

Research Article

Container Ship Carbon and Fuel Estimation in Voyages Utilizing Meteorological Data with Data Fusion and Machine Learning Techniques

Feiyang Ren ¹, Shaohan Wang ^{1,2,3}, Yuanzhe Liu ^{1,3} and Yi Han ³

¹Shanghai Ship and Shipping Research Institute, Shanghai, China

²School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, Shanghai, China

³COSCO Shipping Technology Co., Ltd., Shanghai, China

Correspondence should be addressed to Shaohan Wang; wanghetongtt@sjtu.edu.cn

Received 20 July 2022; Revised 10 October 2022; Accepted 10 November 2022; Published 25 November 2022

Academic Editor: Aditya Rio Prabowo

Copyright © 2022 Feiyang Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The International Maritime Organization (IMO) had made effort to reduce the ship's energy consumption and carbon emission by optimizing the ship's operational measures such as speed and weather routing. However, existing fuel consumption models were relatively simple without considering the quantified effect of weather conditions. In this paper, a knowledge-based ridge regression-based algorithm is presented for enabling automated fuel consumption estimation under varying weather conditions during voyages. Wind speed, wave height, ship speed, draught, AIS segment distance, and ship's heading (HDG) are used as input to predict the fuel consumption value from the MRV report. In this work, 3 types of models are tested: AIS-based model, MRV-based model, and MRV-based normalized model. In AIS based model, weather conditions are divided into nine categories based on wind speed, wave height, and wind directions then trained separately. In MRV-based mode, the daily weather condition was used, and the MRV-normalized model used the normalized daily weather data. The proposed ridge regression models (11 models total) were tested with 4 container ships for a period of one year, and the result shows that compared to real fuel consumption, MRV-based model could achieve the best result with an average error less than 3% comparing to real MRV report.

1. Introduction

In the past few years, one of the most serious challenges faced by the world is the global warming caused by greenhouse gases (GHGs). Based on the record of the United Nations Conference on Trade and Development (UNCTAD), seaborne contributes over 80% of the estimated volume of international trade, and over 90,000 commercial ships are sailing on the ocean with a total tonnage of 1.86 billion deadweight tones [1, 2]. Furthermore, heavy fuel oil, as the main power source for the shipping industry, can create a lot of carbon footprint because of its emission of GHGs and other pollution materials [3, 4]. Thus, the International Maritime Organization (IMO) adopted an initial strategy for the reduction of GHG emissions from ships, setting out a vision that confirms IMO's commitment to

reducing GHG emissions from international shipping and to phasing them out as soon as possible: reducing the CO₂ emission by at least 40% by 2030, with the target of 70% of reduction by 2050 and total annual GHG reduction of 50% by 2050 comparing to the year of 2008 [5]. But despite the growing number of strict environmental protection policies implemented by the IMO and effort made by individual countries and organizations, emissions due to fuel usages are still in a general upward trend.

The reduction of ship fuel usage and carbon emission is a complex and multifaceted problem [6]. For that, numerous research has been conducted, and those methods can be classified into 2 categories: using a new type of fuels such as biofuel, electric fuel, hydrogen fuel, and other synthetics or analyzing fuel consumption and carbon emissions relative to a particular shipping activity. The first

method involves the impact on many areas such as politics, economics, research, and engine replacement [6]. Therefore, the second method will have the advantage of lower cost and is more feasible in near future.

The second method involves the detailed ship activities such as ship speed, ship conditions, and other outside sailing conditions such as the meteorological information. The rapid development of computational power of computers and the speed of 5G network make the use of AIS for identification of ship behavior possible. The automatic identification system (AIS) is an automatic ship self-reporting system for maritime transportation safety purposes [7]. Autonomously broadcasted AIS messages contain kinematic information (including ship location, speed, heading, rate of turn, destination, and estimated arrival time) and static information (including ship name, ship MMSI ID, ship type, ship size, and current time), which can be transformed into useful information for intelligent maritime traffic manipulations [8]. Thus, many researches focus on the data fusion of different data sources such as voyage report, AIS, ship sail weather data, and sensor data for more accurate fuel/carbon estimation [9–11]. With merged data, different data mining techniques are applied to find the hidden relationship between vessel engine power systems, weather conditions, numerous ship-related internal factors, fuel consumptions, and ship carbon footprint.

The white-box models (WBM) come from the traditional computational fluid dynamics method to calculate the resistances faced by the ship from multiple sources based on the physical principles and dynamics laws. The model will first estimate the resistance in calm water and add the additional resistance caused by external factors such as wind and waves. Then, the total engine power needed to overcome the total resistance can be calculated along with the corresponding fuel consumptions and carbon emissions [12]: Medina et al. [13] developed an analytical model with an empirical formula by considering wind values along the sailing route to find the fuel consumption of container ship. Tillig et al. [14] proposed the model for a speed reduction optimization plan during sail by considering weather impact for minimizing fuel consumption over voyages. The advantages of WBMs are that they can be applied at the initial stage of ship design, and the parameters inside the model can be easily interpreted since those parameters are mainly from dynamics laws. But, despite the strength of WBMs, they have some obvious disadvantages: WBMs model's parameters are indeed explainable, but many assumptions must be made before the constructions of the model, and the prior assumptions made can have a large impact on the model's performance and whether it can be applied on real situations. Besides, in order to find the parameters needed, ship structures, and building information, complete sail activities must be complete and fixed, i.e., missing data cannot be applied to such model, and no randomness can be included to allow data uncertainties [15]. Therefore, though the WBMs can be a relatively accurate result, the performance is strict to "fixed" and "absolute" conditions, which make it hard to be applied to real-world shipping operations.

The black-box models (BBMs) are widely used machine learning models for fuel consumption predictions with AIS and weather data. The process of using BBMs begins with data extractions with preprocessing followed by some reasonable assumptions with suitable regression models. And finally, the model's fit and generalization nature is tested on the validation data set. For example, Wang and Meng [16] investigate the optimal sailing speed of container ships on each leg of each ship route in a liner shipping network while considering transshipment and container routing formulated as a mixed-integer nonlinear programming model. Adland et al. [17] develop a flexible framework for the estimation of the fuel consumption-speed curve for ships which allows for speed-dependent elasticity with endogenous thresholds. The problem is, conventionally, that the ship performance is shown by the speed-power curve of the specific type of vessel obtained from real sea trails. But most of the time, the single curve obtained from the sea trails is not enough to study the fuel consumption of the whole range of operating conditions such as weather. Besides, though the cubic law relation between ship sailing speed and fuel consumption rate is widely accepted, it is very hard and expensive since the cubic law requires the vessel's rotational speed, daily main and auxiliary engine fuel consumption rate, design speed, operating speed, and other ship information. Thus, the research gaps can be summarized as follows:

- (1) Existing studies did not have a valid data fusion process to combine AIS data, meteorological data, and voyage report data
- (2) Current studies on fuel consumption did not have a thorough comparison of models based on different data sources (i.e., AIS, MRV)
- (3) Despite the speed-power-fuel curve, with data set constructed by one single vessel, it is not enough to obtain the fuel consumption curve to cover the whole range of weather conditions (i.e., not enough data for every single weather condition)
- (4) Huge errors occur with certain weather conditions, especially in severe weather conditions

The remaining of this paper is organized as follows:

In Section 2, we will briefly discuss the data source used and the data preprocessing method. Section 3 includes the weather classification method: for turning weather numerical values into fuzzy values for the later regression model, and the vessel clustering method: based on ship's dimensions, engine power, and 7 other important features to classify each ship into a specific vessel type. Each vessel type (instead of one single vessel) will share the same fuel consumption model to make sure that data are sufficient for regression model building and result validation. In Section 4, a comparison of the mathematical model used with other commonly used models is presented. In Section 5, results will be shown for the comparison of model performance with voyage report result. In the end, a summarized conclusion will be given along with the discussion of possible future work.

2. Data

2.1. AIS Data. COSCO Shipping Technology has provided a set of AIS data for container vessels. The dataset ranges from 2020-8-1 to 2022-2-1 and contains 801665 ship voyage record data. The dataset includes information such as vessel name, vessel number, time of AIS signal from the vessel, vessel's position, vessel's draft, vessel's track direction, and vessel's sailing speed.

2.2. Meteorological Data. The Shanghai Meteorological Service has provided weather data from 2020-8-1. The weather data include information on wind and wave magnitude, wind and wave direction, current magnitude, current direction, and the extent of the area. Due to the small meteorological range delineated, it is assumed that the meteorological conditions experienced by vessels within that range are represented by the weather data presented for that area.

2.3. MRV Daily Fuel Data. COSCO Shipping Group has provided daily fuel consumption data in MRV from 2020-8-1 to 2022-2-1. Measurement, reporting, and verification (MRV) provide requirements for the monitoring, reporting, and verification of carbon dioxide (CO₂) from ships arriving at, within, or departing from ports. Figure 1 shows an example of the daily fuel consumption data for some of the container vessels studied over a 12-month period.

3. Quantification of Container Ships' Fuel Consumption Model Based on Ridge Regression

3.1. Approach Overview. As the AIS data, weather data, and MRV daily data are not in correspondence, all data need to be matched and processed to obtain data containing weather, sea conditions, navigation information, and fuel consumption, use this data and ridge regression to construct a quantitative fuel consumption model to process, and finally make fuel consumption estimation and prediction. The process is shown in Figure 2.

3.2. Meteorological Classification. The weather at sea is complex and changeable, and ships often encounter different weather conditions during navigation. When facing different weather conditions, the sailing status and behavior of the ship will change to a certain extent, thus affecting the fuel consumption of the ship. In previous studies, weather data are generally collected in the form of daily newspapers, and when building the corresponding fuel consumption model, the weather data are large in granularity, and a single fuel consumption model cannot respond to the different fuel consumption effects brought by weather changes during ship navigation [18–23]. Since the Shanghai Meteorological Bureau provides real-time maritime weather data, it is beneficial to classify different weather conditions. According to the performance of ships under each type of weather

conditions, different fuel consumption prediction estimation models are built separately to adapt to the complex and changing weather conditions at sea.

3.3. Ship-Type Clustering. Due to the fact that single vessel does not have enough data for using the regression method, in this work, each type of vessel (which will contain several similar vessels) will be used. For that, the main engine power, operating ship service speed, deadweight, vessel type, length overall, displacement, draught, and engine rotational speed will be used as input for clustering. However, for the open access dataset, it is common to have some parameters missing; thus, a two-step clustering process is performed: (1) Apply K -means clustering algorithm on all ships with all 9 complete inputs and generate the label for each ship. (2) Train an XG-boost clustering model with labels in (1) and apply ships with incomplete input. The two-step clustering process ensures that all ships can be clustered into one group. And ships inside the same group will be used as a dataset for that the fuel consumption model of a specific vessel type.

3.4. Fusion Data Processing. Since a single data source cannot meet the subsequent modelling and analysis requirements, the AIS data, meteorological data, and MRV daily data need to be fused according to the process shown in Figure 3 to obtain more comprehensive ship data for subsequent modelling and analysis.

3.5. Ridge Regression Model. When there is multicollinearity between the independent variables of a regression equation, the ordinary least squares method can no longer be used to analyze the regression equation accurately. In 1962, Heer proposed an improved least squares estimation method called ridge regression, which is a relatively stable method if there are multiple correlations between the independent variables, and the standard deviation of the regression coefficients estimated by ridge regression is also smaller. And in 1970, Hoerl and Kennard gave a detailed discussion [24].

