Received December 13, 2019, accepted December 21, 2019, date of publication December 25, 2019, date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962210

An Expected Utility-Based Optimization of Slow Steaming in Sulphur Emission Control Areas by Applying Big Data Analytics

YUZHE ZHAO [1], JINGMIAO ZHOU [1,2], YUJUN FAN [1], AND HAIBO KUANG [1]
[1]Collaborative Innovation Center for Transport Studies, Dalian Maritime University, Dalian 116026, China
[2]Business School, Dalian University of Foreign Languages, Dalian 116044, China

Corresponding author: Yuzhe Zhao (zhao_yzz@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71403035, Grant 71902016, and Grant 71831002, in part by the Foundation for Humanities and Social Sciences of the Ministry of Education of China under Grant 18YJC630261, in part by the Key Research and Development. Program for Soft Science Project of Liaoning Province of China under Grant 2018401030, in part by the Key Social Science Foundation of Liaoning Province of China under Grant L19AGL012, in part by the Key Research Program of Dalian Academy of Social Sciences under Grant 2019dlsky065, and in part by the Program for Innovative Research Team in University of Ministry of Education of China under Grant IRT_17R13.

ABSTRACT This paper analyses the operator's risk-based decision (RBD) company for slow steaming, and creates a sailing speed optimization model for slow steaming (SSOM-SS), aiming to balance the expected utility-based objectives (EUO) of fuel consumption, SOx emissions and delivery delay. Considering the limitations of existing theoretical fuel consumption functions under uncertainties in voyages, the authors applies big data analytics (BDA) techniques like data fusion and feature selection to provide the SSOM-SS with accurate and suitable data on fuel consumption. In addition, a solver is built based on the genetic algorithm (GA) to solve the SSOM-SS. The effectiveness of the SSOM-SS is verified through a case study on the RBD for slow steaming of an Orient Overseas Container Line (OOCL) containership sailing across the sulphur emission control areas (SECAs) in Chinese coastal regions. The results show that the SSOM-SS can facilitate the RBD for slow steaming, and provide a novel tool for sailing speed optimization.

INDEX TERMS Big data analytics (BDA), slow steaming, sailing speed optimization, fuel consumption, genetic algorithm (GA), risk aversion.

I. INTRODUCTION

The health of marine industry hinges on the environmentally sustainable operations in maritime shipping [1]. To control the air pollution from ships, the International Maritime Organization (IMO) issued the *MARPOL Annex VI–prevention of air pollution from ships* in 2005, and delineated several sulphur emission control areas (SECAs), such as the Baltic Sea, the North Sea and the North American Area (coastal areas of the United States and Canada) [2]. In these SECAs, the sulfur content of fuel oil used must be reduced from 1% to 0.1% by 2015. In addition to the SECAs, the IMO has decided that the global fuel sulphur limit of 0.50% will enter into force in 2020 [3]. Similarly, the Chinese Ministry of Transport defined the Chinese coasts as an emission control area (ECA) and required ships to burn marine fuels with a sulfur limit of 0.5%, which came into force in 2019 [4]. However, despite best efforts it may not always be possible to obtain compliant fuel oil. The unavailability of low sulfur fuel may happen if its production and logistics capability is hard to meet the rising demand stimulated by the SECA regulations. In such a situation, the ship operator is allowed to continue using high sulfur fuel with a necessity to provide a Fuel Oil Non-Availability Report (FONAR) [5]. Nevertheless, the ship operator that has to use high sulfur fuel may still try to reduce SO_x emissions in order to prevent the use of high sulfur fuel from damaging the public image. In addition, the efforts to reduce SO_x emissions may help the submission of FONAR being accepted. Against this backdrop, it can be considered as a risk control method for the ship operators in marine industry to reduce SO_x emissions.

An effective solution to SO_x emissions reduction is an environment-friendly strategy known as slow steaming. This is because SO_x emissions are proportional to the fuel burned,

The associate editor coordinating the review of this manuscript and approving it for publication was Dalin Zhang.

and its consumption ratio depends on the sailing speed [6], [7]. Various types of commercial ships have adopted this solution, including tankers, bulk carriers and containerships [8]. For example, a Maersk Triple E-class containership emits only half the average amount of SO_x on the Asia-Europe trade lane, if it moves at a slower-than-normal speed (17.8knots) [9]. Nevertheless, the slow sailing speed of slow steaming may delay shipment delivery and increase the round-trip time [10]. The inventory cost thus incurred is essentially a kind of loss, which should be considered in the pursuit of punctuality [11]. To optimize the sailing speed, the shipping company must make a trade-off, or risk-based decision (RBD), between different operational objectives, namely, fuel consumption, SO_x emissions, and delivery delay.

Most studies on sailing speed optimization are based on a theoretical fuel consumption function [12]–[15]. In addition to the sailing speed, the fuel consumption is affected by such factors as hydrology, weather, the time in the seas, as well as the states of the sea and the ship. The effects of these factors can be illustrated by the difference between theoretical and actual fuel consumptions [16]. Figure 1 shows the relationship between actual and theoretical fuel consumptions with the growing number of days in the sea [17]. Despite that the theoretical model fit very well when the time in the sea is within 1 day, it is obvious that there are some points with large deviation from the actual fuel consumption when the number of days in the sea is longer.

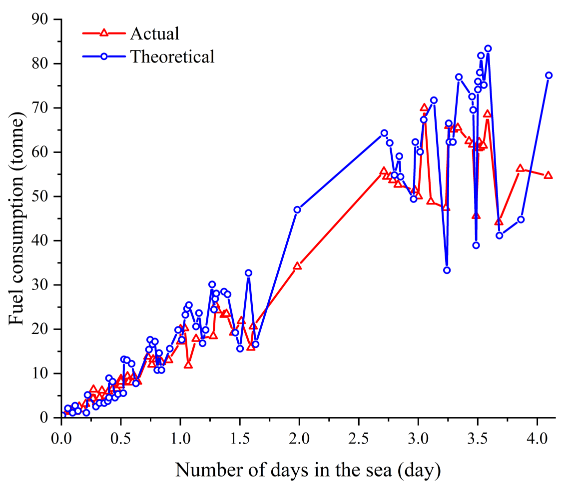

FIGURE 1. **The relationship between actual and theoretical fuel consumptions with the growing number of days in the sea.**

The fuel efficiency of ships can be effectively improved based on big data [18]–[20]. The big data generally refers to the extremely large datasets that may be analysed computationally through big data analytics (BDA) to reveal patterns, trends, and associations. Big data on fuel consumption are available, because most ships today have an automatic identification system (AIS), which continuously monitors the operational performance [21]–[24]. Through the BDA, it is possible to predict the fuel consumption accurately from the AIS data, and compare it with the actual fuel consumption in the captain's log [25].

Nevertheless, the application of the BDA in sailing speed optimization faces two challenges: First, the data from the AIS and the log are not easy to use, because of the typical features of big data, namely, high variety, sheer volume and fast velocity [26]. Second, it is difficult to describe the relationship between fuel consumption and its influencing factors with a single mathematical formula. To cope with these challenges, this paper applies a data mining technique to explore the fuel consumption based on the BDA.

