sciendo

POLISH MARITIME RESEARCH 3 (115) 2022 Vol. 29; pp. 128-140
10.2478/pomr-2022-0033

MONITORING THE PERFORMANCE OF A SHIP'S MAIN ENGINE BASED ON BIG DATA TECHNOLOGY

Meng Liang*
Shanghai Dianji University, Business School, Shanghai, China
Mingzhi Chen
Shanghai Maritime University, Merchant Marine College and Shanghai Engineering Research Center of Ship Intelligent Maintenance and Energy Efficiency, China

* *Corresponding author: liangmeng02@163.com (Meng Liang)*

ABSTRACT

Under the recent background of 'Green Shipping' and rising fuel prices, it is very important to reduce the fuel consumption rate of ships, which is directly affected by the performance of the main engine. A reasonable maintenance schedule can optimise the performance of the main engine. However, a traditional maintenance schedule is based on the navigation distance and time, ignoring many other factors, such as a harsh working environments and frequently changing operating conditions, which will lead to faster performance degradation. In this study, a real-time evaluation method combing big data of ship energy efficiency with physics-based analysis is proposed to judge the degradation of main engine performance and assist in determining the maintenance schedule. Firstly, based on the developed ship energy efficiency big data platform, the distribution statistics and comparison of different operating states are carried out. Gaussian mixture model (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) are used to cluster the data and the high-density data areas are obtained as the analysis points. Then, the data of the analysis points are polynomial fitted, by the least square method, to obtain the propulsion characteristics curves, load characteristic curves, and speed characteristic curves, which can be used to observe the performance degradation of the main engine. The results show that this method can effectively monitor the degradation degree of the main engine performance, and is of great significance to fuel efficiency improvements and greenhouse gas (GHG) emissions reduction.

Keywords: Big data of ship energy efficiency, Main engine, Performance evaluation, Cluster analysis, Mechanism analysis

INTRODUCTION

Ocean transport is vital for global trade and more than 90% of international cargo is transported by the international shipping industry [1]. Despite its significance, however, the great amount of fuel consumption and air pollutants caused by the shipping industry are more and more severe. The main air pollutants caused by shipping include nitrogen oxides (NOx), sulphur oxides (SOx), harmful particulate matter (PM), and greenhouse gas (mainly CO_2). According to the publications of the International Maritime Organisation (IMO), international shipping accounts for 14-31%, 4-10%, and 2-3% of the total worldwide global emissions of NOx, SOx, and CO_2, respectively [2].

The IMO has issued a series of regulations to restrict the emission of air pollutants from shipping because they are aware of the seriousness of the problem. The shipping industry has also taken a series of measures to reduce emissions. The scrubber system is an alternative measure to reduce SOx emissions, although it requires a large amount of cabin space, is complex and expensive to operate [3]. An SCR (Selective Catalytic Reduction) reactor can be used for reducing NOx emissions. However, carbon capture and other CO_2 emission reduction equipment are far less mature than scrubber systems and SCR reactors and not commonly used. Alternative fuels, such as biofuels and hydrogen, are promising developments but there are still some concerns about their storage, technology maturity, and safety mitigation measures in the current situation. Therefore, reducing fuel consumption remains the top priority in controlling greenhouse gas (GHG) emissions for ships in service.

Meanwhile, IMO has put forward more and more regulations concerning CO_2 emissions from ships, prompting international shipping companies to put in place technological and operational measures. In 2011, the Energy Efficiency Design Index (EEDI) [4] was proposed to urge shipping companies to meet basic energy efficiency requirements when building new ships and the Ship Energy Efficiency Management Plan (SEEMP) [5] was put forward in the same year (specific ship operational measures to reduce emissions). In late 2020, IMO approved the MARPOL Annex VI amendment, which requires all existing ships to satisfy both the Efficiency Existing Ship Index (EEXI) and Carbon Intensity Indicator (CII) requirements, as well [6]. This represents the IMO's follow-up action plan for reducing GHG emissions from international shipping vessels up to 2050. Since exhaust gas emission is difficult to measure directly, and is generally proportional to fuel consumption, the regulation for European ships' Monitoring, Reporting, and Verification (MRV) provides indirect monitoring through the ship's fuel consumption [7].

For a ship, the consumption of fuel mainly depends on two factors: the operating efficiency of the main propulsion system and the motion resistance [8]. Król summarised the current technical status of propeller system design and operation with the installation of energy-saving devices [9]. Puzdrowska paid attention to the problem of low controllability of marine medium and high-speed engines during operation, and proposed measures for quickly controlling the temperature of exhaust gas [10]. Król presented a simplified lifting surface method in propeller design [11]. Rudzki et al. proposed a decision-making system to select commands for the ship's propulsion system with a controllable pitch propeller [12]. Varbanets et al. provided a method to accurately measure the top dead centre (TDC) [13]. All of the above studies have laid an important foundation for performance improvements in ship propulsion systems.

It is widely known that the main engine is the main propulsion system overcoming the total resistance from the cargo dead weight, draft, trim, sea weather, etc. [14], to propel the ship forward. The main engine produces the original thrust through fuel consumption and pushes the propeller through a series of transmission devices to overcome the resistance [15]. As the original power of a ship, the performance of the main engine directly affects the fuel consumption rate. Customising a scientific maintenance schedule for the main engine can guarantee its good performance. However, a traditional maintenance schedule usually relies on the total navigation distance or time, without the consideration of many other environmental factors such as humidity, oscillatory working conditions, the switching of operating conditions, or weather changes on the voyage, etc.

This study aims to evaluate the performance degradation of the main engine in real-time through big data of navigation processes, such as ship fuel consumption, marine main engine status, ship status, and cargo loading status. The maintenance schedule, based on the results analysed, considers more factors and is more in line with the actual circumstances. To analyse the behaviour of marine diesel engines in different unsteady states (e.g. determining fuel consumption rate), Ghaemi proposed a goal-based mathematical model and provided a method for determining the model parameters through the available data provided by the engine manufacturer [16]. However, data are still limited.

As one of the most traditional industries, the shipping industry still relies more on intuition than on data [17]. A small amount of original voyage data comes from the ship's noon report, but this data has many flaws, such as manual recording errors, long time intervals, and small amounts of data acquisition, which is not conducive to subsequent analysis [18]. Fortunately, the Internet of Things (IoT), data transmission technology, and big-data technology under the background of 'Industry 4.0', provide a promising approach for voyage data collection and transmission [19]. Fan et al. designed a multi-source information system to collect data related to a ship's energy consumption and navigation environment [20]. Deng et al. analysed and predicted the energy efficiency of ships, based on 6G communication technology, which can access the data in real-time [21]. Both of them focused on inland river ships. Tacjana et al. monitored the heat exchanger in a steam power plant through machine learning algorithms, which shows that machine learning is a measure worthy of research and application [22]. Witkowska et al. presented a multi-dimensional nonlinear dynamic positioning (DP) controller, which adopts the adaptive vector back-stepping method and Radial Basis Function (RBF) artificial neural network [23]. Facing a huge amount of data, it is necessary to choose the most appropriate data handling technology. Data cleaning is the first and most important step, avoiding the waste of analysis resources, or even the 'Garbage in, garbage out' phenomenon [24]. Perera et al. proposed a new digital model and built a data handling framework with pre-processing and post-processing units, based on the proposed digital model [25]. Raptodimos et al. proposed an integrated method based on an artificial neural network (ANN), which applies cross-clustering and self-organising mapping to cluster data, and then realises the main engine fault diagnosis [26]. Vanem et al. used unsupervised machine learning technology to analyse sensor data and monitor marine diesel engine anomalies [27]. Perera et al. introduced the expectation maximisation (EM) algorithm and Gaussian Mixture Models (GMMs) for the main engine speed-power clustering; analysis showed that data could be divided into three clustering centres [28]. Perera et al. studied the clustering methods again and obtained great relationships between ship performance and navigation information by principal component analysis (PCA) [29]. Yan et al. proposed the application of a distributed parallel K-means clustering algorithm for path division. Later, they used the Map Reduce-based k-average algorithm to analyse the environmental factors of different route segments [30, 31]. Adland et al. evaluated the effect of hull fouling on fuel consumption, by regression analysis of daily fuel consumption and ship speed [32]. However, the influence of the decrease in diesel engine performance on the increase in fuel consumption was ignored.

