UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Ingeniería

https://revistas.udistrital.edu.co/index.php/reving/issue/view/1320

DOI: https://doi.org/10.14483/23448393.22627

Research

Regression Analysis of Principal Ship Attributes for Purse Seiners in the Concept Design Stage

Análisis de regresión de los atributos principales de barcos pesqueros de cerco en la etapa de diseño conceptual

Dennys De La Torre[1], and Dheybi Cervan[2]

[1]U1Grupo de Investigación de Transporte Multidisciplinario, Universidad Nacional de Ingeniería, Av. Túpac Amaru 210 – Lima, Perú
[2]Facultad de Ingeniería Mecánica, Av. Túpac Amaru 210 – Lima, Perú

Abstract

Context: In the ship concept design stage, naval architects must estimate the principal attributes of a vessel before the final definition of hull proportions and detailed arrangements. These estimates are essential for assessing feasibility, performance, and operational capability. However, reliable vessel-specific predictive tools are often unavailable for certain vessel categories.
Method: A database of 130 American-type semi-industrial purse with RSW systems was employed to develop practical regression-based power-law models for estimating key ship attributes directly from principal hull dimensions.
Results: The results show that the main volumetric and power attributes exhibit strong and consistent scaling relationships with the principal dimensions, yielding moderate to high determination coefficients, while the fuel and water capacities report a weak geometric dependence. The proposed regression equations demonstrate improved predictive performance when compared to existing formulations reported in the literature for semi-industrial purse seiners.
Conclusions: This work provides a set of vessel-specific, empirically grounded regression tools tailored to American-type semi-industrial purse seiners, which can be directly applied during the ship concept design stage to support rapid and informed decision-making. This study complements previous analyses focused on hull geometric ratios by addressing a distinct and application-oriented design problem that deals with attribute estimation rather than geometric proportion selection.

Keywords: ship dimensions, fishing vessels, hull statistics, ship concept design stage, regression analysis

Article history

Received:
October 28th, 2024

Modified:
October 28th, 2025

Accepted:
December 22nd, 2025

Ing, vol. 31, no. 1, 2026, e22627

* Correspondence: ddelatorrec@uni.edu.pe

Resumen

Contexto: En la etapa de diseño conceptual del buque, los arquitectos navales deben estimar los atributos principales de una embarcación antes de la definición final de las proporciones del casco y de realizar arreglos detallados. Estas estimaciones son esenciales para evaluar la factibilidad, el desempeño y la capacidad operativa. Sin embargo, no se suele disponer de herramientas predictivas confiables y específicas para determinadas categorías de embarcaciones.
Método: Se empleó una base de datos de 130 embarcaciones cerqueras semiindustriales de tipo americano con sistemas RSW para desarrollar modelos prácticos basados en regresión de tipo ley de potencia, a fin de estimar los atributos clave del buque directamente de las dimensiones principales del casco.
Resultados: Los resultados muestran que los principales atributos volumétricos y de potencia presentan relaciones de escalamiento fuertes y consistentes con las dimensiones principales, con coeficientes de determinación moderados a altos, mientras que las capacidades de combustible y agua reportan una débil dependencia geométrica. Las ecuaciones de regresión propuestas demuestran un mejor desempeño predictivo en comparación con las formulaciones existentes que figuran en la literatura para cerqueros semiindustriales.
Conclusiones: Este trabajo proporciona un conjunto de herramientas de regresión específicas para este tipo de embarcaciones, las cuales han sido empíricamente fundamentadas y adaptadas a los cerqueros semiindustriales tipo americano, y pueden aplicarse directamente durante la etapa de diseño conceptual del buque para una toma de decisiones rápida e informada. Al abordar un problema de diseño distinto y orientado a la aplicación, este estudio complementa análisis previos que emplean las relaciones geométricas del casco, centrándose en la estimación de atributos más que en la selección de proporciones geométricas.

Palabras clave: dimensiones del buque, embarcaciones pesqueras, estadística del casco, etapa de diseño conceptual del buque, análisis de regresión

Contents

1. Introduction

Ship concept design documents present typical data for various vessel types, in addition to defining their main linear dimensions (length, breadth, and depth) as functions of owner-specified attributes—*e.g.*, the fish hold capacity for fishing vessels. Hull statistics provide meaningful estimated regression equations to predict ship attributes as a function of linear ship dimensions. Investigations on hull statistics have been a prominent research topic in recent years.

In the literature, [1] analyzed a database of 260 non-sister container ships built between 1979 and 2022. They developed simple regression formulas and compared them against existing equivalents in the literature, demonstrating superior trend approximation. This study updated a set of simple and multivariable regression models, achieving a higher goodness of fit than previous works. These formulas were designed for ease of use, allowing designers to apply them directly in ship concept design and multi-attribute design procedures.

[2] utilized an open dataset of 54 in-service cruise ships delivered since 2000. Their study demonstrated the application of supervised statistical learning methods to analyze key cruise ship characteristics. Furthermore, it established a practical formula to support naval architects in the preliminary design phase for estimating gross tonnage.

[3] presented preliminary design formulas developed using a database of container ships built since 2015. Their formulas establish associations between the main dimensions—*i.e.*, length between perpendiculars (*Lpp*), breadth, draught, and depth—and key operational attributes: deadweight, TEU capacity, and ship speed.

While our previous study, titled *Effect of Ship Size on the Hull Dimension Ratios of Purse Seiners* [4] focused on the influence of vessel size on geometric hull proportions (L/B, L/D, B/D) from an explanatory perspective, this work addresses a fundamentally different design problem: the prediction of the main ship attributes required at the ship concept design stage. In [4], hull dimension ratios were treated as response variables and analyzed to interpret stability, speed, and seakeeping qualities in the preliminary design stage. In this work, hull ratios are not treated as primary outcomes; instead, the response variables are operational and volumetric vessel attributes directly required by designers at the ship concept design stage. Consequently, the statistical analysis in this manuscript serves a predictive and instrumental role rather than an explanatory one.

[5] introduced a procedure for determining the optimal characteristics of multi-purpose cargo vessels (MPCVs) at the preliminary design stage. Their method is based on the statistical analysis of a database compiled from reliable MPCV designs built over the past three decades. Using the resulting diagrams and formulas, designers can derive ship dimensions directly from a required deadweight (key input).

[6] developed an artificial neural network (ANN) model to predict the main particulars of chemical tankers during the preliminary design stage. This model utilizes deadweight and vessel speed as inputs, generating predictions for the overall length, the length between perpendiculars, the breadth, the draught, and the freeboard as outputs. Their results demonstrate that the model determines the initial main particulars with a high degree of accuracy when compared to actual ship data.

[7] employed statistical methods to analyze representative ship samples, establishing the relationship between gross tonnage (GT) and vessel size. This analysis yielded a regression equation that quantifies the GT-size relationship for typical Grand Canal vessels, in accordance with the specifications outlined in the Standard for the Main Dimensions of Ships in the Grand Canal.

[8] presented a plot of depth *vs.* breadth for several vessel types, which corresponds to lines of constant breadth-to-depth ratio. The analysis identified a distinct group—comprising fishing vessels and cargo ships where depth is constrained by stability—exhibiting a characteristic breadth/depth ratio of approximately 1.65.

[9] presented a collection of design diagrams derived from statistical and regression analyses of principal ship design attributes across various vessel types. This compilation synthesizes formulations and regressions developed by thesis students from the National Technical University of Athens (NTUA) and the staff of the NTUA Ship Design Laboratory, utilizing data sourced from the IHS Fairplay World Shipping Encyclopedia. For fishing vessels, the diagrams contrast power against length between perpendiculars, reefer capacity volume (in ft^3) against length between perpendiculars, reefer capacity volume against breadth, and reefer capacity volume against depth. The regression formulas used the determination coefficient (R^2) to indicate the goodness of fit [10].

[11] investigated the main attributes of seiners, fishing tackle, and trawlers by conducting regression analysis based on data from more than 2000 fishing vessels provided by the Ministry of Agriculture and Rural Affairs of China. For seiners, they obtained relationships of attributes against the main dimensions, or between GT and length and power and length, with an R^2 higher than 0.69.

[12] analyzed the principal dimensions and attributes of fishing vessels certified by the Russian Maritime Register of Shipping (RMRS). Their study examined over a thousand vessels of varying specifications. By applying analytical methods, they established relationships between key dimensions and attributes, such as GT *vs.* displacement and breadth *vs.* displacement.