The least squares method commonly used in regression analysis is an unbiased estimate. For a posed problem, X is usually a column of full rank.

$$X\theta = y. \quad (1)$$

Using the least squares method, the loss function is defined as the square of the residuals, minimizing the loss function.

$$\|X\theta - y\|^2. \quad (2)$$

The above optimization problem can be solved directly using the following equation:

$$\theta = (X^T X)^{-1} X^T y. \quad (3)$$

When X is not of full rank, or when the correlation between some columns is relatively large, the determinant of $X^T X$ approaches zero, i.e., $X^T X$ approaches singularity, and

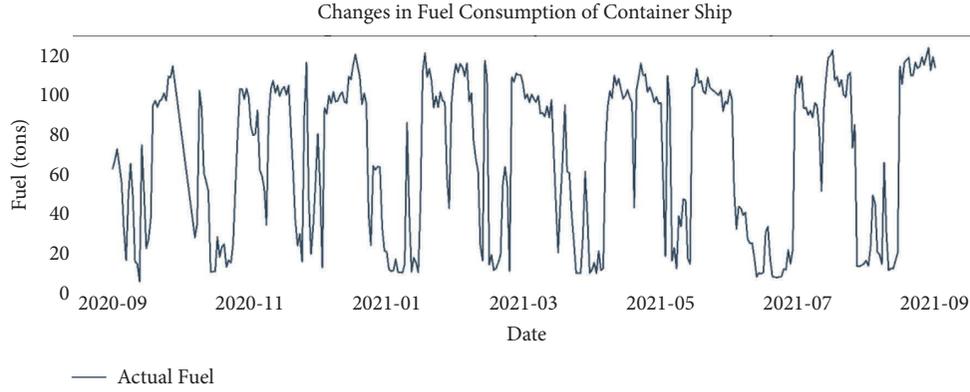


FIGURE 1: Container ship' MRV daily fuel.

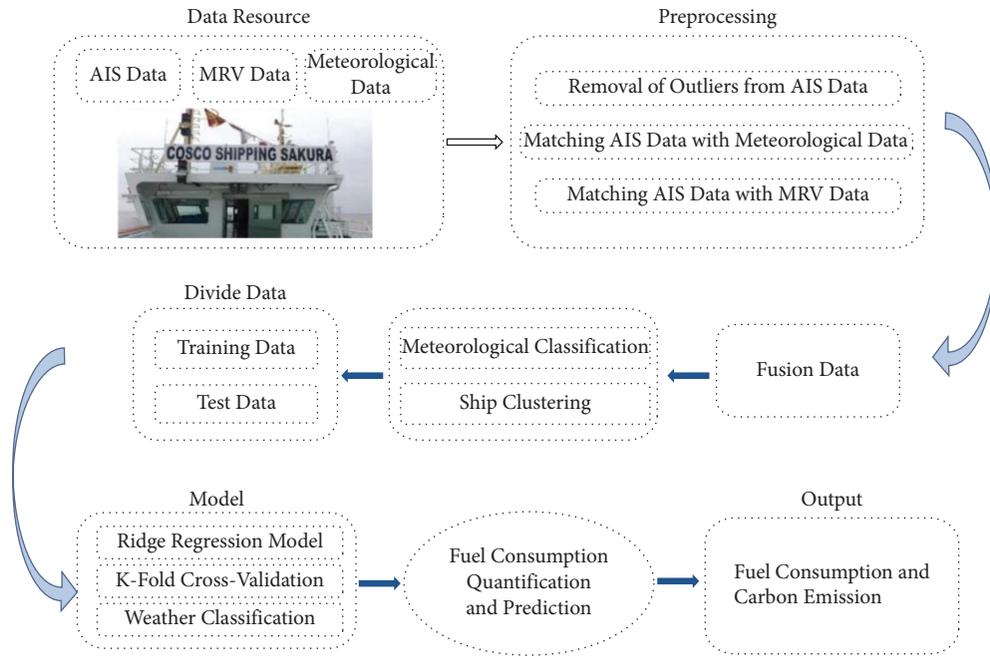


FIGURE 2: Methodological overview of the research.

the above problem becomes an ill-posed problem, at which time the error in computing $(X^T X)^{-1}$ will be large, and the traditional least squares method lacks stability and reliability. In addition, for some matrices, a small change in one of the elements of the matrix can cause a large error in the final computation, which is known as an ill-conditioned matrix. There are also times when incorrect computational methods can cause a normal matrix to behave in a pathological manner in an operation. For the Gaussian elimination method, if the elements on the principal element (i.e., the element on the diagonal) are very small, they will exhibit a pathological character in the calculation [25].

To solve the above problem, it is necessary to transform ill-posed problem into a posed one: for this purpose, a regularization term is added to the loss function.

$$\|X\theta - y\|^2 + \|\Gamma\theta\|^2, \quad (4)$$

where $\Gamma = \alpha I$; then,

$$\theta(\alpha) = (X^T X + \alpha I)^{-1} X^T y, \quad (5)$$

where I is the unit matrix.

As α increases, the absolute values of each element of $\theta(\alpha)$ tend to get smaller and smaller, and their deviation from the correct value of θ_i becomes larger and larger. α tends to infinity when $\theta(\alpha)$ tends to zero. Ridge regression is a complement to least squares regression, trading loss of unbiasedness for higher numerical stability and thus higher computational accuracy [26].

3.6. K-Fold Cross-Validation. K-fold cross-validation is commonly used to estimate the regularization parameter λ for regression models [27–29]. In this study, 10-fold cross-validation is used to estimate the regularization parameter

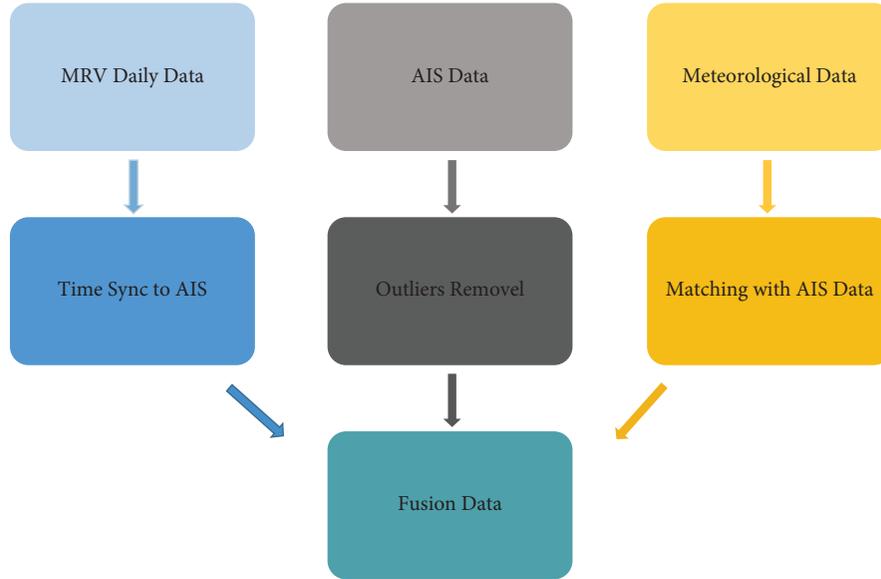


FIGURE 3: Fusion data processing.

λ . The parameter λ is estimated by minimizing the loss function according to equation (5), where the regression coefficient vector θ is estimated using equation (3). In k -fold cross-validation, the training set was partitioned into k data sets of approximately equal size. Then, one of the sets is used in turn for validation, performing k training runs. The error of the regression model trained on the $N - N/k$ samples is taken as the average error of the k cross-validation trainings. Finally, a range of relevant λ is obtained, and the one with the smallest error is then selected.

3.7. Fuel Consumption Prediction. The processed data set mainly includes the following: the maritime mobile service identify (MMSI), the time of AIS data collection, the ship's draft, the ship's track direction, the ship's sailing speed, the wind and wave size, the wind and wave direction, the time interval between two adjacent AIS data collections and the sailing distance, the fuel consumption at the time of each AIS point collection, the MRV daily filling time and the MRV daily fuel consumption, etc., as shown in Table 1.

Table 1 shows that each data field has different criteria, for example, some feature variables contain positive and negative values, which may reduce the accuracy of the model predictions if used directly for model construction and training. Therefore, to improve the prediction accuracy of the model, the data will be normalized using normalization. There are two common normalization methods: linear normalization and zero-score normalization (Z -score normalization). As linear normalization requires the data set to satisfy a Gaussian normal distribution and to reduce the difficulty of modelling, this study will normalize the processed data set using Z -score normalization [30].

For dataset $X = \{x_i\}$, $i = 1, 2, \dots, n$, μ is the mean value of the dataset, and σ is the standard deviation of the dataset. Then, Z -score normalization is defined as follows:

$$z = \frac{X - \mu}{\sigma}, \quad (6)$$

where

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}. \quad (7)$$

After Z -score normalization, the mean is 0, and the standard deviation is 1.

This study uses ridge regression to develop a regression model for the quantitative prediction of ship fuel consumption, which will be represented using a multiple linear regression equation based on Ober [31], Nosal and Miranda [32], and Wu and Liao [33]. The fuel consumption model is defined as follows:

$$\text{fuel} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon, \quad (8)$$

where $\beta_1, \beta_2, \dots, \beta_n$ is the quantification factor, and ε is the error term.

3.8. Carbon Emission Estimation. Since there is no significant carbon emission difference between heavy fuel oil (HFO) and light fuel oil (LFO), the carbon emission can be estimated as one single equation [34]:

$$\text{CarbonEmission} = 3.114 * \text{fuel}. \quad (9)$$

3.9. Database and Evaluation Criteria. In this study, AIS data, meteorological data, and MRV daily data will be used to construct a quantitative prediction model for ship fuel consumption. The original dataset consists of all AIS data, meteorological data, and MRV daily data for a total of eight container vessels from 2020-8-1 to 2022-2-1, consisting of a total of 801665 samples and 14 characteristic variables. Among them, the characteristic variables affecting the fuel

TABLE 1: Data information.

MMSI	Postime	Time_diff (s)	Distance (km)	Speed (kn)	HDG	...	MRV_Daily_Fuel (tons)	AIS_Fuel (tons)
477269400	2020/07/31 16:07:39 + 00:00	904	0.00	0	354	...	11.707	0.0607
477194800	2020/07/31 16:08:41 + 00:00	828	4.51	18.54	50	...	91.995	46.422
477183900	2020/07/31 16:08:56 + 00:00	850	3.38	14.56	60	...	41.950	35.177
...

consumption of the vessels are sailing time, sailing distance, sailing speed, wind and wave magnitude, wind and wave direction, and draft and sailing direction. The correlation between the characteristic variables was assessed for each degree of covariance before the modelling was carried out. Figure 4 illustrates the degree of correlation between the characteristic variables of the raw data.

It can be seen from Figure 4 that the size of the waves, sailing distance, and speed have significant effects on the fuel consumption of the ship. The draught reflects the cargo carried by the ship to a certain extent, and usually, the larger the cargo carried, the deeper the draught of the ship will be, and the fuel consumption of the ship will increase. Table 2 explains all the characteristic variables involved in the discussion and use. From Table 3, most of the characteristic variables have strong co-collinearity problems among them, which can easily cause the traditional regression model constructed to be unstable, resulting in less accurate prediction results. Therefore, ridge regression will be used to build a quantitative prediction model for ship fuel consumption.

Based on AIS data, meteorological data, and MRV daily data, this study establishes daily ship fuel consumption prediction and ship fuel consumption prediction under different weather conditions based on MRV daily fuel consumption and AIS fuel consumption, respectively.

The effects of the different models will be evaluated using mean absolute error (MAE), root mean squared error (RMSE), and R square method. MAE refers to the mean absolute error between the predicted outcome and the true value of the test dataset and is defined as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2. \quad (10)$$

RMSE is used to measure the deviation between the predicted and true values and is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}. \quad (11)$$

R square reflects the extent to which the regression model explains the variation in the dependent variable, or how good the fit to the observations is, and is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (12)$$

where y_i is a predicted value, y'_i is a real value, \bar{y} is a mean value, and n is the number of samples.