In summary, most of the previous studies focused on the application of big data to consider the impact of environmental factors on ship energy efficiency and adopted intelligent algorithms, to assist ship fuel-saving decisions. However, the energy efficiency caused by the degradation of diesel engine performance and the impact of operating conditions on diesel engine performance is widely ignored and most of the black box models were adopted, so the analysis results were difficult to explain clearly. For energy efficiency monitoring and optimisation models, we need to balance the complexity of the model with the interpretability of the results. Based on this, models are usually classified as black box models or white box models. The black box model focuses on high precision of input and output. The white box model is characterised by physical characteristics, focusing on the interaction and logical relationship between various factors. The grey box model lies between the white box model and the black box model and has both advantages [33].

This study demonstrates a ship energy efficiency big data platform, based on the Beidou system, and some functional modules of the platform have also been described in previous work [18, 34]. This paper focuses on the functional modules, including data statistics, clustering, and engine performance analysis. A method of main engine performance evaluation is proposed, based on two years' of monitoring data from an operational ship. Using data statistics, machine learning, and physics-based methods, the comparison curves of diesel engine performance under different working conditions were obtained as the basis of the main engine performance evaluation.

Compared with previous studies, the contributions of this paper are:

- This study makes up for the fact that there are few applications of big data to evaluate the ship's main engine performance in previous studies.
- Previous studies have typically used only one of two approaches: unsupervised machine learning or physics-based analysis. This paper adopts the idea of the grey box model and combines the two methods to consider the accuracy and interpretability of the analysis results.
- The ship energy efficiency big data platform constructed in this study is universal and can also provide information support for energy-saving decisions of other diesel propulsion ships.

The rest of this paper is organised as follows: the description of the platform framework, techniques, and approaches is presented in Section II; data distribution statistics for the main engine operating situation are presented in Section III; the data cluster for the operational data is presented in Section IV; and the analysis results are presented in Section V. The conclusion is provided in Section VI.

DESCRIPTION OF PLATFORM FRAMEWORK, TECHNIQUES, AND APPROACHES

OVERALL FRAMEWORK OF BIG DATA PLATFORM

The framework of the big data platform is shown in Fig. 1. The on-board database records the main engine status, ship status, and ship loading status in real-time. In addition, it records the fuel consumption of the main engine from the high-precision mass flow meter. On the user interface of the on-board software, relevant monitoring data and curves are displayed, as well as the energy flow. Through the Beidou system, the data is transmitted to the onshore database in real time. Ship managers can carry out statistical analysis, clustering, polynomial curve fitting, and physics-based analysis with the collected data. The analysis results can be used as the basis for evaluating the energy efficiency of the ship and the performance of the main engine, which helps to make fuel-saving decisions and maintenance plans.

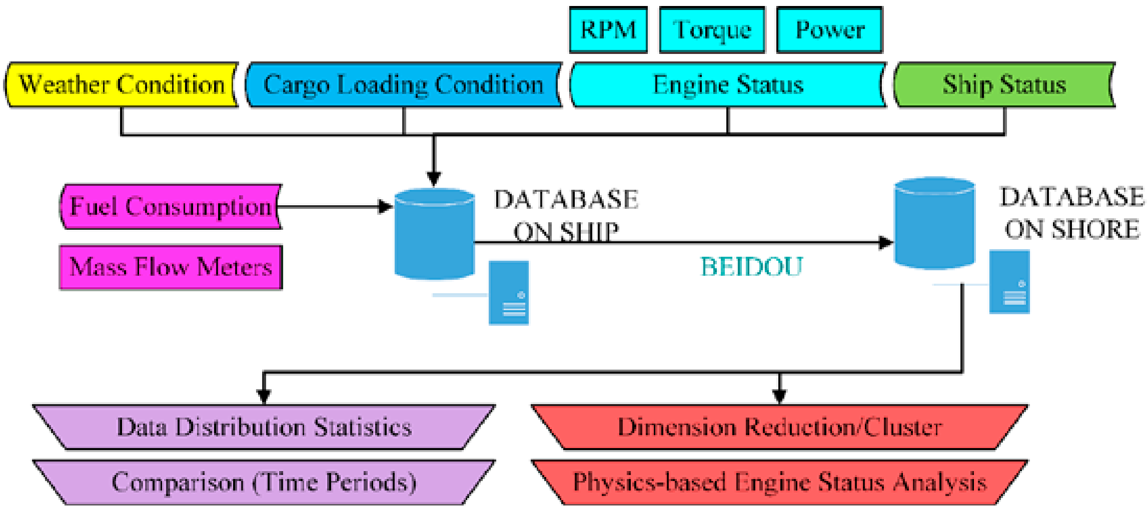

Fig. 1. Framework of big data platform

The on-board sensors measure the fuel consumption, main engine status, ship status, and ship loading status in real-time. Table 1 details some relevant data descriptions. The data comes from different systems and their acquisition frequency is not the same. The fuel consumption is read from the mass flow meter each second, using the RS485 communication device. The data on the engine status is sent out from the monitoring and alarm system every 30 seconds, via the Ethernet communication. Data such as ship speed and draft, relative wind speed and direction are read from the system on the bridge with the serial communication, and their frequency is about 10 seconds. To integrate all the systems with the same intervals, our platform records data every minute. Data from the sensors with a frequency greater than one minute are averaged, which also acts as a filter. In addition to the mean data, the standard deviation of high-frequency data is also calculated and recorded. The data is flagged as being abnormal if the standard deviation is higher than a certain threshold.

Tab. 1. Data Description

Categories	Items	Unit	Description
Fuel Consumption Rate	MEActFOCon	kg/h	fuel consumption rate
Main Engine Status	MERPM	r/min	main engine shaft speed
	METorque	kN*m	main engine shaft torque
	MEShaftPow	kW	main engine shaft power
Ship Status	ShipSpdToWater	knot	ship's water-referenced speed
	ShipSlip	%	slip ratio of propeller
Cargo Loading Status	ShipDraft	m	ship's draft

The MESFOC_kw is the specific fuel consumption of the main engine, see Eq. (1). Its unit is g/kWh. It shows how much fuel a diesel engine needs to produce 1 kW of power in one hour. This index can also reflect the economic performance of the engine and can be used for the analysis of engine performance degradation. Eq. (1) is affected by the lower calorific value (LCV) of the fuel. The ship studied in this work used marine diesel oil (MDO) and heavy fuel oil (HFO), which have different LCVs. However, when using HFO, marine engineers use the same type of fuel. For this concern, the platform automatically distinguishes two kinds of oil, with their different operating temperatures, to reduce the operations of engineers. Density is also recorded from the flow meter, for manual inspection. When the engine was running on HFO, this accounts for most of the available data. It is assumed that the used HFO has the same LCV in the current study, and the platform may also add a field to record the LCV for further consideration in future studies.

$$MESFOC\_kw = \frac{MEActFOCon}{MEShaftPow} * 1000 \tag{1}$$

The propeller's slip rate is an important factor that characterises the main engine load; it is called ShipSlip in the system. The ShipSlip is calculated by Eq. (2), which is determined by the water reference speed v_s and the shaft speed n_{ME}. In Eq. (2), H is the pitch of the propeller, which is the distance that the propeller advances when it rotates one revolution.

$$ShipSlip = (1 - v_s * \frac{1852}{n_{ME}} * H * 60) * 100 \tag{2}$$

Since the vast majority of vessels use diesel engines, the matter of their efficient and safe operation is an ongoing issue [36]. A big data platform was installed on a 2400 TEU container ship in October 2019 and the data obtained from January 1st to July 31st, 2020, as well as the same period in 2021. These are taken as an example to illustrate the effectiveness of the proposed method. The ship uses a two-stroke low-speed diesel engine (MAN B&W 7S60ME-C10.5, MCR: 13,700 kW * 97 r/min) as the main engine. The main engine directly drives the fixed pitch propeller. The propeller has four blades and an average pitch of 6.134 m. Three diesel generators, one exhaust gas boiler, and one oil-fired boiler are installed on the ship.