Larger fishing vessels are appearing in the world's fishing fleet, in response to the need to save more time for catches, improve the use of energy, fish larger species, and stay longer at sea. Ship designers necessitate exhaustive and extensive hull statistics and estimated regression equations to obtain principal attributes and dimensions. However, these statistics and equations are commonly unavailable for a given vessel type. This research was motivated by a pronounced gap in the systematic documentation of principal dimensions and key attributes for American-type semi-industrial purse seiners within the existing literature. Such application-driven statistical analyses have been increasingly recognized as valuable when tailored to specific vessel categories, particularly those that are underrepresented in global design databases. The objective of this study is not to analyze hull ratios as design outcomes, but to develop practical, vessel-specific regression tools that allow naval architects to estimate power, tonnage, and fish hold capacity directly from principal dimensions during the first loop of the design spiral.

This work reports statistics, conducts regression analyses, and makes comparisons with existing formulas while using a database of 130 vessels from the Peruvian fishing fleet, which was published by PRODUCE [13] and DICAPI [14]. This work aimed to determine the relationships between the main ship dimensions of a purse seiner, such as length (L), breadth (B), and depth (D), and attributes like power, fuel, water, gross tonnage (GT), net tonnage (NT), and reefer volume or fish hold capacity (FHC).

Although our regression methodology is intentionally classical and based on simple power-law models, the contribution is not methodological but applicative. The novelty lies in the use of an updated and homogeneous database of American-type semi-industrial purse seiners for the systematically deriving vessel-specific regression formulas, in addition to a quantitative comparison with existing design equations. This provides naval architects with empirically grounded, fleet-representative relationships tailored to fishing vessel categories that have been scarcely addressed in the statistical hull-statics literature [15].

2. Methods

2.1 Vessel database

This research used data on purse seiners obtained from the vessel registry of the Peruvian Government's Ministry of Production (PRODUCE [13]), as well as from the General Directorate of Captaincies and Coast Guard (DICAPI [14]). To perform statistical analyses, we used a collection of 130 semi-industrial purse seiners (Table I) with a section plan approximately as described in [16]. The ships, registered from 1996 to 2020, have a refrigerated seawater (RSW) conservation system. These semi-industrial purse seiners are of the American type, with the bridge and accommodation forward and the working deck aft. We conducted regressions, correlation analyses, and comparisons to investigate the interaction between ship attributes and dimensions. It should be acknowledged that the database used herein is the same as that used in [4]. This reuse follows the international editorial guidelines for derived publications, as the research questions, dependent variables, analytical objectives, and engineering applications are clearly differentiated. This manuscript does not present duplicate results; it reutilizes the data to address a separate and practically relevant design problem.

Table I. Main attributes and dimensions in the purse seiner database

Main characteristic	Symbol	SI Units	Minimum	Maximum
Length	L	m	21.7	77.0
Breadth	B	m	6.0	14.5
Depth	D	m	2.6	8.6
Power	Power	kW	230	2289
Fuel capacity	Fuel	gal.	360	52 500
Water capacity	Water	gal.	80	29 800
Gross tonnage	GT		21.82	1006
Net tonnage	NT		72.45	224
Fish hold capacity	FHC	m^3	89.83	868

The initial dataset was extracted from the complete registry of the Peruvian fishing fleet. The vessels were filtered based on whether they included RSW systems. P*ower*, NT, *GT*, and *FHC data* were fully available for all selected vessels. The fuel and water tank capacities, which are not mandatory fields in the Peruvian fishing vessels database, were manually completed using information retrieved from the public records of the DICAPI. As a result, some vessels lack entries for *fuel* and *water capacity*. These missing values were maintained to reflect the actual availability of documented information, *i.e.*, only 59% of the records for fuel capacity and 54% for water capacity.

No observations were removed. To assess data quality, an outlier screening was performed using standardized Z-scores ($|z|>3$). All detected outliers correspond to real vessels with extreme but plausible physical characteristics (larger tonnage, higher installed power, or unusually high tank capacities), as shown in Fig. 1. Since these values represent the true variability in the fleet rather than measurement errors, they were preserved in the analysis.

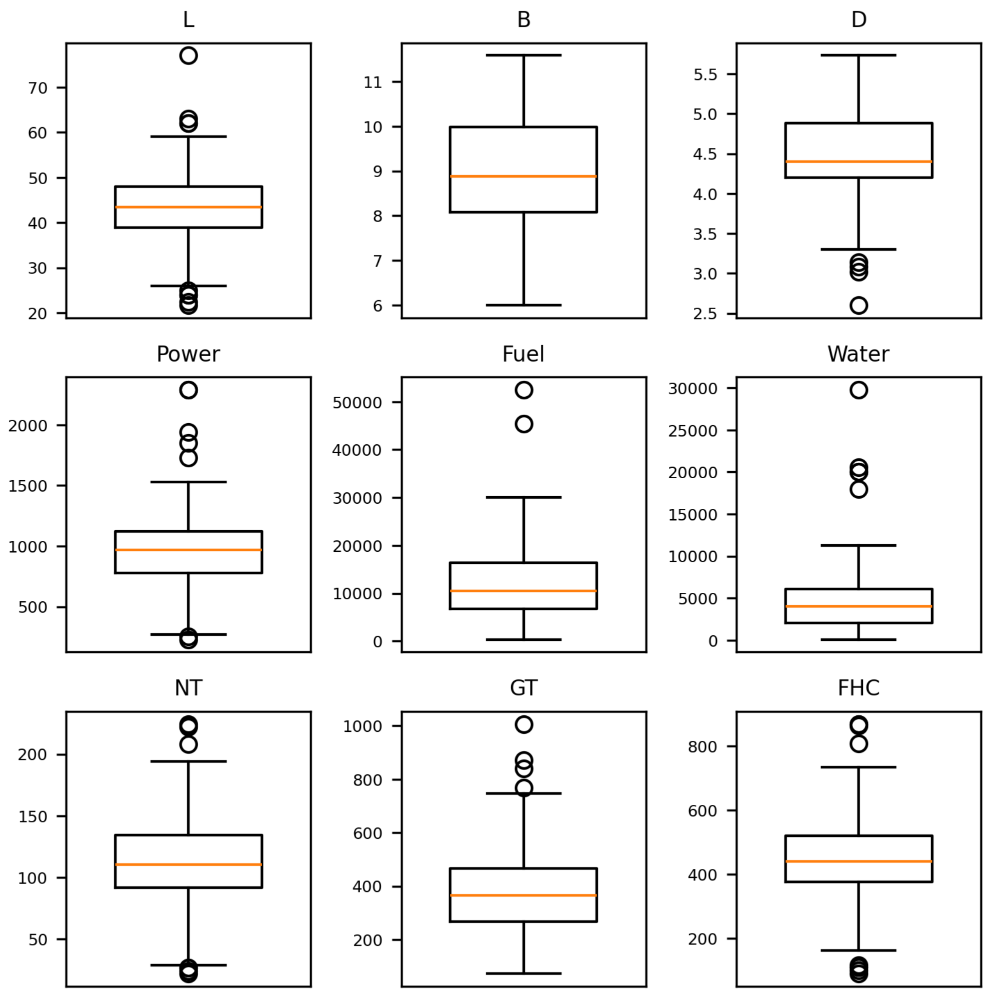

Figure 1. Boxplot for all variables showing outliers

2.2 Regression analysis methodology

Owner requirements are translated into technical characteristics during the ship concept design stage. This is the first loop of the design spiral. The design stage corresponds to the preliminary estimation of the main attributes (power, GT, NT, FHC, and range, as expressed by the fuel and water capacity, among others). To aid the designer in choosing the main attributes, a compilation of hull statistics for the most relevant dimensions must be available. For fishing vessels, this would include the principal dimensions *vs.* the GT, the NT, the FHC, and the fuel and water capacity.

The ship's linear dimensions are L, B, and D. L is the main linear dimension of the ship, which includes the overall length L, the length between perpendiculars Lpp, and the water line length Lwl. All these lengths are roughly the same, but L is generally found in public registries. B is the molded vessel breadth, measured at the midship station. D is the height between the molded baseline and the molded depth line of the uppermost deck at the side of the midship section [17]. Naval architects choose these linear dimensions while considering the geometric restrictions of the port. Then, they use them to select the ship's attributes by means of estimated regression equations. Similarly, when the designer requires a linear dimension for a given volume attribute (*e.g.*, GT), he or she can use the same estimated regression equations to determine it.