This research makes the following contributions: First, a sailing speed optimization model for slow steaming (SSOM-SS) is established to examine the operator's RBD for slow steaming, aiming to balance the EUOs of fuel consumption, SO_x emissions and delivery delay. Second, the authors develop a novel approach to estimate fuel consumption, which can effectively use the AIS data and log data based on BDA, and create a solver based on the genetic algorithm (GA) for the SSOM-SS.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature on slow steaming and fuel consumption estimation; Section 3 presents the EUO function that reflects the RBD for slow steaming and establishes the SSOM-SS; Section 4 details the BDA techniques of data parser and data miner, and proposes a GA-based solver for the SSOM-SS; Section 5 verifies the SSOM-SS through a case study on the RBD for slow steaming of an Orient Overseas Container Line (OOCL) containership sailing between Dalian and Kaohsiung across the SECAs in Chinese coastal regions; Section 6 sums up the findings of this research.

II. LITERATURE REVIEW

A. SLOW STEAMING BASED ON RISK AVERSION

In maritime shipping, slow steaming is the most popular and effective decision-making method for environmentally sustainable operations [27]. Three operational objectives are often addressed in existing slow steaming models, including minimal fuel consumption, minimal SO_x emissions and minimal delivery delay. However, there are few models that aims to minimize the total cost of the conflicting objectives [28].

As mentioned before, the sailing speed is a key decision variable in the expressions of fuel consumption, SO_x emissions and delivery delay. Psaraftis and Kontovas [29] summed up the existing optimization models for the sailing speed, pointing out that most models take the sailing speed as a decision variable in the decision-making problem [30], [31]. Wang *et al.* [32] and Wen *et al.* [33] adopted the sailing speed as an implicit input to the decision-making problem, and derived speed-related explicit inputs from this implicit input.

The inner feeling of the operator, as a decision-maker, is critical to slow steaming, yet has been largely neglected. Before making the decision of slow steaming, the operator often needs to determine an empirical sailing speed. In real-world RBDs, risk aversion is ubiquitous due to the various uncertainties. This contradicts the rational agent hypothesis in expected utility hypothesis and axioms of preference [34]. During the RBD process, the features of expected utility-based objective (EUO) are obvious in the human nervous system. It could be enlightening to include the risk aversion features into the EUO function, before exploring the operator's RBD for slow steaming with different degrees of personal preference [35]. However, there is little report that considers the EUO of slow steaming in sailing speed optimization.

B. FUEL CONSUMPTION ESTIMATION BASED ON THE BDA

The sailing speed is a main determinant of fuel consumption. The relationship between sailing speed and fuel consumption is traditionally described as a third-order exponential curve. However, the fuel consumption thus obtained is a rough estimate, and may deviate greatly from the actual fuel consumption, if the ship has a large capacity or sails at a very low speed [15], [29].

With the development of data collection systems, more and more scholars started to estimate fuel consumption based on various determinants [36]. Thanks to the advanced sensors onboard, the fuel consumption can now be estimated based on massive real-time data, using data-driven statistical models [37]. The existing data-driven statistical models can estimate the fuel consumption with a high precision, under various operating conditions and environments [38]–[42]. Nonetheless, there is still ample room to improve their interpretability and accuracy.

Besides the estimation of fuel consumption, the BDA also facilitates many other aspects of shipping studies. For example, the navigation safety, ship behaviour and shipping environment can be evaluated accurately, using various types of data (e.g. location, environment, and performance of antifouling coating) [18], [43]. Recently, the BDA techniques of data parser and data miner were adopted to create data-driven models on the decision support system for sailing speed. These models could effectively learn the impact of weather on fuel consumption based on the massive data on historical weather conditions [44]. Despite the above studies, few researches have applied the BDA to guide the slow steaming operations.

III. PROBLEM DESCRIPTION

The proposed SSOM-SS focuses on a single ship and applies to the most common scenario in marine logistics, the fixed-route scenario: a ship sails directly from port A to port B carrying the same cargoes onboard, or sails along a multi-leg route between the two ports visiting intermediate ports in a pre-set sequence, with the cargoes onboard changing at each intermediate port. In this scenario, the route is pre-set in route design and fleet deployment.

Let $N = \{1,2,\ldots,n\}$ be the set of all the ports on the route, $D_{ij(i\in N, j\in N)}$(nautical mile) be the inter-port distances,

TABLE 1. The issues affecting the three risk factors.

Risk factors	Potential failure effects	Possible failure mode
Fuel consumption	Increasing operation cost	Risky decision-making for slow steaming; Improper decision on sailing speed
SO_x emissions	Emitting more SO_x and damaging the public image	
Delivery delay	Delaying cargo delivery and causing consumer dissatisfaction	

Q_{ij} (ton) be the cargoes onboard from port i to port $j (j \neq i)$, and the T_{ij}(day) be the agreed delivery schedule from port i to port j. Under these conditions, the operator should determine the proper range of sailing speed V(knots) for slow steaming.

A. THE EUO

In the RBD for slow steaming, the EUO is measured by three risk factors: fuel consumption, SO_x emissions and delivery delay (Table 1). The type of gains or losses induced by each risk factor is called the potential failure effect, while the possible failure scenarios for each risk factor is known as the potential failure mode. In this section, the EUO is evaluated based on the three risk factors, as well as their priority, importance and impact.

According to previous research, the RBD for slow steaming can be characterized by the EUO, if it carries the features of risk aversion. Considering its decreasing marginal utility, the logarithmic function is adopted here to describe the EUO for risk-averse decision-maker in different areas [45]–[47]. Sheng [48] proved that this utility function is suitable for various problems, as well as the same problem with the risk-averse decision-maker having different degrees of personal preference. Our EUO function $U(*)$ can be defined as:

$$U = \ln\left(\frac{x}{x_{\min}}\right), \tag{1}$$

where $x_{\min}$ is the minimum value of risk factor x. Then, the utility of the operator for the risk factor k can be described as:

$$U_k(x) = \frac{1}{\ln\left(2 - \frac{x_{\min}}{x_{\max}}\right)} \ln\left(2 - \frac{x}{x_{\max}}\right), \tag{2}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the risk factor k, respectively. Obviously, $U_k(x_{\min}) = 1$ and $U_k(x_{\max}) = 0$ as the first term of Equation (2) is the normalising factor.

Let F, E and S be the function of the variation in fuel consumption (ton), SO_x emissions (ton) and delivery delay (day) through slow steaming, respectively. As mentioned before, the sailing speed of slow steaming should be controlled within a proper range. Then, the EUO functions of the three risk factors are denoted as $U_1(F)$, $U_2(E)$ and $U_3(S)$, respectively.

