LJMU Research Online

Yuan, Z, Liu, J, Zhang, Q, Liu, Y, Yuan, Y and Li, Z

Prediction and optimisation of fuel consumption for inland ships considering real-time status and environmental factors

http://researchonline.ljmu.ac.uk/id/eprint/14272/

Article

Prediction and optimisation of fuel consumption for inland ships considering real-time status and environmental factors

Zhi Yuan[1,2,3], Jingxian Liu[1,2], Qian Zhang[3*], Yi Liu[1,2*], Yuan Yuan[4] and Zongzhi Li[5]

[1] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, 1040 Heping Avenue, Wuhan, Hubei 430063, PR China.
[2] National Engineering Research Centre for Water Transport Safety (WTSC), Wuhan University of Technology, 1040 Heping Avenue, Wuhan, Hubei 430063, PR China.
[3] Department of Electronics and Electrical Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK.
[4] ChangJiang Shipping Science Research Institute CO. Ltd., Wuhan 430060, China
[5] Department of Civil, Architectural and Environmental Engineering, Illinois Institute of Technology, Chicago, IL 60616, USA

*corresponding author, email: Qian Zhang Q.Zhang@ljmu.ac.uk
                                 Yi Liu liuyi_hy@whut.edu.cn;

Abstract: The information about ships' fuel consumption is critical for condition monitoring, navigation planning, energy management and intelligent decision-making. Detailed analysis, modelling and optimisation of fuel consumption can provide great support for maritime management and operation and are of significance to water transportation. In this study, the real-time status monitoring data and hydrological data of inland ships are collected by multiple sensors, and a multi-source data processing method and a calculation method for real-time fuel consumption are proposed. Considering the influence of navigational status and environmental factors, including water depth, water speed, wind speed and wind angle, the Long Short-Term Memory (LSTM) neural network is then tailored and implemented to build models for prediction of real-time fuel consumption rate. The validation experiment shows the developed model performs better than some regression models and conventional Recurrent Neural Networks (RNNs). Finally, based on the fuel consumption rate model and the speed over ground model constructed by LSTM, the Reduced Space Searching Algorithm (RSSA) is successfully used to optimise the fuel consumption and the total cost of a whole voyage.

Key words: inland ship; fuel consumption; data-driven modelling; optimisation; LSTM; RSSA

1 Introduction

Ships' fuel consumption occupies a major part of ship operating costs (Leifsson et al., 2008). It also reflects the operating status of ship engines (Schaub et al., 2019), which makes it an important monitoring variable in modern intelligent ship systems (Li et al., 2019) and a key parameter to control in unmanned ship navigation systems (Wright, 2019). On the other hand, the ship fuel consumption is also closely related to exhaust emissions and is an important indicator for pollution research and environmental monitoring (Van et al., 2019; Hansen et al., 2020). Therefore, the ship fuel consumption has become an important research topic for many scholars and practitioners.

Over the past decade, ship speed optimisation has been considered as an effective approach to improve the energy efficiency and reduce the fuel consumption. Therefore, some researchers have focused on the speed optimisation of the whole voyage for reducing the fuel consumption (Fagerholt et al., 2010; Wang and Meng, 2012; Psaraftis et al., 2014; Fagerholt et al., 2015; Wen et al., 2017; Li et al., 2018; Du et al., 2019). However, the speed of a ship depends to a large extent on the speed of the ship engines, the engines' running conditions and the environmental conditions. All of these factors have a significant effect on the ship fuel consumption and fleet cost during the voyage. Wang et al. (2016) established an approach for real-time optimisation of ship energy efficiency during the working condition in a short distance ahead of the ship and achieved real-time optimisation under different navigation conditions. Sheng et al. (2019) developed a mixed-integer convex cost-minimisation method for determination of optimal vessel speeds and fleet size. However, they all took the ship speed as the decision variable, where the ship speed is provided by the output power of engines, and is also closely related to navigational conditions. To achieve a suggested speed of a ship in the varying navigational conditions, one may need to constantly change the engine running speed, which may cause more fuel consumption than expected.

In recent years, intelligent sensing devices with high acquisition rates are more and more widely used in modern ships, and many real-time and continuous data collection systems have been developed. Using the new systems, a large number of multi-source monitoring data, including longitude, latitude, Speed Over Ground (SOG), Course Over Ground (COG), engine speed, engine temperature, voyage mileage, reserve fuel and bunker fuel, have been collected. This provides abundant fundamental data for fuel consumption prediction,

energy efficiency optimisation and emission reduction (Satpathi et al., 2017; Huang et al., 2018; Wang et al., 2019).

For fuel consumption prediction, there have been some methods and models in the existing literature. Beşikçi et al. (2016) tried to reduce the ship speed and predict ship fuel consumption for various operational conditions through employment of an Artificial Neural Network (ANN), where seven input variables, including ship speed, revolutions per minute, mean draft, trim, cargo quantity on board, wind and sea effects, were used. Coraddu et al. (2017) compared three different approaches, a White Box Model, a Black Box Model and a Gray Box Model (GBM), in the prediction of the fuel consumption based on data measured by on-board automation systems. Wang et al. (2018) proposed a prediction model for the ship fuel consumption on the basis of the Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm. It used the dataset of ship reports, which includes the information about length of overall, beam, SOG, Beaufort scale and swell height. Yuan and Nian (2018) developed a Gaussian Process (GP) meta-model to predict the ship fuel consumption for different scenarios in consideration of the operational conditions' effects, which involves speed, draft, trim, wind speed, wind direction, wave height and wave direction. Yang et al. (2019) proposed a genetic-algorithm-based GBM for the ship fuel consumption prediction using ship speed and Beaufort number. Gkerekos et al. (2019) presented a comparison of multiple data-driven regression algorithms for predicting the main engine fuel oil consumption, including Support Vector Machines, Random Forest Regressors, Extra Trees Regressors and ANNs. They considered vessel speed, engine speed and sea conditions as input variables.

However, the existing research has some limitations. (1) Limited input variables were used for fuel consumption prediction. There is a lack of some important variables, such as engine temperature, water speed and wind direction. (2) There is a lack of detailed analysis about the trajectory characteristics and geographic environment when predicting fuel consumption. (3) In the optimisation of fuel consumption and total cost, only the vessel speed was used as the decision variable, while the more directly-related and controllable variable, engine speed, was not employed. (4) The environmental factors were not taken into consideration when optimising the fuel consumption and the total running cost.

To solve the above problems, this work collects various monitoring data of ship sailing by multi-source sensors. After specific data processing and analysis, the real-time fuel consumption rate of ships is calculated

and the feature variables that are most correlated with fuel consumption are obtained. The prediction model for real-time fuel consumption rate is then constructed based on the Long Short-Term Memory (LSTM) network, which is verified by the measured data and compared with some traditional regression methods, Back Propagation Neural Networks (BPNNs) and other Recurrent Neural Networks (RNNs). The prediction model of SOG will be used for fuel consumption and cost optimisation, which is also built by the LSTM network. Finally, an optimisation algorithm Reduced Space Searching Algorithm (RSSA) is used to minimise the fuel consumption and the total cost of a voyage. RSSA is a nature-inspired heuristic technique that tries to switch and zoom in/out the targeted search space to speed up the searching process and jump out from local optima. It has been verified to be able to find optimal solutions fast and accurately, and outperform some other well-known heuristic optimisation algorithms (Zhang and Mahfouf, 2010). The whole research framework is as shown in Fig. 1.
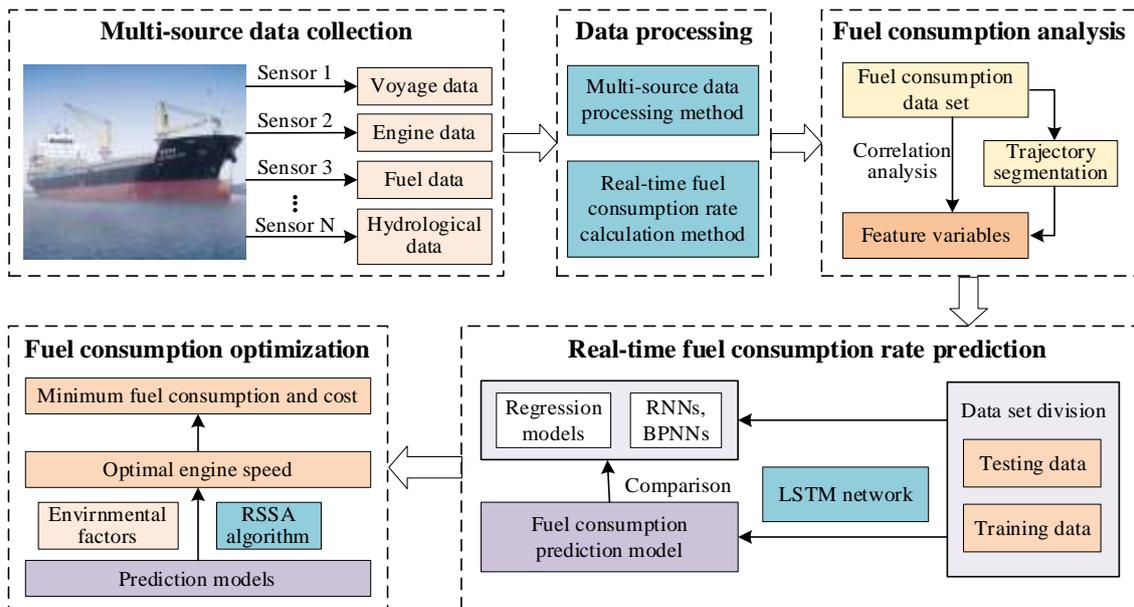
Fig. 1. The framework for fuel consumption prediction and optimisation

The rest of the paper is organised as follows. First, the detailed multi-source data processing and analysis are presented in Section 2, where a real-time fuel consumption calculation method is proposed and the correlation between multiple variables is analysed. Then, the prediction model of real-time fuel consumption rate is constructed in Section 3. Detailed experiments are carried out to optimise the fuel consumption and the total voyage cost of inland ships in Section 4. Finally, conclusions are drawn in Section 5. Table 1 summarises the abbreviations used in the paper.