Fundamentally, a fishing vessel is a volume carrier, not a deadweight carrier [18]. The volume attributes related to fishing vessels are FHC, GT, and NT. FHC refers to the refrigerated fish hold volume, GT denotes the total enclosed volume of the vessel, and NT only represents the cargo volume and the passenger spaces.

Estimated regression equations are used to predict the response variable values as a function of a predictor variable. These equations are usually the result of simple regressions. We conducted simple regressions with the following form:

$$Y = aX^b \tag{1}$$

All power-law regressions were obtained using the classical log-log linearization approach [19], *i.e.*, applying an ordinary least-squares fit to the transformed variables, or $\ln y$ and $\ln x$. This procedure is consistent with the methodology used in hull-scaling studies and preliminary design formulations in naval architecture. The coefficient of determination is often quantified and compared to assess the accuracy of simple regressions:

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

where y_i denotes the values of n observations, y_i* the predicted values, and $\overline{y}$ the mean value.

3. Results

3.1 Ship statistics

Normality tests (Shapiro-Wilk, D'Agostino-Pearson, and Anderson-Darling, with a 95% confidence level) showed that no variables follow a normal distribution. Breadth is the closest to normality. This behavior, characterized by skewness, kurtosis, and long-tailed histograms, is typical of heterogeneous fishing fleets, where technological variation, vessel age dispersion, and operational heterogeneity produce non-symmetric distributions. As a consequence, the median was considered to be a more representative measure of central tendency than the mean for all variables. It should be noted that the lack of normality does not invalidate the regression analysis, since normality is required for model residuals rather than for the raw variables themselves [19], and simple potential regression remains appropriate for uncovering structural scaling relationships in conceptual ship design.

Table II shows that the ship attributes do not exhibit a normal distribution. This table also includes the *p*-value of the Shapiro-Wilk normality test, rejecting normality of the variables for a 95% confidence level. The distribution of B is the closest to normality.

Fig. 2 reports that, in this vessel category, power ranges between 750 and 1250 kW. There are a few ships with levels higher than 1500 kW in the Peruvian fishing fleet. The average power is about 1000 kW. Fig. 3 shows that the fuel ranges between 800 and 1100 gallons. This variable does not have a normal distribution, but the distribution peak marks a high concentration near its median. A high concentration of ships is also found around the medians for L, power, fuel, water, NT, GT, FHC, and D.

The Peruvian fishing fleet ranges from about 33 to 55 m in length, 8.8 to 9.2 m in breadth, 4.4 to 4.6 m in depth, 750 to 1250 kW, 8000 to 11000 gal fuel capacity, 1000 to 5000 gal water capacity, 300 to 500 GT, 100 to 120 NT, and 300 to 500 m^3 FHC. In addition, a few large vessels belong to another category, with lengths over 60 m, power values over 1500 kW, fuel capacities over 20 000 gal, water capacities over 9000 gal, NT over 175, GT over 600, and FHC over 600 m^3.

Table II. Descriptive statistics for the ship attributes of American-type semi-industrial purse seiners in the Peruvian fleet

Statistic	Length (m)	Breadth (m)	Depth (m)	Power (kW)	Fuel (gal)	Water (gal)	Net tonnage	Gross tonnage	FHC (m^3)
Mean	43.5	9.0	4.4	965	12939	5196	113.7	376.5	446
Standard error	0.8	0.1	0.05	30.4	1091	642	3.62	14.5	12.3
Median	43.3	8.9	4.4	969	10500	4000	110.5	366.5	440
Mode	40.5	8.9	4.4	969	12000	5000	127.0	248.3	450
Standard deviation	8.7	1.1	0.55	346	9572	5371	41.36	164.72	140
Sample variance	75.9	1.2	0.3	11985	91E6	28E6	1710.9	27134	1974
Kurtosis	1.55	-0.34	0.9	2.65	4.00	7.4	4.17	1.6	1.49
Skewness	0.21	-0.01	-0.4	0.99	1.64	2.51	0.94	0.83	0.13
Shapiro-Wilk	0	0.035	0	0	0	0	0	0	0
Range	52.3	6.6	3.1	2059	52140	29720	290.81	934.2	779

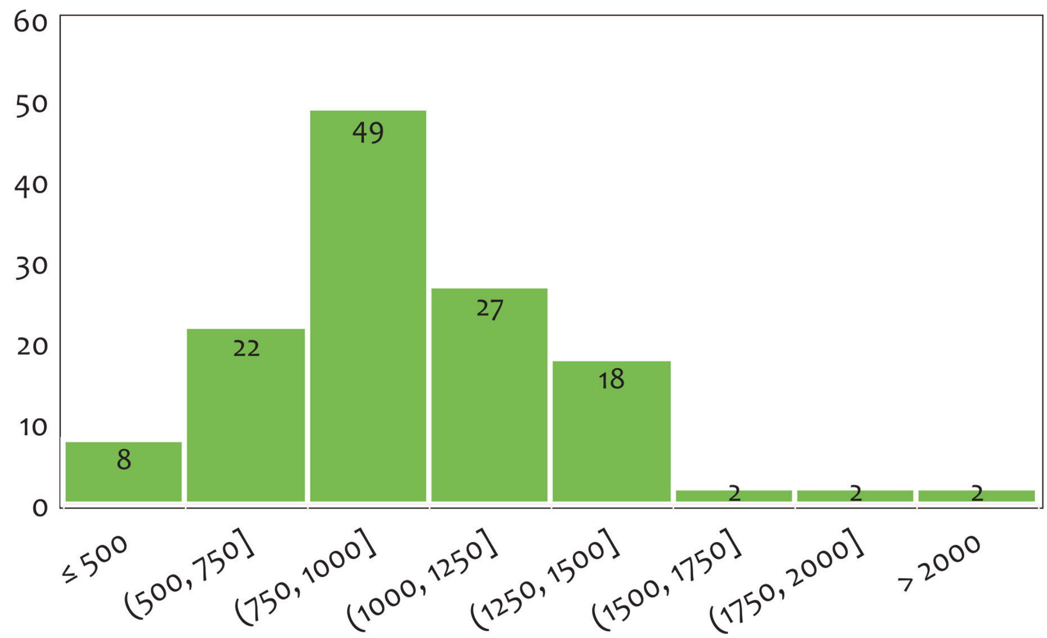

Figure 2. Power histogram of American-type, semi-industrial purse seiners

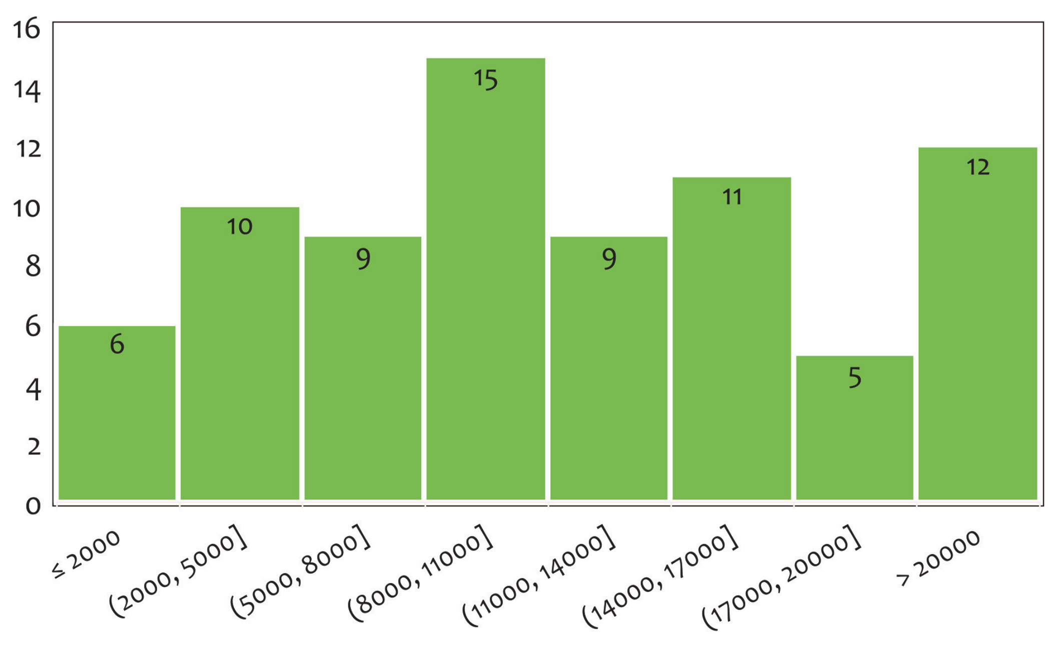

Figure 3. Fuel histogram of American-type, semi-industrial purse seiners

3.2 Regression analysis

Note that the response variables are power, fuel capacity, water capacity, GT, NT, and FHC, and the predictor variables are L, B, and D. Table III presents the estimated regression equations and the determination coefficients (R^2) for all variables. The regression equations rely on simple potential regression. The determination coefficients are higher than 0.64 for all response variables except fuel and water. Since $R^2 > 0.64$, it could be said that the predictor variables explain 64% of the variance in power, GT, NT, and FHC. The estimated regression equations for fuel and water capacity yield values with high variation.