Through optimization of the sailing speed, the RBD for slow steaming aims to maximize the cost-effectiveness and greenness in terms of EUO. Thus, the objective function of the RBD process is the weighted average of $U_1(F)$, $U_2(E)$

and $U_3(S)$:

$$\mathrm{Max}U = w_1 U_1 (F) + w_2 U_2 (E) + w_3 U_3 (S), \tag{3}$$

where w_1, w_2 and w_3 are the weights for the trade-off between the three objective functions ($w_1 + w_2 + w_3 = 1$).

B. SSOM-SS

1) FUEL CONSUMPTION FUNCTION

The fuel consumption function describes the cubic relationship between fuel consumption and sailing speed [49]. The daily fuel consumption of a ship considering the sailing speed, $F_d(V)$, (ton/day) can be computed by:

$$F_d (V) = F_M \left(\frac{V}{V_d}\right)^3 + F_A, \tag{4}$$

where F_M is the daily fuel consumption of main engine at design speed (ton), V_d represents the design speed (knot) and F_A is the daily fuel consumption of auxiliary engine (ton).

In the literature, the effects of several parameters (e.g. displacement, cargoes onboard, wind, waves and currents) on fuel consumption in slow steaming is widely ignored. Let $\nabla = \{\nabla_1, \nabla_2 \ldots, \nabla_K\}$ be the set of these parameters. Then, the influence of sailing speed V and these parameters are quantified, and used to improve the Equation (4) into the function for the daily fuel consumption considering both sailing speed and other parameters, $F_d(V,\nabla)$. On this basis, the total fuel consumption of a ship sailing at $24V$ nautical miles per day across the route can be expressed as:

$$F(V, \nabla) = \sum_{i \in N, j \in N} \left(F_d (V, \nabla) \times \frac{D_{ij}}{24V}\right), \tag{5}$$

2) SO_x EMISSIONS FUNCTION

There is linear proportionality between SO_x emissions and fuel consumption, which is measured by the actual sulfur content $\sigma(\%)$. Here, for a ship using high sulfur fuel, the SO_x emissions that exceed the sulfur limit between ports i and j can be expressed as:

$$E_{ij} (V, \nabla) = (\sigma - \varpi) \times F_d (V, \nabla) \times \frac{D_{ij}^{SECA}}{24V}, \tag{6}$$

where ϖ is the sulfur limit of SECA, D_{ij}^{SECA} is the distance within the SECA between the two ports. Then, the total SO_x emissions exceeding the sulfur limit of SECA across the route $E(V)$ can be described as:

$$E(V, \nabla) = \sum_{i \in N, j \in N} \left(E_{ij} (V, \nabla)\right). \tag{7}$$

3) DELIVERY DELAY FUNCTION

Being an indicator of the service level, delivery delay equals the total time delay (day) multiplying the cargoes onboard. The time delay refers to the deviation of the actual sailing time $D_{ij}/24V$ from the agreed delivery schedule T_{ij} between ports i and j. Therefore, the delivery delay function can be expressed as:

$$S(V) = \sum_{i \in N, j \in N} \left(\left(\frac{D_{ij}}{24V} - T_{ij}\right) \times Q_{ij}\right). \tag{8}$$

The time delay is obviously positive. Without considering the port operation time and time delay, the agreed delivery schedule T_{ij} can be described as:

$$\frac{D_{ij}}{24V} \geq T_{ij} = \frac{D_{ij}}{24V_d}. \tag{9}$$

Thus, the sailing speed V can be controlled as:

$$V_{\min} \leq V \leq V_{\max} = V_d. \tag{10}$$

where $V_{\max}$ is the maximum sailing speed (knots), i.e. the designed sailing speed V_d(knots); $V_{\min}$ is the minimum sailing speed (knots).

4) MATHEMATICAL MODEL

The objective function of the RBD for slow steaming is derived from the EUO functions to optimize the sailing speed. Hence, the SSOM-SS can be formulated as the nonlinear programming model below:

$$\begin{aligned} \underset{V}{\mathrm{Max}}\ U(V) &= w_1 U_1 (F(V, \nabla)) + w_2 U_2 (E(V, \nabla)) \\ &\quad + w_3 U_3 (S(V)) \\ \text{s.t. } & V_{\min} < V \leq V_{\max}. \end{aligned} \tag{11}$$

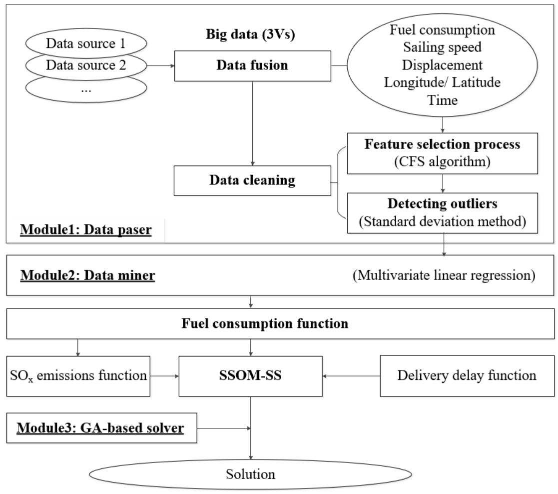

FIGURE 2. Big data-based solution framework for the SSOM-SS.

IV. BDA TECHNIQUES AND GA-BASED SOLVER

The framework in Lamba and Singh [50], the SSOM-SS is solved based on big data through the procedure in Figure 2. The dataset is generated in the light of the volume, variety and velocity (3Vs) of big data.

To ensure the quality of data inputted to the SSOM-SS, the authors develop a data reconciliation method. Hence, the first and second modules of our model-solving framework are data

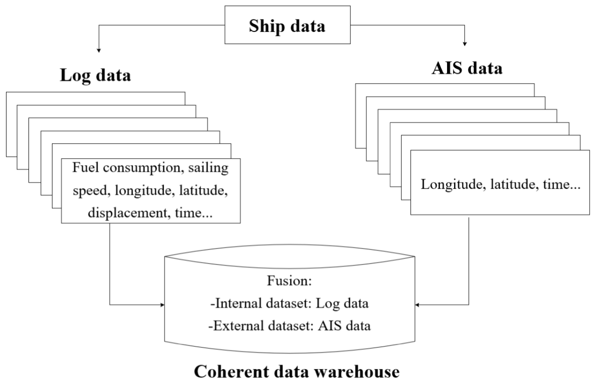

FIGURE 3. **Data fusion model.**

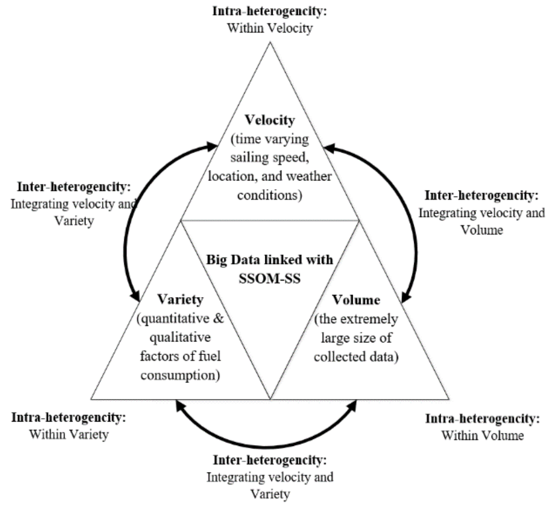

FIGURE 4. **The 3Vs of big data linked with SSOM-SS.**

parser and data miner, respectively. After the input data have been processed and analysed in these two modules, the fuel consumption function is developed by BDA techniques.