4

Table 1. A list of abbreviations

Abbreviation	Meaning	Abbreviation	Meaning
AIS	Automatic Identification System	GPS	Global Positioning System
ANN	Artificial Neural Network	IMO	International Maritime Organization
BPNN	Back Propagation Neural Networks	IR	Interaction Linear Regression
BRNN	Bidirectional RNN	LR	Linear Regression
COG	Course Over Ground	LSTM	Long Short-Term Memory
CYN	China Yuan	MAE	Mean Absolute Error
DBPNN	BPNN with two hidden layers	PQR	Pure Quadratic Regression
DRNN	Deep RNN	PSO	Particle Swarm Optimisation
ES	Engine Speed	RLR	Robust Linear Regression
ET	Engine Temperature	RMSE	Root Mean Square Error
FCRM	Fuel Consumption Rate Model	RNN	Recurrent Neural Network
FTR	Fine Tree Regression	RSSA	Reduced Space Searching Algorithm
GA	Genetic Algorithm	SOG	Speed Over Ground
GBM	Gray Box Model	SOGM	SOG model

2 Multi-source data analysis

2.1 Data collection

The data studied in this work came from the cargo ship sailing on the Yangtze River trunk, which was equipped with two engines rated at 735 kW. The main parameters of the ship are shown in Table 2. The raw data were collected by the multi-source sensors installed on the ship, such as Global Positioning System (GPS), Automatic Identification System (AIS), fuel sensor, speed sensor, temperature sensor and others. The collected data include IMO, ship name, date, time, longitude, latitude, SOG, COG, voyage mileage, engine speed, engine temperature, bunker fuel, tank fuel, and hydrological data such as water depth, water speed, wind speed and wind direction, as shown in Fig. 2. The multi-source data set includes 32,143 data records, collected from the Yangtze River trunk from Wusongkou to Chongqing, between September 11, 2019 (11:00:58) and October 7, 2019 (23:59:23), which belongs to the dry season. The example of original data during September 11, 2019 is shown in Table 3.

Table 2. Main parameters of the ship

| Parameter name	Value	Unit
Designed length	110.00	m
Moulded breadth	19.20	m
Moulded depth	5.60	m
Deadweight	7028	Mt
Designed draught (full load)	4.65	m
Designed speed (calm water)	18	km/h
Main engine rated power	735×2	kW
Main engine rated speed	830	rpm
Propeller diameter	2.50	m
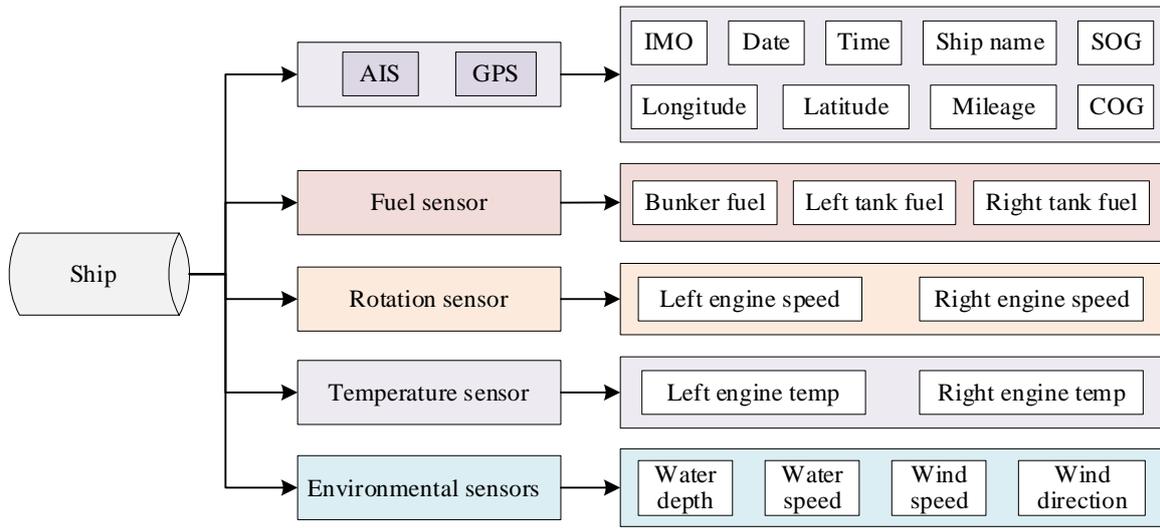
Fig. 2. Multi-source monitoring data collection

Table 3. Original monitoring data of the ship (during 2019/09/11)

Time	Longitude	Latitude	SOG (km/h)	COG (°)	Reserve fuel (L)		Barge oil (L)	Engine speed (rpm)		Engine temper(°C)		…
					Left	Right		Left	Right	Left	Right	
11:00:58	121.5690	31.4025	0.078	0	1428.6	1562.9	0	0	0	36.4	35.4	…
11:02:08	121.5690	31.4025	0.085	0	1428.6	1562.7	0	0	0	36.1	35.8	…
11:03:15	121.5690	31.4025	0.067	0	1429.1	1562.9	0	0	0	36.2	35.7	…
11:04:24	121.5690	31.4025	0.057	0	1429.0	1562.2	0	0	0	36.4	35.7	…
……	……	……	……	……	……	……	……	……	……	……	……	…
17:00:11	121.1897	31.6794	7.456	300.16	1274.9	1395.7	0.0294	438.8	459.3	60.5	56.6	
17:01:41	121.1880	31.6801	7.734	299.41	1273.2	1394.3	0.0294	438.9	459.2	60.7	56.4	
17:03:01	121.1870	31.6806	7.658	298.94	1274.4	1391.2	0.0294	437.6	459.6	60.9	56.8	
17:04:08	121.1860	31.6812	7.549	296.70	1272.2	1393.4	0.0294	439.3	459.6	60.7	56.6	
……	……	……	……	……	……	……	……	……	……	……	……	…

2.2 Data pre-processing

The multi-source monitoring data were obtained by continuous time sampling. Therefore, the raw data usually includes some errors and anomalies due to ship docking and swaying, data transmission delay and/or other reasons, as shown in Fig. 3. For example, the normal range of the engine speed is from 350 to 780 rpm, while in the collected data set there are some abnormal engine speed readings between 100 and 350 rpm; for longitude, the data samples with value 0 are obviously erroneous data.

6
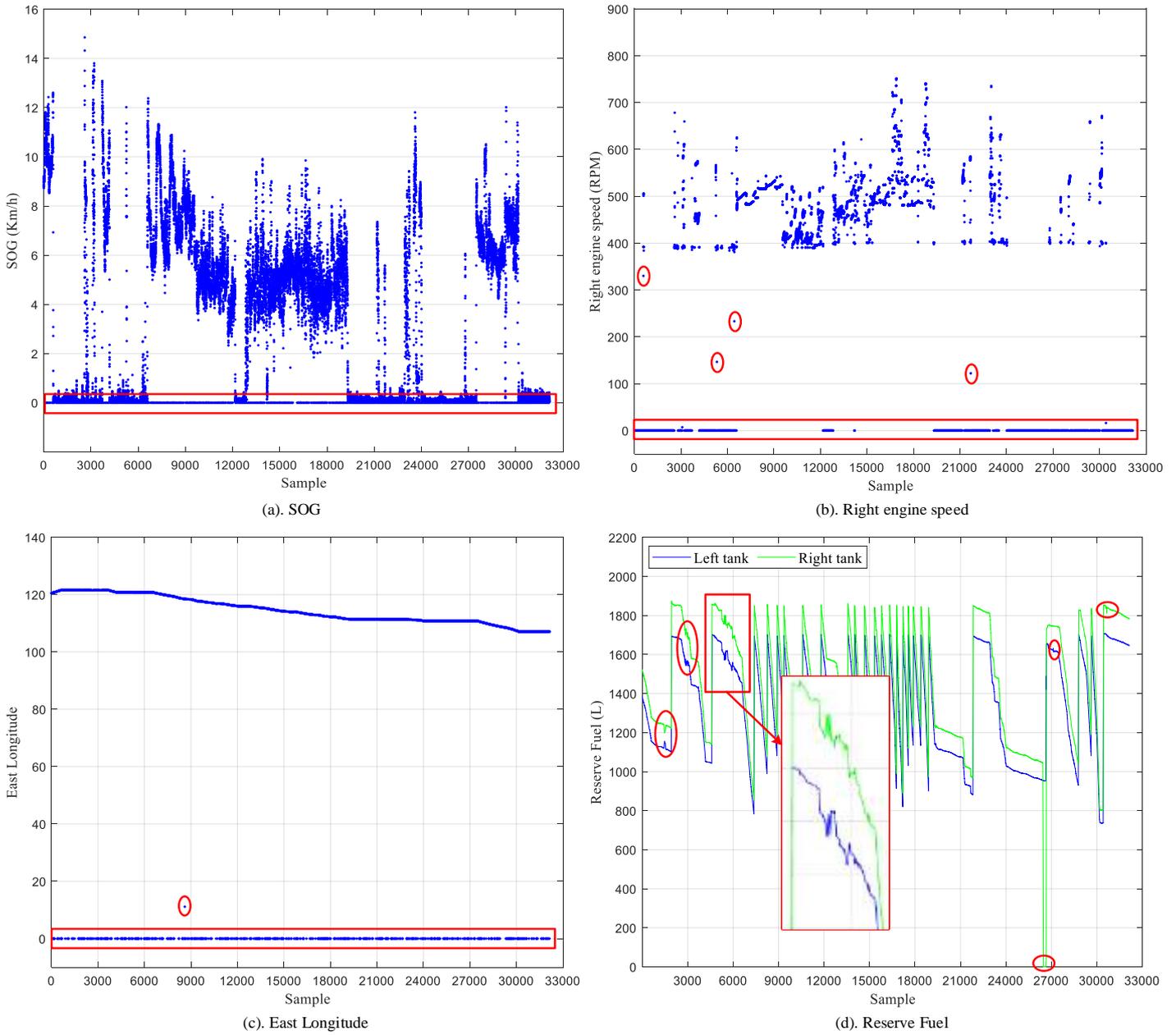
Fig. 3. Anomalies and errors in raw data: (a) SOG with anomalies caused by ship docking; (b) right engine speed with several erroneous data; (c) longitude with a large number of erroneous data; (d) reserve fuel of the left and right tanks with outliers caused by ship shaking.