Based on these energy efficiency-related data, approaches such as numerical distribution statistics, machine learning, and physics-based analysis were used for engine performance analysis.

In the data pre-processing, the data were removed while the vessel was at berth. The system on the bridge sometimes does not send data because the system may be shut down. These data were also removed. In this work, the performance of the engine under stable operation was studied, so only the data of specific fuel consumption in the range of 160-260 g/(kW*h) was kept. Table 2 shows a sample of the values of the variables.

Tab. 2. A sample of the values

PCDate	PCTime	MEActFOCons	MESFOC_kw	MESFOC_nmile	ShipSpdToWater
2020/5/24	9:01:00	1496.718	190	83.76997	17.867
MERpm	METorque	MEShaftPow	WindSpd	WindDir	Latitude
86	875.833	7877.5	29.017	239	10.50 N
ShipSpd	ShipHeel	ShipTrim	ShipDraft	ShipSlip	Longitude
17.651	0.29	0.382	8.651	-2.79775	126.00 E

DESCRIPTION OF TECHNIQUES

The platform applied data statistics, clustering, and curve fitting to fulfil the study.

Numerical Distribution Statistics

Statistics aim to extrapolate the ships' overall running and operating situations, based on the data. The analysis and comparison of data distribution are helpful to understanding the difference in working conditions of different ships or the same ship at different times. In numerical distribution statistics, the probability mass function (PMF) is a normalised histogram that shows the frequency of occurrence of each value, typically for discrete variables. The good performance of PMF needs to deal with the grouping interval, while the cumulative distribution function (CDF) does not, and it can explain the difference between the distributions more clearly. PMF and CDF are both applied to the data distribution. As in Eq. (3) and Eq. (4), the PMF of the value x in the set X is the probability of x in X, while the CDF of the value x in the set X is the cumulative probability of the values in X that are less than x.

$$pmf_X(x) = \mathrm{P}(X = x) \tag{3}$$

$$cdf_X(x) = \mathrm{P}(X <= x) \tag{4}$$

Data Clustering

The basic idea of clustering is to construct k data clusters for a given data set with n samples. Generally, $k \leq n$; each cluster contains at least one sample and each sample only belongs to one cluster. Clustering methods divide data into k clusters, based on data distance or probability density models. In this paper, two different methods were used in the clustering process: the Gaussian mixture model (GMM) and density-based noise application spatial clustering (DBSCAN). The GMM is based on the assumption that the data conforms to the Gaussian statistical model and iteratively improves the parameters of the estimated model by using the Expectation-Maximisation (EM) algorithm. The algorithm can obtain the mean and covariance matrix of the clusters and get the weights of different clusters. When predicting the cluster category of a sample, the probability that the sample belongs to different clusters can be estimated. The GMM algorithm needs to input the number of clusters first, and then iteratively converges to the centre of the data cluster.

DBSCAN is a density-based clustering algorithm [35]. It can divide the region with a high-enough density into clusters and find cluster shapes in the noise data. DBSCAN describes the compactness of the sample set based on data density. Parameters (EPS, min_samples) are used to describe the compactness of the sample distribution in the neighbourhoods. Wherein, 'EPS' describes the small distance of neighbourhoods belonging to the same cluster, and 'min_samples' describes the minimum number of samples in the cluster.

If the number of samples included in the 'EPS' neighbourhood of sample x_i is not less than 'min_samples', i.e. $|N_{EPS}(x_i)| \geq \text{min\_samples}$, then x_i is called the core point. Otherwise, if $|N_{EPS}(x_i)| < \text{min\_samples}$, then x_i may be in the neighbourhood of other core points, which is called a boundary point. If x_i is neither a core point nor a boundary point, it is a noise point. The DBSCAN algorithm also defines the concepts of density direct, density reachable, and density connected. The basic process of the DBSCAN clustering algorithm is to determine all the core points according to the parameters (EPS, min_samples). It needs to find the sample with the highest density to generate a cluster and repeat the process to divide the data samples.

In this paper, we used Python-based Scikit-Learn to perform DBSCAN and GMM clustering algorithms on the pre-processed data, to find 'clusters' [37].

Polynomial curve fitting

Polynomial fitting is used to find the unknown parameters in the empirical formula dominated by a known law or model hypothesis through several known points. For example, given a function $f(\boldsymbol{x};\, a_0, a_1, a_2, \ldots, a_m) = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m$. $\boldsymbol{x}$ represents an m-dimensional sample vector. $a_0, a_1, \ldots, a_m$ are the parameters to be determined. In general, given a set of samples $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ whose measured function values are $(y_1, y_2, \ldots, y_n)$, it is possible to determine $(a_1, a_2, \ldots, a_m)$ by the least square method. A polynomial curve fit, as in Eq. (5) and (6), finds the parameter that minimises the squared value of the deviation between the output of the function and the measured value.

$$f(\boldsymbol{x}; a_0, a_1, \cdots, a_m) = a_0 + a_1x_1 + \cdots + a_mx_m \tag{5}$$

$$\underset{a_0, a_1, \cdots, a_n}{\arg\min} \sum_{i=1}^{n} [f(\boldsymbol{x}_i; a_0, a_1, \cdots, a_m) - y_i]^2 \tag{6}$$

The least square method is used to perform polynomial curve-fitting on the relationship between ME shaft speed - ME shaft power, ME shaft power – MESFOC_kw, and ME shaft speed - MESFOC_kw to further study the characteristics of the engine.

Physics-based Analysis

The effective power and specific fuel consumption of the main engine changes with the operating conditions of the diesel engine. The system carries out the performance evaluation of the main engine from three kinds of curves, which are the propulsion characteristics curve, the load characteristics curve, and the speed characteristics curve. The relationship between the main engine shaft power MEShaftPow and the main engine shaft speed MERPM during a ship's operation is cubic. This relationship is the propulsion characteristic of the main engine. The data of main engine shaft power, shaft speed, and specific fuel consumption rate MESFOC_kw are filtered in the cluster centres. The relationships between MEShaftPow - MESFOC_kw and MERPM - MESFOC_kw, which are the load and speed characteristics of the main engine, are obtained through data query and curve fitting by a self-developed software, written in C#.

DATA DISTRIBUTION STATISTICS FOR THE MAIN ENGINE OPERATING SITUATION

Using the method of distribution statistics, the two periods from January 1 to July 31 in 2020, and the same period in 2021, are taken as the comparison periods to compare the main engine status and the cargo loading status.

COMPARISON OF THE OPERATING STATUS OF THE MAIN ENGINE

Fig. 2 to Fig. 4 show the PMF and CDF of main engine shaft speed MERPM, main engine shaft power MEShaftPow and ship water reference speed ShipSpdToWater during the observation period. It can be seen that the MERPM is often set at 47, 62, 77, 80, and 85 r/min on board and the MEShaftPow is also concentrated in the corresponding section. While the ship is sailing at sea, most of the load of the main engine comes from overcoming the navigation resistance under still water and part of the power is affected by wind, current, wave, and cargo loading, etc. The ship usually needs to attain a certain speed to transfer goods in time. However, at the same ship speed, the ship resistance also changes with the weather and loading conditions. Therefore, the shaft power of the main engine fluctuates with the shipload. As shown in Fig. 4, the ship water reference speed ShipSpdToWater has three obvious operation centres: 9.0, 12.5, and 16.0 knots.

It can also be seen from Fig. 2 to Fig.4 that the recorded main engine shaft speed MERPM and main engine shaft power MEShaftPow are more often in the high load range in 2021, than in the same period in 2020, so the ship's speed also increases. This is due to the increased volume of containers in 2021 and the need for vessels to speed up their cargo turnover.