The third module of our framework is the solving algorithm of the nonlinear SSOM-SS. Since the exact algorithms are unable to deal with the 3Vs of big data, a GA-based solver is designed to generate the approximate optimal solutions, which reflect the trade-off between fuel consumption, SO_x emissions and delivery delay for sailing speed.

A. DATA PARSER
Data parser is a vital BDA technique that screens out outliers and thus eliminates the generation of misleading results. For the quality of input and output data, the Data Parser is provided with data fusion and data cleaning functions [51].

1) DATA FUSION
The original data are varied in type and from multiple internal or external sources. For instance, the AIS data and log data may differ in recording frequency, which could be an obstacle for further analysis. This calls for data fusion in the coherent data warehouse [51], [52]. The necessity of data fusion also arises from the inevitable missing entries in the raw data. The most direct way to fill the missing data is to look for the corresponding entries in the database from other sources. Our data fusion model is illustrated in Figure 3 below.

As shown in Figure 4, the 3Vs of the data linked with the SSOM-SS show obvious heterogeneity:

Variety: The fuel consumption function involves multiple variables like location, sailing distance/time and engine speed, reflecting the variety of the collected data.

Volume: The fuel consumption function considers many volume features of the collected data, such as the number of ships, the number of available sailing speeds for each ship, the number of ports visited by each ship, and the number of legs.

Velocity: The velocity depends on the tendency of the data to change in real time. The difference of our data sources comes from the fact that the AIS data are updated hourly, while the log data are reported daily.

2) DATA CLEANING
There are still redundant features or outliers in the fused dataset, which may suppress the solving efficiency or fitting accuracy. High-dimensional databases usually contain similar features, and inaccuracy is inevitable for the manually collected log data. The redundant features and outliers must be controlled on an acceptable level.

Drawing on Abbasian *et al.* [51], the correlation-based feature selection (CFS) algorithm is introduced to the feature selection process, aiming to create a subset of the features that closely related to our objective. This algorithm is an efficient tool to reduce the number of features in a dataset with weak interactions [53]. In addition, the Spearman's rank correlation coeficient is selected to deal with the nonlinear relationship in the features.

Furthermore, the standard deviation method, a classical model-free nonparametric method, is employed to screen out the outliers [51]. According to Chebyshev's Theorem, at least 15/16 (94%) of the data lie within 4 standard deviations of the mean; any number whose deviations from the average and trend line greater than 4 standard deviations is an outlier.

B. DATA MINER
Data mining is a process to examine the collected database and generate new information. In this paper, a data miner is designed to evaluate the degree of impact from each influencing factor (e.g. sailing speed) on fuel consumption, based on the data outputted by the data parser. These data are subjected to multivariate nonlinear regression (Figure 5) before being imported to the fuel consumption function [29].

C. GA-BASED SOLVER
The SSOM-SS is solved by a solver based on the GA, an algorithm mimicking the natural selection process [54]. To ensure its feasibility, the solution is generated through steps like selection, crossover and mutation. Besides, the parameters of the fitness function are finetuned through repeated tests.

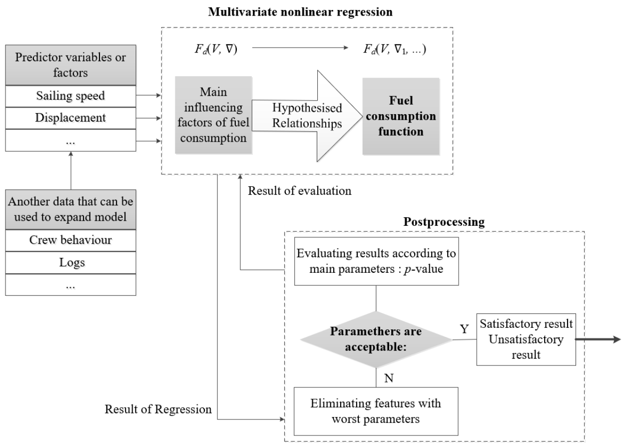

FIGURE 5. Multivariate nonlinear regression of the parsed data.

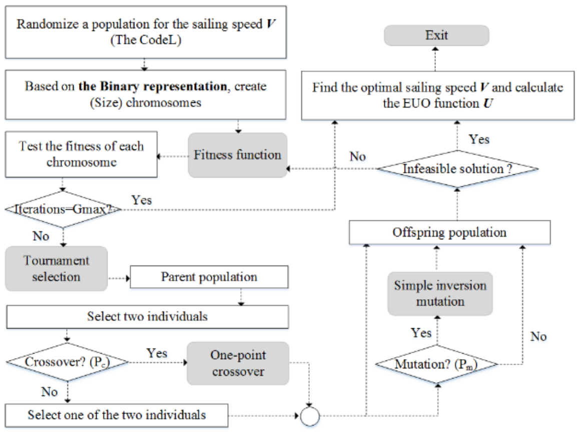

FIGURE 6. Flowchart of the GA-based solver.

As shown in Figure 6, the workflow of the GA-based solver includes solution representation, fitness calculation, selection, crossover, mutation and infeasible solution adjustment.

Step 1 (Solution Representation): Based on the features of the decision variable, binary representation is selected for the chromosome of the SSOM-SS. First, the solution is encoded as strings of zeros and ones; Next, each binary string is converted into a decimal number and normalized to a real number V in the specified interval, respectively by the following equations:

$$(b_1 b_2 \cdots b_{\text{CodeL}})_2 = \left(\sum_{s=1}^{\text{CodeL}} 2^s b_s\right)_{10} = y^t, \tag{12}$$

$$V = \text{Min}V + y^t \frac{(\text{Max}V - \text{Min}V)}{2^{\text{CodeL}} - 1}. \tag{13}$$

Step 2 (Fitness Calculation): Each solution satisfying the constraints is treated as a chromosome. The selection operator of the GA-based solver may be weakened, if the reciprocal of the objective function is directly used as the fitness function. If so, different chromosomes will have basically the same chance of being selected. Since the SSOM-SS is a maximization problem, the fitness function is linearly calibrated as follows to keep the fitness a positive number:

$$U'(V) = U(V) - \min(U(V)) + \xi, \tag{14}$$

where $\xi = 1$.

FIGURE 7. The containership service between Dalian and Kaohsiung.

Step 3 (Selection): Selection is the driving force of genetic search. If fitness proportionate selection, a.k.a. roulette wheel selection, is adopted, the fitness function must be transformed. To avoid the transformation, the tournament selection is adopted to identify the best individuals in the current population and keep them in the next generation.