4. Fuel Consumption Model Constructions

4.1. Fuel Consumption Model Based on MRV Daily Fuel. As the MRV daily reports can only provide the daily fuel consumption of the vessel, there is no AIS data point-by-point fuel consumption corresponding to each AIS point and meteorological data and MRV fuel consumption. Therefore, the ship voyage data and meteorological data were averaged, and the average of each feature vector for each day was used as a representative of the day's meteorological and voyage data and input into the ridge regression model. The processed data set was divided into a training set and a test set. 80% of the data were used for training the model, and the remainder was used for testing the model, which was generated by random selection from the original data set.

In this study, ship fuel consumption was used as the model response, and the variables with high correlation and importance to ship fuel consumption were selected as the input to the model by combining the knowledge of ship navigation operations. All the characteristic variables will be directly input into the ridge regression model or pre-processed with Z-score standardization before inputting into the ridge regression model to compare the quantified results, and the comparison of data before and after standardization is shown in Tables 4 and 5. The regularization parameter α of the ridge regression model was estimated by 10-fold cross-validation according to 3.4. The training dataset was divided into 10 datasets of approximately equal sample size, and then, one of the datasets was used for 10 training sessions. The quantification results are shown in Table 6.

Table 6 shows the quantitative results and the results of the evaluation of model effects for Models 1 and 2. Model 1 uses unnormalized data, and the quantified results show that wave size, sailing distance, and speed magnitude have a significant effect on fuel consumption, which is consistent with the degree of correlation between the different characteristic variables and fuel consumption as shown in Section 3.5. From the MAE, RMSE, and R^2 values, the model that uses unstandardized daily average values to quantify and estimate fuel consumption for the MRV daily newspaper works well; the standardized data used in model 2 give similar results to model 1 from the quantified results. However, the values of MAE and RMSE for model 2 are significantly lower compared to model 1, which reflects the optimization of the model effect from the data point of view, but due to the large change in the size of the data set after normalization, it is not possible to conclude which model is more effective from the two evaluation indicators, MAE, and MRSE, and also, the R^2 of model 2 has decreased compared

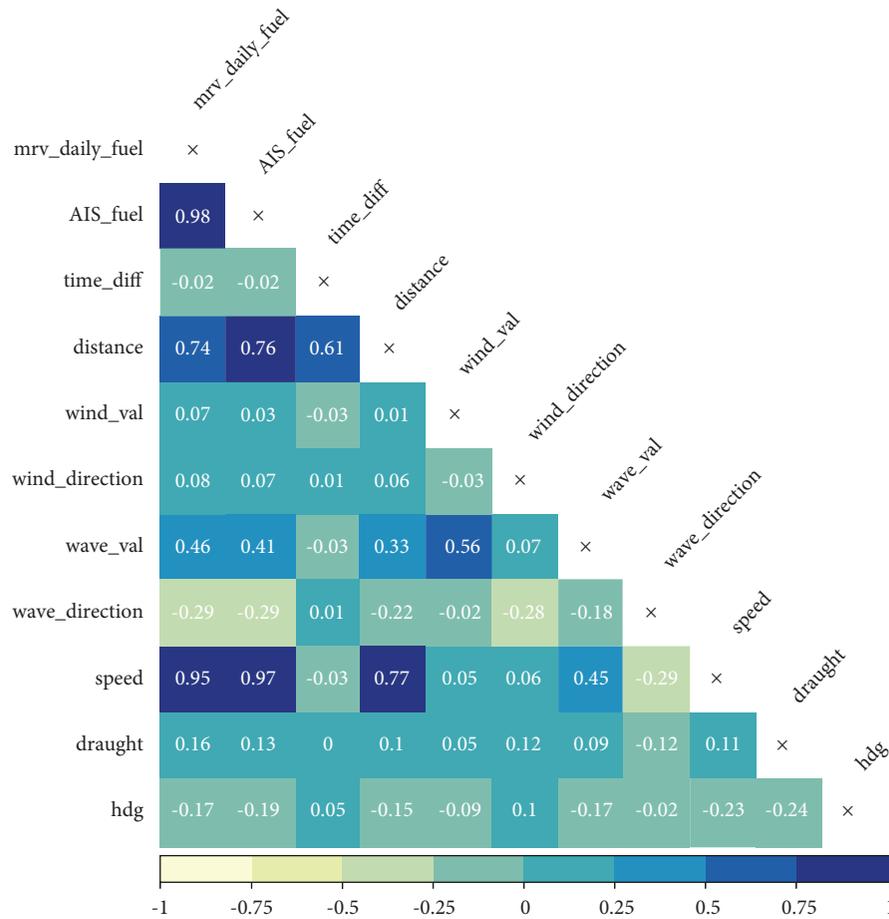


FIGURE 4: The correlation of the feature variables.

TABLE 2: Explanation of the characteristic variables used.

The characteristic variables	Explanation
Mrv_daily	Vessel’s daily sailing fuel consumption recorded by MRV daily report, in tons/day
AIS_fuel	Vessel’s sailing fuel consumption during the time interval between a certain AIS message and its previous AIS message, in tons
Time_diff	The time difference between the upload time of an AIS message and its previous AIS message, in seconds
Distance	The distance travelled by a ship during the time interval between an AIS message and its previous AIS message upload, in kilometers
Wind_val	Wind size in m/s
Wind_direction	Wind direction
Wave_val	Wave height in meters
Wave_direction	Wave direction
Speed	Vessel’s sailing speed, in knot
Draught	Draught of the ship, in meters
Hdg	The direction corresponding to the bow of the ship when it is sailing

TABLE 3: Colinearity among the feature variables.

Distance (km)	Time_diff (s)	HDG	Speed (kn)	Draught (m)	Wind_val (m/s)	Wind_direction	Wave_val (m)	Wave_Direction
79.35079	113.3095	5.21142	65.50859	82.15992	9.34322	5.57543	5.24874	6.35917

TABLE 4: Data before standardization.

MRV_Daily_Fuel (tons)	Distance (km)	Wind_val (m/s)	Wave_val (m)	Speed (kn)	Draught (m)	HDG
78.9	263.28	10.32	2.66	15.97	12.3	106.64
76.3	392.84	8.78	2.27	15.98	12.3	106.73
74.4	374.03	7.46	1.70	15.85	12.3	109.45
80.8	393.33	9.00	1.75	16.05	12.3	92.78
78.6	386.69	8.24	2.06	15.74	12.3	89.22
...

TABLE 5: Data after standardization.

MRV_Daily_Fuel (tons)	Distance (km)	Wind_val (m/s)	Wave_val (m)	Speed (kn)	Draught (m)	HDG
0.2523	-0.1090	1.4259	1.6455	0.5670	-0.8699	-0.7439
0.1917	0.4704	0.8820	1.2152	0.5686	-0.8699	-0.7429
0.1474	0.3863	0.4157	0.5944	0.5506	-0.8699	-0.7128
0.2966	0.4726	0.9597	0.6481	0.5789	-0.8699	-0.8973
0.2453	0.4429	0.6899	0.9871	0.5364	-0.8699	-0.9367
...

TABLE 6: Quantitative results for model 1, model 2, MAE, RMSE, R^2 .

	Wind_val (m/s)	Wave_val (m)	Speed (kn)	Draught (m)	Distance (km)	HDG	Error	MAE	RMSE	R^2
Model 1	0.0981	1.8071	5.3721	1.8292	0.0465	0.0267	-29.5268	10.6234	13.2521	0.9042
Model 2	0.0064	0.0384	0.9170	0.0591	0.0053	0.0563	-0.0012	0.2470	0.3104	0.9034

to that of model 1. Therefore, model 1 is more effective compared to model 2.

4.2. Fuel Consumption Model Based on AIS Fuel Data

4.2.1. Meteorological Classification. Model 1 and Model 2 use the average values of daily sailing and meteorological data as input to the model, which is more granular and does not provide a more detailed description of the impact of sailing and meteorological data on the fuel consumption of a ship in different weather conditions. The AIS fuel consumption data, on the other hand, shows the fuel consumption of the ship at the time each AIS data is received and contains more detailed and granular sailing and meteorological data. By using the more granular AIS navigational, meteorological, and fuel consumption data as input, a quantitative model for estimating the fuel consumption of a ship under different weather conditions will be obtained, improving the accuracy of the quantitative estimation.

This study uses the K -means clustering algorithm to classify meteorological data. As K -means requires the number of clusters to be set manually, the optimal number of clusters for clustering the meteorological data was explored in advance using the elbow method and silhouette score before clustering the meteorological data [35]. From Figure 5, it can be seen that for using the elbow rule, when the K value increases from 3 to 4, the rate of decline is slower, indicating that increasing the value of K at this time can no longer make the clustering effect significantly better, and there is no need to continue to increase the clustering; for the silhouette score, the maximum value of k for the silhouette score is 2, which indicates that the optimal number of clusters is 2. However, from the elbow rule of k -value and

SSE, we can see that when the k -value is 2, the SSE is still very large, so this is a less reasonable number of clusters. Therefore, the second largest K -value of the silhouette score is chosen -3, when the SSE is already at a relatively low level. Therefore, the number of clusters was set to 3, and the weather data were clustered using K -means. As can be seen in Figure 5, after using K -means, the wind and waves were effectively divided into three categories. The wind size was classified into three categories: smaller than 5 m/s, 5–11 m/s, and bigger than 11 m/s. The wave size was classified into three categories: smaller than 1 m, 1–2.5 m, and bigger than 2.5 m. In conjunction with the wind and wave classification table (Table 7) [36], the meteorological data were classified into three categories of meteorological conditions: smaller than level 4 wind-level 3 waves, between level 4 wind-level 3 waves and level 6 wind-level 5 waves, and bigger than level 6 wind-level 5 waves.

When a ship is sailing, the direction of wind and waves is also changing at any time. COSCO shipping group uses the classification in Figure 6 to divide the direction of wind and waves into three divisions: downwind, sideways headwind, and upwind. $0^\circ - 30^\circ$ is downwind, $30^\circ - 150^\circ$ is sideways headwind, and $150^\circ - 180^\circ$ is upwind. Therefore, the weather size classification is combined with the weather direction classification in this study to build a quantitative estimation model for ship fuel consumption in different weather conditions.