Fig. 4 shows a strong positive relationship between power and L. The data points are uniformly dispersed about the trend line. The data points concentrate between 30 and 55 m, and between 750 and 1250 kW. Fig. 5 shows a strong positive association between GT and L. This association has an R^2 of nearly 1. L explains 91% of the variation in GT. A power function closely approximates the GT-L relationship, as the data points adhere to the regression curve. Fig. 6, the scatter plot of NT *vs.* L, also reveals a strong relationship. L explains 75% of the variation in NT. The data points align closely with the regression line. Moreover, Fig. 7 shows a strong positive association between FHC and L. L explains 74% of the variation in FHC. The data samples closely fit the regression curve.

Table III. Regression analysis coefficients for the attributes of American-type semi-industrial purse seiners

	Power (kW)		**Fuel (gal)**		**Water (gal)**	
	Estimated equation	**R^2**	**Estimated equation**	**R^2**	**Estimated equation**	**R^2**
Length (m)	$3.44 \cdot L^{1.48}$	0.68	$2.73 \cdot L^{2.15}$	0.24	$2.76 \cdot L^{1.86}$	0.13
Breadth (m)	$3.78 \cdot B^{2.5}$	0.69	$7.67 \cdot B^{3.23}$	0.18	$14.12 \cdot B^{2.47}$	0.08
Depth (m)	$29.1 \cdot D^{2.32}$	0.64	$39.3 \cdot D^{3.69}$	0.25	$35.5 \cdot D^{3.03}$	0.12
	Net tonnage		**Gross tonnage**		**FHC (m^3)**	
	Estimated equation	**R^2**	**Estimated equation**	**R^2**	**Estimated equation**	**R^2**
Length (m)	$0.19 \cdot L^{1.69}$	0.75	$0.11 \cdot L^{2.16}$	0.91	$1.09 \cdot L^{1.59}$	0.74
Breadth (m)	$0.30 \cdot B^{2.67}$	0.67	$0.20 \cdot B^{3.39}$	0.80	$1.59 \cdot B^{2.55}$	0.67
Depth (m)	$2.00 \cdot D^{2.67}$	0.72	$3.24 \cdot D^{3.14}$	0.74	$9.34 \cdot D^{2.57}$	0.74

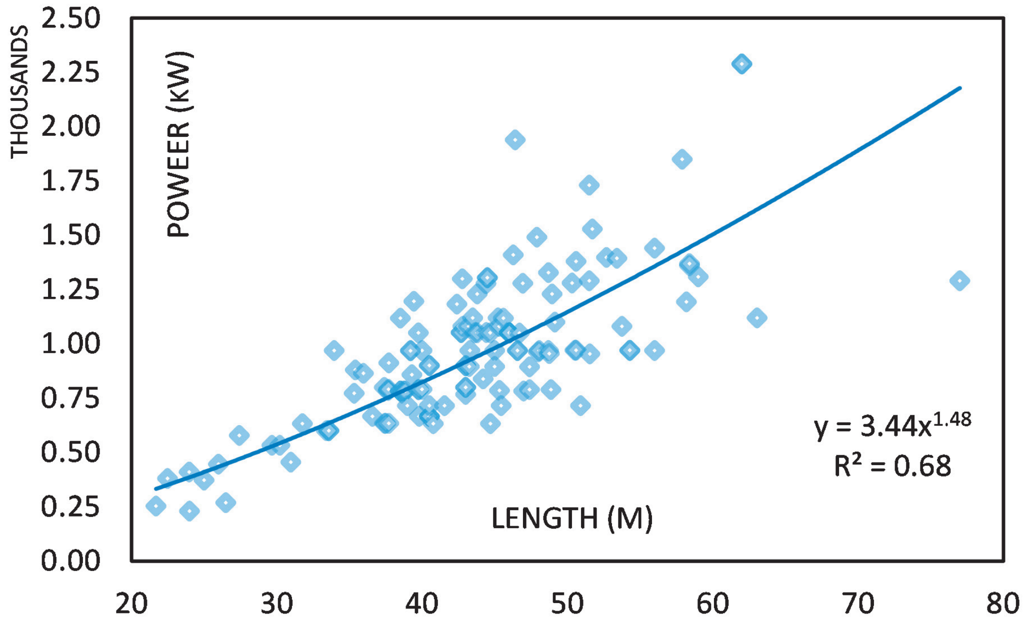

Figure 4. Power *vs.* length scatter plot and regression equation of for the American-type semi-industrial purse seiners in the Peruvian fleet

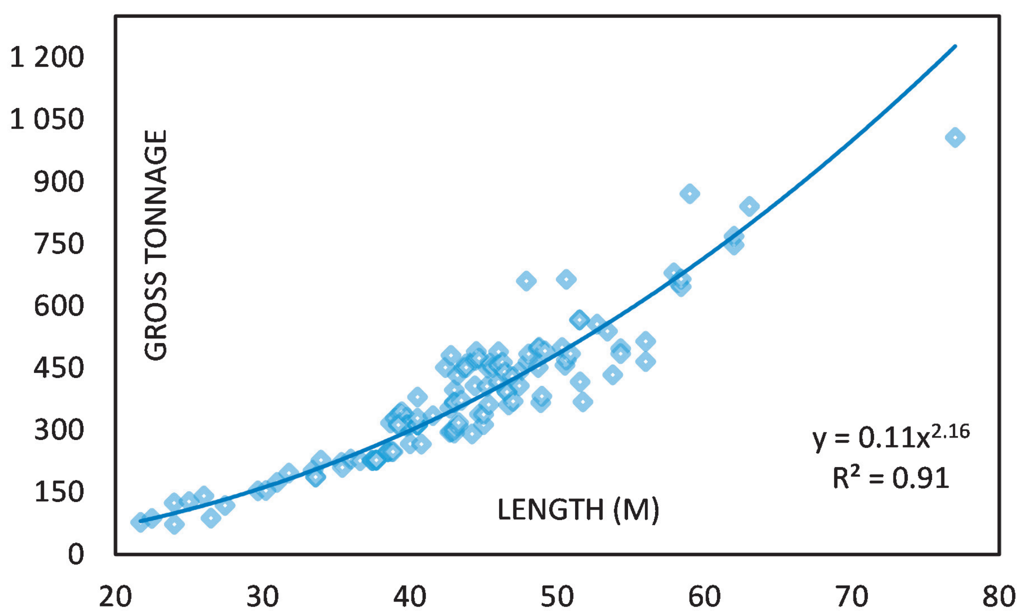

Figure 5. GT *vs.* L scatter plot and regression equation of for the American-type semi-industrial purse seiners in the Peruvian fleet

