brodogradnja 70 (Issue 3), 1–23.

Padhma, M., 2023. A Comprehensive Introduction To Evaluating Regression Models. Data Science Blogathon updated On October 31st.

Papanikolaou, A.D., 2014. Ship Design, Methodologies Of Preliminary Design. Springer. https://doi.org/10.1007/978-94-017-8751-2.

Park, N.K., Suh, S.C., 2019. Tendency toward mega container ships and the constraints of container terminals. J. Mar. Sci. Eng. 7 (No. 5), 131. https://doi.org/10.3390/jmse7050131.

Romero-Tello, P., Gutiérrez-Romero, J.E., Serván-Camas, B., 2022. Prediction of seakeeping in the early stage of conventional monohull vessels design using artificial neural network. J. Ocean Eng. Sci. https://doi.org/10.1016/j.joes.2022.06.033.

Saxon, S., Stone, M., 2017. Container Shipping: the Next 50 Years, Travel, Transport & Logistics. Mc Kinsey & Company.

Trincas, G., Žanić, V., Grubišić, I., 1994. Comprehensive Concept of Fast Ro-Ro Ships by Multiattribute DecisionMaking, IMDC'94. Proceedings of 5th International Marine Design Conference, Delft.

Uyanık, T., Karatuğ, C., Arslanoğlu, Y., 2020. Machine learning approach to ship fuel consumption: a case of container vessel. Transport. Res. Part D 84, 102389. https://doi.org/10.1016/j.trd.2020.102389.

World Shipping Council, 2023. Containers Lost at Sea – 2023 Update.

Yao, Y., Yang, Y., Wnag, Y., Zhao, X., 2019. Artificial intelligence-based hull structural plate corrosion damage detection and recognition using convolutional neural network. Appl. Ocean Res. 90 (1), 101823 https://doi.org/10.1016/j.apor.2019.05.008.

Žanić, V., Grubišić, I., Trincas, G., 1992. Multiattribute decision making system based on random generation of nondominated solutions: an application to fishing vessel design. In: Proceedings of PRAD, 5th Intl Symp on the Practical Design of Ships and Mobile Units, pp. 17–22. May 1992; Newcastle upon Tyne, U.K. Vol 2, p 2.1443.