4.2.2. Fuel Consumption Model for Different Weather Classifications. A total of nine weather categories can be arranged according to the classification of weather size and direction. Using the data under each category of weather, modelling is carried out to obtain a quantitative estimation model of fuel consumption corresponding to each category

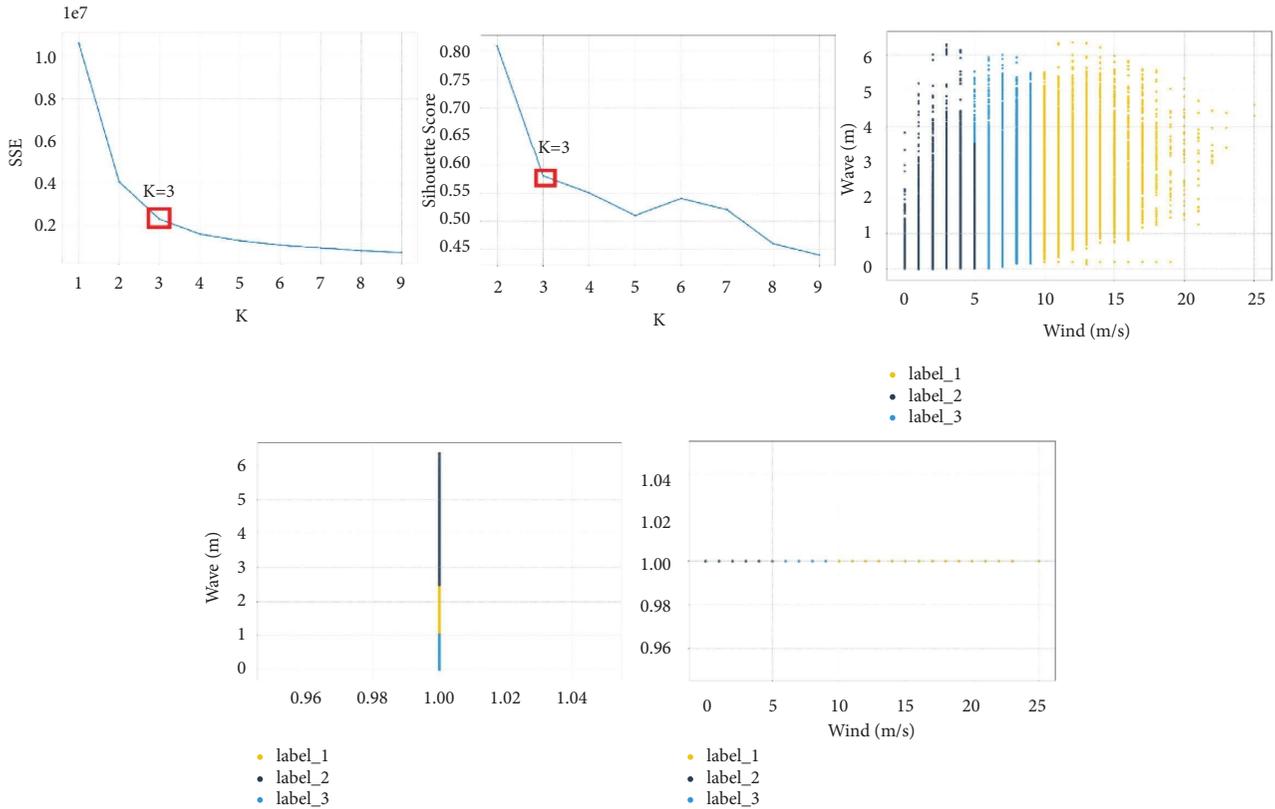


FIGURE 5: Meteorological data classification.

TABLE 7: Wind/wave scale.

Level	Wind_val (m/s)	Wave_val (m)
0	[0–0.3]	0
1	(0.3–1.6]	[0–0.1]
2	(1.6–3.4]	(0.1–0.3]
3	(3.4–5.5]	(0.3–1.0]
4	(5.5–8.0]	(1.0–1.5]
5	(8.0–10.8]	(1.5–2.5]
6	(10.8–13.9]	(2.5–4.0]
7	(13.9–17.2]	(4.0–5.5]
8	(17.2–20.8]	(5.5–7.0]
9	(20.8–24.5]	(7.0–10.0]
10	(24.5–28.5]	(10.0–16.0]

of weather, which is used to estimate the fuel consumption corresponding to each AIS point.

Among them, model 3, model 4, and model 5 belong to the category of good weather. Model 3, when the meteorological size falls within level 4 winds-level 3 waves and the wind and wave zoning is downwind, is regarded as good weather. The AIS data, meteorological data, and AIS fuel consumption that satisfy the conditions of model 3 are entered into the model, and the optimal α value is further determined by 10-fold cross-validation to obtain the quantified coefficients corresponding to the characteristic variables. As can be seen from Table 8, the MAE for model 3 is 0.08, the RMSE is 0.306, and the R^2 is 0.89, indicating that the model for ship fuel consumption in good weather corresponding to model 3 works well.

Model 4, when the meteorological size falls within level 4 winds-level 3 waves and the wind and wave zoning falls within the side headwinds, is also regarded as good weather. The results are shown in Table 8, with MAE of 0.23, RMSE of 0.88, and R^2 of 0.79, indicating that the model for fuel consumption in good weather for model 4 is effective. Model 5, when the meteorological size belongs to level 4 winds-level 3 waves and the wind and wave zoning belongs to upwind, is regarded as the second best weather; the data satisfying the conditions of model 5 are input into the model, and the results as shown in Table 8 can be obtained, the MAE is 0.12, the RMSE is 0.34, and the R^2 is 0.81, which indicates that the ship fuel consumption model under the second best weather corresponding to model 5 works well.

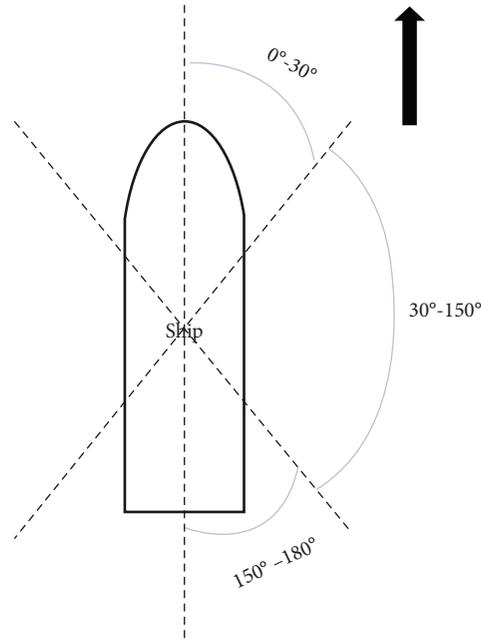


FIGURE 6: Wave/wind direction relative to ship's movement.

TABLE 8: Fuel consumption model in good weather.

	Wind_val (m/s)	Wave_val (m)	Speed (kn)	Draught (m)	Distance (kn)	HDG	Error	MAE	RMSE	R^2
Model 3	0.0014	-0.0007	0.0008	0.0139	0.2041	0.00008	-0.1779	0.08	0.306	0.89
Model 4	0.0041	-0.0117	0.0139	-0.0044	0.1412	0.00009	0.0487	0.23	0.880	0.79
Model 5	0.0061	-0.0233	0.0129	0.0091	0.1700	0.00016	-0.1676	0.12	0.340	0.81

Model 6, model 7, and model 8 fall into the general weather category. Model 6, when the meteorological size is between level 4 winds-level 3 waves and level 6 winds-level 5 waves and the wind and wave partition belong to downwind, is regarded as general weather, and the data satisfying model 6 is input into the model, and the same 10-fold cross-validation is used to further determine the value of the best α . As can be seen from Table 9, the MAE of model 6 is 0.09, the RMSE is 0.216, and the R^2 is 0.94, indicating that the model 6 corresponds to the corresponding model for ship fuel consumption in general weather works well. Model 7, when the meteorological size is between level 4 winds-level 3 waves and level 6 winds-level 5 waves and the wind and wave partition is a side headwind, is also regarded as general weather. The results shown in Table 9 are obtained when the data for model 7 are entered into the model. The MAE is 0.08, the RMSE is 0.23, and the R^2 is 0.903, indicating that the ship fuel consumption model for model 7 works well in normal weather. Model 8, when the meteorological size is between level 4 winds-level 3 waves and level 6 winds-level 5 waves and the wind and wave partition is upwind, is also regarded as general weather; the data satisfying the conditions of model 8 are input into the model, and the results shown in Table 9 can be obtained, with MAE of 0.09, RMSE of 0.305, and R^2 of 0.873, indicating that the ship fuel

consumption model under general weather corresponding to model 8 is effective.

Model 9, model 10, and model 11 belong to the category of severe weather. Model 9, when the meteorological size exceeds level 6 winds-level 5 waves and the wind and wave partition belongs to downwind, is regarded as severe weather. The data satisfying model 9 is input into the model, and the best α value is further determined by the same 10-fold cross-validation, as can be seen from Table 10, the MAE of model 9 is 0.04, the RMSE is 0.063, and the R^2 is 0.93, indicating that the severe weather corresponding to model 9 ship fuel consumption model works well. Model 10, when the meteorological magnitude exceeds between level 6 winds and level 5 waves, and the wind and wave partition is a side headwind, is also regarded as severe. The results are shown in Table 10, with MAE of 0.04, RMSE of 0.079, and R^2 of 0.923, indicating that the fuel consumption model for model 10 in severe weather is working well. Model 11, when the weather size exceeds between level 6 winds and level 5 waves and the wind and wave partition is upwind, is also regarded as bad weather; the data to meet the conditions of model 11 are input into the model, and the results shown in Table 10 can be obtained, with MAE of 0.04, RMSE of 0.086, and R^2 of 0.887, indicating that the ship fuel consumption model under bad weather corresponding to model 11 works well.

TABLE 9: Fuel consumption model in normal weather.

	Wind_val (m/s)	Wave_val (m)	Speed (kn)	Draught (m)	Distance (km)	HDG	Error	MAE	RMSE	R ²
Model 6	-0.0044	0.0103	0.0060	0.0080	0.1749	-0.00020	-0.0382	0.09	0.216	0.94
Model 7	-0.0032	0.0159	0.020	0.0131	0.2089	-0.00010	-0.1333	0.23	0.880	0.79
Model 8	0.0014	-0.0021	0.0079	-0.0058	0.2063	0.00012	0.0643	0.09	0.305	0.87

TABLE 10: Fuel consumption model in bad weather.

	Wind_val (m/s)	Wave_val (m)	Speed (kn)	Draught (m)	Distance (km)	HDG	Error	MAE	RMSE	R ²
Model 9	0.0028	0.00008	0.0041	-0.0203	0.1424	0.00011	0.2069	0.04	0.063	0.93
Model 10	0.0056	0.00083	0.0020	-0.0112	0.1861	0.00015	0.0856	0.04	0.079	0.92
Model 11	-0.0026	-0.0347	-0.0003	0.0199	0.2226	0.00005	-0.2134	0.04	0.086	0.89

5. Experimental Results

To demonstrate the performance of the proposed ridge regression-based fuel consumption prediction model, it was compared with typical regression methods such as ANN.

As shown in Figure 7, the actual fuel consumption, the prediction results based on ridge regression, and the ANN-based prediction results are shown by black, yellow, and blue lines, respectively. It is obvious that the ridge regression-based fuel consumption prediction model can accurately fit the actual values in most cases due to the ANN. Figure 7 shows that the prediction results of the ridge regression-based fuel consumption prediction model can accurately fit the actual values in most cases.

Table 11 shows the RMSE and MAE results for the ridge regression model and ANN with the same test data set. As expected, the ridge regression model has a lower RMSE and MAE of 13.25 and 10.62, respectively. The lower RMSD implies that the proposed model has a better fitting performance and can accurately predict the fuel consumption variation under different navigation conditions.

Finally, these models were completed in Python 3.9.7 and with normal PC specifications (i5 processor). As shown in Table 11, the efficiency of the two models was evaluated based on training time. The time cost of the ridge regression-based model is lower than the time cost of the other model. In addition, the ridge regression model has a better interpretation compared to other regression models such as ANN, which is convenient for subsequent studies to explore the potential association of each different characteristic variable with fuel consumption and carbon emissions.

6. Discussion

6.1. Fuel Consumption Comparison. To obtain a more accurate ship fuel consumption prediction model, a ship fuel consumption quantification estimation model based on MRV daily fuel consumption and a ship fuel consumption quantification model based on AIS fuel consumption were constructed, respectively. The corresponding ship fuel consumption quantification equations under the two models can be obtained from Section 4.

$$\text{fuel} = \sum_{i=1}^6 \beta_i x_i + \varepsilon, \text{ model1}, \tag{13}$$

$$\text{fuel} = \begin{cases} \sum_{i=1}^6 \beta_{1i} X_i + \varepsilon_1, \text{ model3}, \\ \sum_{i=1}^6 \beta_{2i} X_i + \varepsilon_2, \text{ model4}, \\ \sum_{i=1}^6 \beta_{3i} X_i + \varepsilon_3, \text{ model5}, \\ \sum_{i=1}^6 \beta_{4i} X_i + \varepsilon_4, \text{ model6}, \\ \sum_{i=1}^6 \beta_{5i} X_i + \varepsilon_5, \text{ model7}, \\ \sum_{i=1}^6 \beta_{6i} X_i + \varepsilon_6, \text{ model8}, \\ \sum_{i=1}^6 \beta_{7i} X_i + \varepsilon_7, \text{ model9}, \\ \sum_{i=1}^6 \beta_{8i} X_i + \varepsilon_8, \text{ model10}, \\ \sum_{i=1}^6 \beta_{9i} X_i + \varepsilon_9, \text{ model11}. \end{cases} \tag{14}$$

For equation (13), $x_i (i = 1, 2, \dots, 6)$ is the characteristic variables in model 1, $\beta_i (i = 1, 2, \dots, 6)$ is the characteristic coefficients, and ε is the error term. Equation (13) shows the quantitative estimation model based on the MRV daily fuel consumption.