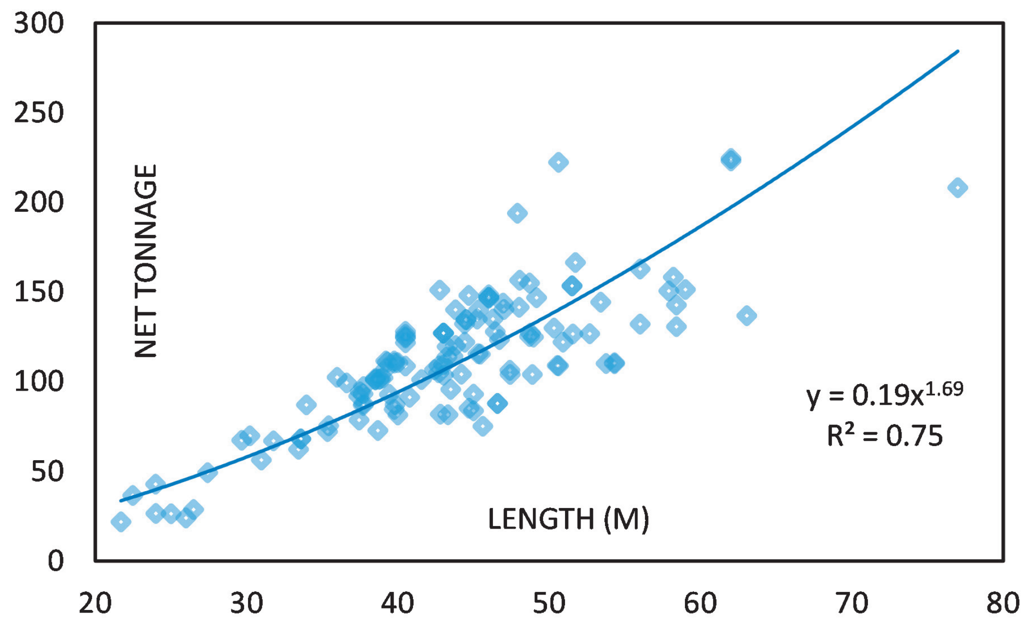

Figure 6. NT *vs.* L scatter plot and regression equation of for the American-type semi-industrial purse seiners in the Peruvian fleet

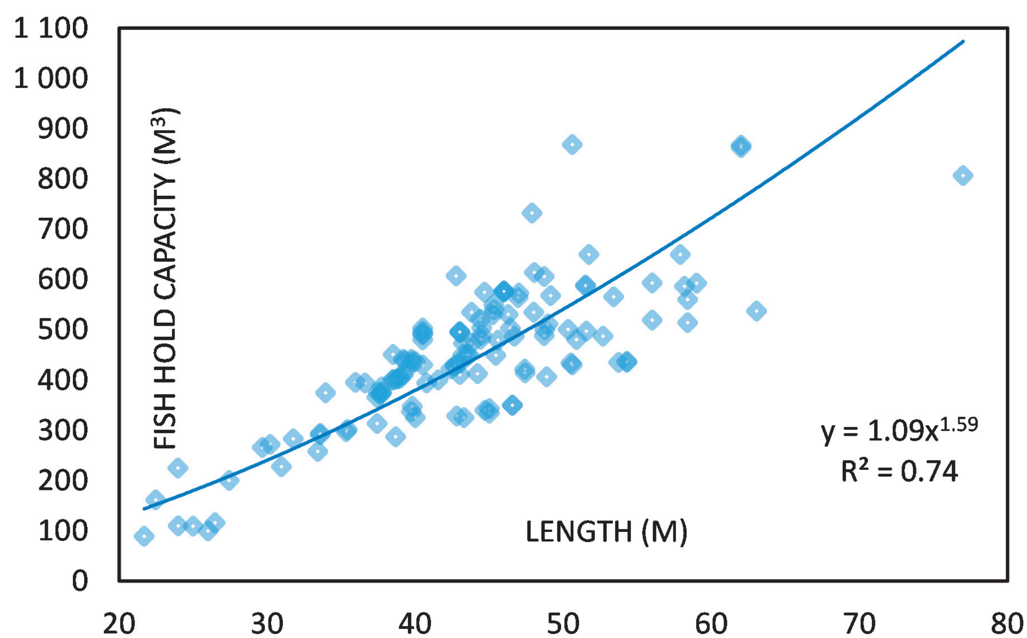

Figure 7. FHC vs. L scatter plot and regression equation of for the American-type semi-industrial purse seiners in the Peruvian fleet

Although L, B, and D strongly influence structural, volumetric, and power-related attributes, the fuel and water capacities of purse seiners are less dependent on the ship's linear dimensions. These tanks are typically sized according to operational requirements—such as desired autonomy, fishing grounds, and mission profile—rather than hull geometry alone. As discussed in standard naval design literature, consumables like fuel and freshwater are mission-driven capacities defined primarily by endurance criteria, propulsion plant characteristics, fuel consumption [21], and the expected operational range, which may vary regardless of the vessel's main dimensions. Therefore, it is reasonable that the fuel and water capacities exhibit weak correlations with L, B, and D in analyzed dataset.

Table IV demonstrates that power capacity, GT, NT, and FHC are strongly dependent on vessel dimensions. Across the three predictors (length, breadth, and depth), these variables consistently show high coefficients of determination ($R^2 \approx 0.64 - 0.91$) and comparatively low error metrics, with root mean square error (RMSE) and mean absolute error (MAE) values corresponding to their data ranges. This indicates that their scaling behavior is largely governed by hull dimensions. In particular, tonnage measures and the hold capacity follow expected dimensional relationships, while the installed power increases predictably with vessel size, as a consequence of propulsion requirements and hydrodynamic resistance.

In contrast, the fuel and water capacities show very low R^2 values, combined with high RMSE and MAE, regardless of whether L, B, or D is used as the predictor variable. These error patterns demonstrate that the fuel and water capacities are not directly determined by principal dimensions but instead depend on mission-specific, operational, and regulatory factors, which vary regardless of hull scaling. As a result, their variability cannot be captured through simple geometric regression models.

Table IV. Power-law regression coefficients and error metrics (RMSE, MAE) for principal vessel attributes regressed against length, breadth, and depth

	Length (m)		**Breadth (m)**		**Depth (m)**		
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	
Power (kW)	237.52	171.21	203.82	157.86	226.51	174.68	
Fuel (gal)	9660.40	6326.25	9117.37	5927.89	8967.28	5795.70	
Water (gal)	5459.11	3241.92	5439.20	3194.88	5345.40	3149.84	
Net tonnage	23.92	18.62	21.18	16.85	19.85	14.18	
Gross tonnage	62.20	45.42	85.43	58.26	98.48	70.29	
FHC (m^3)	88.97	69.79	78.07	61.55	70.85	52.13	

3.3 Model validation results

A statistical validation of the power-law models was performed to assess the three classical regression assumptions: (i) error normality, (ii) homoscedasticity, and (iii) residuals independence. The Shapiro-Wilk test was used for normality, the Breusch-Pagan test for homoscedasticity, and the Durbin-Watson statistic for independence.

The results indicate that residuals normality is not strictly satisfied for any of the variables. This does not affect the validity of the models, since normality is required for parametric inference rather than for predictive relationships, and power regressions in naval architecture are not typically used for hypothesis testing. For example, [22] and [23] use power regressions for predictions but not for hypothesis testing. Residuals independence was confirmed for all variables, with Durbin-Watson values close to 2.0. Heteroscedasticity was more pronounced for the fuel and water capacities, which is consistent with their weaker geometric dependence.

4. Discussion

The methodological approach adopted in this study is intentionally simple, relying on classical potential regressions that ensure interpretability and direct use in the concept ship design stage. The innovation, therefore, does not reside in the statistical technique itself, but in the development of updated, vessel-specific regression formulas for American-type semi-industrial purse seiners, a category for which systematic design data are largely absent. By integrating a recent fleet-wide database and benchmarking the resulting regressions against existing formulations in the literature, this study provides a refined and more accurate set of predictive relationships. These results aid in closing a documented gap in design knowledge and offer practical benefits for naval architects working with this class of fishing vessels.

4.1 Modern statistical techniques

Although modern statistical and machine learning-based techniques—such as multiple regression, artificial neural networks, or ensemble-based models—can offer predictive improvements, their use is not a standard requirement in ship *concept* design. At this early stage, naval architects rely primarily on low-dimensional empirical relationships derived from hull statistics, where proportionality laws, geometric similarity, and scaling behavior dominate the governing trends. The foundational literature emphasizes the long-established use of simple models for estimating ship main particulars and attributes, given their interpretability, robustness, and compatibility with classical naval architecture principles [19], [25]. Moreover, conceptual design decisions typically require transparent and physically meaningful formulations that can be applied quickly during iterative design loops, which favors simple power-law regressions over more complex black-box approaches [10]. Therefore, our use of simple regression methods aligns with standard and widely accepted practices in preliminary ship design.