It should be noted that the raw data containing errors and noise were collected by different sensors. That is to say, for the data collected at the same time, if the longitude is wrong, it does not mean that the engine speed is also wrong. Moreover, not both engines work all the time. Therefore, we propose a data processing method, as shown in Method 1 to clean the original monitoring data.

Method 1. Multi-source data processing
Input: *RawDataSet* (Raw data set)
Output: *CleanDataSet* (Data set after processing)

[1] Initialise, $CleanDataSet \leftarrow \emptyset$

[2] Sort $RawDataSet$ by $Time$

[3] Delete the records of time duplication in $RawDataSet$

[4] Find records with $Longitude$ is wrong

[5] Repair the wrong longitude value using spline interpolation

[6] Find records with SOG is zero, $ID\_1 = find\ (RawDataSet\ [SOG] == 0)$

[7] Find records with COG is zero, $ID\_2 = find\ (RawDataSet\ [COG] == 0)$

[8] Find records with $LeftES$ is below $Threshold$,
$\quad ID\_3 = find\ (RawDataSet\ [LeftES] < ThresholdES)$

[9] Find records with $RightES$ is below $Threshold$,
$\quad ID\_4 = find\ (RawDataSet\ [RightES] < ThresholdES)$

[10] Find records with engine speed abnormal, $ID\_5 = ID\_3 \cap ID\_4$

[11] Find records with abnormal data, $ID\_6 = ID\_1 \cup ID\_2 \cup ID\_5$

[12] Delete abnormal data, $CleanDataSet \leftarrow RawDataSet\ [ID\_6] = [\ ]$

In Method 1, $LeftES$ and $RightES$ are left and right engine speeds. $ThresholdES$ is the threshold of engine speed, which is set to 350 rpm. $ID\_1,\ ID\_2,\ ID\_3,\ ID\_4,\ ID\_5$ and ID_6 are temporary index variables. First, the data are sorted by time, where the raw data are sampled at one-minute intervals. Then, according to the characteristics of SOG, COG and engine speed, one finds out the indices of their outliers. Then, one finds out the indices of the abnormal engine speed samples by intersection operation. Finally, all the abnormal data found by union operation are deleted. After the data processing by Method 1, a clean and ready-to-use data set is obtained.

2.3 Calculation of real-time fuel consumption rate

Normally, the collected multi-source monitoring data only include the amount of fuel in different fuel containers, including the bunker fuel, left reserve fuel and right reserve fuel of the ship, during navigation. There is no real-time fuel consumption rate, which needs to be calculated using the recorded fuel information. However, there are three difficulties in the calculation. First of all, during a voyage, the bunker fuel is irregularly refilled for many times, and the change in the amount of bunker fuel directly relates to the reserve fuel, which can be seen from Fig. 4. Fig. 4 (b) is an enlarged view of the black-boxed area in Fig. 4 (a). As shown in Fig. 4 (b), when the volume of bunker fuel remains unchanged, the left and right reserve fuel gradually decrease, which reflects the normal fuel consumption of ship sailing. When the volume of bunker fuel increases, the volume of the left and right reserve fuel increase rapidly, because the cumulative amount of banker fuel is the amount of fuel added to the left and right tanks. When the refuelling process is over, the

reserve fuel returns to its normal charge. Second, the ship may dock for many times during the voyage, but the data during the docking time are not stopped recording, which has a great impact on the calculation of time and rate. Third, the ship sways during the voyage, which causes large fluctuations in the readings of the reserve fuel. Considering these issues, we propose a method to calculate the real-time fuel consumption rate as Method 2.


(a). Reserve fuel and Bunker fuel

(b). Enlargement of the black box in figure (a)
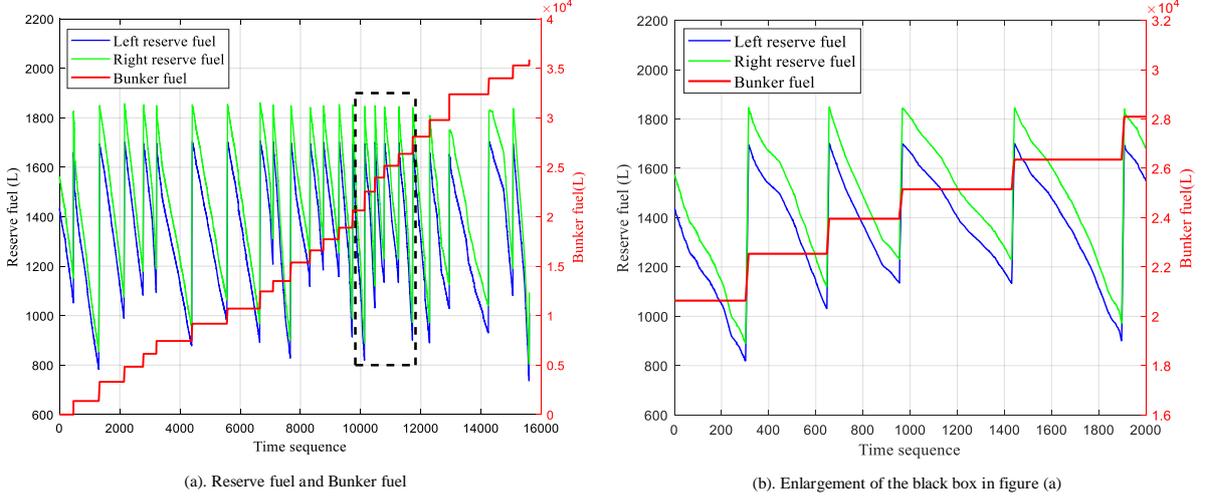
Fig. 4. Raw reserve fuel and bunker fuel.

In Method 2, i and j are loop variables; $TimeStart, TimeEnd$ and $TimeFirst$ are time variables. $ThresholdSOG$ is the threshold of SOG and we set it to 0.5 km/h, where a SOG less than 0.5 km/h indicates that the ship may be in the starting state before sailing; $ThresholdTime$ is the threshold of time used to identify if a continuous sampling is interrupted and we set it to 10 minutes; $DeltaTime$ is the time difference between adjacent data records; $TempD1$ and $TempD2$ are the temporary data. $DeltaBunkerF$ is the difference between adjacent bunker fuel; $LeftRF$ and $RightRF$ are fuel of left and right tanks; RF is the total reserve fuel; $Fuel$ is fuel consuming; $DeltaFuel$ is the difference between adjacent $Fuel$; $FuelCR$ is the real-time fuel consumption rate (L/min). First, we delete the SOG less than $ThresholdSOG$ to ensure the validity of the data set. Then, two judgments are made and the two cases of each judgment are treated differently. The first judgment is to determine whether there is an increase in the amount of bunker fuel according to its difference value. The second judgment is to determine whether there is berthing during the period, so as to calculate the fuel consumption according to the time period. In addition, before calculating the real-time fuel consumption rate, it adds the smooth processing for the total reserve fuel, eliminating the reading fluctuations of fuel caused by ship shaking and improving the accuracy of the calculation results.

9

Method 2. Real-time fuel consumption rate calculation
Input: $CleanDataSet$ (Data set after processing)
Output: $FuelDataSet$ (Fuel consumption data set)

[1] Initialise: $FuelDataSet \leftarrow \emptyset, i, j, TimeStart, TimeEnd, TimeFirst = 1$
[2] Delete the records with SOG is below $ThresholdSOG$
[3] Calculate the $DeltaBunkerF$
[4] **for** $DeltaBunkerF$ **do**
    [5] **if** $DeltaBunkerF(i) > 0$
        [6] Record current time: $TimeEnd = i$
        [7] Extract data from $CleanDataSet(TimeStart: TimeEnd)$ to $TempD1$
        [8] Extract $Time$ from $TempD1$, and calculate the $DeltaTime1$
        [9] **for** $DeltaTime1$ **do**
            [10] **if** $DeltaTime(j) > ThresholdTime$
                [11] Record current time: $TimeLast = j$
                [12] Extract data from $TempD1(TimeStart: TimeEnd)$ to $TempD2$
                [13] Extract $LeftRF$ and $RightRF$ from $TempData\_2$
                [14] Calculate RF, $RF = LeftRF + RightRF$
                [15] Smooth the RF, $RF = \boldsymbol{moving}(RF)$
                [16] Calculate $Fuel$ and $DeltaFuel$, $Fuel = BunkerF - RF$
                [17] Extract $Time$ from $TempD2$, and calculate the $DeltaTime2$
                [18] Calculate $FuelCR$: $FuelCR = DeltaFuel/DeltaTime2$
                [19] Save $FuelCR$, obtain the $FuelDataSet$
                [20] Update the $TimeFirst$, $TimeFirst = j + 1$
            [21] **else**
                Update the $TimeLast$, $TimeLast = j$
            [22] **end if**
            [23] Repeat the steps [12] to [19]
        [24] **end for**
        [25] Update the $TimeStart$, $TimeStart = i{+}1$
    [26] **else**
        Update the $TimeEnd$, $TimeEnd = i$
    [27] **end if**
    [28] Repeat the steps [7] to [24]
[29] **end for**

2.4 Trajectory segmentation

By the use of Method 1 and Method 2, we successfully obtained high-quality real-time fuel consumption

data. However, the whole voyage has a long mileage and spans many waterways, which may cause troubles in

real-time fuel consumption prediction. In addition to the geographical location of these waterways, the biggest

difference is the water depth. In fact, the environmental factors have an impact on the fuel consumption.

Generally, the wind and waves in inland rivers are relatively stable, and the water depth of the waterway is one

of the biggest influential factors on the fuel consumption of ships. Therefore, we divided the trajectory of whole voyage into 27 segments by (1) referring to the geographical characteristics information of the Yangtze River trunk line, (2) analysing the water depth and water speed of the main waterways, (3) considering the restrictions from navigation rules and the navigation experience of shipmasters. The segmentation is shown in Fig. 5. As shown in Fig. 5 (b), the waterway from Yuanshi to Yichang is divided into 4 segments, due to the existence of Three Gorges Dam and Gezhouba Dam, which cause large changes in water depth in a relatively short distance. As a consequence of segmentation, a new variable segment number is added to the fuel consumption data set.

