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

A multi-task analysis and modelling paradigm using LSTM for multi-source monitoring data of inland vessels

A Multi-task Analysis and Modelling Paradigm using LSTM for Multi-source Monitoring Data of Inland Vessels

Zhi Yuan[1,2,3], Jingxian Liu[1,2], Yi Liu[1,2*], Qian Zhang[3*] and Ryan Wen Liu[1,2]

[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.*

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

Abstract: The vessel monitoring data provide important information for people to understand the vessel dynamic status in real time and make appropriate decisions in vessel management and operations. However, some of the essential data may be incomplete or unavailable. In order to recover or predict the missing information and best exploit the vessels monitoring data, this paper combines statistical analysis, data mining and neural network methods to propose a multi-task analysis and modelling framework for multi-source monitoring data of inland vessels. Specifically, an advanced neural network, Long Short-Term Memory (LSTM) was tailored and employed to tackle three important tasks, including vessel trajectory repair, engine speed modelling and fuel consumption prediction. The developed models have been validated using the real-life vessel monitoring data and shown to outperform some other widely used modelling methods. In addition, statistics and data technologies were employed for data extraction, classification and cleaning, and an algorithm was designed for identification of the vessel navigational state.

Key words: vessel monitoring; multi-source data; LSTM; trajectory repair; engine speed modelling; fuel consumption

1. Introduction

In waterborne transportation, the vessel dynamic and static information obtained in real time is essential for maritime management and operational tasks, including vessel traffic monitoring, traffic management, risk assessment, safety management, route planning, etc. The data containing such dynamic and static information can be collected from different monitoring systems, such as the Automatic Identification System (AIS), the Vessel Traffic Service (VTS) system and a vessel-borne radar system. To derive accurate and practical information from the data, one needs to employ data-related techniques, such as statistical analysis, noise filtering, data compression, data mining, data-driven modelling and optimisation.

AIS is a conventional data source for vessel information, which contains a vessel's static information, such as its International Maritime Organization (IMO) code, name, length and width, as well as the vessel's dynamic information, such as real-time latitude, longitude, Speed Over Ground (SOG), Course Over Ground (COG) and heading angle (Last et al., 2014). The AIS message is broadcasted via a Very High Frequency (VHF) channel in the form of digital codes, which can provide basic data for vessel collision avoidance and trajectory visualisation (Willems et al., 2009). The VTS system integrates the AIS data, the shore-based radar data and weather data to provide information for area-wide vessel tracking and provide video to facilitate the monitoring and management of water traffic (Jordan et al., 2001; Robards et al., 2016). Ship-borne radar data contain real-time vessel coordinates, azimuth, weather information, etc., which can be used for distance measurement and weather prediction (Xie et al., 2017). Although the data sources mentioned above can process mass data acquisition in discrete vessel navigation status and necessary environmental information, they still lack some other important information such as fuel consumption and engine operational conditions. This may affect the comprehensive accomplishment of navigation status judgment, risk assessment, green shipping and autonomous navigation.

In recent years, with the wide application of information technology, smart devices and sensor networks, modern waterway transportation is developing towards high density, wide area, high complexity, systematisation and intelligence (An et al., 2011; Yan et al., 2012; Train et al., 2018; Shi et al., 2018). Sensing technology has been widely implemented to the integrated monitoring systems on vessels, as shown in Fig. 1. Such a system converts different forms of signals into a digital format for further processing and analysing, which provides abundant fundamental data for information fusion, intelligent navigation, energy saving and emission reduction, and unmanned shipping. However, the use of different types of sensors leads to heterogeneity of data with various sampling frequencies, and this sometimes causes network transmission failures. In this case, the stored data may involve problems relating to noise interference, redundancy and partial data loss, which would greatly reduce usability and reliability of the data. Therefore, data processing and repair technologies for vessel monitoring systems have become important research topics.