Step 4 (Crossover): Crossover refers to the swap between two individuals on different chromosomes. In this paper, the crossover is implemented in three steps: First, the parent population meeting the crossover probability P_c is identified; second, the crossover segments are randomly selected from in the parent population; third, the genes in these segments are swapped by one-point crossover, creating the offspring population.

Step 5 (Mutation): Mutation determines the local search ability of the GA-based solver and ensures the diversity of the population. Here, the simple inversion mutation is performed to generate new individuals at the mutation probability P_m.

Step 6 (Infeasible Solution Adjustment): If the solution (chromosome) is infeasible after crossover and mutation, Steps 2∼5 should be repeated until the termination condition is satisfied.

V. CASE STUDY AND RESULTS ANALYSIS

A. FUEL CONSUMPTION BASED ON BDA TECHNIQUES

In this section, the BDA-based fuel consumption estimation approach and the SSOM-SS with GA-based solver are verified with the RBD for slow steaming of an OOCL containership sailing between Dalian and Kaohsiung.

The fixed route of the containership includes the following ports: 1 Ningbo, 2 Dalian, 3 Tianjin, 4 Qingdao,

TABLE 2. Agreed delivery schedule and cargoes onboard.

T_{ij}(day)	T_{12}	T_{23}	T_{34}	T_{45}	T_{45}	T_{56}	T_{67}	T_{78}
Value	1.926	0.634	1.256	0.301	2.646	0.357	0.321	0.985
Q_{ij}(TEU)	Q_{12}	Q_{23}	Q_{34}	Q_{45}	Q_{56}	Q_{67}	Q_{78}	Q_{81}
Value	7,000	8,000	9,000	9,000	10,000	6,000	8,000	10,000

TABLE 3. The log data of the containership.

Date (day)	Time (hour)	Location	Fuel consumption (ton)	Sailing speed (knots)	Displaceme nt (ton)	Wave direction (compass card)	Wave height (m)	Wind force (Beaufort scale)	Temperatur e (°C)	…
07/16/2 016	4,706	29°17'8''N; 122°21 '9''E	94.2	16.9	117,905	N	1.2	4	33	…
07/17/2 016	4,730	24°42'8''N; 120°15'3''E	98.4	17.1	118,056	NE	2.1	3	34	…
07/18/2 016	4,754	22°14'4''N; 122°19'3''E	100.6	17.2	118,006	E	4.2	6	32	…
	…	…	…	…	…	…	…	…	…	…

5 Lianyungang, 6 Kaohsiung, 7 Taichung and 8 Keelung. The inter-port distances D_{ij} and the distances within the China's SECA D_{ij}^{SECA} for the ports are shown in Figure 7 below.

Assuming a possible situation where low sulfur fuel is unavailable, the actual sulfur content σ has to be 3.50%. The sulfur limit ϖ for the SECA in China is 0.5%. The agreed delivery schedule T_{ij} is calculated based on the $D_{ij}/24V_d$ and cargoes onboard Q_{ij} from port i to port j (Table 2).

A comprehensive data warehouse is needed to set up the fuel consumption function based on BDA techniques. Here, the AIS data and log data of the containership in the full operation period (340 days) in 2016 are collected to construct the data warehouse. In total, there are 248,200 records, covering 30 features for hourly AIS data, and 10 features of daily log data. The hourly AIS data records features mostly regarding geographic positions including latitude, longitude, date and time. The daily log data includes features reflecting the ship's sailing behaviour, such as location, fuel consumption, sailing speed and displacement and so on. The data fields of the daily updated logs of the containership are displayed in Table 3.

The raw data are fused to reconcile their difference in recording frequency: the daily log data are converted into hourly format, the AIS data are integrated into the log data, the frequency distribution of the features in the logs is computed to fill the missing data (30% for fuel consumption and 10% for displacement). The fusion between log data and AIS data is explained in Figure 8.

As mentioned in Subsection 4.1(2), the redundant features are eliminated by the CFS algorithm, and the Spearman's rank correlation coefficient [55] are computed between the fuel consumption and each influencing factor, e.g. sailing speed, displacement, wave direction, and wave height. The computed results (Table 4) show that three features are the most significant ones: fuel consumption, sailing speed and displacement. Hence, the daily fuel consumption per ship is described as $F_d(V, \nabla_1)$, where ∇_1 is displacement.

The raw dataset contains some extreme outliers due to incorrect recordings. The potential outliers are detected and

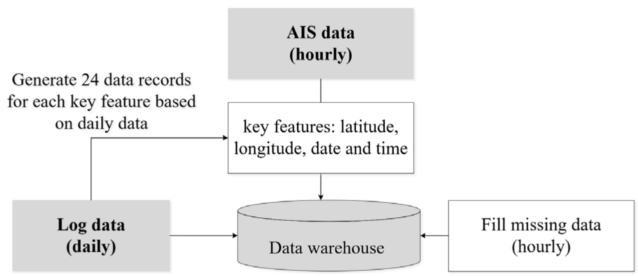

FIGURE 8. Fusion between log data and AIS data.

eliminated through a statistical hypothesis test, in which the upper control limit (UCL) and lower control limit (LCL) are set as ±4 times the standard deviation, respectively. The volume of the cut-out data is within 6% of the raw dataset [56], [57].

Figure 9 is the timeseries of sailing speed and displacement of the containership throughout the operation period. The red and blue lines are UCL and LCL, respectively. Any data point not falling on the lines are regarded as outliers and deleted.

Based on the fused and cleaned dataset, the daily fuel consumption per ship is obtained, $F_d(V, \nabla_1)$. The data are regressed by the following model:

$$F_d\left(V, \nabla_1\right) = c_1 V^{c_2} \nabla_1^{c_3}. \tag{15}$$

The p-values of F-tests and t-tests are recorded in Table 5. The p-value of c_3 is smaller than 0.05, indicating the significant impact of the actual displacement on fuel consumption; the exponent of the sailing speed (c_2) for is smaller than 3, for the maximum sailing speed always fall between 8 and 16 knots.

B. SENSITIVITY ANALYSIS OF THE SSOM-SS

This subsection evaluates the SSOM-SS performance through the sensitivity analysis of EUO functions $U_1(F)$, $U_2(E)$ and $U_3(S)$ with their weights w_1, w_2, and w_3.

The first step is to explore the SSOM-SS performance. The EUO functions were solved by the GA-based solver on

TABLE 4. Spearman's rank correlation coefficients between fuel consumption and its influencing factors.