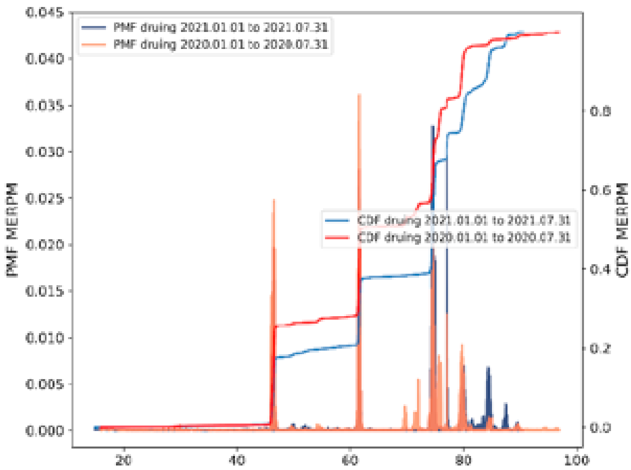

Fig. 2. PMF and CDF of main engine shaft speed (comparison between 2020 and 2021)

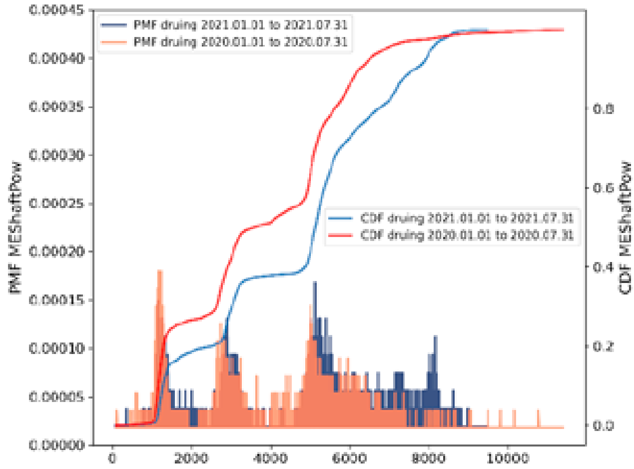

Fig. 3. PMF and CDF of main engine shaft power (comparison between 2020 and 2021)

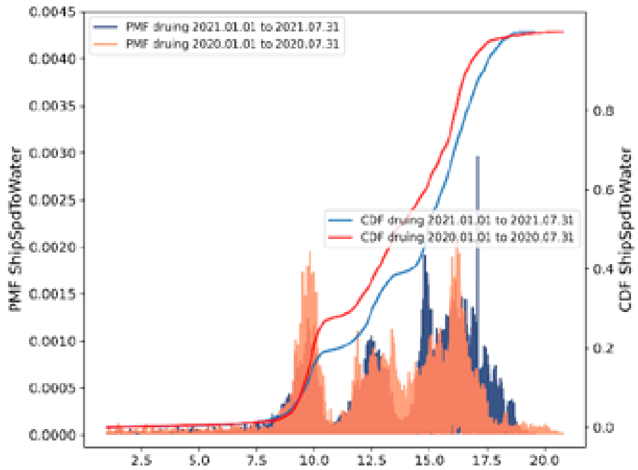

Fig. 4. PMF and CDF of water referenced ship speed (comparison between 2020 and 2021)

However, as shown in Fig. 5, the specific fuel consumption of the main engine MESFOC_kw in 2021 is more inclined to the left than that in the same period in 2020. That is to say, the specific fuel consumption is relatively low in 2021 because the main engine runs more in the high-power area in 2021 than that in 2020. It is hard to identify the engine performance degradation by the data statistics, which needs to be further studied from the engine load characteristics.

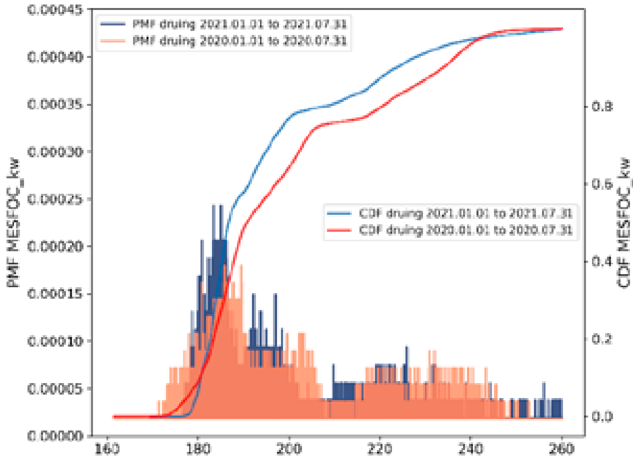

Fig. 5. PMF and CDF of main engine specific fuel consumption (comparison between 2020 and 2021)

COMPARISON OF SHIP BALLAST STATUS AND SHIP LOADING STATUS

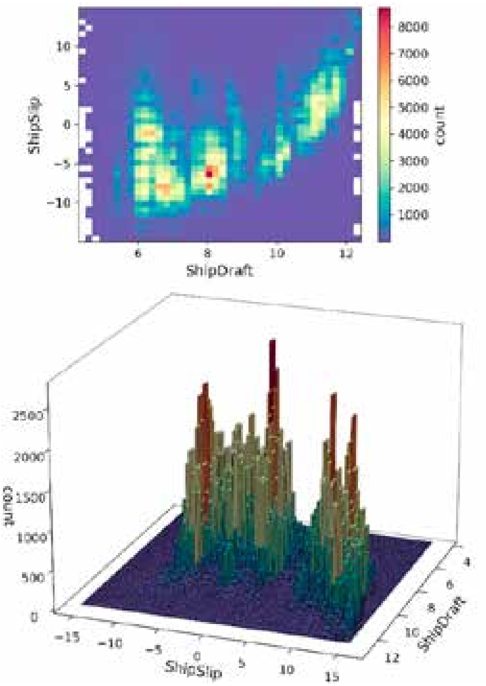

Fig. 6. Ship draft and propeller slip ratio (January 1st-July 31st, 2020)

Fig. 6 and Fig. 7 show the ship's draft ShipDraft and propeller slip ratio ShipSlip from January 1 to July 31 2020, and 2021, respectively. They show that the region of the ship's draft with the most data in 2020 is 8 m, while the region with the most data in 2021 is 11 m. The ship's draft significantly increases because the cargo volume of the ship in 2021 is significantly higher than that in the same period in 2020. In addition, both Fig.6 and Fig.7 show that, with the increase of the ship's draft, the propeller slip ratio increases as well, due to the increase of the ship's resistance.

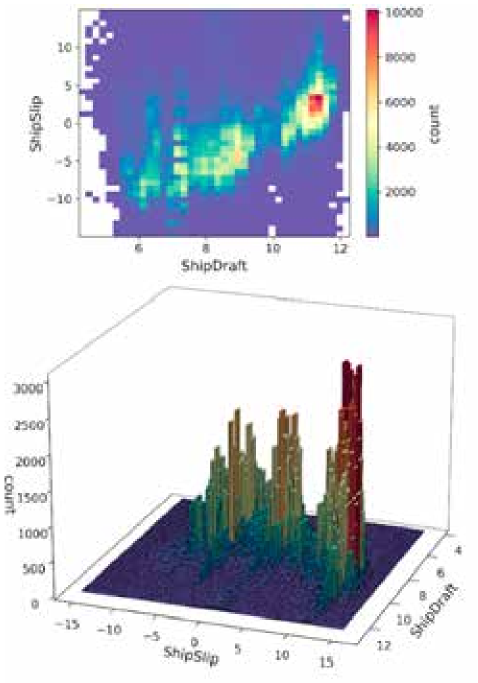

Fig. 7. Ship draft and propeller slip ratio (January 1st-July 31st, 2021)

The data in June 2020 is screened to show the ship draft effect on the main engine's load. In a month, the affect of the fouling of the ship's hull on the result is low. In the same period, the weather condition is similar. In addition, the relative wind speed is limited to the range 12-18 knots. The relationship between the main engine speed and the shaft power of the ship under ballast and full load conditions was studied. The data when the main engine runs in a stable manner is obtained. Under the ballast condition, the data of the ship's draft range is 6-7 m, and the data of the ship's draft range is 10.5-11.5 m under the full load condition. If two or more of the same data are at nearly the same engine speed, they are averaged at intervals of 0.2 *r*/*min* of engine speed. Fig. 8 shows the results. When the ship is in a ballast state, the data points of the main engine shaft speed and power are displayed as purple triangles; when the ship is fully loaded, they are displayed as yellow circles. On the same horizontal axis, the yellow circles are slightly higher than the purple triangles onźthe vertical axis. Loading cargo will increase the propeller slip ratio, leading to an increase in the load of the main engine. The average shaft speed of the ship at full load is 65.71 r/min and the required average shaft power is 4529.73 kW. The average shaft speed of the ship in ballast is 67.30 r/min and the required average shaft power is 4247.38 kW. Cargo loading causes the average shaft speed to decrease by 1.59 r/min, while the average power increases by 282.34 kW.