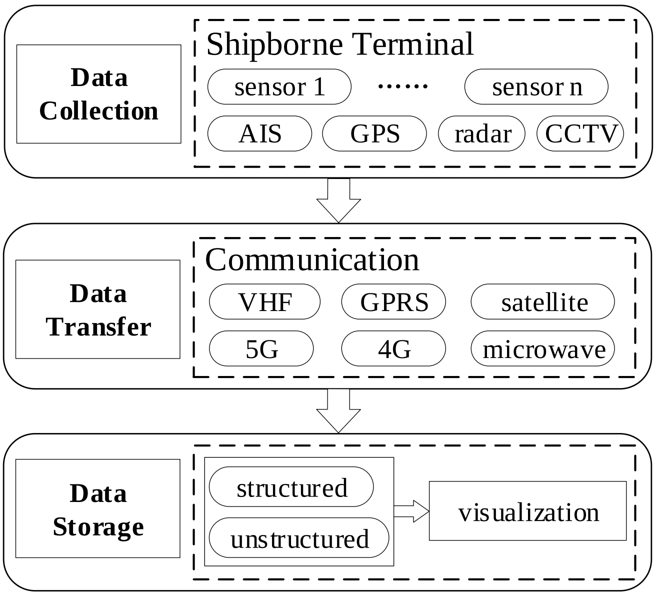


Fig. 1. The architecture of a vessel monitoring system

The Yangtze River, known as the "golden waterway", is the main artery of water transportation across western and eastern china. Its cargo volume ranks first in the world's inland rivers (Notteboom et al., 2020). However, the navigation environment of the Yangtze River trunk is complex where natural conditions along the waterway vary greatly. The vessel traffic flow is complicated at certain segments. This paper considers the Yangtze River trunk as the research subject. It aims to break through the limitations of a single data acquisition source, collect data from multiple monitoring systems, and make full use of advanced theories and technologies to develop a multi-task data analysis and modelling framework. The methods of data partitioning, abnormality determination and data cleaning are proposed to prepare high-quality vessel data sets. Based on the pre-processed data, further studies on vessel navigational state recognition, trajectory repair, engine speed modelling and fuel consumption prediction are carried out, which provide a solid basis for the construction of a large-scale waterway transportation database and the development of intelligent transportation systems.

The remaining of this paper is organised as follows: In Section 2, the methods relating to noise filtering, AIS data processing and Artificial Neural Networks (ANNs) are reviewed. The proposed multi-task framework is introduced in details in Section 3. Experiments based on field data are discussed in Section 4. Finally, conclusions and perspectives are presented in Section 5.

2. Related work

In this section, related work in the field of data processing and analysis for water transportation is reviewed. We especially focus on the topics of AIS data cleaning and analysis, vessel trajectory mining and fuel consumption analysis.

AIS is an important data source for water transportation, which can produce 2,000 reports per minute. It provides a vast amount of near-real-time information that can be used to support maritime management and operational decision-making. With the development of information technology, big data and data mining methods are increasingly used in maritime data processing. Mao et al. (2018) constructed a standard AIS database for vessel trajectory learning, prediction and data mining. Arguedas et al. (2017) proposed a two-layer network to automatically produce synthetic maritime traffic representations from AIS data. Filipiak et al. (2018) used Hadoop-compliant processing framework and a data mining technology, capable of handling big data in a fast and efficient manner, to describe movement of tankers worldwide in 2015. Wu et al. applied fuzzy logic based approaches to the AIS data to design a ship-bridge collision alert system (Wu et al., 2019) and an intelligent navigation strategy (Wu et al., 2020).

In data cleaning, accurate identification of abnormal data is especially important. Riveiro et al. (2018) reviewed the maritime anomaly detection from four aspects, including data, methods, systems and users. Researchers initially detected anomalies implicitly by creating normalcy models. For instance, Rhodes et al. (2005) divided water into small zones and used normalcy box to detect abnormal vessel speed in each zone. Kazemi et al. (2013) defined a large number of anomalous behaviours and the corresponding events and validated their anomaly detection method using a published open dataset. Gaussian Mixture Models (GMMs) and Kernel Density Estimation (KDE) methods were also explored for anomaly detection (Laxhammar, 2008; Ristic et al., 2008). Subsequently, some time-series analysis techniques, such as Gaussian process (Will, 2011; Kowalska and Peel, 2012) and Bayesian Networks (BNs) (Mascaro et al., 2014), were used to capture the abnormal sequence structure of the AIS data streams.