(a). The waterway of the Yangtze River

(b). The actual trajectory of the ship

(c). The waterway of Segments 18-21
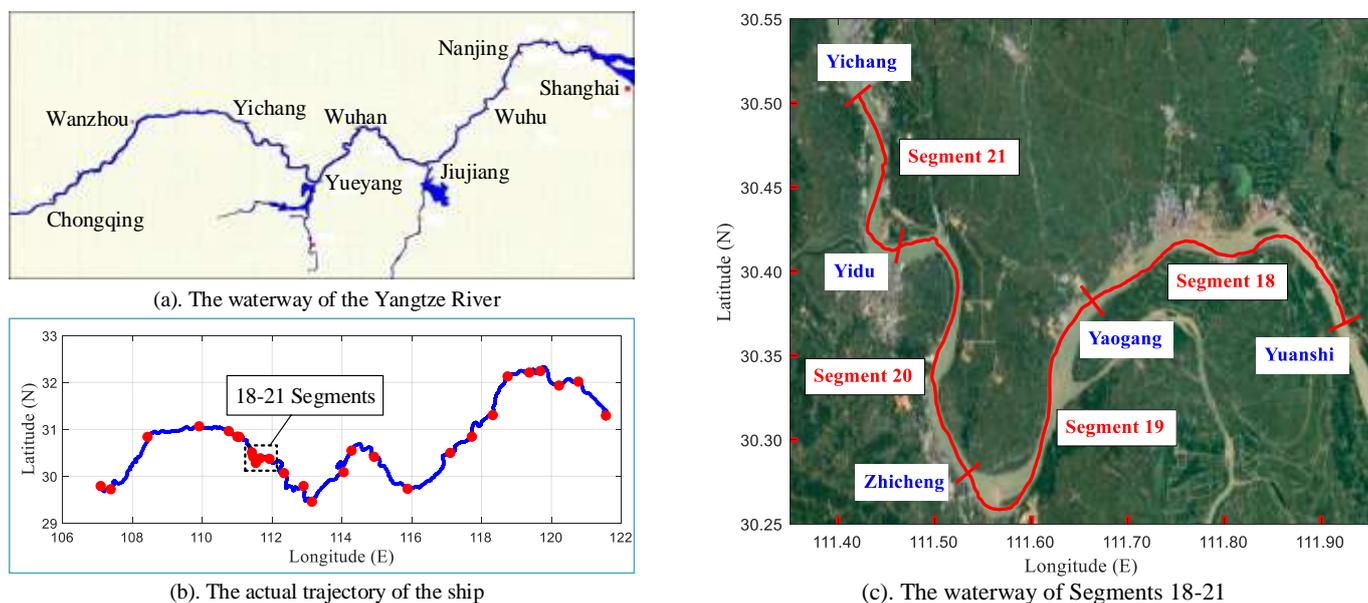
Fig. 5. The waterway of the Yangtze River, the ship trajectory and segments division

2.5 Correlation analysis

It is worth noting that many feature variables may affect the real-time fuel consumption. Wang et al. (2016) selected the engine speed as the input variable of a fuel consumption model, and Coraddu et al. (2017) used the ship speed and direction as inputs of a fuel consumption model. In fact, the fuel consumption of inland ships is not only related to navigation state variables such as SOG, COG, engine speed and engine temperature, but also affected by the environmental factors such as water depth, water speed, wind speed and wind direction.

In this study, we collected another important variable, engine temperature, which reflects the running state of the engine. We fully consider the comprehensive impact of the ship's real-time navigation status variables and environmental factors on the real-time fuel consumption rate. The angle between COG and the wind

11

direction is calculated and named the wind angle, which ranges from -180° to 180°. 0° means that the heading of ship is consistent with the wind direction and the ship is sailing downwind. -180° or 180° means that the ship is sailing upwind. Furthermore, the Pearson correlation coefficient r was used to analyse the correlation between each variable and the real-time fuel consumption, as shown in Equation (1).

$$r = \frac{\sum_{i=1}^{n}(V_i - \bar{V})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{n}(V_i - \bar{V})^2}\sqrt{\sum_{i=1}^{n}(F_i - \bar{F})^2}} \qquad (1)$$

where V indicates different feature variables, F indicates real-time fuel consumption, $\bar{V}$ and $\bar{F}$ represent their mean values respectively, and n represents the length of variables. The correlation coefficient r and the corresponding significance value p values between multi-source variables and the real-time fuel consumption rate were obtained, as shown in Table 4. For wind angel, we take its absolute value which makes sure it is from 0° to 180°; for wind speed, we take its component in the ship's heading direction, that is, the wind speed multiplied by the cosine of the wind angle. We then divide the correlation into 5 grades: very strong ($r > 0.65$), strong ($0.5 < r < 0.65$), moderate ($0.35 < r < 0.5$), weak ($0.2 < r < 0.35$) and very weak ($r < 0.2$). It is not difficult to find that, except for the engine speed, the engine temperature has the highest correlation with the fuel consumption, and it will play an important role in the subsequent experiments in real-time fuel consumption prediction. The significance levels of these correlations are all above 99%.

Table 4. Correlation between real-time fuel consumption rate and multi-source variables

Variable name	r	p	Grade
SOG	0.4918	0.0008	Moderate
COG	0.1398	0.0012	Very weak
Left ES (Left engine speed)	0.7930	0	Very strong
Right ES (Right engine speed)	0.6643	0	Very strong
Left ET (Left engine temperature)	0.5993	0	Strong
Right ET (Right engine temperature)	0.5574	0	Strong
WaD (Water depth)	-0.3817	0.0003	Moderate
WaS (Water speed)	0.4191	0.0002	Moderate
WiS (Wind speed)	0.1359	0.0005	Very weak
WiA (Wind angle)	0.2349	0.0002	Weak

3 Modelling for fuel consumption

3.1 Modelling method

The LSTM (Long Short-Term Memory) network is an advance RNN model, which was proposed to solve the problem of gradient dispersion in the conventional RNN model. As shown in Fig. 6, LSTM has two transmission states C^t and h^t, one gate control signal Z and three gate control states Z^i, Z^f and Z^o, where i, f and o represent three different control gates: input gate, forget gate and output gate. t is time variable, x^t is the input vector and y^t is the output vector. C^t is named "cell state", which memorises information; h^t is named "hidden state", which is the output of the hidden node. C^{t-1} is the state of C^t at time $t-1$ and h^{t-1} is the state of h^t at time $t-1$. Z, Z^i, Z^f and Z^o are obtained by transforming the operation results of vector and weight matrix W, W^i, W^f and W^o using activation functions "*sigmoid*" and "*tanh*". The operation equations are shown as Equations (2)-(5). Fig. 7 shows the structure of the LSTM network, where U is the weight matrix from the input layer to the hidden layer, V is the weight matrix from the hidden layer to the output layer, and W is a self-looping weight matrix in the hidden layer.

$$Z = \tanh(W\ x^t + W\ h^{t-1}) \tag{2}$$

$$Z^i = \tanh(W^i\ x^t + W^i\ h^{t-1}) \tag{3}$$

$$Z^f = \tanh(W^f x^t + W^f\ h^{t-1}) \tag{4}$$

$$Z^o = \tanh(W^o x^t + W^o\ h^{t-1}) \tag{5}$$
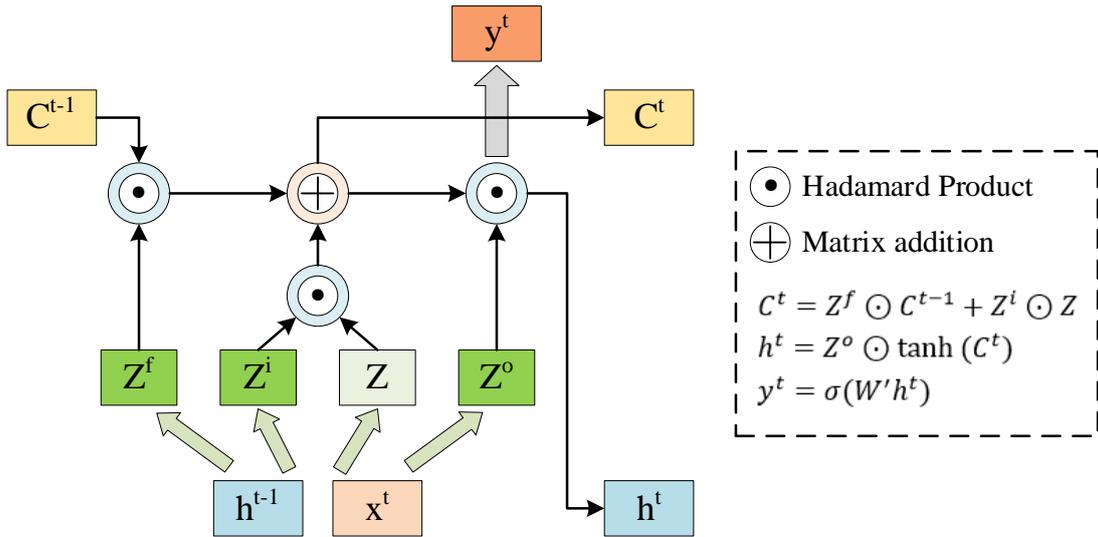
**Fig 6.** The four gate control states of the LSTM module
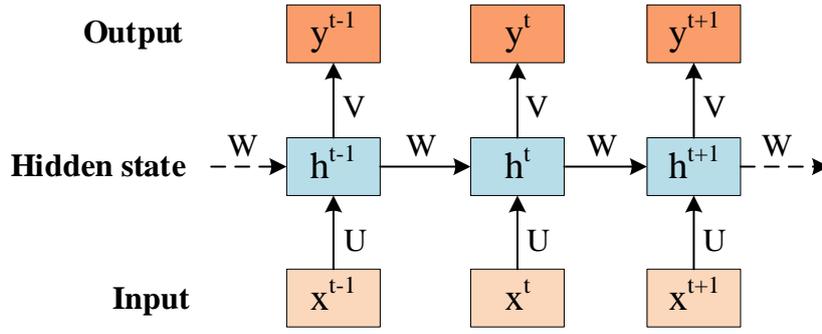
Fig. 7 The diagram of LSTM network

Because of the strong self-learning capabilities (Alahi et al., 2016; Zhao et al., 2017), the LSTM network is suitable for constructing predictive models for monitoring data. It has been widely used in data processing and modelling in water transport. Lin et al. (2019) proposed a RNN with convolution for online obstacle avoidance in unmanned underwater vehicles. Mu et al. (2019) proposed Hybrid RNNs that use unidirectional and bi-directional LSTMs to handle different sensor data. In this paper, an improved LSTM network is used to build a predictive model of real-time fuel consumption for an inland ship. The specific modelling steps are designed as follows.