	Fuel consumption	Sailing peed	Displacement	Wave direction	Wave height	Wind force	Temperature	Salinity
Fuel consumption	1.000							
Sailing speed	**0.851**	1.000						
Displacement	**0.452**	0.321	1.000					
Wave direction	0.033	-0.292	-0.210	1.000				
Wave height	0.033	-0.261	0.191	0.991	1.000			
Wind force	0.042	0.850	0.125	-0.111	-0.101	1.000		
Temperature	0.079	0.051	0.035	0.004	0.019	0.083	1.000	
Salinity	0.011	0.041	0.065	-0.059	-0.061	0.007	0.121	1.000

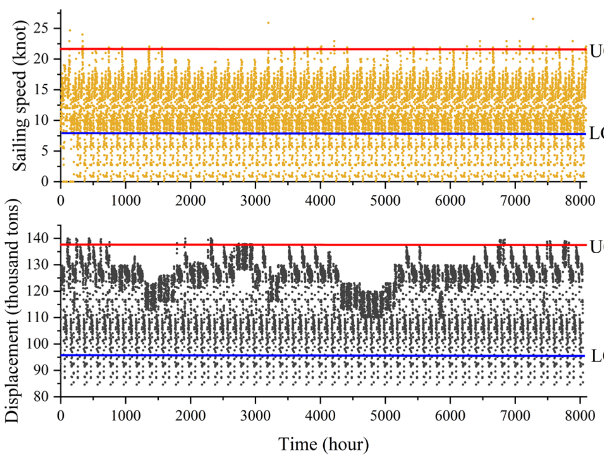

FIGURE 9. Timeseries of sailing speed and displacement.

TABLE 5. Regression results of model.

	c_1	c_2	c_3
Value	0.0051	2.7803	0.2102
p-value (t-test)	0.0000	0.0000	0.0032
R^2	0.9456	0.9387	0.9532
Adjusted R^2	0.9348	0.9267	0.9458
p-value (F-test)	0.0000	0.0000	0.0000

a personal computer (Hexa Core 3.7GHz Processor; 16 GB RAM). The values of EUO functions $U_1(F)$, $U_2(E)$ and $U_3(S)$ at different sailing speeds V are shown in Figure 10.

As shown in Figure 10, the values of $U_1(F)$ and $U_2(E)$ for fuel consumption and SO_x emissions both increase, while the value of $U_3(S)$ for delivery delay declines, with the sailing speed V falling from the maximum $V_{\max} = 22$knots to the minimum $V_{\min} = 8$knots. Meanwhile, the objective function $U(V)$ increases to the maximum of 0.874 at the sailing speed $V = 14.015$knots, before entering the downward phase. Hence, the optimal sailing speed is determined as $V^* = 14.015$ knots. If the containership operates at this reduced sailing speed, the fuel consumption and the SO_x emissions will be reduced by 38.62% from the level at the designed speed.

In the above analysis, $w_1 = w_2 = 1/4$ and $w_3 = 1/2$. In other words, delivery delay is treated as the most important factor in order to reflect the realistic concerns of the ship operator. In actual operation, the operator can adjust the weights based on factors like fuel price, limits on SO_x emissions,

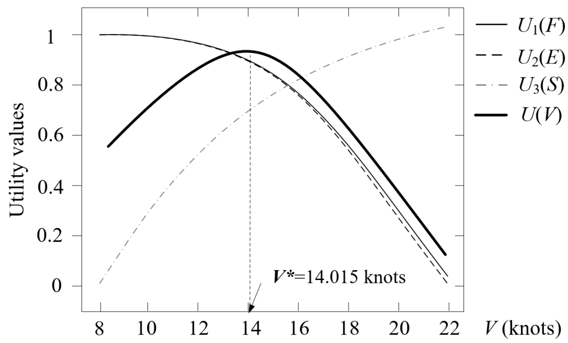

FIGURE 10. The values of $U_1(F)$, $U_2(E)$ and $U_3(S)$ and $U(V)$.

TABLE 6. Sensitivity analysis on different weight combinations of the EUO functions.

(w_1, w_2, w_3)	V	$U(V)$
(1/4, 1/4, 1/2)	14.015	0.874
(1/4, 1/2, 1/4)	11.835	0.922
(1/2, 1/4, 1/4)	11.882	0.913
(1/3, 1/3, 1/3)	12.326	0.901
(0, 0, 1)	22.000	1.000
(0, 1, 0)	8.000	1.000
(1, 0, 0)	8.000	1.000

delivery schedule and cargoes onboard. Next, the objective function $U(V)$ and the optimal sailing speed V^* are determined at different combinations of w_1, w_2, and w_3 for EUO functions $U_1(F)$, $U_2(E)$ and $U_3(S)$. According to the results in Table 6, the optimal sailing speed V^* is 8 knots when w_1 and w_2 reach the maximum of 1, indicating that the operator enjoyed the greatest EUO. Besides, when w_3 reaches the maximum of 1, the optimal sailing speed V^* equals the maximum speed V_{max}, i.e. the design speed $V_d = 22$knots. Comparing the objective function $U(V)$ values of the three weight sets (1/2, 1/4, 1/4), (1/4, 1/2, 1/4), and (1/4, 1/4, 1/2), it is clear that the optimal sailing speed is negatively correlated with the weight of fuel consumption or SO_x emissions, and positively with the weight of delivery delay.

VI. CONCLUSION

Slow steaming is an effective way for ships to achieve environmentally sustainable operations. It is essentially an RBD under three risk factors: fuel consumption, SO_x emissions and delivery delay. To facilitate the EUO-based RBD for slow steaming, this paper proposes the SSOM-SS model to optimize the sailing speed, and applies a series of BDA techniques to process the big data from the AIS and logs before estimating fuel consumption. In addition, a GA-based solver is designed to solve the SSOM-SS, aiming to determine the optimal sailing speed based on fuel consumption, SO_x emissions and delivery delay.

This research mainly makes two contributions: First, the BDA-based estimation of fuel consumption can effectively fuse and analyse the AIS data and log data collected through the operation period, providing SSOM-SS with precise and suitable inputs. Second, the SSOM-SS helps the risk-averse decision-maker, i.e. the operator, to clarify the relationship between various decision elements.

A limitation of the research is that it doesn't integrate the weather factors into the regression of fuel consumption model due to their minor impacts on fuel consumption. However, our research fully demonstrates how the cutting-edge BDA techniques should be used flexibly and adaptively to achieve meaningful results in different phases of the BDA, and enable the operator to estimate fuel consumption in different routes and complete the RBD for slow steaming. The methodology of dealing with maritime big data can be used in the followed research and applied in other cases with multi-objective concerns. For example, although shipping pollutants such as NO_x and particle matters are not considered as objectives in this paper because they are currently not limited by the SECA regulations in China and thus not focused by most shipping companies. Nevertheless, if there is a scenario where one can regard NO_x or particle matters as a practical objective, the objective of SO_x emissions and the related functions can be easily substituted without changing the methodology too much.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] Y. H. V. Lun, K. H. Lai, C. W. Y. Wong, and T. C. E. Cheng, "Green shipping management," *Transp. Res. Part E*, vol. 55, no. 7, pp. 1–2, 2013.