Vessel trajectory is vital in analysing the features of vessel behaviour and vessel traffic flow; therefore it has become a common research interest of many scholars. Sang et al. (2015) proposed a method to restore the trajectory of an inland waterway vessel based on AIS data. They also developed three rules to identify and remove inaccurate data based on the reception range of data and the manoeuvring characteristics of inland waterway vessels. Zhang et al. (2018) proposed a multi-regime approach for vessel trajectory reconstruction through a three-step procedure using AIS data, which allowed for vessel trajectory reconstruction in different navigational states, namely hoteling, manoeuvring and normal-speed sailing. Pan et al. (2014) introduced a trajectory clustering algorithm, developed based on sampling and density information, to group similar movement tracks of cars, vessels and airplanes. Li et al. (2016) used the Douglas-Peucker (DP) algorithm to simplify massive AIS trajectories and visualised vessel trajectory density based on the KDE method. Li et al. (2017) proposed a multi-step trajectory clustering method for robust vessel trajectory clustering. In the work of Li et al. (2018), an improved method combining density-based spatial clustering with a noise algorithm was proposed to group spatial points to acquire the optimal clusters.

Estimation of fuel consumption and emission for vessels are essential for the realisation of green shipping and autonomous navigation. However, limited research has been conducted to study the relevant topics. Lou et al. (2017) tried to find the relationship among fuel consumption, emission and cruise speed of tugboats, and established an optimisation model for cruise speed. Coraddu et al. (2017) analysed and compared three different models, including a white-box model, a black-box model and a grey-box model, in the prediction of vessel fuel consumption.

In summary, there are several drawbacks in the existing research. First, the existing research relied on the AIS data and lacked the adoption of other valuable information, such as engine status and fuel consumption. Second, few studies investigated the navigational status of vessels, which is very valuable for waterway traffic management. Last, the existing research normally tackled a single data-related task and there are very few frameworks that process multi-source data and consider multiple tasks.

To address the above issues, we propose a multi-task analysis and modelling framework for exploitation of multi-source monitoring data of inland vessels. In details, we first collect multi-source monitoring data of real vessels and pre-process them with statistical analysis and data cleaning methods. An algorithm for identification of vessel navigational state is then designed. Finally, an advanced ANN, the Long Short-Term Memory (LSTM) network (Gers et al., 2000), is tailored to solve multiple tasks, including vessel trajectory repair, engine data modelling and fuel consumption prediction. To the best of our knowledge, for the first time, LSTM is employed into the applications relating to inland vessel monitoring and some data-driven models, such as the engine speed model, are developed for the first time.

3. The proposed multi-task analysis and modelling framework

The data studied in this paper came from the vessel-borne monitoring terminals of bulk cargo vessels sailing on the Yangtze River trunk. Like many other real-life data sets, the vessel monitoring data also have problems of noise interference, data redundancy and partial data missing. The vessel-borne monitoring data were acquired from different sensors and consist of various types of information, which makes the data set heterogeneous and asynchronous. To best utilise the available data to improve the vessel performance monitoring, this paper proposes a multi-task data processing framework as shown in Fig. 2. It mainly consists of 1) data collection; 2) data preparation, including data extraction and classification; 3) data pre-processing, including data partition, data sorting and abnormal data detection and removal; 4) vessel navigational state identification; and 5) data analysis and modelling, including vessel trajectory repair, engine speed modelling, fuel consumption prediction, etc.