Step 1: Feature variables selection. The feature variables of the fuel consumption model needs to be analysed and selected using the correlation analysis described in Section 2.5.

Step 2: Data normalisation. In order to eliminate the dimensional influence between multi-source variables, the data are normalized to be between 0 and 1.

Step 3: Data set partitioning. In this step, the data set is randomly divided into a training set and a testing set.

Step 4: Input and output data reshaping. In the LSTM network, the input of each network layer must be three-dimensional, which needs the original data to be reshaped according to the total number of samples and the number of feature variables. The format is [samples, time steps, features].

Step 5: Network initialisation. This step is to set the parameters of the LSTM network, such as the number of neurons, activation functions, loss functions, the number of epochs (training times), the batch size (the size of each input data sample) and the number of hidden layers.

Step 6: Model training. The LSTM network is trained according to the reshaped data and parameters.

Step 7: Result analysis. The prediction result needs to be denormalised, and the prediction accuracy is then analysed by calculating some performance measures, such as Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*) and R-square (R^2), as shown in Equations (6)-(8).

$$RMSE = \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \right)^{1/2} \tag{6}$$

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t| \tag{7}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} (y_t - \bar{y}_t)^2} \tag{8}$$

where t represents the index of a datum and T represents the number of data; y_t and $\hat{y}_t$ are the real values and the predicted values of the tth datum, respectively; $\bar{y}_t$ is the mean of y_t, where $t = 1,2,3 \dots T$.

3.2 Case study

In this study, the host platform was a desktop computer, of which the Central Processing Unit was Intel (R) Core (TM) i5-8500, the main memory was 16GB memory and the operating system was 64-bit Windows 10. The programming language was Python 3.7 and MATLAB 2019a, where a Python Integrated Development Environment Spyder and an open-source ANN library Keras were employed.

Using the methods of multi-source data processing and fuel consumption calculation proposed in Section 2, we obtained a high-quality fuel consumption data set, which contains 15,521 valid records. Then, we use the modelling method proposed in Section 3.1 to predict the real-time fuel consumption rate of an inland ship.

First of all, we select $LeftES$ (left engine speed) and $RightES$ (right engine speed) with the strongest correlation (as shown in Table 4) as input feature variables to select and optimise the parameters of the prediction model. The original data set is randomly divided into two parts, 80% as the training set and the remaining 20% as the testing set. Two-dimensional matrix of input data $\{LeftES, RightES\}$ and one-dimensional of output data $\{FCR\}$ (real-time fuel consumption rate) are presented to the LSTM network with the time step being 1. The number of epochs is set to 3000; the "*mae*" (mean absolute error) function is selected as the loss function; the "*tanh*" function is selected as the activation function and the "*rmsprop*" (root mean square propagation optimiser) is selected as the optimisation function. After a number of preliminary

15

experiments as shown in Table 5, the following parameter settings are found to be appropriate and are used: the

number of neurons is set to 150 and the batch size is set to 100.

Table 5. Some training experiments with different parameter settings

Variable	Neurons number	Batch size	Epochs	*RMSE*	*MAE*	R^2
Neurons number	120	80	3000	0.2071	0.1183	0.8861
	130	80	3000	0.2094	0.1219	0.8836
	140	80	3000	0.2083	0.1209	0.8848
	150	80	3000	0.2062	0.1176	0.8872
	160	80	3000	0.2091	0.1230	0.8839
Batch size	150	90	3000	0.2009	0.1089	0.8928
	150	**100**	3000	0.1994	0.1066	0.8945
	150	110	3000	0.2097	0.1228	0.8832
	150	120	3000	0.2010	0.1090	0.8928

In the modelling experiments, we developed several LSTM network models, which used different

combination of feature variables as inputs according to the correlation in Table 4. The details of models' inputs

and the prediction accuracy against training data and testing data are shown in Table 6. It is easy to see from

Table 6 that the engine temperatures as input variables greatly improve the accuracy of the fuel consumption

model. Moreover, considering the environmental factors, including the water depth, water speed, wind speed

and direction, can improve the prediction performance to some extent. In particular, when monitoring attributes

and hydrological factors are combined as system inputs, the prediction result reaches the best.

Table 6. Prediction accuracy of models with different input variables

Input	Training			Testing		
	RMSE	*MAE*	R^2	*RMSE*	*MAE*	R^2
2	0.2040±0.0007	0.1145±0.0009	0.8895±0.0008	0.2140±0.0006	0.1171±0.0009	0.8794±0.0007
4	0.1890±0.0019	0.1083±0.0034	0.9052±0.0019	0.2012±0.0024	0.1105±0.0035	0.8934±0.0026
6	0.1674±0.0015	0.0919±0.0014	0.9256±0.0013	0.1838±0.0032	0.0975±0.0014	0.9110±0.0031
7	0.1685±0.0039	0.0937±0.0043	0.9232±0.0037	0.1847±0.0047	0.1009±0.0045	0.9102±0.0048
10	0.1437±0.0016	0.0827±0.0028	0.9452±0.0012	0.1806±0.0080	0.0930±0.0032	0.9182±0.0076

2 inputs: Left ES and Right ES.

4 inputs: Left ES, Right ES, Left ET and Right ET.

6 inputs: Left ES, Right ES, Left ET, Right ET, SOG and COG.

7 inputs: Left ES, Right ES, Water depth, Water speed, Wind speed, Wind angle and Segment ID.

10 inputs: Left ES, Right ES, Left ET, Right ET, SOG, Water depth, Water speed, Wind speed, Wind angle and Segment ID.

Figures 8-9 demonstrate the prediction performance of two example models under different input variables.

The above prediction results show that: (1) the multi-source monitoring data we processed can be well used for

fuel consumption modelling; (2) the models we constructed have good performance in prediction; (3) the navigation environment and state of a ship affect the ship's fuel consumption, and considering these factors in modelling can effectively enhance the accuracy of the predictive model.

(a) Training data   (b) Testing data
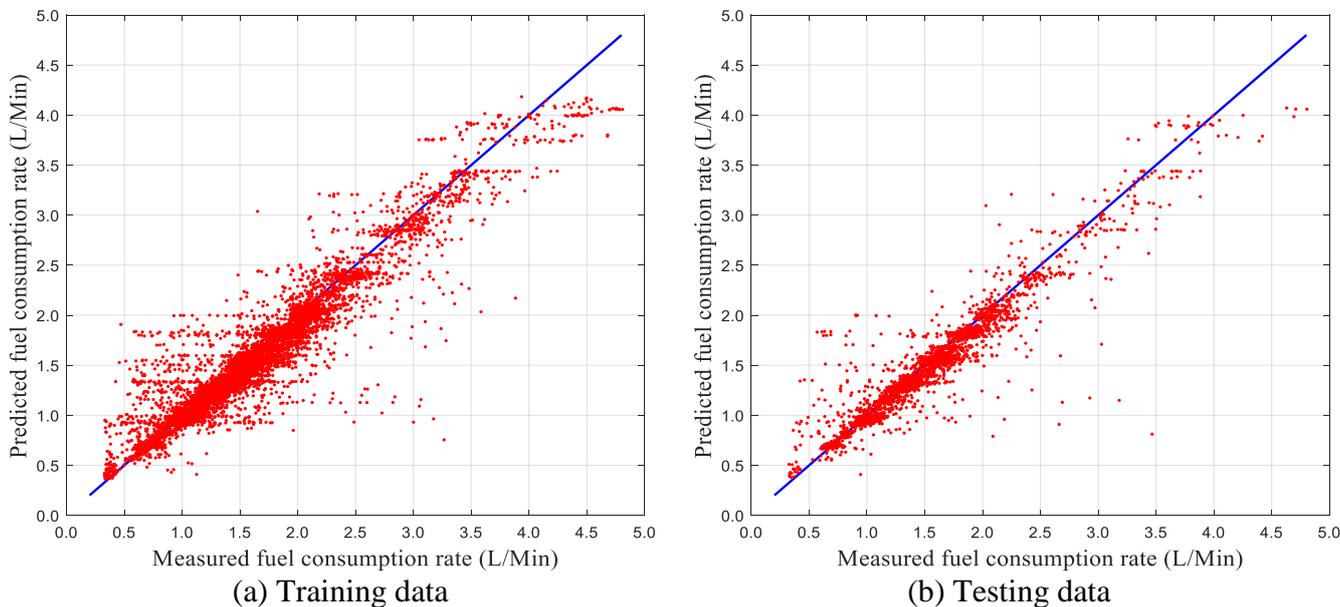
Fig. 8. The measured data vs. predicted data with 4 inputs

(a) Training data   (b) Testing data
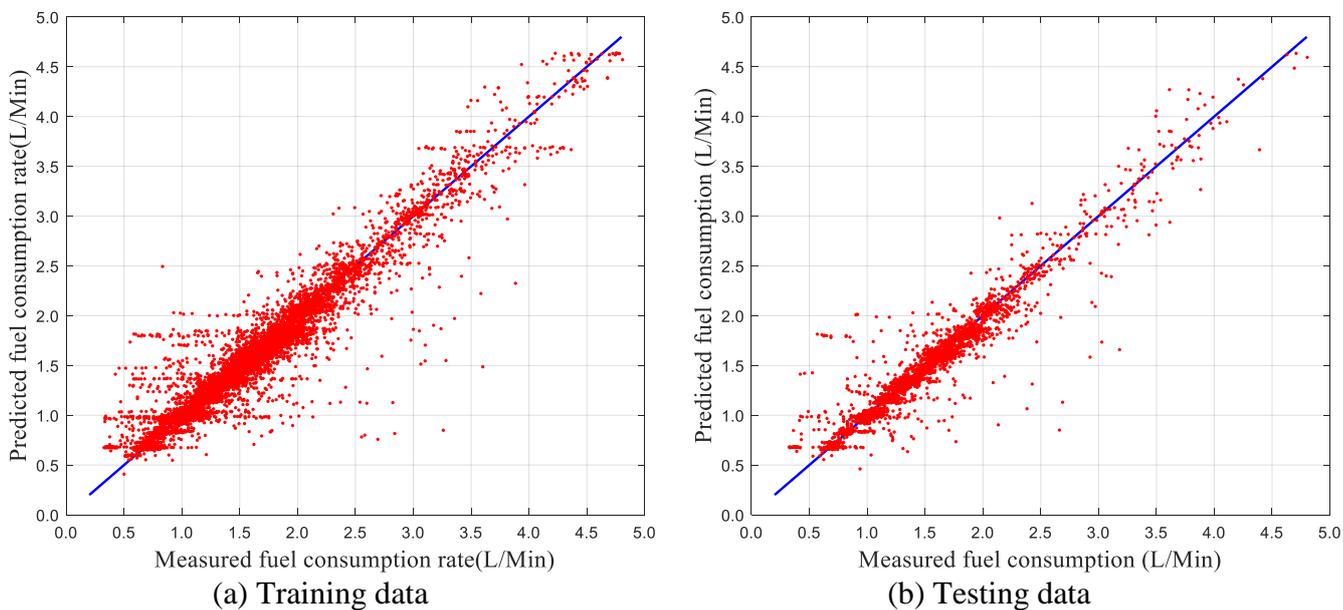
Fig. 9. The measured data vs. predicted data with 7 inputs