4.2 Model validation

Power-law regression models of the form $Y = aX^b$ were estimated using log-log linear regression, following the standard procedure adopted in preliminary ship design tools. After fitting the models, their performance was assessed by means of the R^2, the RMSE, and the MAE on the original scale of the variables. For the main design attributes (power, GT, NT, and FHC), the resulting models exhibited moderate to high goodness of fit, with R^2 values ranging from approximately 0.64 to 0.91 and error levels compatible with the accuracy that is typically expected during the conceptual design stage. In addition, the residuals were inspected to check for systematic patterns, revealing no clear structural bias with respect to the predictor and indicating that the power-law form is adequate for capturing the dominant trends in the data. In contrast, the models for fuel and water capacity exhibited very low R^2 values and comparatively high dispersion, confirming that these attributes are poorly explained by principal dimensions alone and should be used with caution.

These validation steps confirm that power-law regressions are suitable as first-order predictors for key volumetric and tonnage attributes, while also highlighting the limitations of simple models for range-related variables such as fuel and water capacity.

4.3 Assessment of scaling models for preliminary design parameters

In addition to the classical power model $y = ax^b$, a nonlinear regression was also performed using the extended formulation, *i.e.*, $y = ax^b + c$, following the approach employed by [26]. Although this model introduces an additional degree of freedom, the estimated constant c resulted in a large negative value, revealing two issues:

i. The parameters lose physical interpretability, as the model predicts unrealistic offsets for small values of the independent variable.

ii. The constant term acts as a numerical compensator instead of representing a meaningful contribution to the vessel's geometric or scaling behavior.

Conversely, the classical power-law model produced stable coefficients and a monotonic trend consistent with well-established naval architectural scaling patterns [20], where hull-related attributes often follow proportionality laws of the form $y \propto x^b$. The absence of a constant term also preserves the log-linear transformability of the model, ensuring statistical robustness and allowing direct utilization in preliminary design contexts. Therefore, the power model $y = ax^b$ is considered to be an appropriate and physically meaningful representation for the analyzed characteristics.

4.4 Comparison with previous studies

Table V presents the estimated regression equations, derived from the database and formulations found in the literature [10], [11], for *power, FHC, and GT*. Figs. 8-10 compare the regression formulas for FHC from this study's database against those from [10]. For all dimension ranges, FHC's dependence on B and D follows a similar trend in both regressions. However, for L values exceeding 35 m, the estimated and referenced regressions for FHC exhibit two different trends. The database in [10], which included ships up to 130 m and 400 000 ft^3, yields a regression equation that overestimates values for the shorter vessels in our database. Fig. 11 shows that the estimated formulation for GT *vs.* L fits the present database well, but the equation in [11] overestimates the values of of GT when L is greater than 35 m. Fig. 12 shows that the equations in [10] and [11] can overestimate power values, but our regression equation provides a good approximation across all dimension ranges. Therefore, the regression equations for the analyzed database yield more precise estimations. Table VI reports the calculated R^2 for FHC, GT and power as a functions of ship dimensions. Although the R^2 values of the cited regression equations are higher than those of our database—except *FHC*=f(*D*)—, our estimated regression equations are appropriate for predicting the ship attributes of American-type semi-industrial purse seiners.

Table V. Comparison of estimated formulas of attributes of semi-industrial purse seiners

Present Database	Units	Papanikolau [10]	Bin *et al.* [11]
$Power = 3.44 \cdot L^{1.48}$	kW *vs.* m	$Power=65.945 \cdot L_{pp} - 1266.22$	$Power=50.19 \cdot L-971$
$FHC = 38.40 \cdot L^{1.59}$	ft^3 *vs.* m	$FHC = 7.9873 \cdot L_{pp}^{2.0778}$	
$FHC = 56.33 \cdot B^{2.55}$	ft^3 *vs.* m	$FHC = 15.144 \cdot B^{3.0943}$	
$FHC = 329.68 \cdot D^{2.57}$	ft^3 *vs.* m	$FHC = 774.89 \cdot D^{1.9626}$	
$GT = 0.11 \cdot L^{2.16}$	GT *vs.* m		$GT = 0.0057 \cdot L^{2.994}$

Table VI. Comparison of the R^2 values for the attributes of semi-industrial purse seiners

R^2	Power=f(L)	FHC=f(B)	FHC=f(D)	FHC=f(L)	GT=f(L)
Our database	0.68	0.67	0.74	0.74	0.91
Papanikolaou [10]	0.8481	0.8001	0.6557	0.8059	---
Bin *et al.* [11]	0.6929	---		---	0.9613

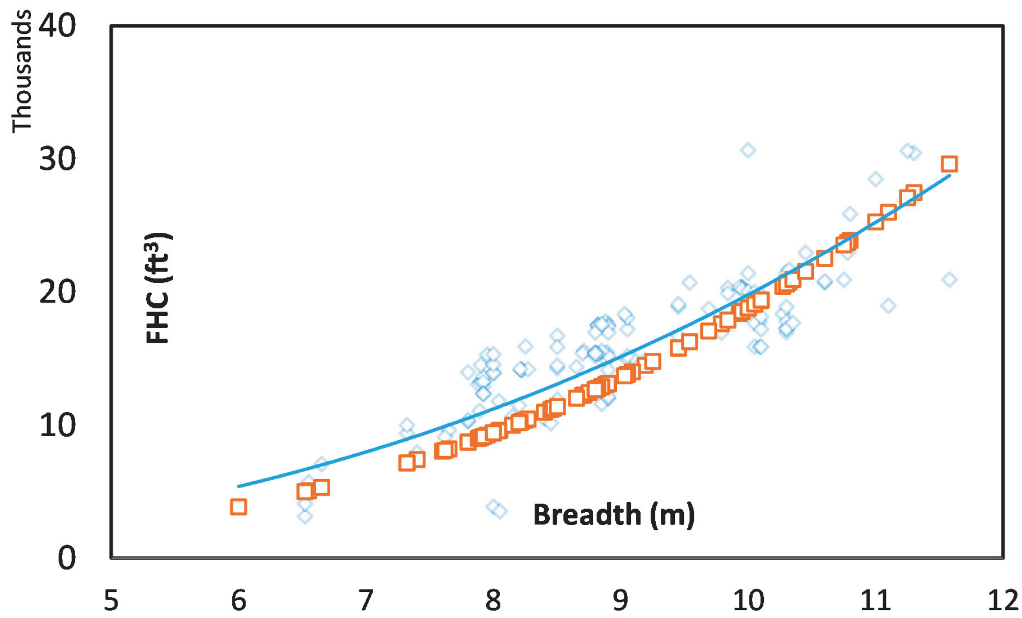

Figure 8. FHC *vs.* B comparison against previous studies. The blue line corresponds to our database, and the orange points to [10].

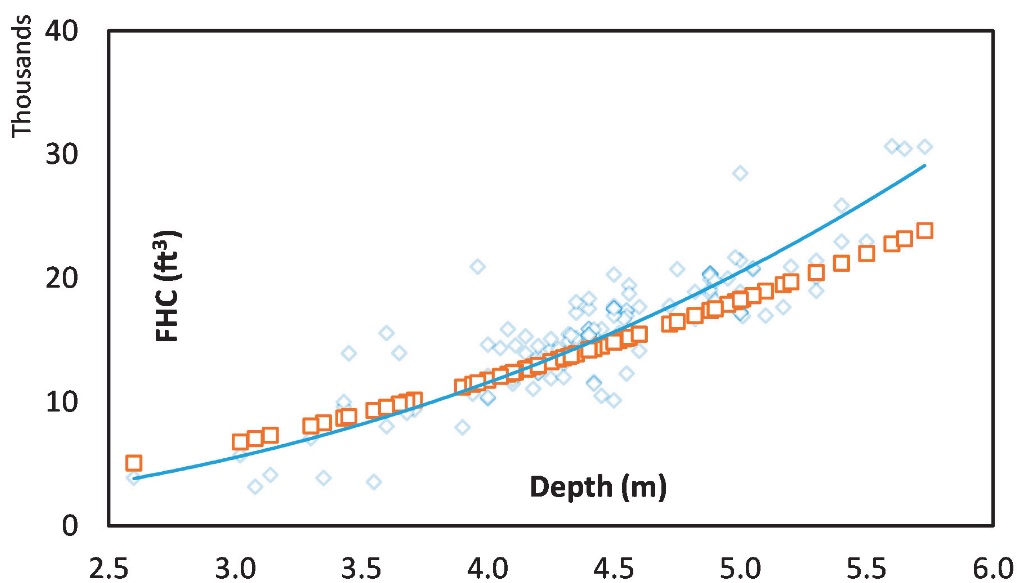

Figure 9. FHC *vs.* D comparison against previous studies. The blue line corresponds to our database, and the orange points to [10].