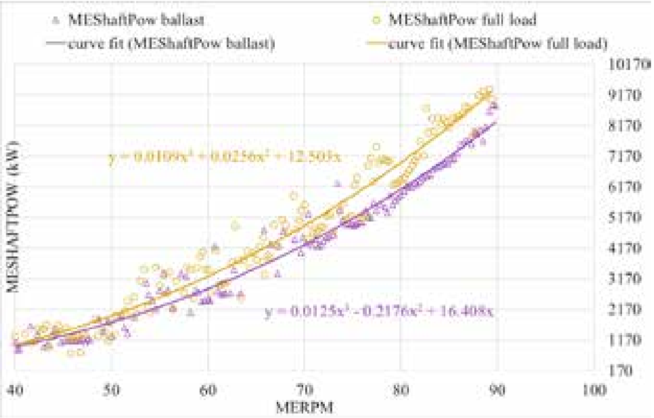

Fig.8 . Propulsion characteristics curves of the main engine (comparison under the ship in ballast and full load conditions)

CLUSTER OF OPERATION DATA

The pre-processed data distribution can be displayed as variable histograms. Fig. 9, Fig. 10, and Fig. 11 show the histograms of energy efficiency-related variables of the main engine with all data, ballast state data, and full load state data, respectively. In the figures, the horizontal axis is the variables' area and the vertical axis is the variables' count. The histogram results show that all variables are within a reasonable range. Moreover, the main engine mainly operates in the three power centres.

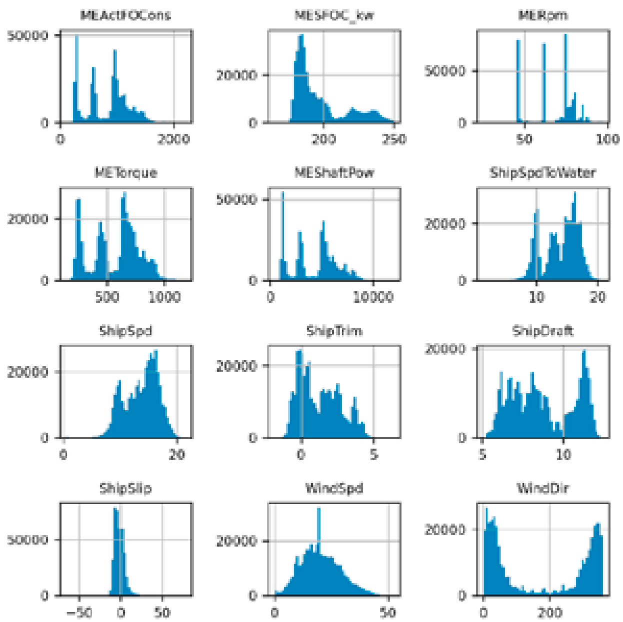

Fig. 9. Histogram of variables after data pre-processing (all data)

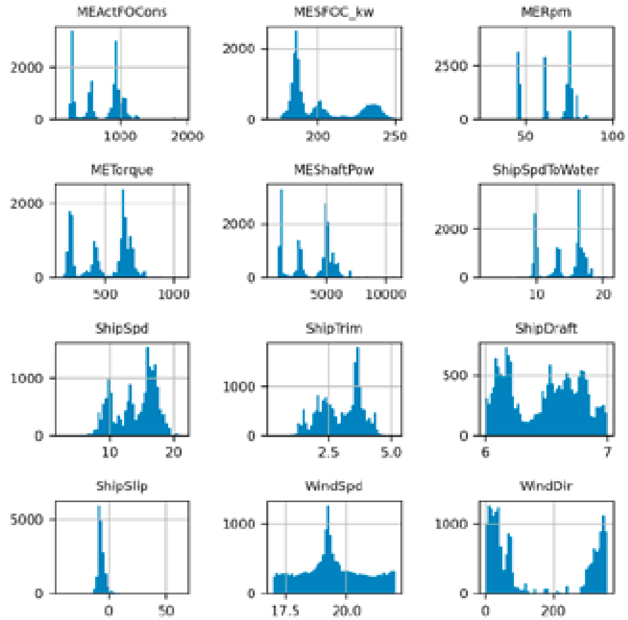

Fig. 10. Histogram of variables after data pre-processing under ballast state (relative wind speed in 12-18 knots and ship draft in 6-7 m)

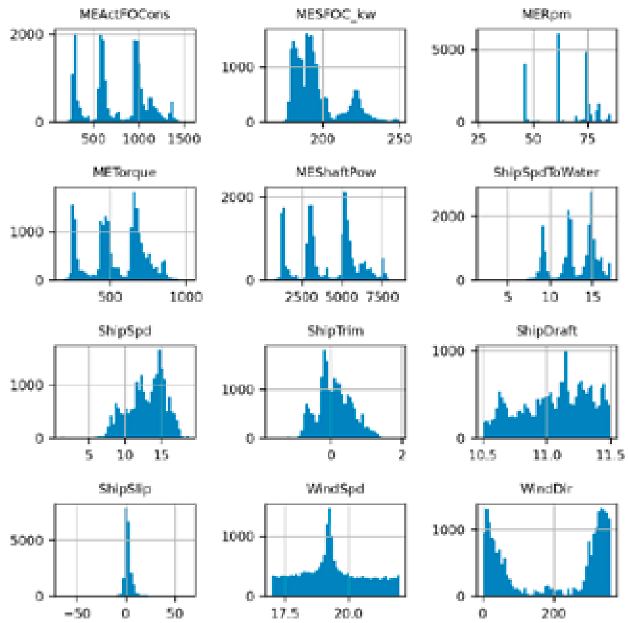

Fig. 11. Histogram of variables after data pre-processing under full load state (relative wind speed in 12-18 knots and ship draft in 10.5-11.5 m)

To explore the correlation between variables, we conducted Pearson correlation tests on them. Fig. 12 shows the Pearson correlation between all variables. It can be seen that the fuel consumption rate of the main engine is not only closely related to the state of the main engine but also has a certain correlation with the ship's draft, trim, and relative wind speed. Although this correlation is small, it should also be considered in the study. The ship's trim is highly related to the ship's draft. To further consider factors such as ship loading status and weather conditions, the study further selects the same relative wind speed range to cluster similar weather. Data are also constrained under full load and ballast conditions, to remove the effects of vessel loading variations.