[2] IMO, London, U.K. *Emission Control Areas (ECAs) Designated Under MARPOL Annex VI*. Accessed: Jul. 2, 2018. [Online]. Available: http://www.gard.no/Content/26411326/IMO%20MEPC1-Circ778-Rev3_Special%20Areas%2C%20ECAs%20and%20PSSAs%20under%20MARPOL.pdf.

[3] DNV GL, Oslo, Norway. *Global Sulphur Cap 2020—Extended and Updated*. Accessed: 2019. [Online]. Available: https://www.dnvgl.com/maritime/publications/global-sulphur-cap-2020.html

[4] CCS, Beijing, China. *Technical Notice on Implementation Plan for Marine Air Pollutant Emission Control Areas*. Accessed: Dec. 11, 2018. [Online]. Available: https://www.nepia.com/media/965885/CCS-Technical-Notice-No57-Implementation-Marine-Air-Pollutant-Emissi.pdf

[5] IMO, London, U.K. *Development of Draft Guidelines for Consistent Implementation of Regulation 14.1.3 of MARPOL ANNEX VI*. Accessed: 2018. [Online]. Available: http://www.ics-shipping.org/docs/default-source/Submissions/IMO/draft-standard-for-reporting-on-fuel-oil-non-availability.pdf?sfvrsn=2

[6] Y. Qian, W. Wang, L. R. Leung, and D. P. Kaiser, "Variability of solar radiation under cloud—Free skies in China: The role of aerosols," *Geophys. Res. Lett.*, vol. 34, no. 12, pp. 2111–2121, 2007.

[7] C. Wang and J. Chen, "Strategies of refueling, sailing speed and ship deployment of containerships in the low-carbon background," *Comput. Ind. Eng.*, vol. 114, pp. 142–150, Dec. 2017.

[8] H. N. Psaraftis and C. A. Kontovas, *Slow Steaming in Maritime Transportation: Fundamentals, Trade-Offs, and Decision Models. Handbook of Ocean Container Transport Logistics*. Cham, Switzerland: Springer, 2015, pp. 315–358.

[9] H. N. Psaraftis and C. A. Kontovas, "Ship speed optimization: Concepts, models and combined speed-routing scenarios," *Transp. Res. C*, vol. 44, no. 4, pp. 52–69, 2014.

[10] Y. Du, Q. Meng, and Y. Wang, "Budgeting fuel consumption of container ship over round-trip voyage through robust optimization," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2477, no. 1, pp. 68–75, 2016.

[11] K. G. Gkonis and H. N. Psaraftis, "Modeling tankers' optimal speed and emissions," *Trans.-Soc. Naval Architects Mar. Eng.*, vol. 120, pp. 90–109, Jan. 2013.

[12] K. Fagerholt, G. Laporte, and I. Norstad, "Reducing fuel emissions by optimizing speed on shipping routes," *J. Oper. Res. Soc.*, vol. 61, no. 3, pp. 523–529, 2010.

[13] R. Fukasawa, Q. He, and Y. Song, "A disjunctive convex programming approach to the pollution-routing problem," *Transp. Res. B*, vol. 94, pp. 61–79, Dec. 2016.

[14] I. Norstad, K. Fagerholt, and G. Laporte, "Tramp ship routing and scheduling with speed optimization," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 5, pp. 853–865, 2011.

[15] S. Wang and Q. Meng, "Sailing speed optimization for containerships in a liner shipping network," *Transp. Res. E*, vol. 48, no. 3, pp. 701–714, 2012.

[16] L. Huang, Y. Wen, X. Geng, C. Zhou, and C. Xiao, "Integrating multi-source maritime information to estimate ship exhaust emissions under wind, wave and current conditions," *Transp. Res. D*, vol. 59, pp. 148–159, Mar. 2018.

[17] Z. Yao, S. H. Ng, and L. H. Lee, "A study on bunker fuel management for the shipping liner services," *Comput. Oper. Res.*, vol. 39, no. 5, pp. 1160–1172, 2012.

[18] D. Yang, L. Wu, S. Wang, H. Jia, and K. X. Li, "How big data enriches maritime research—A critical review of automatic identification system (ais) data applications," *Transp. Rev.*, vol. 39, no. 6, pp. 755–773, 2019.

[19] C. Ponsard, M. Touzani, and A. Majchrowski, "How to conduct big data projects: Methods overview and industrial feedback," *Ingénierie Des Systémes d'Inf.*, vol. 23, no. 1, pp. 9–33, 2018.

[20] Q. Baert, A. C. Caron, M. Morge, and J. C. Routier, "Fair task allocation for large data sets analysis," *Revue d"Intell. Artificielle*, vol. 31, no. 4, pp. 401–426, 2018.

[21] H. D. Gerritsen, C. Minto, and C. Lordan, "How much of the seabed is impacted by mobile fishing gear? Absolute estimates from vessel monitoring system (VMS) point data," *ICES J. Mar. Sci.*, vol. 70, no. 3, pp. 523–531, 2013.

[22] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, "The complex network of global cargo ship movements," *J. Roy. Soc. Interface*, vol. 7, no. 48, pp. 1093–1103, 2010.

[23] L. P. Perera, P. Oliveira, and C. Guedes Soares, "Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1188–1200, Sep. 2012.

[24] M. Winther, J. H. Christensen, M. S. Plejdrup, E. S. Ravn, F. Eriksson, and H. O. Kristensen, "Emission inventories for ships in the Arctic based on satellite sampled ais data," *Atmos. Environ.*, vol. 91, pp. 1–14, Jul. 2014.

[25] J. Carlton, *Marine Propellers and Propulsion*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2012.

[26] H. Kaur and S. P. Singh, "Heuristic modeling for sustainable procurement and logistics in a supply chain using big data," *Comput. Oper. Res.*, vol. 98, pp. 301–321, Oct. 2018.

[27] S. Mander, "Slow steaming and a new dawn for wind propulsion: A multi-level analysis of two low carbon shipping transitions," *Mar. Policy*, vol. 75, pp. 210–216, Jan. 2017.

[28] S. A. Mansouri, H. Lee, and O. Aluko, "Multi-objective decision support to enhance environmental sustainability in maritime shipping: A review and future directions," *Transp. Res. E*, vol. 78, pp. 3–18, Jun. 2015.

[29] H. N. Psaraftis and C. A. Kontovas, "Speed models for energy-efficient maritime transportation: A taxonomy and survey," *Transp. Res. C*, vol. 26, pp. 331–351, Jan. 2013.

[30] T. Bektaş and G. Laporte, "The pollution-routing problem," *Transp. Res. B, Methodol.*, vol. 45, no. 8, pp. 1232–1250, 2011.

[31] S. Erdoğan and E. Miller-Hooks, "A green vehicle routing problem," *Transp. Res. E, Logistics Transp. Rev.*, vol. 48, no. 1, pp. 100–114, 2012.