For equation (14), $\beta_{ji} (i = 1, 2, \dots, 6; j = 1, 2, \dots, 9)$ is the characteristic coefficients corresponding to the characteristic variables in each model, and $\varepsilon_j (j = 1, 2, \dots, 9)$ is the error term of each model. Equation (14) shows the quantitative estimation model based on the AIS fuel consumption. The values and characteristic variables corresponding to each parameter are shown in Table 12.

In order to validate the proposed work, ship fuel consumption models based on MRV data and AIS data will be

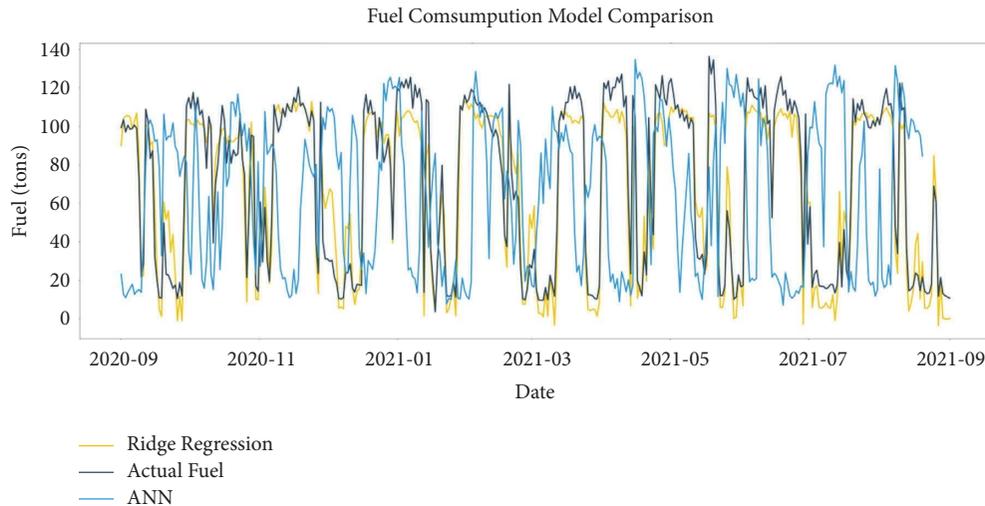


FIGURE 7: Fuel consumption model comparison.

TABLE 11: RMSE, MAE results and running time for the ridge regression model and ANN model.

Model	RMSE	MAE	Running time (s)
ANN	21.5	17.36	30.2
Ridge regression	13.25	10.62	15.9

compared. Using the obtained quantitative formulae, the fuel consumption of four container vessels will be estimated, and the results of the two models will be compared with the predicted fuel consumption and the real fuel consumption demonstrated by COSCO Haikē, respectively.

Figures 8–11 show the results of the quantitative estimation model for ship fuel consumption based on MRV daily fuel consumption, shown by the yellow line. Compared to the actual fuel consumption (black line), there is roughly the same trend, and for most of the time, the difference with the real fuel consumption is smaller; compared to COSCO's predicted fuel consumption (blue line), MRV has a smoother trend for some of the time and, for some of the time, is closer to the real fuel consumption. Figures 12–15 show the results of the quantitative estimation model for ship fuel consumption based on AIS fuel consumption. The model based on AIS fuel consumption is closer to the real fuel consumption than the model based on MRV daily fuel consumption. The trend is more reasonable and more accurate than COSCO's predicted fuel consumption. As the model considers different weather conditions, it is more adaptable to different weather conditions than the model based on MRV daily fuel consumption and COSCO's prediction model and more adaptable to the changing weather conditions at sea.

Table 13 shows the annual (2020/09–2021/09) actual cumulative fuel consumption of the four container ships and the cumulative predicted fuel consumption of the two models, as well as the absolute error between the two models and the annual actual cumulative fuel consumption.

For the quantitative estimation model based on MRV daily fuel consumption, when selecting the daily weather and navigation data, the mean value of each characteristic variable is used as the daily representative, so there are many cases where the original characteristics of the weather and navigation data are lost, for example, no wind and fast speed in the first few hours, and gusty wind and little speed in the second few hours, so there are still many shortcomings in the performance of the estimation results; in addition, the model utilizes daily average data, which behaves differently in the face of different weather conditions, and cannot more accurately express the correspondence between the characteristic variables and fuel consumption through each characteristic coefficient.

For the quantitative estimation model based on AIS fuel consumption, although each type of weather condition is considered separately, there is a large amount of unreasonable data in the original dataset that is difficult to process because AIS fuel consumption is filled in manually; in addition, the consideration of weather conditions is not refined enough, and the weather classification in this study cannot be better adapted to the estimation of fuel consumption under various different weather variations. Therefore, the estimation results of the model still have some errors with the actual values.

6.2. Carbon Emissions Comparison. Figures 16–19 show the results of the ship's carbon emissions based on MRV daily fuel consumption, indicated by the yellow dashed line. There is approximately the same trend compared to the real (black

TABLE 12: Equations (13) and (14) each parameter explanation.

	Model 1	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Wind_val	$\beta_1 = 0.0981$	$\beta_{11} = 0.0014$	$\beta_{21} = 0.0041$	$\beta_{31} = 0.0061$	$\beta_{41} = -0.0044$	$\beta_{51} = -0.0032$	$\beta_{61} = 0.0014$	$\beta_{71} = 0.0028$	$\beta_{81} = 0.0056$	$\beta_{91} = -0.0026$
Wave_val	$\beta_2 = 1.8071$	$\beta_{12} = -0.0007$	$\beta_{22} = -0.0117$	$\beta_{32} = -0.0233$	$\beta_{42} = 0.0103$	$\beta_{52} = 0.0159$	$\beta_{62} = -0.0021$	$\beta_{72} = 0.00008$	$\beta_{82} = 0.00083$	$\beta_{92} = -0.0347$
Speed	$\beta_3 = 5.3721$	$\beta_{13} = 0.0008$	$\beta_{23} = 0.0139$	$\beta_{33} = 0.0129$	$\beta_{43} = 0.0060$	$\beta_{53} = 0.02$	$\beta_{63} = 0.0079$	$\beta_{73} = 0.0041$	$\beta_{83} = 0.002$	$\beta_{93} = -0.0003$
Draught	$\beta_4 = 1.8292$	$\beta_{14} = 0.0139$	$\beta_{24} = -0.0044$	$\beta_{34} = 0.0091$	$\beta_{44} = 0.0080$	$\beta_{54} = 0.0131$	$\beta_{64} = -0.0058$	$\beta_{74} = -0.0203$	$\beta_{84} = -0.0112$	$\beta_{94} = 0.0199$
Distance	$\beta_5 = 0.0465$	$\beta_{15} = 0.2041$	$\beta_{25} = 0.1412$	$\beta_{35} = 0.1700$	$\beta_{45} = 0.1749$	$\beta_{55} = 0.2089$	$\beta_{65} = 0.2063$	$\beta_{75} = 0.1424$	$\beta_{85} = 0.1861$	$\beta_{95} = 0.2226$
Hdg	$\beta_6 = 0.0267$	$\beta_{16} = 0.00008$	$\beta_{26} = 0.00009$	$\beta_{36} = 0.00016$	$\beta_{46} = -0.0002$	$\beta_{56} = -0.0001$	$\beta_{66} = 0.00012$	$\beta_{76} = 0.00011$	$\beta_{86} = 0.00015$	$\beta_{96} = 0.00005$
Error term	$\varepsilon = -29.5268$	$\varepsilon_1 = -0.1779$	$\varepsilon_2 = 0.0487$	$\varepsilon_3 = -0.1676$	$\varepsilon_4 = -0.0382$	$\varepsilon_5 = -0.0382$	$\varepsilon_6 = 0.0643$	$\varepsilon_7 = 0.2069$	$\varepsilon_8 = 0.0856$	$\varepsilon_9 = 0.2134$

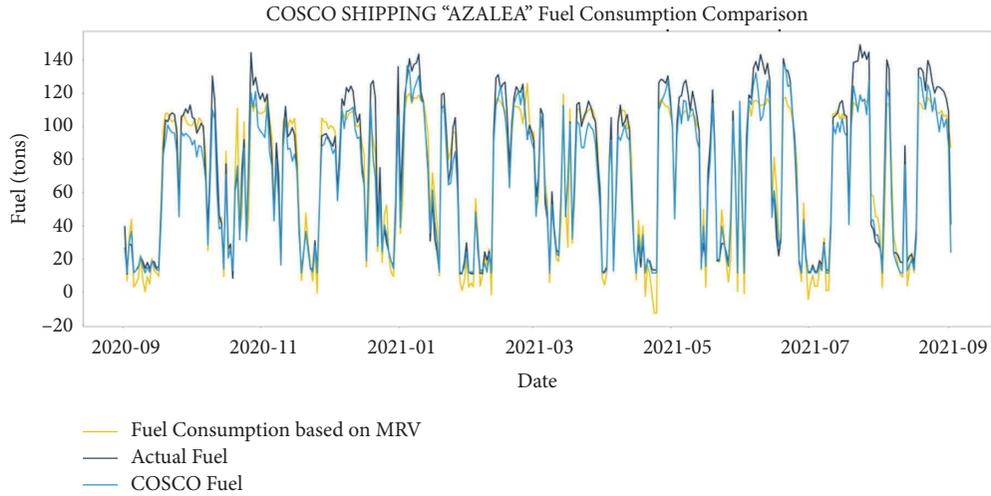


FIGURE 8: “AZALEA” fuel consumption comparison based on regression model (MRV-based), actual fuel(MRV report), and COSCO fuel (AIS-based).

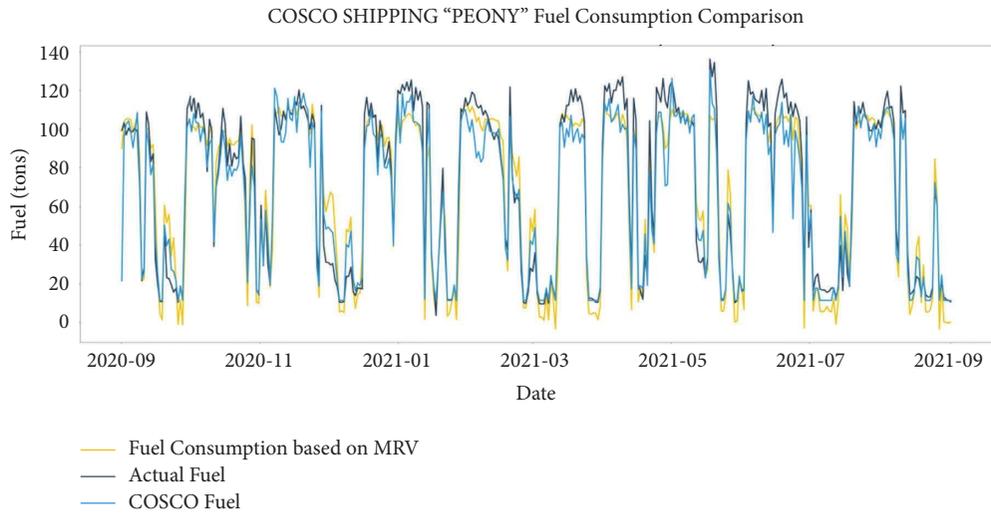


FIGURE 9: “PEONY” fuel consumption model based on MRV daily fuel compare.