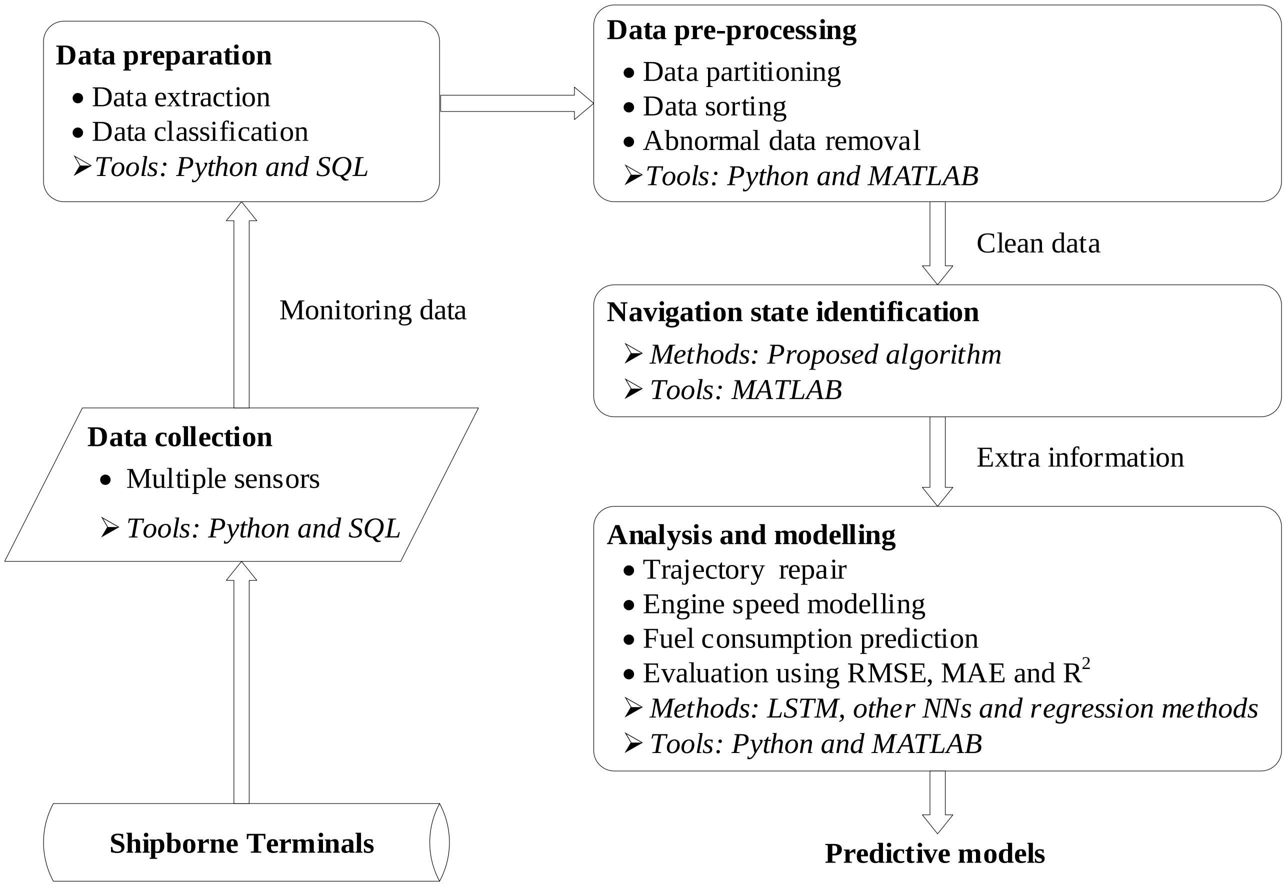


Fig. 2. The multi-task data processing framework for performance monitoring of inland vessels

3.1 Data preparation

The vessel monitoring data that are commonly studied nowadays can be classified into three categories: navigational status data, voyage data and performance data, as shown in Fig. 3.

In the navigational status data, longitude and latitude record the discrete locations of a vessel and provide a basis for subsequent segment division and trajectory reconstruction. Some dynamic information, including SOG, COG and draft, is the basis for vessel navigational state identification. They reflect environment and conditions of the navigation water to a certain extent. Draft data also provide valuable information for vessel overload and safety monitoring. In the voyage data, the voyage number and voyage time information helps in voyage division, and the navigation port information helps in the determination of navigable region and the calculation of navigation mileage. In the vessel performance data, the engine speed and temperature data reflect the operating state of engines. The data of fuel and reserve fuel facilitate the calculation of fuel consumption. In many cases, both engine speed and fuel level have multiple readings for multiple engines and fuel conservators. In this study, the data sources include the

AIS system and the sensors used to measure engine speed and fuel consumption. Some examples of the raw data collected on February 21, 2019 are shown in Table 1.