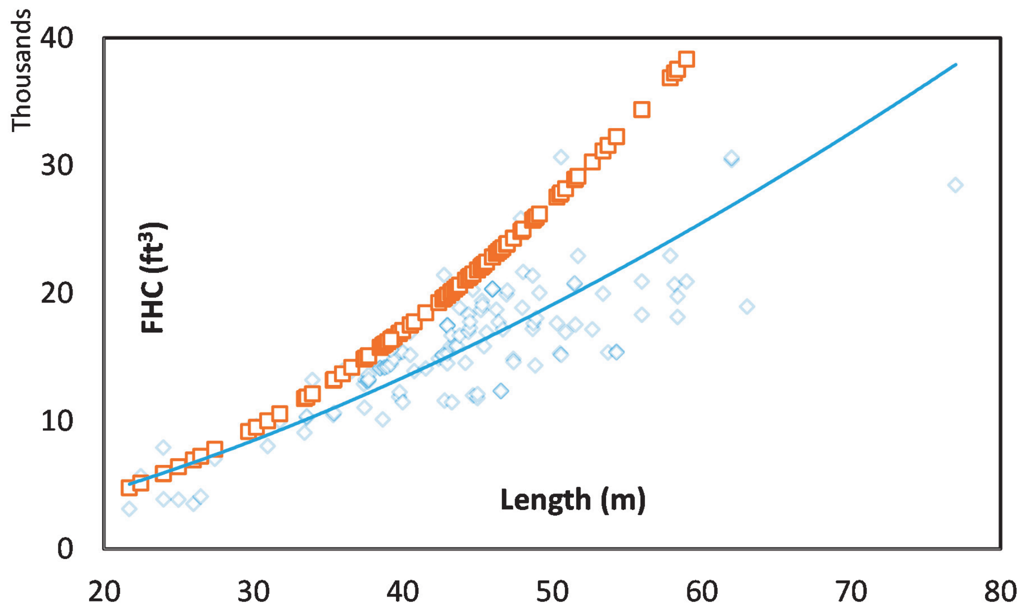

Figure 10. FHC *vs.* L comparison against previous studies. The blue line corresponds to our database, and the orange points to [10].

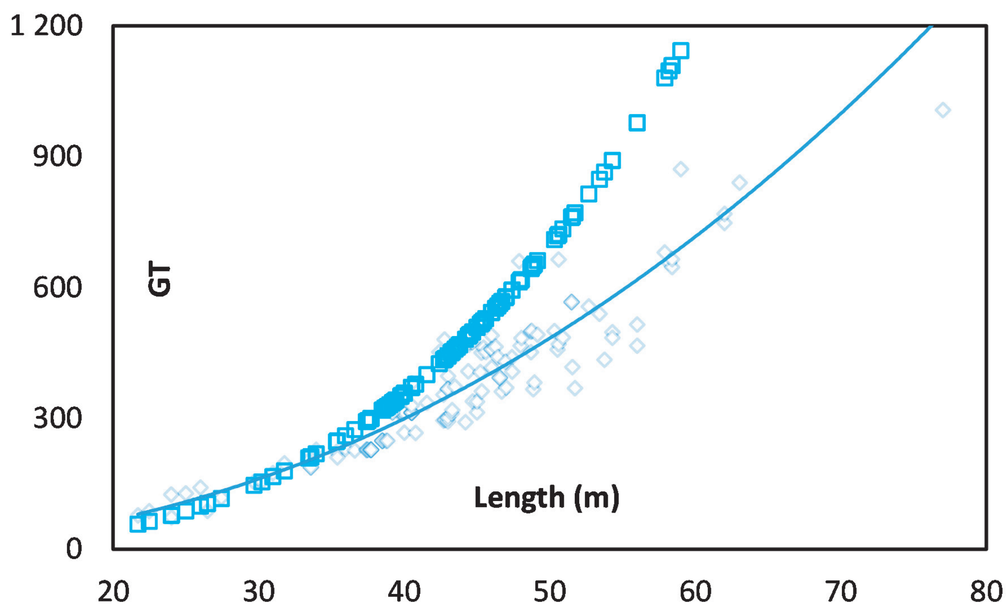

Figure 11. GT *vs.* L comparison against previous studies. The blue line corresponds to our database, and the orange points to [11].

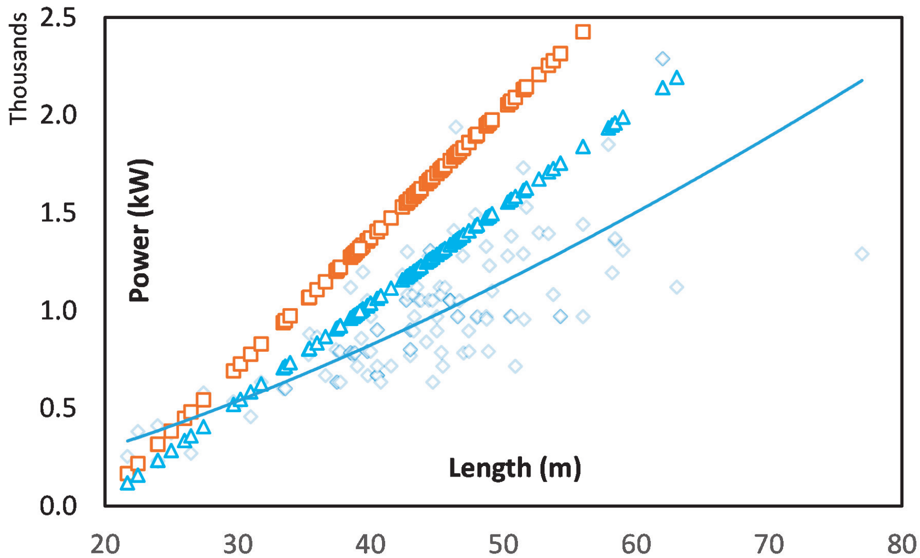

Figure 12. Power comparisons against previous studies. The blue line corresponds to our database, the orange points to [10], and the light blue points to [11].

The results of our regression analysis show that, when previously published formulas [10], [11] are applied to our dataset of Peruvian American-type semi-industrial purse seiners with RSW systems, the predicted key attributes are systematically higher than observed. This discrepancy can be attributed to the differences in the underlying populations: earlier works utilized broader fishing-vessel fleets, encompassing diverse gear types and operational profiles, whereas our dataset was limited to a highly specialized vessel class with distinct operational requirements and design optimizations (*e.g.*, mid-size RSW purse seiners). Therefore, the over-estimation is not indicative of a methodological error, but rather of the need to derive tailored regression models for specific vessel populations. These findings emphasize the value of focused empirical datasets in ship design regressions and the potential limitations faced when generalizing from broader fleets.

5. Conclusions

Designers can use the median for all linear dimensions and attributes during the early design of American-type semi-industrial purse seiners with an RSW systems. Although the medians and means of L, B, and D are close, statistically, B is the variable with the closest distribution to normality.

The Peruvian fishing fleet, specifically when it comes to American-type semi-industrial purse seiners with an RSW system, ranges from about 33 to 55 m in length, 8.8 to 9.2 m in breadth, 4.4 to 4.6 m in depth, 750 to 1250 kW, 8000 to 11 000 gal fuel capacity, 1000 to 5000 gal fuel capacity, 300 to 500 GT, 100 to 120 NT, and 300 to 500 m^3 FHC. Ships with lengths over 60 m, installed power over 1500 kW, fuel capacity over 20 000 gal, and water capacity over 9000 gal would belong to another category of ships: larger purse seiners with a tall tower to spot tuna or industrial purse seiners.

The combined interpretation of the RMSE, MAE, and R^2 confirms that the power-law regressions provide reliable estimations for GT, NT, FHC, and power. The fuel and water capacities exhibit very high RMSE and MAE values and should therefore be used with caution, as they are not primarily governed by geometric scaling.

The regression models developed in this study are specifically tailored to American-type semi-industrial purse seiners with RSW, a vessel class of high regional relevance. Because they are derived from a homogeneous and representative dataset, these models provide more accurate early-design predictions than some broader formulas in the literature. Their application supports more reliable dimensioning and capacity estimation in the conceptual design stage of this type of fleet.