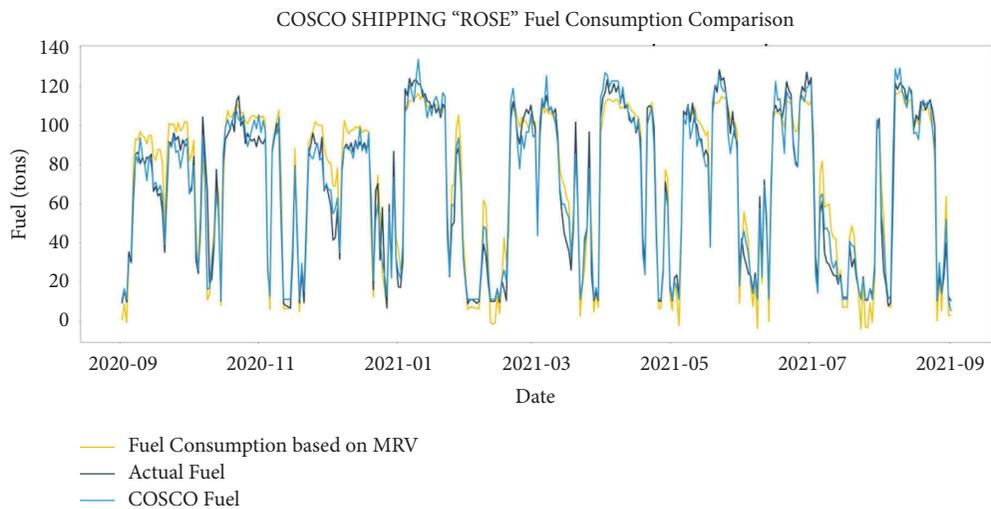


FIGURE 10: “ROSE” fuel consumption model based on MRV daily fuel compare.

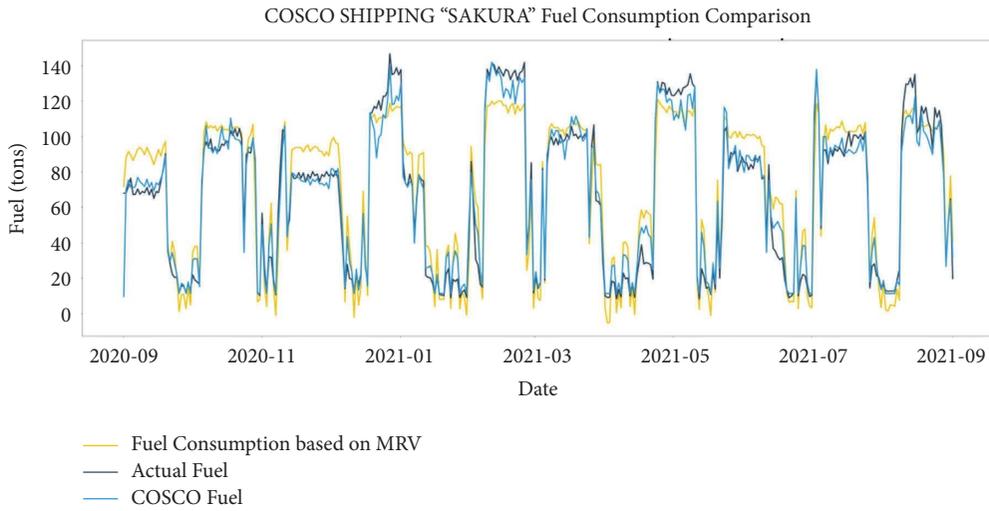


FIGURE 11: "SAKURA" fuel consumption model based on MRV daily fuel compare.

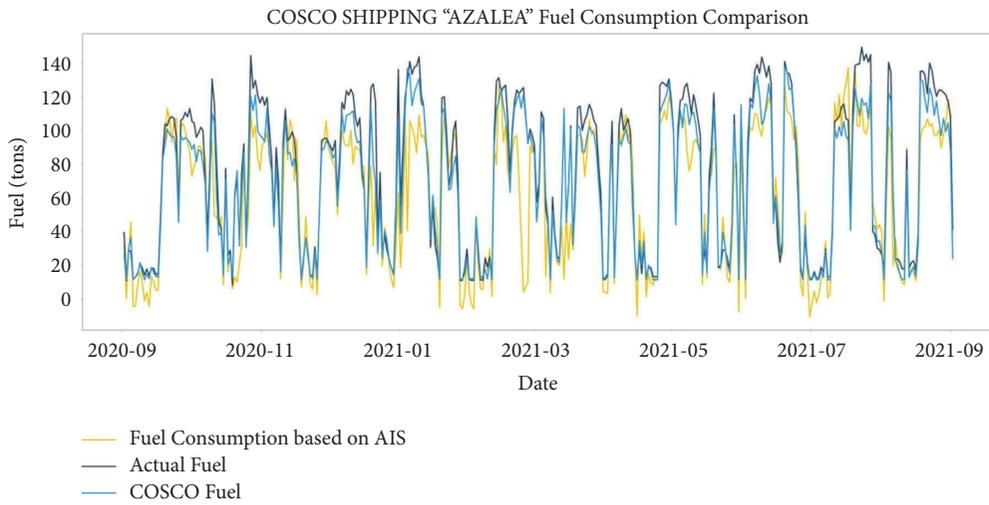


FIGURE 12: "AZALEA" fuel consumption comparison based on regression model (AIS-based), actual fuel(MRV report), and COSCO fuel (AIS-based).

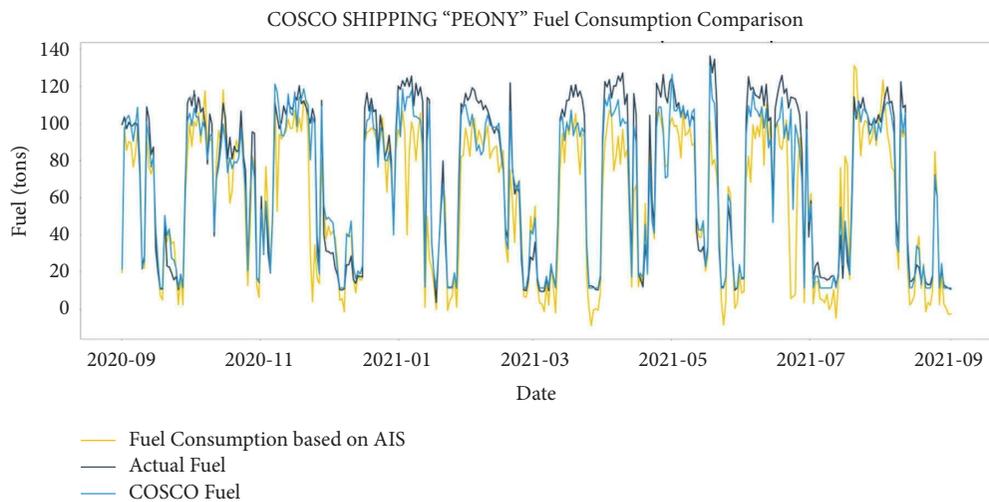


FIGURE 13: "PEONY" fuel consumption model based on AIS fuel compare.

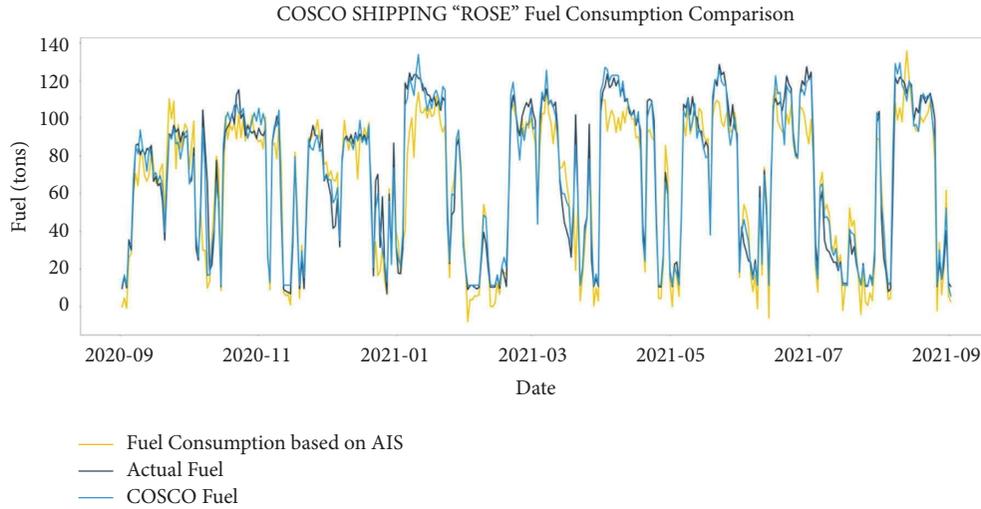


FIGURE 14: “ROSE” fuel consumption model based on AIS fuel compare.

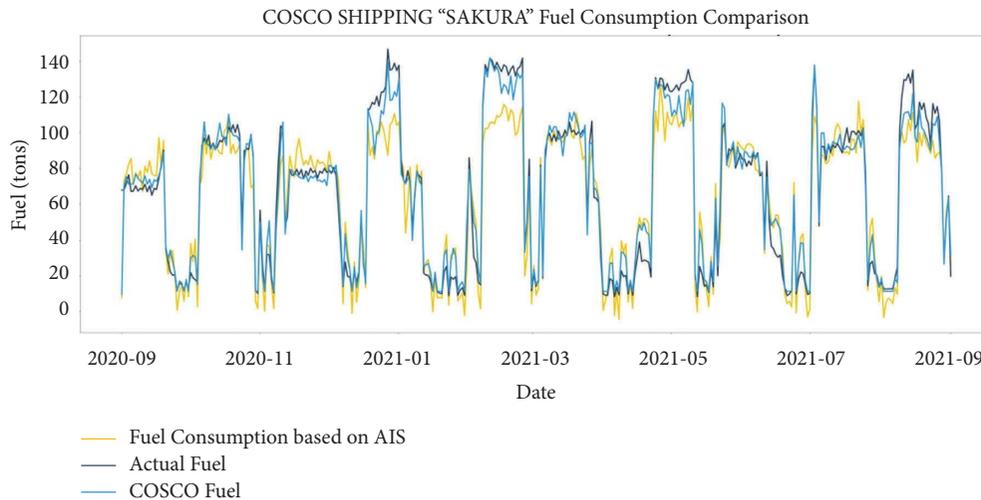


FIGURE 15: “SAKURA” fuel consumption model based on AIS fuel compare.

TABLE 13: Fuel consumption analyzation.

Ship name	Actual cumulative fuel consumption throughout the year (tons)	Cumulative prediction of fuel consumption estimation model based on MRV daily fuel consumption (tons)	Cumulative absolute error (%)	Cumulative prediction of fuel consumption estimation model based on AIS fuel consumption (tons)	Cumulative absolute error (%)
AZALEA	27339	26558	2.9	24655	9.4
PEONY	26619	25621	3.7	24144	8.9
ROSE	25652	26476	3.2	24108	6
SAKURA	25066	26235	4.7	23700	5.4

line), with a difference less than 10% compared to the real carbon emissions for most of the time; Figures 20–23 show the results for ship carbon emissions based on AIS carbon emissions. The model based on MRV is closer to the true CO₂ emission (less than 5%) than the model based on AIS.