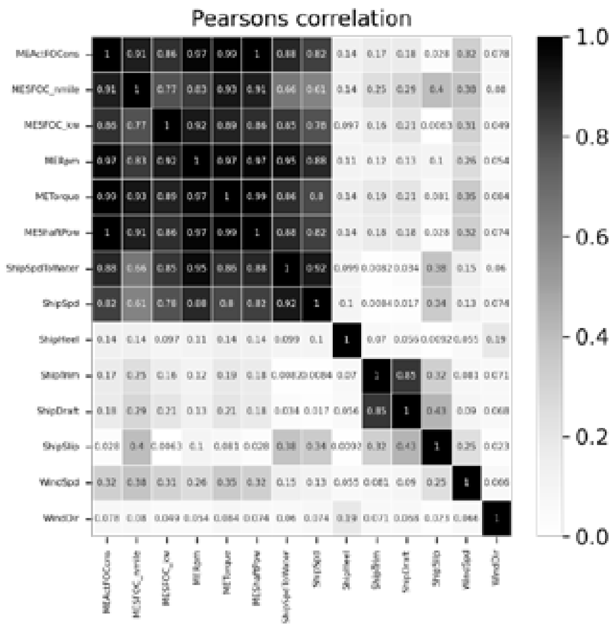

Fig. 12 Pearson correlation of the data (all data)

In this study, GMM and DBSCAN were combined for data clustering analysis. The general idea is to set the neighbourhood size (EPS) of DBSCAN to 3 and the minimum number of samples in the neighbourhood (min_samples) to 6, in order to obtain the frequent operation area of the main engine. Then GMM is applied to the cluster analysis of five variables (engine shaft speed MERPM, engine shaft power MEShaftPow, engine fuel consumption rate MEActFOCon, the specific fuel consumption of the main engine MESFOC, and water reference speed ShipSpdToWater) and three cluster centres are obtained. To consider the effects of the ship's loading status, weather conditions, and other factors, on the performance of the main engine, the cluster analysis was carried out on all data, ballast state data, and full load data, respectively.

Cluster analysis of all data

The cluster analysis results of all data are shown in Table 3 and Fig. 13-16. The three dark blue data clusters in Fig. 13 show the areas where the main engine often runs. Table 3 shows the three clustering centres and the clustering results are very consistent with the statistical analysis of data distribution. For a more intuitive display, Fig. 14 is a visualisation of three-dimensional data clusters of shaft speed, power, and fuel consumption rate, Fig. 15 displays two-dimensional data clusters of shaft speed and shaft power, and Fig. 16 is a visualisation of two-dimensional data clusters of water reference speed and shaft power.

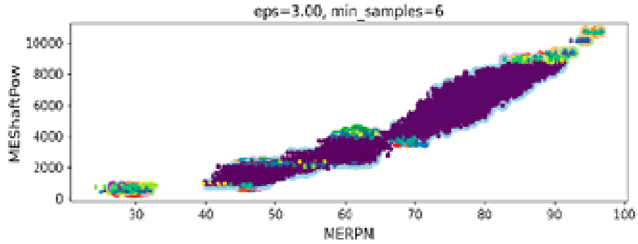

Fig. 13. Clustering analysis of shaft speed and power by DBSCAN in all states

Tab. 3. Clustering centre (All data)

Cluster centre	MERPM (r/min)	MEShaftPow (kW)	MEActFOCon (kg/h)	MESFOC (g/kw*h)	ShipSpdToWater (knot)
Cluster centre 1	46.56	1294.62	293.04	227.80	9.53
Cluster centre 2	78.10	6096.09	1123.29	184.68	16.07
Cluster centre 3	61.48	2970.05	591.04	199.43	12.60

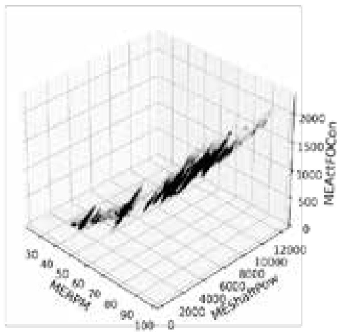

Fig. 14. Clustering analysis of shaft speed, power and fuel consumption rate by GMM in all states

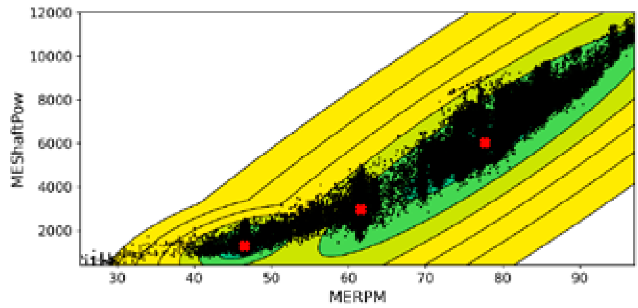

Fig. 15. Clustering analysis of shaft speed and power by GMM of all data

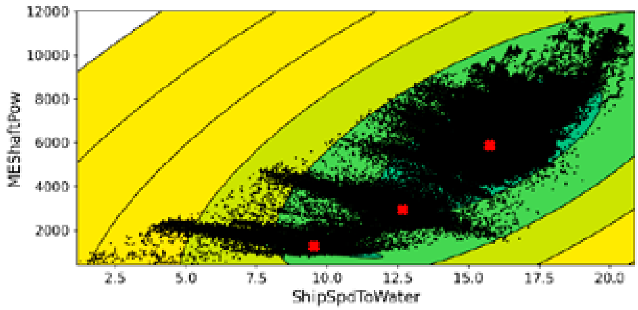

Fig. 16. Clustering analysis of water referenced ship speed and shaft power by GMM of all data

Clustering analysis of data under ballast and full load conditions

With the same steps (DBSCAN is used to determine three clusters and GMM is used to determine the cluster centres of key variables) and settings (EPS = 3, min_samples = 6), cluster analysis was performed on the data of ships with ballast and full load state, respectively. The cluster centres obtained from the analysis are shown in Tables 4 and 5.

Tab. 4. Clustering centre (data under relative wind speed in 12-18 knots and ballast state)

Cluster centre	MERPM (r/min)	MEShaftPow (kW)	MEActFOCon (kg/h)	MESFOC (g/kw*h)	ShipSpdToWater (knot)
Cluster centre 1	46.42	1194.62	293.04	234.04	9.84
Cluster centre 2	76.83	5469.11	1013.31	185.63	16.80
Cluster centre 3	60.46	2708.24	548.33	203.84	12.73

Tab. 5. Clustering centre (data under relative wind speed in 12-18 knots and full load state)

Cluster centre	MERPM (r/min)	MEShaftPow (kW)	MEActFOCon (kg/h)	MESFOC (g/kw*h)	ShipSpdToWater (knot)
Cluster centre 1	46.86	1433.69	314.81	221.08	9.10
Cluster centre 2	76.37	5779.90	1072.69	186.00	15.06
Cluster centre 3	61.57	3085.68	604.20	196.02	12.14

These cluster centres are areas with large amounts of data. They can be used to set the fuel consumption rate and the shaft speed in the mechanism analysis of the speed characteristic curve and the load characteristic curve.

MECHANISM ANALYSIS RESULTS

In the three clustering centres, we applied the polynomial curve fitting to fit the relationship between the key indicators of the main engine. Firstly, we took the three cluster centres of shaft speed MERPM generated by GMM as the set value for the data query and fitted the curve to obtain the relationship between MEShaftPow - MESFOC_kw of the main engine, which is the load characteristic curve. Then we took the fuel consumption rate MEActFOCon of cluster centre 2 as the set value for the data query, and fitted the curve to obtain the relationship between MERPM - MESFOC_kw, i.e. the speed characteristic curve. These characteristic curves help us to understand the changing trends of the main engine performance.

COMPARISON OF MAIN ENGINE LOAD CHARACTERISTIC CURVES IN TWO PERIODS

Fig. 17 and 18 are the correlation diagrams of MEShaftPow - MESFOC_kw at the cluster centres under ballast and full load states, respectively. Fig. 17 and 18 both show that the main engine's specific fuel consumption MESFOC_kw in 2021 was higher than that in 2020. Tables 6 and 7 list the average main engine shaft power (average MEShaftPow) and the average specific fuel consumption (average MESFOC_kw) of the load characteristic curves in 2020 and 2021. In summary, the data of both ballast and full load conditions show that the specific fuel consumption rate of the engine has increased due to engine degradation. The parent engine test data in the manual shows that the specific fuel consumption rate at 50% MCR is 167.8 g/kWh, and the testing fuel LCV is 42009.20 kJ/kg. The measured data of the engine in actual operation is higher than the engine test. The measured data is reasonable since the actual operating conditions are more severe.