[32] S. Wang, Q. Meng, and Z. Liu, "Bunker consumption optimization methods in shipping: A critical review and extensions," *Transp. Res. E, Logistics Transp. Rev.*, vol. 53, pp. 49–62, Jul. 2013.
[33] M. Wen, D. Pacino, C. A. Kontovas, and H. N. Psaraftis, "A multiple ship routing and speed optimization problem under time, cost and environmental objectives," *Transp. Res. D*, vol. 52, pp. 303–321, May 2017.
[34] R. S. Wilson, J. L. Arvai, and H. R. Arkes, "My loss is your loss. . .Sometimes: Loss aversion and the effect of motivational biases," *Risk Anal., Off. Publication Soc. Risk Anal.*, vol. 28, no. 4, pp. 929–938, 2010.
[35] U. Schmidt and H. Zank, "What is loss aversion?" *J. Risk Uncertainty*, vol. 30, no. 2, pp. 157–167, 2005.
[36] S. Wang, B. Ji, J. Zhao, W. Liu, and T. Xu, "Predicting ship fuel consumption based on LASSO regression," *Transp. Res. D, Transp. Environ.*, vol. 65, pp. 817–824, Dec. 2018.
[37] J. P. Petersen, O. Winther, and D. J. Jacobsen, "A machine-learning approach to predict main energy consumption under realistic operational conditions," in *Proc. 10th Int. Conf. Comput. IT Appl. Maritime Ind. (COMPIT)*, 2011, pp. 305–316.
[38] J. Holtrop and G. G. J. Mennen, "An approximate power prediction method," *Int. Shipbuilding Prog.*, vol. 29, no. 335, pp. 166–170, Jul. 1982.
[39] J. Holtrop, "A statistical re-analysis of resistance and propulsion data," *Int. Shipbuilding Prog.*, vol. 31, pp. 272–276, Nov. 1984.
[40] V. M. Krishna, "Emissions control and performance evaluation of spark ignition engine with oxy-hydrogen blending," *Int. J. Heat Technol.*, vol. 36, no. 1, pp. 118–124, 2018.
[41] J. P. Petersen, D. J. Jacobsen, and O. Winther, "Statistical modelling for ship propulsion efficiency," *J. Marine Sci. Technol.*, vol. 17, no. 1, pp. 30–39, Mar. 2012.
[42] Q. Meng, Y. Du, and Y. Wang, "Shipping log data based container ship fuel efficiency modeling," *Transp. Res. B, Methodol.*, vol. 83, pp. 207–229, Jan. 2016.
[43] R. Ramsden, P. Lelliot, J. Thomason, and D. V. Roermund, "Project helm: Insights from AIS, fouling control and big data," in *Proc. Compit*, Lecce, Italy, 2016, pp. 439–447.
[44] H. Lee, N. Aydin, Y. Choi, S. Lekhavat, and Z. Irani, "A decision support system for vessel speed decision in maritime logistics using weather archive big data," *Comput. Oper. Res.*, vol. 98, pp. 330–342, Oct. 2018.
[45] A. E. Abbas, "Decomposing the cross derivatives of a multiattribute utility function into risk attitude and value," *Decis. Anal.*, vol. 8, no. 2, pp. 103–116, 2011.
[46] C. Johns, K. Mak, G. Hu, and W. Feng, *Effects of Utility Functions on Network Response Time and Optimization. Software and Network Engineering*. Berlin, Germany: Springer, 2012, pp. 77–88.
[47] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, Jan./Feb. 1997.
[48] C. L. Sheng, "A general utility function for decision-making," *Math. Model.*, vol. 5, no. 4, pp. 265–274, 1984.
[49] J. J. Corbett, H. Wang, and J. J. Winebrake, "The effectiveness and costs of speed reductions on emissions from international shipping," *Transp. Res. D*, vol. 14, no. 8, pp. 593–598, 2009.
[50] K. Lamba and S. P. Singh, "Big data analytics in supply chain management: Some conceptual frameworks," *Int. J. Autom. Logistics*, vol. 2, no. 4, pp. 279–293, 2016.
[51] N. S. Abbasian, A. Salajegheh, H. Gaspar, and P. O. Brett, "Improving early OSV design robustness by applying 'multivariate big data analytics' on a ship's life cycle," *J. Ind. Inf. Integr.*, vol. 10, pp. 29–38, Jun. 2018.
[52] W. S. van der Loeff, J. Godar, and V. Prakash, "A spatially explicit data-driven approach to calculating commodity specific shipping emissions per vessel," *J. Cleaner Prod.*, vol. 205, pp. 895–908, Dec. 2018.
[53] M. A. Hall, *Correlation-Based Feature Selection for Machine Learning*. Hamilton, New Zealand: The Univ. of Waikato, 1999.
[54] J. H. Holland, "Genetic algorithms," *Sci. Amer.*, vol. 267, no. 1, pp. 66–72, Jul. 1992.
[55] G. W. Corder and D. I. Foreman, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*, vol. 264. Hoboken, NJ, USA: Wiley, 2009.
[56] D. C. Montgomery, *Introduction to Statistical Quality Control*. Hoboken, NJ, USA: Wiley, 2013.
[57] B. T. Hazen, C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications," *Int. J. Prod. Econ.*, vol. 154, pp. 72–80, Aug. 2014.

YUZHE ZHAO received the Ph.D. degree from the Dalian University of Technology (DUT), in 2012. He joined Dalian Maritime University (DMU), in 2012. From March 2015 to July 2018, he conducted collaborative research at DUT, as a Postdoctoral Fellow. He is currently an Associate Professor with the School of Maritime Economics and Management, DMU, where he is also the Chief Operating Officer of the Collaborative Innovation Center for Transport Studies. His research focuses on green shipping management, shipping big data, and maritime logistics optimization. He has published more than 30 reviewed articles in international scientific journals, five reviewed articles on international conferences. He has received six provincial and industrial awards of China in the recent five years.

JINGMIAO ZHOU received the Ph.D. degree from the Dalian University of Technology (DUT), in 2018. She joined the Dalian University of Foreign Languages (DUFL), as an Assistant Professor, in 2018. She is currently conducting collaborative research with the Collaborative Innovation Center for Transport Studies, Dalian Maritime University, as a Postdoctoral Fellow. She has published more than 20 reviewed articles in international scientific journals. Her research focuses on maritime economics and logistics, green growth management, and big data analysis.

YUJUN FAN received the bachelor's degree in business administration from the Dalian University of Technology, in June 2017. He is currently pursuing the master's degree with the School of Maritime Economics and Management, Dalian Maritime University, under the supervision of Prof. Y. Zhao. His research focuses on green shipping management and shipping big data.

HAIBO KUANG received the Ph.D. degree in management from the Dalian University of Technology, China. He is currently a Professor with the School of Maritime Economics and Management, Dalian Maritime University (DMU), where he is also the President of the Collaborative Innovation Center for Transport Studies. He was awarded by Program of Innovative Research Team in the University of Ministry of Education of China. He has served as the Team Leader on Port Synergistic Development and Green Growth. He has published more than 100 reviewed journal articles and earned more than 20 provincial and industrial awards of China. His research focuses on maritime policy and management, green supply chain management, and maritime big data.