In light of the present findings, our proposal for incorporating artificial intelligence and other advanced predictive techniques arises directly from the limitations identified in this analysis. Attributes such as the fuel and water capacity exhibited low determination coefficients under classical power-law regression, suggesting that their behavior may depend on nonlinear or multi-factor interactions not captured by single-dimension predictors. This motivates the exploration of data-driven approaches—such as multiple regression, robust regression, or machine learning models—to assess whether these methods can better represent complex operational or design dependencies. Thus, the recommendation for future integration of said techniques is not external to the study; it emerges naturally from the statistical patterns and modeling gaps observed in the analyzed dataset.

6. Acknowledgements

We would like to thank the Vice-Rectorate for Research of Universidad Nacional de Ingeniería, Lima, Peru, as well as the reviewers for their helpful comments.

7. Author contributions

Dennys de la Torre: conceptualization, methodology, data curation, writing (original draft)
Dheybi Cervan: visualization, research, revision.

8. Funding

This research received no external funding.

9. Data availability

The data will be made available upon request.

10. Conflict of 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.

11. Use of artificial intelligence

During this work, the authors did not employ any technologies associated with artificial intelligence.

References

[1] B. Rinauro, E. Begovic, F. Mauro, and G. Rosano, "Regression analysis for container ships in the early design stage," *Ocean Eng.*, vol. 292, art. 116499, Jan. 2024, https://doi.org/10.1016/j.oceaneng.2023.116499

[2] A. Lepore, L. Mocerino, B. Palumbo, E. Rizzuto, E. Rossi, and L. Vitiello, "A supervised statistical learning approach to the preliminary design of cruise ship Gross Tonnage," *Appl. Ocean Res.*, vol. 143, art. 103837, Feb. 2024, https://doi.org/10.1016/j.apor.2023.103837

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

[4] D. de la Torre, "Effect of ship size on the hull dimension ratios of purse seiners," *Ing. Investig.*, vol. 45, no. 1, art. e116364, 2025, https://doi.org/10.15446/ing.investig.116364

[5] M. Kalajdžić and N. Momčilović, "A step toward the preliminary design of seagoing multi-purpose cargo vessels," *Brodogradnja*, vol. 71, no. 2, pp. 75–89, Jun. 2020, https://doi.org/10.21278/brod71205

[6] S. Gurgen, I. Altin, and M. Ozkok, "Prediction of main particulars of a chemical tanker at preliminary ship design using artificial neural network," *Ships Offshore Struct.*, vol. 13, no. 5, pp. 459–465, Jul. 2018, https://doi.org/10.1080/17445302.2018.1425337

[7] X. Peng, H. Lei, and W. Zhong, "Research on estimation method of gross registered tonnage of ships in Grand Canal," in *Modeling Risk Management for Resources and Environment in China*, First., D. D. Wu and Y. Zhou, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 101–108, https://doi.org/10.1007/978-3-642-18387-4

[8] D. G. . M. Watson, *Practical ship design*, vol. 1. Oxford: Elsevier Science LTD, 1998.

[9] S&P Global, "IHS Fairplay World Shipping Encyclopedia version 12.01," 2011. [Online]. Available: https://www.spglobal.com/marketintelligence/en/mi/products/maritime-portal-desktop.html

[10] A. Papanikolaou, *Ship design. Methodologies of preliminary design*, 1st ed. Dordrecht: Springer Netherlands, 2014, https://doi.org/10.1007/978-94-017-8751-2

[11] H. Bin, L. Xiuwen, and R. Yuqing, “Statistics and analysis of main parameters of ocean fishing vessels in China,” in *2020 5th Int. Conf. Electromech, Control Tech. Trans. (ICECTT)*, IEEE, May 2020, pp. 594–598, https://doi.org/10.1109/ICECTT50890.2020.00135

[12] O. E. Surov, M. V. Kitaev, E. E. Solovieva, V. A. Veselov, and D. V. Tyuftyaev, “Analysis of main dimensions and characteristics of fishing vessels,” *Russ. J. Water Transp.*, vol. 2, no. 72, pp. 41–53, Sep. 2022, https://doi.org/10.37890/jwt.vi72.273

[13] Ministerio de la Producción, Gobierno del Perú, “Embarcaciones pesqueras.” [Online]. Available: https://consultasenlinea.produce.gob.pe/ConsultasEnLinea/consultas.web/embarcacion

[14] Dirección General de Capitanías y Guardacostas del Perú, “Consulta datos de la nave.” [Online]. Available: https://www.dicapi.mil.pe/consultas/naves-artefactos-navales

[15] H. M. Gaspar, “Data-driven ship design,” in *17th Int. Conf. Comput. IT Appl. Marit. Ind.*, 2015, pp. 426–439. http://www.shiplab.hials.org/wp-content/uploads/2019/05/40_Gaspar.pdf

[16] L. J. C. Valerio Mena, V. N. Acosta Pastor, and V. A. Loarte Vicuña, “Determinación numérica de la resistencia al avance de buques,” *TECNIA*, vol. 9, no. 2, pp. 61–68, Dec. 1999, https://doi.org/10.21754/tecnia.v9i2.324

[17] N. Hamlin, “Ship geometry,” in *Principles of Naval Architecture*, 2nd ed., vol. I, E. V. Lewis, Ed. Jersey City, NJ, USA: The Society of Naval Architects and Marine Engineers, 1988, p. 19.

[18] B. Tibbitts, “Naval surface ships,” in *Ship Design and Construction*, T. Lamb, Ed. Jersey City, NJ, USA: The Society of Naval Architects and Marine Engineers, 2003, pp. 28_1-28_11.

[19] J. Fox, *Applied regression analysis & generalized linear models*, 3rd ed. Thousand Oaks, CA, USA: SAGE Publications, 2016.

[20] D. Doane and L. Seward, *Applied statistics in business and economics*, 3rd ed. New York: McGraw-Hill, 2011.

[21] N. Bialystocki and D. Konovessis, “On the estimation of ship’s fuel consumption and speed curve: A statistical approach,” *J. Ocean Eng. Sci.*, vol. 1, no. 2, pp. 157–166, Apr. 2016, https://doi.org/10.1016/j.joes.2016.02.001

[22] H. Schneekluth and V. Bertram, *Ship design for efficiency and economy*, 2nd ed. Oxford, UK: Butterworth-Heinemann, 1998.

[23] C. B. Barrass, *Ship design and performance for masters and mates*. Oxford, UK: Elsevier Butterworth-Heinemann, 2004.

[24] P. Gale, “Concept design,” in *Ship Design and Construction*, J. Lamb, Ed. Jersey City, NJ, USA: The Society of Naval Architects and Marine Engineers, 2003, pp. 5–9.

[25] K. Rawson and E. Tupper, *Basic ship theory. Vol. I*, 5th ed. Oxford, UK: Butterworth-Heinemann, 2001.

[26] M. L. Hakim, B. Nugroho, I. K. Suastika, and I. K. A. P. Utama, “Alternative empirical formula for predicting the frictional drag penalty due to fouling on the ship hull using the design of experiments (DOE) method,” *Int. J. Technol.*, vol. 12, no. 4, art. 829, Oct. 2021, https://doi.org/10.14716/ijtech.v12i4.4692

Dennys De La Torre

He received a Naval Engineering degree from Universidad Nacional de Ingeniería, Lima, Peru, in 2004. He is a MSc in Investment Projects from the same university, as well as a PhD in Energy. He is a professor of ship dynamics with more than 15 years of teaching experience. He has experience in projects on board oil platforms and national and international shipyards. He is the manager of a boat design company and a research professor of the Vice-Rectorate of Research of Universidad Nacional de Ingeniería (lines of research: wave energy and ship dynamics).

Email: ddelatorrec@uni.edu.pe

Dheybi Cervan

He received his Mechanical-Electrical Engineering degree from Universidad Nacional de Ingeniería, Lima, Peru, as well as a MSc and PhD in Sciences with a mention in Energy. His main research interests include power systems reliability, data science, and artificial intelligence applications for industrial systems. He is a consultant with ten years of experience who provides specialized services to various companies and organizations in Peru's electrical and energy sectors. He is currently qualified as a researcher by CONCYTEC (National Council for Science, Technology and Technological Innovation). Additionally, he has teaching experience in Master's programs in Energy and Maintenance Engineering.

Email: dcervanp@uni.pe