Tab. 5. Average MEShaftPow and MESFOC_kw (load characteristic curves under ballast state)

Cluster centre	2020		2021	
	MEShaftPow	MESFOC_kw	MEShaftPow	MESFOC_kw
Cluster centre 1	11.66% MCR	222.28 g/kWh	11.01% MCR	224.01 g/kWh
Cluster centre 2	39.93% MCR	185.90 g/kWh	39.84% MCR	186.73 g/kWh
Cluster centre 3	22.00% MCR	199.66 g/kWh	22.04% MCR	205.08 g/kWh

Tab. 6. Average MEShaftPow and MESFOC_kw (load characteristic curves under full load state)

Cluster centre	2020		2021	
	MEShaftPow	MESFOC_kw	MEShaftPow	MESFOC_kw
Cluster centre 1	14.07% MCR	213.43 g/kWh	13.98% MCR	215.80 g/kWh
Cluster centre 2	47.66% MCR	183.78 g/kWh	47.27% MCR	185.94 g/kWh
Cluster centre 3	25.03% MCR	194.26 g/kWh	25.00% MCR	195.37 g/kWh

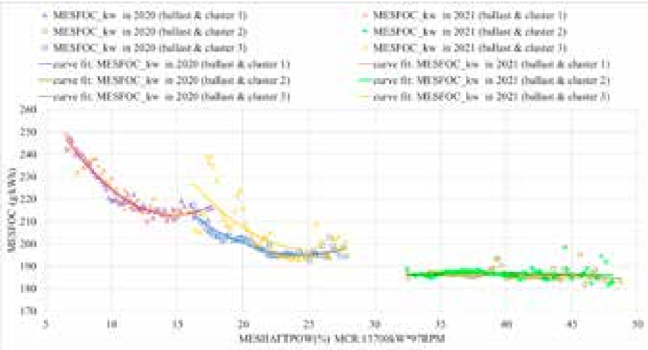

Fig. 17. Load characteristic curve of main engine shaft power - specific fuel consumption at cluster centres of ballast state

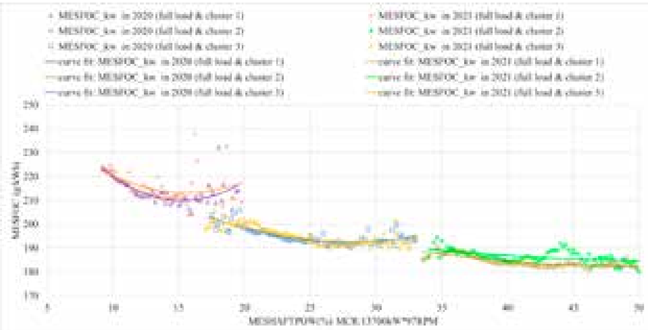

Fig. 18. Load characteristic curve of main engine shaft power - specific fuel consumption at cluster centres of full load state

COMPARISON OF MAIN ENGINE SPEED CHARACTERISTIC CURVES IN TWO PERIODS

The speed characteristic curves of the ship in the ballast and full load states, during the comparison periods, are created at cluster centre 2, as shown in Fig. 19.

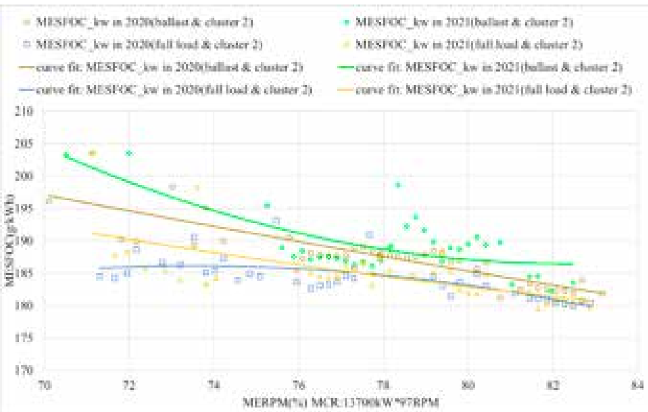

Fig.19. Speed characteristic curve of main engine shaft speed - specific fuel consumption at cluster centre 2 of ballast & full load states

The MEActFOCon of cluster centre 2 in the full load state is 1072.69 kg/h, while that in the ballast state is 1013.31 kg/h. At the query point, the engine load in the full load state is a little higher than in the ballast state, and so MESFOC_kw of the speed characteristic curve in the full load state is a little lower than that in the ballast state. Although the results also show that the specific fuel consumption rate in 2021 was higher than in 2020, the amount of data in the speed characteristic curve is still relatively small, and it is necessary to collect more data for further research.

CONCLUSION

This paper presents a main engine performance monitoring method based on energy efficiency big data. The method combines distribution statistics, clustering, and polynomial curve fitting to obtain the propulsion characteristics and load characteristic curves, as the basis of performance degradation monitoring and maintenance schedule of the main engine. The test results show that this method can effectively monitor the degradation of engine performance.

This research can be used as an important basis for creating a maintenance schedule for main marine engines. In the future, the ship maintenance schedule can be determined automatically, by developing the ship's intelligent energy efficiency system, which should consist of various autonomous technologies, such as machine learning (ML) and artificial intelligence (AI) technologies, assisting in intelligent monitoring, forecasting, and making auxiliary fuel-saving decisions.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (52001195, 61873161) and the Science & Technology Commission of Shanghai Municipality and Shanghai Engineering Research Center of Ship Intelligent Maintenance and Energy Efficiency (20DZ2252300).

REFERENCES

1. B. Qiao, W. He, Y. Tian, Y. Liu, O. Cai, and Y. Li, "Ship emission reduction effect evaluation of air pollution control countermeasures," *Transportation Research Procedia*, vol. 25, pp. 3606-3618, 2017, doi: 10.1016/j.trpro.2017.05.325.

2. F. Jasper, H. Shinichi et al., "Fourth IMO Greenhouse Gas Study 2020," *International Maritime Organisation (IMO)*, London, UK, 2020.

3. V. Kuznetsov, B. Dymo, S. Kuznetsova, M. Bondarenko, and A. Voloshyn, "Improvement of the cargo fleet vessels power plants ecological indexes by development of the exhaust gas systems," *Polish Maritime Research,* vol. 28, pp. 97-104, 2021, doi: 10.2478/pomr-2021-0009.

4. I. Ančić and A. Šestan, "Influence of the required EEDI reduction factor on the CO_2 emission from bulk carriers,"

Energy Policy, vol. 84, pp. 107-116, 2015, doi: 10.1016/j.enpol.2015.04.031.

5. E.K. Hansen, H.B. Rasmussen, and M. Lützen, "Making shipping more carbon-friendly? Exploring ship energy efficiency management plans in legislation and practice," *Energy Research & Social Science*, vol. 65, pp. 101459, 2020, doi: 10.1016/j.erss.2020.101459.

6. M. Kalajdžić, M. Vasilev, and N. Momčilović, "Power Reduction Considerations for Bulk Carriers with respect to Novel Energy Efficiency Regulations," *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike*, vol. 73, pp. 79-92, 2022, doi: 10.21278/brod72205.

7. L. Fedi, "The Monitoring, Reporting and Verification of Ships' Carbon Dioxide Emissions: A European Substantial Policy Measure towards Accurate and Transparent Carbon Dioxide Quantification," *Ocean Yearbook Online*, vol. 31, pp. 381-417, 2017, doi: 10.1163/22116001-03101015.

8. W. Tarełko, "The effect of hull biofouling on parameters characterising ship propulsion system efficiency," Polish Maritime Research, vol. 21, pp. 27-34, 2014, doi: 10.2478/pomr-2014-0038.

9. P. Król, «Hydrodynamic state of art review: rotor–stator marine propulsor systems design,» *Polish Maritime Research,* vol. 28, pp. 72-82, 2021, doi: 10.2478/pomr-2021-0007.