Table 14 shows the annual (2020/09–2021/09) actual cumulative carbon emissions of the four container ships and

the cumulative predicted carbon emissions, as well as the absolute error between the two models and the annual actual cumulative carbon emissions.

It can be seen from Table 14 that the ship carbon emissions estimation model based on MRV daily fuel consumption is better than that based on AIS fuel consumption. However, in practical application, there may be

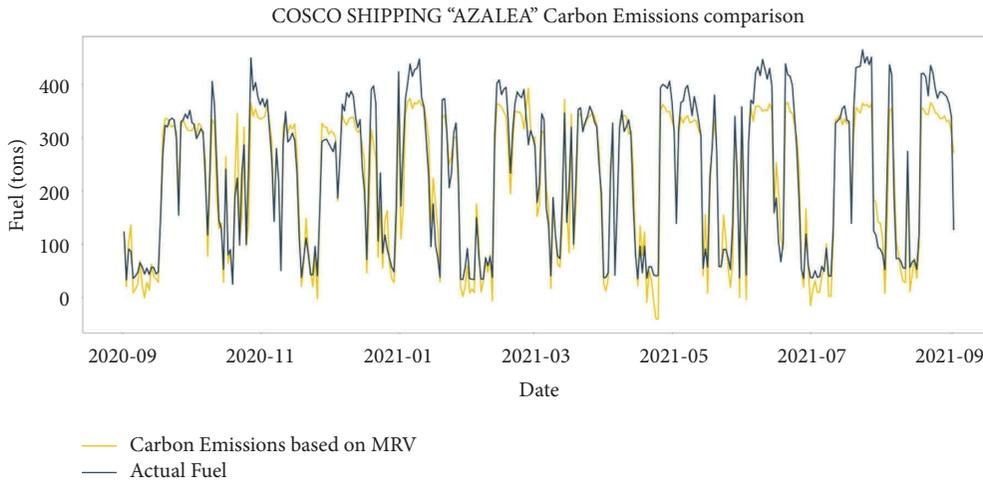


FIGURE 16: "AZALEA" carbon emissions comparison based on MRV.

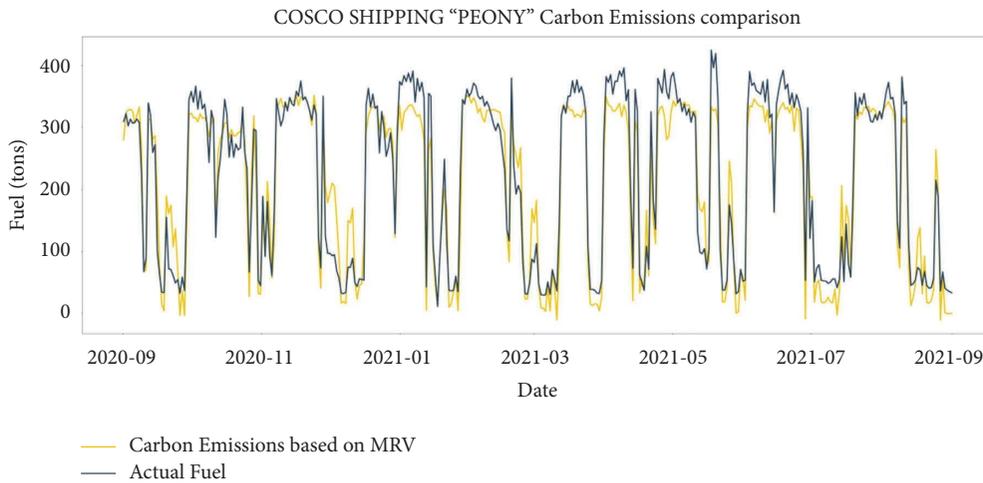


FIGURE 17: "PEONY" carbon emissions comparison based on MRV.

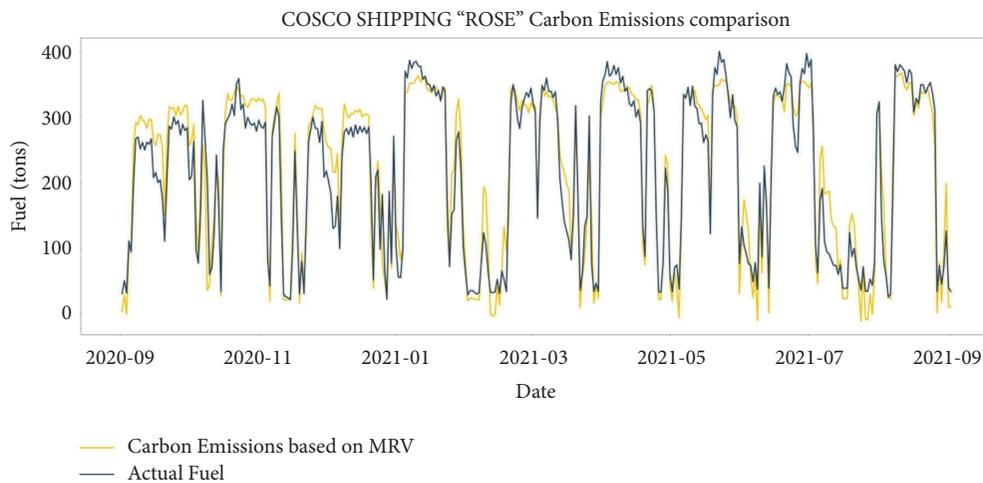


FIGURE 18: "ROSE" carbon emissions comparison based on MRV.

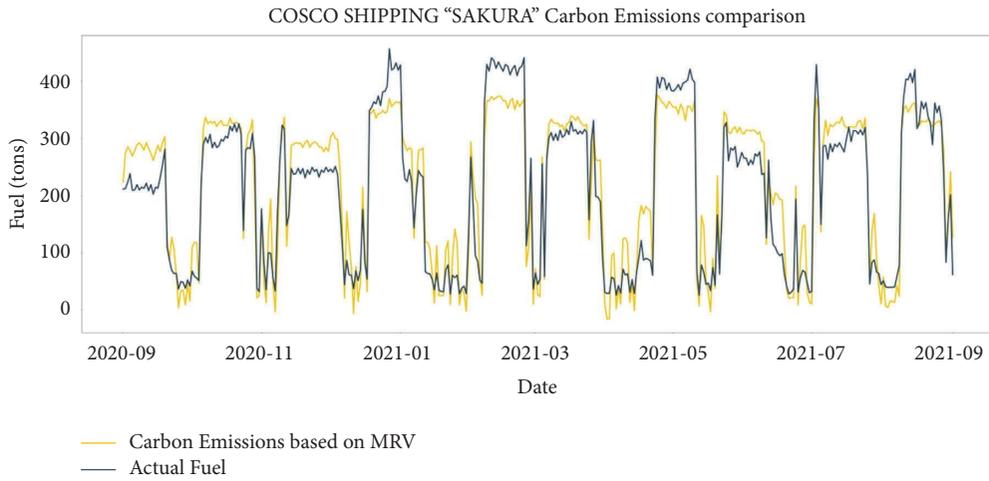


FIGURE 19: "SAKURA" carbon emissions comparison based on MRV.

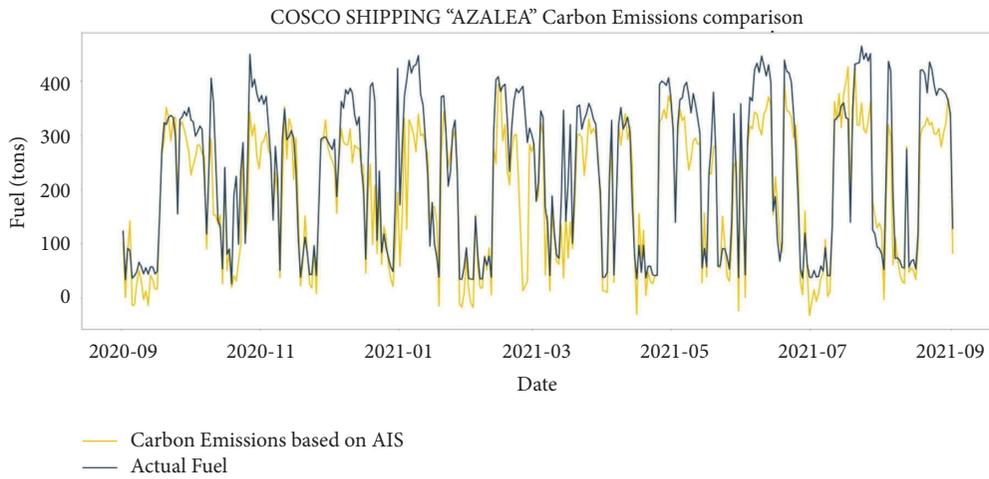


FIGURE 20: "AZALEA" carbon emissions comparison based on AIS.

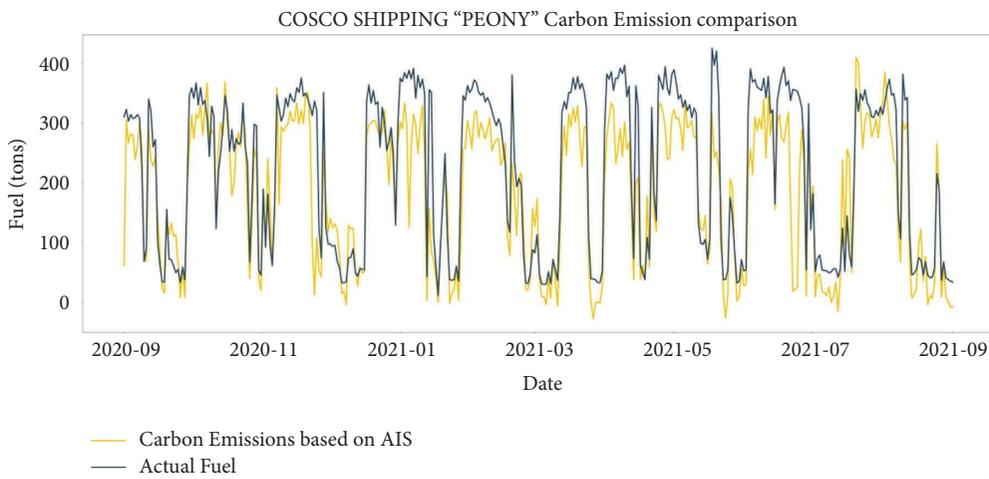


FIGURE 21: "PEONY" carbon emissions comparison based on AIS.

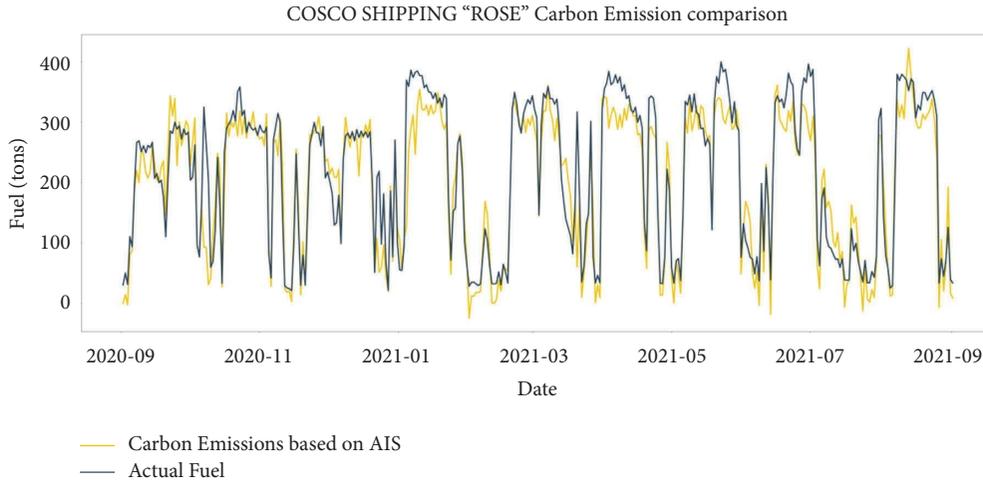


FIGURE 22: “ROSE” carbon emissions comparison based on AIS.