To verify the advantages of the constructed model, we compare it with multiple traditional regression models, including Linear Regression (LR), Robust Linear Regression (RLR), Interaction Linear Regression (IR), Pure Quadratic Regression (PQR) and Fine Tree Regression (FTR) (Yang, 2018; Goebel and Plötz, 2019; Acharya et al., 2019), BPNN (Back Propagation Neural Network) and DBPNN (BPNN with two hidden layers),

and some common RNN networks, including general RNN, Bidirectional RNN (BRNN) and Deep RNN (DRNN) (Cui et al., 2018; Mu et al., 2019). In the comparison experiments, we select 10 feature variables as inputs of models and use the same training (12416 data records) and testing data (3105 data records). In the regression methods, the time step in each input is set to 1. The prediction performance (mean and standard deviation of 10 experiments) of different methods is shown in Table 7. From Table 7, we can find that, compared with the LSTM network, all regression models and basic RNN networks have relatively high prediction errors. DBPNN has higher errors than BPNN, because the training was stopped early to prevent network overfitting. Although FTR, BPNN and DRNN show good performance in modelling, but they are still not as good as the LSTM network. Also, BPNN takes much more training time.

Table 7. Comparison of different methods in fuel consumption rate prediction

Method	Training			Testing		
	$RMSE$	MAE	R^2	$RMSE$	MAE	R^2
LR	0.2224 ± 0	0.1301 ± 0	0.8688 ± 0	0.2342 ± 0	0.1324 ± 0	0.8592 ± 0
IR	0.2175 ± 0	0.1268 ± 0	0.8794 ± 0	0.2243 ± 0	0.1278 ± 0	0.8702 ± 0
PQR	0.1934 ± 0	0.1021 ± 0	0.8997 ± 0	0.2021 ± 0	0.1141 ± 0	0.8916 ± 0
RLR	0.2156 ± 0	0.1273 ± 0	0.8786 ± 0	0.2221 ± 0	0.1301 ± 0	0.8692 ± 0
FTR	0.1993 ± 0	0.1072 ± 0	0.8947 ± 0	0.1904 ± 0	0.1114 ± 0	0.8936 ± 0
BPNN	0.1868 ± 0.0122	0.1010 ± 0.0121	0.9012 ± 0.0113	0.1973 ± 0.0126	0.1142 ± 0.0131	0.8971 ± 0.0114
DBPNN	0.1986 ± 0.0114	0.1079 ± 0.0096	0.8964 ± 0.0104	0.2039 ± 0.0101	0.1154 ± 0.0103	0.8896 ± 0.0142
RNN	0.2020 ± 0.0043	0.1120 ± 0.0041	0.8916 ± 0.0046	0.2102 ± 0.0032	0.1139 ± 0.0037	0.8836 ± 0.0035
BRNN	0.1996 ± 0.0031	0.1087 ± 0.0029	0.8942 ± 00033	0.2095 ± 0.0025	0.1112 ± 0.0029	0.8844 ± 0.0028
DRNN	0.1795 ± 0.0059	0.1073 ± 0.0083	0.9143 ± 0.0056	0.1943 ± 0.0071	0.1116 ± 0.0091	0.9005 ± 0.0073
Proposed LSTM	0.1437 ± 0.0016	0.0827 ± 0.0028	0.9452 ± 0.0012	0.1806 ± 0.0080	0.0930 ± 0.0032	0.9182 ± 0.0076

The parameters of three RNN networks are consistent with our proposed LSTM network, and the DRNN uses a two-layer RNN network with 150 and 15 neurons. The neurons number of BPNN is 150, and the DBPNN uses two hidden layers with 150 and 15 neurons.

4 Optimisation of fuel consumption and voyage cost

On the basis of the developed models, optimisation of fuel consumption and voyage cost of inland ships is then conducted, considering different environmental conditions. In this work, two optimisation problems are targeted, to minimise the fuel consumption of the whole voyage and to minimise the total cost of the whole voyage.

4.1 Cost calculation

It is worth noting that fuel consumption is the accumulation of fuel consumption rate during the travel time, and the travel time can be calculated from travel speed SOG. Therefore, if a voyage is divided into a number (i) of segments, the fuel consumption of each segment can be calculated by Equations (9) and (10). The total cost

of the whole voyage includes fixed cost, fuel consumption cost, personnel cost and maintenance cost. The fixed cost mainly includes the fixed rent, port fee and insurance. The personnel cost refers to the remuneration of crews and their living expenses and the maintenance cost mainly contains the cost of wear and repair fee, which are calculated on a daily basis. The calculation is shown in Equations (11)-(14).

$$FC_i = FCR_i \times T_i \tag{9}$$

$$T_i = \frac{SM_i}{SOG_i} \tag{10}$$

$$C_{total} = C_{fixed} + C_{fuel} + C_{personnel} + C_{maintenance} \tag{11}$$

$$C_{fuel} = FC_{total} \times P_{fuel} = \sum_i FC_i \times P_{fuel} \tag{12}$$

$$C_{personnel} = \frac{T_{total}}{24} \times P_{personnel} = \frac{\sum_i T_i}{24} \times P_{personnel} \tag{13}$$

$$C_{maintenance} = \frac{T_{total}}{24} \times P_{maintenance} = \frac{\sum_i T_i}{24} \times P_{maintenance} \tag{14}$$

where i is the index of a short segment, within which the trajectory characteristics and environmental conditions keep the same; FC_i is the fuel consumption of segment i, FCR_i is the fuel consumption rate in Segment i, T_i is the travel time passing through Segment i, SM_i represents the navigation mileage of Segment i, C_{total} is the total cost, C_{fixed} is the fixed cost of the whole voyage, C_{fuel} is the fuel-related cost, P_{fuel} is the price of fuel consumption per litre, $C_{personnel}$ is the crew-related cost, $P_{personnel}$ is the crew-related cost per day, $C_{maintenance}$ is the maintenance cost, and $P_{maintenance}$ is the maintenance cost per day.

4.2 Optimisation problems

In this paper, we studied two optimisation problems, (1) minimising the fuel consumption of the whole voyage and (2) minimising the total cost of the whole voyage. These two problems are different in objectives, but have the same decision variables and constraints. More details are described as follows.

Objective of Problem 1: Minimise the fuel consumption of the whole voyage:

$$FC_{total} = \sum_i FCR_i \times \frac{SM_i}{SOG_i} \times 60 \tag{15}$$

Objective of Problem 2: Minimise the total cost of the whole voyage:

$$C_{total} = C_{fixed} + \sum_i FCR_i \times \frac{SM_i}{SOG_i} \times 60 \times P_{fuel} + \frac{\sum_i \frac{SM_i}{SOG_i}}{24} \times (P_{personnel} + P_{maintenance}) \qquad (16)$$

Decision variables: Left ES and Right ES. According to the correlation analysis in Table 4, it is known that the greatest impact on the fuel consumption rate comes from the Left ES and the Right ES, which are also the observable and controllable variables. Considering the manoeuvrability of the optimisation results, we set the speed of the left engine and the speed of the right engine to be equal, because they are of the same type and have the same output power.

Constraints: Environmental conditions include Water depth (WaD), Water speed (WaS), Wind speed (WiS) and Wind angle (WiA) and they are different along the voyage. The Left ES and Right ES should be within the working range of the engines.

4.3 Models and optimisation algorithm

In order to solve the above two optimisation problems, two prediction models are constructed and utilised. First, a Fuel Consumption Rate Model (FCRM) is used to predict the real-time fuel consumption rate, which contains 7 input variables as shown in Table 6. Furthermore, using the LSTM network, a SOG model (SOGM) is established to predict the SOG under various conditions. For both the FCRM and the SOGM, the input variables are Left ES, Right ES, WaD, WaS, WiS, WiA and Segment ID. It should be noted that the input variables do not include engines temperatures, which are difficult to be accurately controlled in the actual manipulation and can be largely reflected by engines speeds' values. Their performance is shown in Table 8. It can be seen that both models have good performance. The performance of FCRM is slightly better than SOGM, as the SOG data have bigger variation than the fuel consumption data, which makes SOG more difficult to model.

Table 8. The performance of the developed FCRM and SOGM

Model	Training			Testing		
	RMSE	*MAE*	*R²*	*RMSE*	*MAE*	*R²*
FCRM	0.1685±0.0039	0.0937±0.0043	0.9232±0.0037	0.1847±0.0047	0.1009±0.0045	0.9102±0.0048
SOGM	0.7308±0.0378	0.5017±0.0280	0.8443±0.0172	0.7005±0.0390	0.5022±0.0270	0.8602±0.0172

In this paper, the Reduced Space Searching Algorithm (RSSA) is employed to solve the optimisation problems. RSSA is a nature-inspired algorithm that switches and zooms in/out the targeted search space to

speed up the searching process and jump out from local optima. The algorithm focuses on dividing and transforming the search space to allocate the best sub-space while most of the other optimisation algorithms focus on generation of solutions using proper operators. RSSA has been verified to outperform some salient heuristic optimisation algorithms, such as Covariance Matrix Adaptation Evolution Strategy, Differential Evolution and Generalised Generation Gap model (Zhang and Mahfouf, 2010). It has also been successfully applied to some engineering problems (Zhang et al., 2015; Datta et al., 2016). The Pseudo code of the optimisation algorithm based on RSSA is shown in Method 3.