10. P. Puzdrowska, «Diagnostic information analysis of quickly changing temperature of exhaust gas from marine diesel engine. Part i single factor analysis,» *Polish Maritime Research,* vol. 28, pp. 97-106, 2021, doi: 10.2478/pomr-2021-0052.

11. P. Król, "Blade section profile array lifting surface design method for marine screw propeller blade," *Polish Maritime Research,* vol. 26, pp. 134-141, 2019, doi: 10.2478/pomr-2019-0075.

12. K. Rudzki and W. Tarelko, "A decision-making system supporting selection of commanded outputs for a ship's propulsion system with a controllable pitch propeller," *Ocean Engineering,* vol. 126, pp. 254-264, 2016, doi: 10.1016/j.oceaneng.2016.09.018

13. R. Varbanets, V. Zalozh, A. Shakhov, I. Savelieva, and V. Piterska, "Determination of top dead centre location based on the marine diesel engine indicator diagram analysis," *Diagnostyka,* vol. 21, pp. 51-60, 2020, doi: 10.29354/diag/116585.

14. S. Park, S. W. Park, S. H. Rhee, S. B. Lee, J. E. Choi, and S. H.Kang, "Investigation on the wall function implementation for the prediction of ship resistance," *International Journal of Naval Architecture and Ocean Engineering*, vol. 5, pp. 33-46, 2013, doi: 10.2478/IJNAOE-2013-0116.

15. M.B. Samsul, "Blade cup method for cavitation reduction in marine propellers," *Polish Maritime Research*, 2021, doi: 10.2478/pomr-2021-0021, doi: 10.2478/pomr-2021-0021.

16. M.H. Ghaemi, "Performance and emission modelling and simulation of marine diesel engines using publicly available engine data," *Polish Maritime Research,* vol. 28, pp. 63-87, 2021, doi: 10.2478/pomr-2021-0050.

17. B.D. Brouer, C.V. Karsten, and D. Pisinger, "Big data optimisation in maritime logistics," *Big data optimisation: Recent developments and challenges. Springer, Cham*, vol. 18, pp. 319-344, 2016, doi: 10.1007/978-3-319-30265-2_14.

18. X. Zeng and M. Chen, "A Novel Big Data Collection System for Ship Energy Efficiency Monitoring and Analysis Based on BeiDou System," *Journal of Advanced Transportation*, vol. 2021, pp.1-10, 2021, doi: 10.1155/2021/9914720.

19. I. Zaman, K. Pazouki, R. Norman, S. Younessi, and S.Coleman, "Challenges and opportunities of big data analytics for upcoming regulations and future transformation of the shipping industry," *Procedia engineering*, vol. 194, pp. 537-544, 2017, doi: 10.1016/j.proeng.2017.08.182.

20. A. Fan, X. Yan, and Q. Yin, "A multisource information system for monitoring and improving ship energy efficiency," *Journal of Coastal Research,* vol.32, pp. 1235-1245, 2016, doi: 10.2112/JCOASTRES-D-15-00234.1.

21. J. Deng, J. Zeng, S. Mai, B. Jin, B. Yuan, Y. You. S. Lu, and M.Yang, «Analysis and prediction of ship energy efficiency using 6G big data internet of things and artificial intelligence technology,» *International Journal of System Assurance Engineering and Management,* vol. 12, pp. 824–834, 2021, doi: 10.1007/s13198-021-01116-9.

22. T. Niksa-Rynkiewicz, N. Szewczuk-Krypa, A. Witkowska, K. Cpałka, M. Zalasiński, and A. Cader, "Monitoring regenerative heat exchanger in steam power plant by making use of the recurrent neural network," *Journal of Artificial Intelligence and Soft Computing Research,* vol. 11, pp. 143-155, 2021, doi: 10.2478/jaiscr-2021-0009.

23. A. Witkowska and T. Niksa-Rynkiewicz, «Dynamically positioned ship steering making use of backstepping method and artificial neural networks,» *Polish Maritime Research,* vol. 25, pp. 5-12, 2018, doi: 10.2478/pomr-2018-0126.

24. S. García, J. Luengo, and F. Herrera, "Data preprocessing in data mining," *Cham, Switzerland: Springer International Publishing,* vol. 72, pp. 59-139, 2015, doi: 10.1007/978-3-319-10247-4.

25. L.P. Perera and B. Mo, "Ship performance and navigation information under high-dimensional digital models," *Journal of Marine Science and Technology,* vol. 25(1), pp. 59-139, 2020, doi: 10.1007/978-3-319-10247-4.

26. Y. Raptodimos and I. Lazakis, "Using artificial neural network-self-organising map for data clustering of marine engine condition monitoring applications," *Ships and Offshore Structures,* vol. 13, pp. 649-656, 2018, doi: 10.1080/17445302.2018.1443694.

27. E. Vanem and A. Brandsæter, "Unsupervised anomaly detection based on clustering methods and sensor data on a marine diesel engine," *Journal of Marine Engineering & Technology,* vol. 20, pp. 217-234, 2021, doi: 10.1080/20464177.2019.1633223.

28. L. P. Perera, and B. Mo, "Data analytics for capturing marine engine operating regions for ship performance monitoring," *International Conference on Offshore Mechanics and Arctic Engineering,* American Society of Mechanical Engineers, 2016, Vol. 49989, doi: 10.1115/OMAE2016-54168

29. L. P. Perera, and B. Mo, "Marine engine operating regions under principal component analysis to evaluate ship performance and navigation behaviour," IFAC-PapersOnLine, vol. 49(23), pp. 512-517, 2016, doi: 10.1016/j.ifacol.2016.10.487.

30. X. Yan, K. Wang, Y. Yuan, X. Jiang, and R. R. Negenborn, "Energy-efficient shipping: An application of big data analysis for optimizing engine speed of inland ships considering multiple environmental factors," Ocean Engineering, vol. 169, pp. 457-468, 2018, doi: 10.1016/j.oceaneng.2018.08.050.

31. K. Wang, X. Yan, Y. Yuan, X. Jiang, G. Lodewijks, and R. R. Negenborn, «Study on route division for ship energy efficiency optimisation based on big environment data,» *2017 4th International Conference on Transportation Information and Safety (ICTIS), IEEE,* pp. 111-116, 2017, doi: 10.1109/ICTIS.2017.8047752.

32. R. Adland, P. Cariou, H. Jia, and F. C. Wolff, "The energy efficiency effects of periodic ship hull cleaning," *Journal of Cleaner Production*, vol. 178, pp. 1–13, 2018, doi: 10.1016/j.jclepro.2017.12.247.

33. O. Loyola-Gonzalez, "Black box vs. white-box: Understanding their advantages and weaknesses from a practical point of view, " *IEEE Access,* vol. 7, pp. 154096–154113, 2019, doi: 10.1109/ACCESS.2019.2949286.

34. X. Zeng, M. Chen, H. Li and X. Wu, "A Data-Driven Intelligent Energy Efficiency Management System for Ships," *IEEE Intelligent Transportation Systems Magazine*, doi: 10.1109/MITS.2022.3153491.

35. M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *kdd,* vol. 96, pp. 226-231, 1996.

36. R. Varbanets, V. Klymenko, O. Fomin, V. Píštěk, P. Kučera, D. Minchev, A. Khrulev, and V. Zalozh, "Acoustic method for estimation of marine low-speed engine turbocharger parameters," Journal of Marine Science and Engineering, vol. 9, 2021, doi: 10.3390/jmse9030321.

37. A. GéRon, "Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems". *Sebastopol, CA, USA: O'Reilly Media,* 2017.

CONTACT WITH THE AUTHORS

Meng Liang
e-mail: liangmeng02@163.com

Shanghai Dianji University, Business School,
Shanghai
China

Mingzhi Chen

Shanghai Maritime University,
Merchant Marine College and Shanghai Engineering
Research Center of Ship Intelligent Maintenance
and Energy Efficiency, Shanghai
China