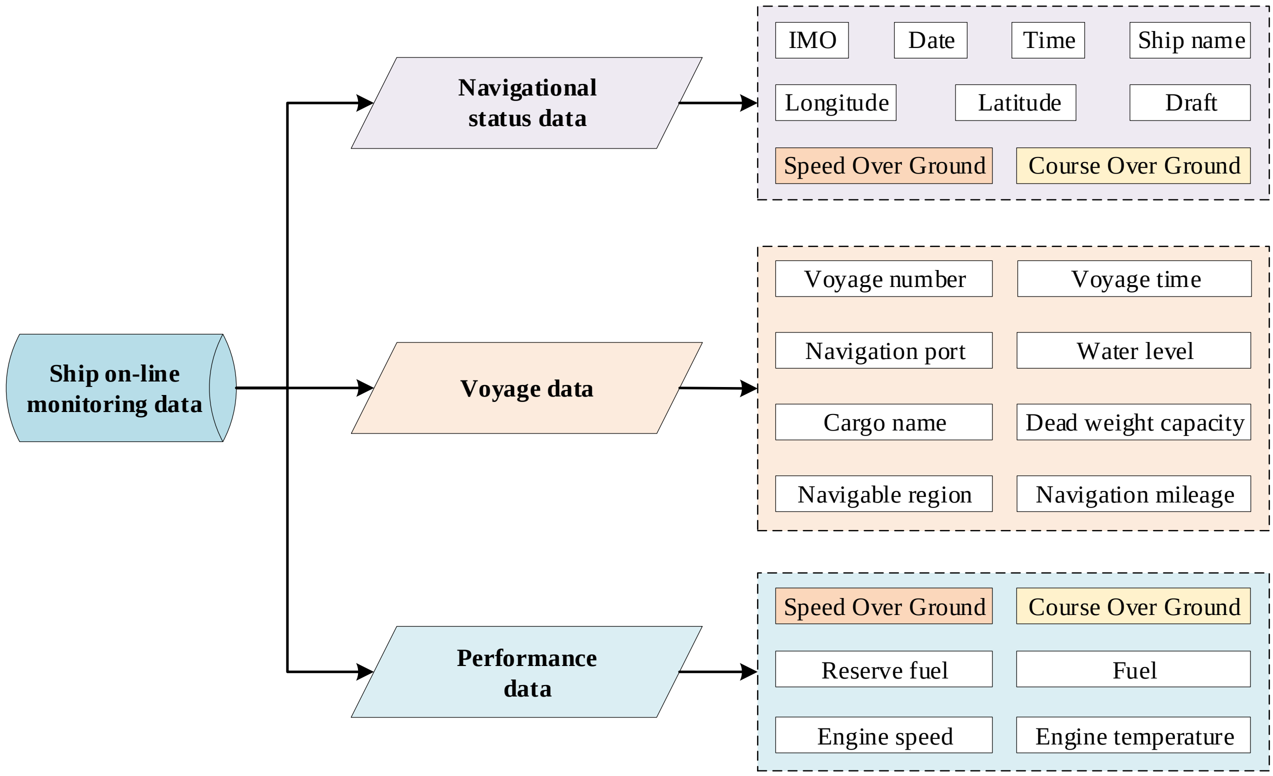


Fig. 3. Classification of vessel monitoring data

Table 1. Examples of raw monitoring data of a vessel on February 21, 2019

Time	Longitude	Latitude	SOG (km/h)	COG (°)	Mileage (km)	Reserve fuel (L)		Engine speed (rpm)		……
						Left	Right	Left	Right	
19:00:22	114.2603	30.5238	1.706	199.69	0.0316	1537.7	1679.5	424.3	446.9	……
19:02:03	114.2602	30.5233	2.767	200.23	0.0984	1536.9	1677.0	483.4	449.0	……
19:03:03	114.2602	30.5233	3.097	209.91	0.0984	1535.6	1676.1	483.6	449.5	……
19:04:03	114.2598	30.5227	3.519	211.37	0.1707	1534.8	1674.9	484.4	509.1	……
19:05:44	114.2592	30.5219	4.137	210.58	0.2823	1533.2	1672.6	484.7	509.9	……
19:06:44	114.2592	30.5219	4.380	213.19	0.2823	1532.4	1671.1	484.9	510.4	……
……	……	……	……	……	……	……	……	……	……	……

3.2 Data pre-processing

There are many errors and anomalies in the original vessel monitoring data. Fig. 4 shows the anomalies and noises within the 10,828 original data samples on No. 1902 voyage. The red boxes and ellipses indicate the obvious abnormal data. In Fig. 4(a) and Fig. 4(b), the SOG and COG samples with zero values are noise data, which may be caused by transmission error or the fact of the vessel laying at an anchor. In Fig. 4(c), the normal range of longitude is between 105 and 115, and the zero values are obviously abnormal. The normal reserve fuel should go less and less, but there are some fluctuations due to vessel shaking, as shown in Fig. 4(d).