Method 3. Optimisation algorithm based on RSSA
Input: $Dim, Xvar, Ob, Nt, Emax, Error$
Output: Pb, Xpb, Eva
1. Initialisation of setting.
2. Randomly select one candidate solution in the original search space, and record it as Xpb. Set $n = 0$, which is used to control the bounds of the search space.
3. Randomly select the candidate solutions in the current search space, record it as Xb. If C_1-continuous Xb satisfies $f(Xb) < f(Xpb)$ and $n > 1$, then $Xpb = Xb$ and $n = n - 1$. If C_2-continuous Xb satisfies $f(Xb) > f(Xpb)$, then $n = n + 1$. If *non*-continuous Xb satisfies $f(Xb) < f(Xpb)$, then $Xpb = Xb$.
4. Change the size of the search space using the ratio K ($0 < K < 1$). Xpb is located at the new space. Set Y_{min_i} being the lower bound of the ith decision variable in the new search space and Y_{max_i} being the upper bound: $Y_{min_i} = \max(X_{min_i}, Xpb(i) - K^n L(i))$, $Y_{max_i} = \min(X_{max_i}, Xpb(i) + K^n L(i))$.
5. Repeat Steps 3 and 4 until $n = m$.
6. Perform the variation operator on Xpb and obtain Xc. If $f(Xc) < f(Xpb)$, then $Xpb = Xc$, $n = 0$, and repeat Steps 3 to 5.
7. Repeat Step 6 until the 'optimal' solution is found or the termination criterion is reached.

In Method 3, $f()$ is the fitness function. Dim is the dimension of the decision variables. $Xvar$ includes the bounds of the decision variables, which has two columns. The first column is the lower bound and the second column is the upper bound. In this work, the lower and upper bounds of the Left ES and Right ES were set 350 and 750. Ob is the preconceived optimum value of the target problem, which was set to 0 in this paper. Nt is the times of division for RSSA, which was set to 50. $Emax$ is the max number of evaluation for termination, which was set to 3000. $Error$ is the accepted error rate for termination, which was set to 0.0001 in this paper.

4.4 Optimisation results

Based on the developed FCRM and SOGM, RSSA is employed to optimise the Left ES and Right ES to reduce the fuel consumption of each segment under given environmental conditions. It is worth noting that the environmental conditions vary from segment to segment, but it keeps the same in a single segment. The results are compared with measured fuel consumption. We also add a comparative experiment, which uses the average Left ES and Right ES of the measured data as input values and calculates the fuel consumption of each segment through FCRM and SOGM under the same environmental conditions. The results are shown in Table 9. Similarly, we have optimised and compared the cost of each segment, and the results are shown in Table 10. In these two optimisation cases, we assume that C_{fixed} is 8000.00 China Yuan (CYN), P_{fuel} is 5.50 CNY per litre, $P_{personnel}$ is 5000.00 CNY per day and $P_{maintenance}$ is 1500.00 CNY per day. These costs assumption comes from the data in the ship benefit report of historical voyages provided by the ChangJiang Shipping Science Research Institute CO. Ltd.

Table 9. Optimisation results of fuel consumption for each segment under given environmental conditions

SID	Environmental conditions				Optimised ES (rpm)		Comparable ES (rpm)		Fuel Consumption (L)		
	WaD (m)	WaS (m/s)	WiS (BS)	WiA (°)	Left ES	Right ES	Left ES	Right ES	Optimised	Comparable	Measured
1	12.50	3.4	2	90	401.1	401.1	432.8	441.3	553	927	1270
2	12.50	3.3	2	10	454.0	454.0	464.1	497.9	418	625	809
3	10.50	3.4	3	60	410.0	410.0	457.3	489.9	554	943	817
4	10.80	3.4	3	-3	486.8	486.8	476.9	497.0	324	337	563
5	10.20	3.3	3	55	414.7	414.7	511.6	511.6	321	824	850
6	10.50	3.4	3	-10	521.0	521.0	517.1	523.4	900	985	1557
7	8.00	3.3	3	-25	537.0	537.0	516.8	518.4	1626	1738	2054
8	7.86	3.2	2	20	412.7	412.7	413.1	424.7	805	816	1028
9	6.15	3.3	3	15	399.0	399.0	412.1	422.8	1832	2141	2387
10	6.65	2.1	4	90	435.7	435.7	450.4	469.1	1764	2992	2509
11	6.14	2.0	3	28	410.8	410.8	481.6	508.3	884	2085	2073
12	5.17	2.2	3	-1	401.1	401.1	447.3	459.5	114	175	931
13	5.34	2.2	3	15	500.0	500.0	494.8	499.2	1309	2551	3244
14	4.72	1.9	2	126	500.0	500.0	611.5	605.6	2152	2870	2575
15	4.48	1.7	2	75	470.5	470.5	504.0	506.4	2160	2969	2610
16	4.27	1.7	2	-160	472.6	472.6	491.6	495.7	1344	1517	1693
17	4.11	1.8	2	-160	535.0	535.0	617.5	601.8	936	1391	1103
18	4.07	1.8	2	90	500.0	500.0	569.5	570.3	498	538	616
19	24.18	1.7	2	-80	470.0	470.0	521.4	503.9	360	372	433
20	48.48	1.6	2	-72	496.0	496.0	604.5	518.0	361	400	787
21	95.67	1.7	2	10	425.0	425.0	525.4	524.8	745	906	996
22	97.81	1.6	2	-115	386.0	386.0	400.0	403.6	207	223	334
23	97.98	1.7	1	-8	390.0	390.0	411.2	413.0	121	156	256

Table 10. Optimisation results of cost for each segment under given environmental conditions

SID	Environmental conditions				Optimised ES (rpm)		Comparable ES (rpm)		Cost (CNY)		
	WaD (m)	WaS (m/s)	WiS (BS)	WiA (°)	Left ES	Right ES	Left ES	Right ES	Optimised	Comparable	Measured
1	12.50	3.4	2	90	520.0	520.0	432.8	441.3	5455	8621	11414
2	12.50	3.3	2	10	545.5	545.5	464.1	497.9	3578	5388	6707
3	10.50	3.4	3	60	414.2	414.2	457.3	489.9	5598	8107	7033
4	10.80	3.4	3	-3	510.0	510.0	476.9	497.0	2793	2855	4717
5	10.20	3.3	3	55	419.4	419.4	511.6	511.6	3166	6640	6852
6	10.50	3.4	3	-10	532.0	532.0	517.1	523.4	7500	7899	12368
7	8.00	3.3	3	-25	540.7	540.7	516.8	518.4	12558	13789	15037
8	7.86	3.2	2	20	452.4	452.4	413.1	424.7	7886	8030	9783
9	6.15	3.3	3	15	402.2	402.2	412.1	422.8	18634	20466	22720
10	6.65	2.1	4	90	435.7	435.7	450.4	469.1	15801	26197	21727
11	6.14	2.0	3	28	411.1	411.1	481.6	508.3	8658	17252	17087
12	5.17	2.2	3	-1	406.2	406.2	447.3	459.5	1138	1543	8092
13	5.34	2.2	3	15	560.0	560.0	494.8	499.2	13090	20687	26354
14	4.72	1.9	2	126	536.5	536.5	611.5	605.6	16847	20210	18093
15	4.48	1.7	2	75	510.1	510.1	504.0	506.4	20094	24917	20823
16	4.27	1.7	2	-160	476.1	476.1	491.6	495.7	11996	12789	13811
17	4.11	1.8	2	-160	580.0	580.0	617.5	601.8	7416	9720	7704
18	4.07	1.8	2	90	550.4	550.4	569.5	570.3	3716	3731	4491
19	24.18	1.7	2	-80	536.5	536.5	521.4	503.9	2912	3027	3434
20	48.48	1.6	2	-72	540.5	540.5	604.5	518.0	2949	3054	6144
21	95.67	1.7	2	10	581.5	581.5	525.4	524.8	6097	7008	7682
22	97.81	1.6	2	-115	406.0	406.0	400.0	403.6	2072	2139	2374
23	97.98	1.7	1	-8	390.0	390.0	411.2	413.0	1496	1772	2612

It can be observed from Table 9 that the optimal engine speed which makes the fuel consumption lower than the measured value has been found in every segment. The optimised solution also performs better than the comparative experiment which uses a moderate and constant engine speed. Similar results also appear in the total cost optimisation, which can be seen in Table 10. It is worth noting that in the optimisation of fuel consumption and cost, the degree of improvement is much different in each segment. This reflects the complexity of the inland navigation environment, and also highlights the need of trajectory segmentation. On the other hand, there may still be some errors in the developed models, which may lead to some improper solutions generated for certain segments.

The total fuel consumption and the total cost are further calculated for the whole voyage, as shown in Table 11. It can be seen that the optimised solutions show lower fuel consumption and cost, compared to the measured and comparable cases. Optimisation Problem 1 focuses on reducing the fuel consumption and the optimised solution addressing it shows the best performance in fuel consumption. Problem 2 considers the total cost of the voyage, which leads to a solution that balances the fuel cost and other time-dependent costs.

Table 11. Comparison of total fuel consumption and total cost for the whole voyage

	T_{total} (h)	FC_{total} (L)		C_{total} (CNY)	
Measured case	381	36100	/	309678	/
Comparable case	354	32619	9.64% less	283181	8.56% less
Optimised Solution 1	342	**23994**	**33.53% less**	232612	24.89% less
Optimised Solution 2	295	25140	30.36% less	**226218**	**26.95%** less

Moreover, we compare RSSA with some other well-known optimisation algorithms, including Genetic Algorithm (GA) (Subramanian and Jyothish, 2020) and Particle Swarm Optimisation (PSO) (Kaloop et al., 2020). The optimisation results of the total fuel consumption and total cost for the whole voyage are shown in Table 12. From Table 12, one can find that, compared with GA and PSO, RSSA finds the solutions with relatively low fuel consumption and cost.