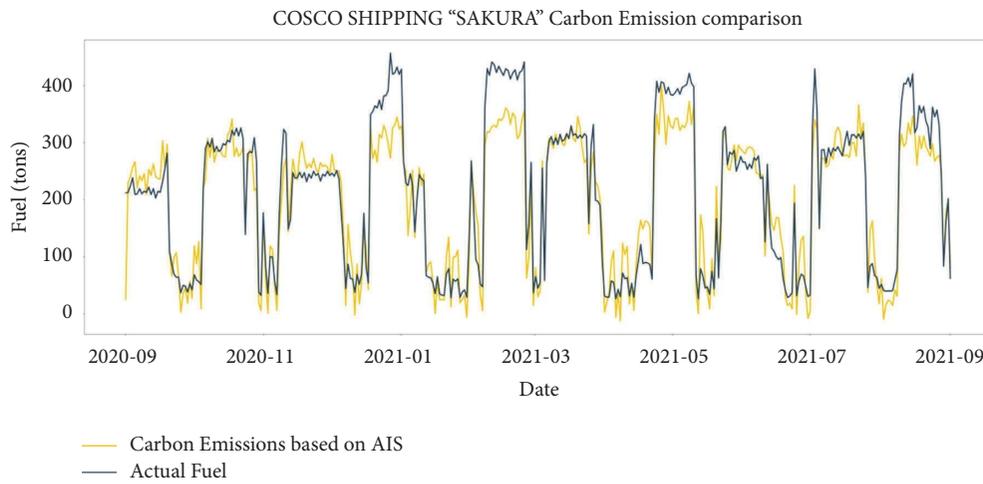


FIGURE 23: “SAKURA” carbon emissions comparison based on AIS.

TABLE 14: Carbon emission comparison based on MRV-based and AIS-based regression model.

Ship name	Actual cumulative carbon emissions throughout the year (tons)	Cumulative prediction of carbon emissions estimation model based on MRV daily fuel consumption (tons)	Cumulative absolute error (%)	Cumulative prediction of carbon emissions estimation model based on AIS fuel consumption (tons)	Cumulative absolute error (%)
AZALEA	85136	82701	2.9	76775	9.4
PEONY	82892	79784	3.7	75184	8.9
ROSE	79880	82446	3.2	75072	6
SAKURA	78055	81695	4.7	73801	5.4

AIS data loss. For the ship fuel consumption estimation model based on AIS fuel consumption, AIS time segment and distance are affected, which can cause the predicted fuel consumption deviation, resulting in inaccurate estimation of carbon emissions. At the same time, the ship fuel consumption estimation model based on AIS can adapt to the changeable weather and navigation conditions at sea and facilitate ship path planning and meteorological navigation.

In addition, some ships lack MRV daily report, using AIS fuel consumption data for prediction can better estimate and monitor ship fuel consumption.

7. Conclusion

In this work, a framework is shown from raw AIS data to ship fuel consumption and carbon emissions in different

weather conditions. Two models were constructed: one with daily MRV data and one with AIS data. Results show that both models can have good performance under each classified weather. At first, this study uses ridge regression to quantify ship fuel consumption, which reduces the impact of covariance between the characteristic variables and enhances the adaptability of the models to data sets with some ambiguity, low accuracy, and limited samples. Then, both models are effective in explaining and analyzing characteristics such as speed, distance travelled, wind and wave size, and ship draft, which have a significant impact on ship fuel consumption, enhancing the interpretability of the models.

The significance of this work is that rather than having the fuel consumption and carbon emission from daily noon report, it can be estimated between AIS segments. Therefore, this research will contribute the maritime transportation and the GHG emission reduction in 2 ways:

- (1) Optimizing the estimation models and applying the ship fuel consumption estimation models to route planning so that ship can have a good estimation of carbon emission before departure so that routes can be chosen nicely
- (2) Current MRV report contains a human error, and this model can be used as an early warning when an unexpected report is written

Data Availability

Data can be obtained upon request.

Consent

Not applicable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, vol. 27, no. 5, pp. 1413–1432, 2017.
- [2] S. N. Sirimanne, "Review of maritime transport 2019," in *Proceedings of the United Nations conference on trade and development*, Geneva, Switzerland, October 2019.
- [3] P. Cariou and A. Cheaitou, "The effectiveness of a European speed limit versus an international bunker-levy to reduce CO2 emissions from container shipping," *Transportation Research Part D: Transport and Environment*, vol. 17, no. 2, pp. 116–123, 2012.
- [4] A. Cheaitou and P. Cariou, "Liner shipping service optimisation with reefer containers capacity: an application to northern Europe–South America trade," *Maritime Policy & Management*, vol. 39, no. 6, pp. 589–602, 2012.
- [5] W. Kenton and E. Estevez, "International maritime organization (IMO)," 2019, <https://www.imo.org/>.
- [6] L. T. Le, G. Lee, K. S. Park, and H. Kim, "Neural network-based fuel consumption estimation for container ships in Korea," *Maritime Policy & Management*, vol. 47, no. 5, pp. 615–632, 2020.
- [7] L. Zhang, Q. Meng, Z. Xiao, and X. Fu, "A novel ship trajectory reconstruction approach using AIS data," *Ocean Engineering*, vol. 159, pp. 165–174, 2018.
- [8] S. Mao, E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G. B. Huang, "An automatic identification system (AIS) database for maritime trajectory prediction and data mining," in *Proceedings of ELM-2016*, pp. 241–257, Springer, Berlin, Germany, 2018.
- [9] X. Li, Y. Du, Y. Chen et al., "Data fusion and machine learning for ship fuel efficiency modeling: Part I–Voyage report data and meteorological data," *Communications in Transportation Research*, vol. 2, Article ID 100074, 2022.
- [10] Y. Du, Y. Chen, X. Li, A. Schonborn, and Z. Sun, "Data fusion and machine learning for ship fuel efficiency modeling: Part II–Voyage report data, AIS data and meteorological data," *Communications in Transportation Research*, vol. 2, Article ID 100073, 2022.
- [11] Y. Du, Y. Chen, X. Li, A. Schonborn, and Z. Sun, "Data fusion and machine learning for ship fuel efficiency modeling: Part III–Sensor data and meteorological data," *Communications in Transportation Research*, vol. 2, Article ID 100072, 2022.
- [12] M. Haranen, P. Pakkanen, R. Kariranta, and J. Salo, "White, grey and black-box modelling in ship performance evaluation," in *Proceedings of the 1st Hull Performance & Insight Conference (HullPIC)*, Castello di Pavone, Italy, April 2016.
- [13] J. R. Medina, J. Molines, J. A. G. Escrivá, and J. Aguilar, "Bunker consumption of container ships considering sailing speed and wind conditions," *Transportation Research Part D: Transport and Environment*, vol. 87, Article ID 102494, 2020.
- [14] F. Tillig, J. W. Ringsberg, H. N. Psaraftis, and T. Zis, "Reduced environmental impact of marine transport through speed reduction and wind assisted propulsion," *Transportation Research Part D: Transport and Environment*, vol. 83, Article ID 102380, 2020.
- [15] R. Yan, S. Wang, and H. N. Psaraftis, "Data analytics for fuel consumption management in maritime transportation: status and perspectives," *Transportation Research Part E: Logistics and Transportation Review*, vol. 155, Article ID 102489, 2021.
- [16] S. Wang and Q. Meng, "Sailing speed optimization for container ships in a liner shipping network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 3, pp. 701–714, 2012.
- [17] R. Adland, P. Cariou, and F. C. Wolff, "Optimal ship speed and the cubic law revisited: empirical evidence from an oil tanker fleet," *Transportation Research Part E: Logistics and Transportation Review*, vol. 140, Article ID 101972, 2020.
- [18] P. Karagiannidis, N. Themelis, G. Zaraphonitis, C. Spandonidis, and C. Giordamalis, "Ship fuel consumption prediction using artificial neural networks," in *Proceedings of the Annual meeting of marine technology conference proceedings*, Hellenic Institute of Marine Technology, Athens, Greece, November 2019.
- [19] T. Uyanık, Ç. Karatug̃, and Y. Arslanođlu, "Machine learning approach to ship fuel consumption: a case of container vessel," *Transportation Research Part D: Transport and Environment*, vol. 84, Article ID 102389, 2020.
- [20] R. Vettor and C. G. Soares, "Reflecting the uncertainties of ensemble weather forecasts on the predictions of ship fuel consumption," *Ocean Engineering*, vol. 250, Article ID 111009, 2022.
- [21] S. Wang, B. Ji, J. Zhao, W. Liu, and T. Xu, "Predicting ship fuel consumption based on LASSO regression," *Transportation*

- Research Part D: Transport and Environment*, vol. 65, pp. 817–824, 2018.
- [22] A. Bourazanis, “Probabilistic Analysis/estimation on Fuel Consumption and Ship Emissions,” Diploma Thesis, National Technical University of Athens, Athens, 2016.
- [23] F. Ren, Y. Han, S. Wang, and H. Jiang, “A novel high-dimensional trajectories construction network based on multi-clustering algorithm,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, p. 18, 2022.
- [24] A. E. Hoerl and R. W. Kennard, “Ridge regression: biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [25] R. Bostanabad, T. Kearney, S. Tao, D. W. Apley, and W. Chen, “Leveraging the nugget parameter for efficient Gaussian process modeling,” *International Journal for Numerical Methods in Engineering*, vol. 114, no. 5, pp. 501–516, 2018.
- [26] X. S. Yuan, D. D. Yang, and I. M. Center, “The research and implementation of excel operating dynamic proposition system,” *Computer Knowledge and Technology*, vol. 11, pp. 85–87, 2015.
- [27] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, no. 2, 2004.
- [28] R. Tibshirani, “Regression shrinkage and selection via the lasso: a retrospective,” *Journal of the Royal Statistical Society: Series B*, vol. 73, no. 3, pp. 273–282, 2011.
- [29] M. R. Osborne, B. Presnell, and B. A. Turlach, “On the LASSO and its Dual,” *Journal of Computational and Graphical statistics*, vol. 9, no. 2, pp. 319–337, 2000.
- [30] D. A. Anggoro, “Improving accuracy by applying Z-score normalization in linear regression and polynomial regression model for real estate data,” *International Journal of Emerging Trends in Engineering Research*, vol. 7, no. 11, pp. 549–555, 2019.
- [31] P. B. Ober, “Introduction to linear regression analysis,” *Journal of Applied Statistics*, vol. 40, no. 12, pp. 2775–2776, 2013.
- [32] T. Nosal and L. F. M. Miranda, “The effect of weather on the use of North American bicycle facilities: a multi-city analysis using automatic counts,” *Transportation Research Part A Policy and Practice*, vol. 66, pp. 213–225, 2014.
- [33] J. Wu and H. Liao, “Weather, travel mode choice, and impacts on subway ridership in beijing,” *Transportation Research Part A: Policy and Practice*, Web, vol. 135, pp. 264–279, 2020.
- [34] Imo, “Resolution MEPC. 1/Cric. 684, Guidelines for the Voluntary Use of the Ship Energy Efficiency Operational Indicator,” *International Maritime Organization*, London, UK, 2009, http://www.imo.org/blast/blastDataHelper.asp?data_id=26531&filename=684.pdf, (accessed on 1 August 2016).
- [35] K. D. Joshi and P. S. Nalwade, “Modified K-means for better initial cluster centres,” *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 7, 2013.
- [36] “Wind Scale,” *National Public Service Platform for Standards Information*, 2012.