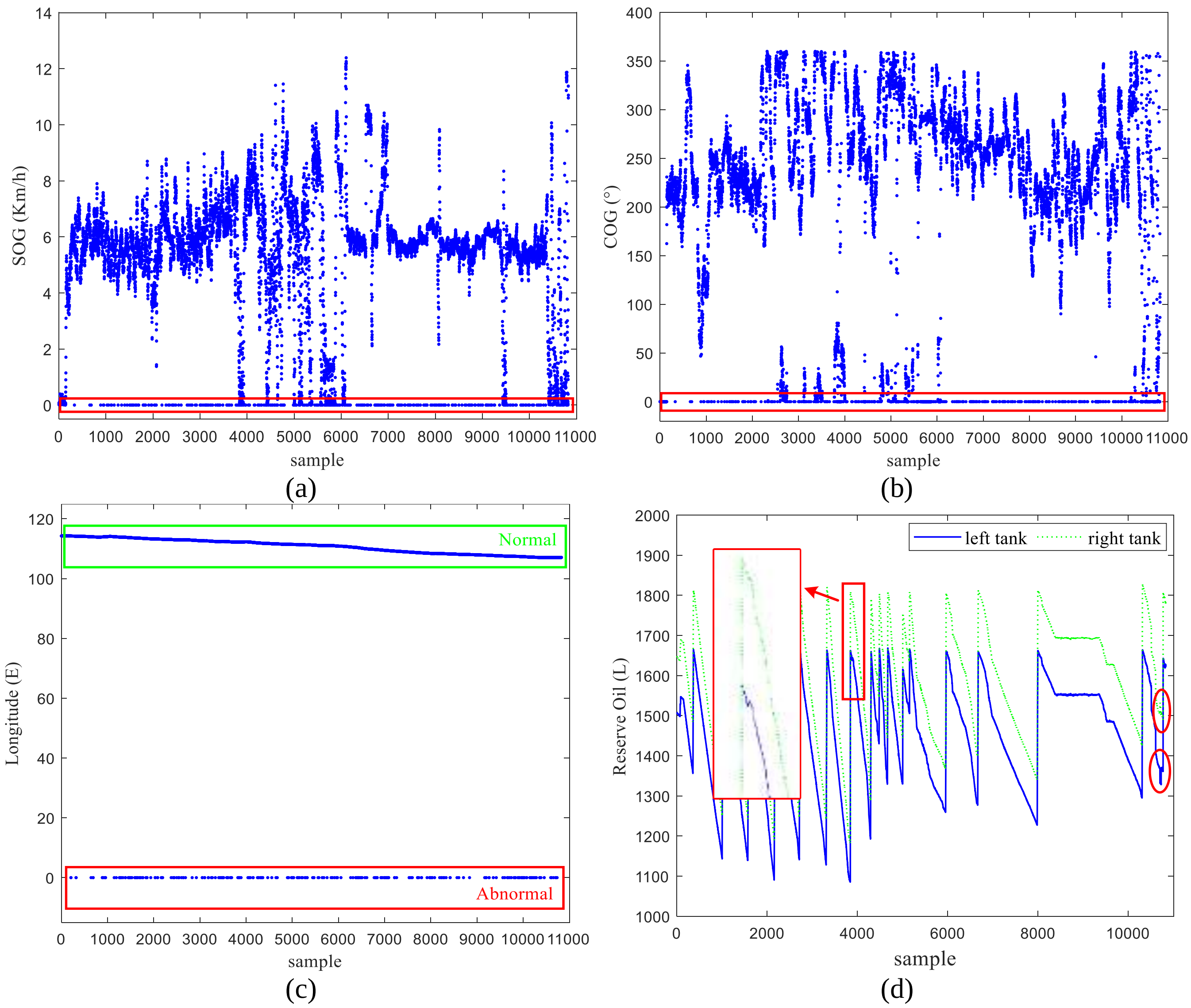


Fig. 4. The original monitoring data: (a) SOG, (b) COG, (c) longitude and (d) reserve fuel

It should also be noted that the monitoring data of different vessels show different characteristics. The data of the same vessel have some degree of change with different voyages under different navigational conditions. Therefore, it is necessary to pre-process the vessel monitoring data using partitioning, sorting and cleaning methods to obtain ready-to-use data. The pre-processing mechanism is designed as Fig. 5.