Table 12. Comparison of different algorithms in optimisation of total fuel consumption and total cost

Algorithm	Optimised Solution 1			Optimised Solution 2		
	T_{total} (h)	FC_{total} (L)	C_{total} (CNY)	T_{total} (h)	FC_{total} (L)	C_{total} (CNY)
GA	349	26724	250123	309	28017	242141
PSO	352	27335	256423	315	28246	246096
RSSA	342	**23994**	232612	295	25140	**226218**

5 Conclusions

In this paper, based on the multi-source data composed of monitoring data and hydrological data, the real-time fuel consumption of inland ships has been analysed and modelled, and the optimisation of fuel consumption and the total cost for a whole voyage has been performed. The multi-source monitoring data have been processed to delete abnormal data and retain the ship's fuel information to the maximum extent. A method has been proposed to calculate the accurate real-time fuel consumption rate from data. Correlation analysis of multiple variables has been made, which facilitates the selection of input feature variables for predictive models. In modelling, the LSTM network has been tailored and utilised in building the fuel consumption rate models. Specifically, the engine temperature was used for the fuel consumption prediction for the first time. Compared with some representative regression methods (LR, RLR, IR, PQR and FTR) and recurrent neural networks (RNN, BRNN and DRNN), the constructed LSTM model consistently outperforms the other methods. Finally, the developed models have been implemented in optimisation of the engine speed to minimise the total fuel consumption and the total cost of the whole voyage, considering different environmental conditions. The

optimal solutions have been compared with experience-based scenarios and it shows that through the optimisation the total fuel consumption is reduced by 33.54 % and the total cost is reduced by 26.95%.

In future, more sophisticated methods will be designed for dynamic trajectory segmentation, such as employing unsupervised clustering analysis. In addition, to enhance the applicability of the proposed prediction and optimisation techniques in different operational scenarios, a multi-objective optimisation framework that considers varying environmental conditions will be developed, where similar research has been carried out in other transport mode (Chen et al., 2016).

Acknowledgements

References:

Acharya, M. S., Armaan. A and Antony, A. S. (2019). "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), pp. 1-5.

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 961-971).

Beşikçi, E. B., Arslan, O., Turan, O., & Ölçer, A. I. (2016). An artificial neural network based decision support system for energy efficient ship operations. Computers & Operations Research, 66, 393-401.

Capezza, C., Coleman, S., Lepore, A., Palumbo, B., & Vitiello, L. (2019). Ship fuel consumption monitoring and fault detection via partial least squares and control charts of navigation data. Transportation Research Part D: Transport and Environment, 67, 375-387.

Chen, J., Weiszer, M., Locatelli, G., Ravizza, S., Atkin, J. A., Stewart, P. & Burke, E. K. (2016). Toward a More Realistic, Cost-Effective, and Greener Ground Movement Through Active Routing: A Multiobjective Shortest Path Approach, IEEE Transactions on Intelligent Transportation Systems, 17(12), 3524-3540.

Coraddu, A., Oneto, L., Baldi, F., & Anguita, D. (2017). Vessels fuel consumption forecast and trim optimisation: a data analytics perspective. Ocean Engineering, 130, 351-370.

Cui, Y., Zhao, S., Wang, H., Xie, L., Chen, Y., Han, J., & Liu, T. (2018). Identifying brain networks at multiple time scales via deep recurrent neural network. IEEE journal of biomedical and health informatics, 23(6), 2515-2525.

Datta, S., Mahfouf, M., Zhang, Q., Chattopadhyay, P. P., & Sultana, N. (2016). Imprecise Knowledge Based Design and Development of Titanium Alloys for Prosthetic Applications. Journal of the Mechanical Behavior of Biomedical Materials, 53, 350-365.

Du, Y., Meng, Q., Wang, S., & Kuang, H. (2019). Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. Transportation Research Part B: Methodological, 122, 88-114.

Gkerekos, C., Lazakis, I., & Theotokatos, G. (2019). Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study. Ocean Engineering, 188, 106282.

Goebel, D., & Plötz, P. (2019). Machine learning estimates of plug-in hybrid electric vehicle utility factors. Transportation research part D: Transport and environment, 72, 36-46.

Fagerholt, K., Laporte, G., & Norstad, I. (2010). Reducing fuel emissions by optimizing speed on shipping routes. Journal of the Operational Research Society, 61(3), 523-529.

Fagerholt, K., Gausel, N. T., Rakke, J. G., & Psaraftis, H. N. (2015). Maritime routing and speed optimization with emission control areas. Transportation Research Part C: Emerging Technologies, 52, 57-73.

Hansen, E. K., Rasmussen, H. B., & Lützen, M. (2020). Making shipping more carbon-friendly? Exploring ship energy efficiency management plans in legislation and practice. Energy Research & Social Science, 65, 101459.

Huang, L., Wen, Y., Geng, X., Zhou, C., & Xiao, C. (2018). Integrating multi-source maritime information to estimate ship exhaust emissions under wind, wave and current conditions. Transportation Research Part D: Transport and Environment, 59, 148-159.

Hu, H., Wang, L., & Qi, G. J. (2019, July). Learning to Adaptively Scale Recurrent Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 3822-3829).

Hu, Z., Jin, Y., Hu, Q., Sen, S., Zhou, T., & Osman, M. T. (2019). Prediction of Fuel Consumption for Enroute Ship Based on Machine Learning. IEEE Access, 7, 119497-119505.

Kaloop, M. R., Kumar, D., Zarzoura, F., Roy, B., & Hu, J. W. (2020). A wavelet-Particle swarm optimization-Extreme learning machine hybrid modeling for significant wave height prediction. Ocean Engineering, 213, 107777.

Leifsson, L. Þ., Sævarsdóttir, H., Sigurðsson, S. Þ., & Vésteinsson, A. (2008). Grey-box modeling of an ocean vessel for operational optimization. Simulation Modelling Practice and Theory, 16(8), 923-932.

Li, G., Deng, X., Zhou, M., Zhu, Q., Lan, J., Xia, H., & Mitrouchev, P. (2019, November). Research on Data Monitoring System for Intelligent Ship. In International Workshop of Advanced Manufacturing and Automation (pp. 234-241). Springer, Singapore.

Li, X., Sun, B., Zhao, Q., Li, Y., Shen, Z., Du, W., & Xu, N. (2018). Model of speed optimization of oil tanker with irregular winds and waves for given route. Ocean Engineering, 164, 628-639.

Lin, C., Wang, H., Yuan, J., Yu, D., & Li, C. (2019). An improved recurrent neural network for unmanned underwater vehicle online obstacle avoidance. Ocean Engineering, 189, 106327.

Mu, X., He, B., Zhang, X., Song, Y., Shen, Y., & Feng, C. (2019). End-to-end navigation for Autonomous Underwater Vehicle with Hybrid Recurrent Neural Networks. Ocean Engineering, 194, 106602.

Psaraftis, H. N., & Kontovas, C. A. (2014). Ship speed optimization: Concepts, models and combined speed-routing scenarios. Transportation Research Part C: Emerging Technologies, 44, 52-69.

Satpathi, K., Balijepalli, V. M., & Ukil, A. (2017). Modeling and real-time scheduling of DC platform supply vessel for fuel efficient operation. IEEE Transactions on Transportation Electrification, 3(3), 762-778.

Schaub, M., Finger, G., Riebe, T., Dahms, F., Hassel, E., & Baldauf, M. (2019, June). Data-based modelling of ship emissions and fuel oil consumption for transient engine operation. In OCEANS 2019-Marseille (pp. 1-5). IEEE.

Sheng, D., Meng, Q., & Li, Z. C. (2019). Optimal vessel speed and fleet size for industrial shipping services under the emission control area regulation. Transportation Research Part C: Emerging Technologies, 105, 37-53.

Subramanian, R., & Jyothish, P. V. (2020). Genetic Algorithm Based Design Optimization of a Passive Anti-Roll Tank in a Sea Going Vessel. Ocean Engineering, 203, 107216.

Tang, Z., Peng, X., Li, T., Zhu, Y., & Metaxas, D. N. (2019). Adatransform: Adaptive data transformation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2998-3006).

Van, T. C., Ramirez, J., Rainey, T., Ristovski, Z., & Brown, R. J. (2019). Global impacts of recent IMO regulations on marine fuel oil refining processes and ship emissions. Transportation Research Part D: Transport and Environment, 70, 123-134.

Wang, K., Yan, X., Yuan, Y., & Li, F. (2016). Real-time optimization of ship energy efficiency based on the prediction technology of working condition. Transportation Research Part D: Transport and Environment, 46, 81-93.

Wang, S., & Meng, Q. (2012). Sailing speed optimization for container ships in a liner shipping network. Transportation Research Part E: Logistics and Transportation Review, 48(3), 701-714.

Wang, S., Ji, B., Zhao, J., Liu, W., & Xu, T. (2018). Predicting ship fuel consumption based on LASSO regression. Transportation Research Part D: Transport and Environment, 65, 817-824.

Wang, Z., Fang, Z., Wu, Y., Liang, J., & Song, X. (2019). Multi-Source Evidence Data Fusion Approach to Detect Daily Distribution and Coverage of Ulva Prolifera in the Yellow Sea, China. IEEE Access, 7, 115214-115228.

Wen, M., Pacino, D., Kontovas, C. A., & Psaraftis, H. N. (2017). A multiple ship routing and speed optimization problem under time, cost and environmental objectives. Transportation Research Part D: Transport and Environment, 52, 303-321.

Wright, R. G. (2019). Intelligent autonomous ship navigation using multi-sensor modalities. TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, 13(3).

Yang, L., Chen, G., Rytter, N. G. M., Zhao, J., & Yang, D. (2019). A genetic algorithm-based grey-box model for ship fuel consumption prediction towards sustainable shipping. Annals of Operations Research, 1-27.

Yang, Y. (2018). "Prediction and analysis of aero-material consumption based on multivariate linear regression model," 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 628-632.

Yuan, J., & Nian, V. (2018). Ship energy consumption prediction with Gaussian process metamodel. Energy Procedia, 152, 655-660.

Zhang, Q., & Mahfouf, M. (2010). A nature-inspired multi-objective optimisation strategy based on a new reduced space searching algorithm for the design of alloy steels. Engineering applications of artificial intelligence, 23(5), 660-675.

Zhang, Q., Mahfouf, M., Panoutsos, G., Beamish, K., & Liu, X. (2015). Multiobjective Optimal Design of Friction Stir Welding Considering Quality and Cost Issues. Science and Technology of Welding and Joining, 20(7), 607-615.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, 11(2), 68-75.