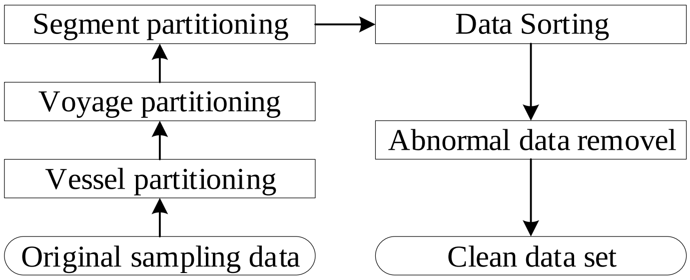


Fig. 5. Pre-processing for inland vessel monitoring data

As shown in Fig. 5, the proposed data pre-processing mechanism for inland vessel monitoring data is divided into the following five steps:

Step 1: Vessel partitioning. According to the IMO code, the original data are divided into groups, which can be validated by the vessel name.

Step 2: Voyage partitioning. Data for each vessel are further divided according to the voyage number. Verification can be performed using the voyage time including the start time and the end time.

Step 3: Segment partitioning. The data of a voyage can be further partitioned into segments in accordance with its longitude and latitude. The navigation region and navigation mileage can be used to verify the accuracy of this partitioning.

Step 4: Data sorting. All data are sorted according to the sampling date and time.

Step 5: Abnormal data removal. The key of this step is the establishment of abnormal determination principles. It relates to not only the characteristics exhibited by the data itself, but also the navigational conditions and vessel operations. After extensive data analysis, on-site investigation and expert consultation, abnormal determination principles for inland vessel monitoring data are formulated as follows:

(1) Too few data records in a certain range. This makes feature extraction inaccurate and leads to poor representativeness and credibility for the analysis and modelling results.

(2) Data duplication. This is one main type of abnormal data. Repeated data can lead to redundancy and increase the complexity of data processing.

(3) Abnormal position. This is the case that the longitude or latitude is not within the normal range of the navigation area.

(4) Abnormal SOG. Speed of a vessel should generally not exceed a certain range and not change too much in a short time.

(5) Abnormal COG. COG of a vessel should maintain in a certain range in a short time.

(6) Abnormal engine speed. Many inland cargo ships are equipped with two engines and have two engine running states: single engine running and double engine running. Regardless of the engine running state, engine speed has an upper limit and a lower limit, and it is abnormal if the engine speed is beyond the normal range.

(7) Abnormal reserve fuel. Fuel reserve readings record the amount of fuel stored in different fuel tanks. Normally, the overall reserve fuel decreases with the increase of sailing time. However, if there is a case of bunkering during the voyage, it will result in an increase in the readings.

(8) Abnormal water level. The water level data show large fluctuations, but it cannot exceed a normal range in the navigation.

It should be noted that the position information, such as longitude and latitude, was obtained by the Global Navigation Satellite System (GNSS) terminal of the vessel, while the data of engine speed, fuel and reserve fuel were derived from the vessel performance sensors. This means that if a data record shows a position abnormality, the engine information at the same time point may be normal. If the whole data vector is directly removed, it may cause the loss of some normal information. Therefore, in the proposed data cleaning step, we consider the abnormality according to the type of data information, and only remove the abnormal piece of a specific datum. The purpose of this operation is to get clean data, while retain real and useful information as much as possible.

3.3 Identification of vessel navigational state

It is crucial to judge the navigational state of vessels, as different navigational states affect the vessel trajectory analysis, fuel consumption calculation, collision risk assessment, traffic accident investigation and so on. Generally speaking, the navigational state of a vessel can be divided into four types, namely berthing, manoeuvring, sailing and temporary stopping. In the berthing state, a vessel is usually located in the service area of a dock or a port, etc., and it is fixed with cables for loading, unloading or other activities such as replenishing water, food and fuel. In this state, except for the slight shaking caused by wind and waves, the position and speed of the vessel are basically not changed, and the engine is also switched off. In the state of manoeuvring, the vessel would move with the help of a tug vessel and thus the vessel trajectory would be different from that of a normal speed sailing. In this study, there is no monitoring data of vessels in the manoeuvring state. In the sailing state, the vessel is sailing in the navigable water area with a normal speed. At this time, the engine runs at a high speed to drive the corresponding SOG, COG, longitude and latitude to change accordingly. Meanwhile, the reserve fuel decreases with the increase of navigation time and navigation mileage. In addition to the above three navigational states, there is a special state of navigation, temporary stopping, which is caused by a vessel failure or avoiding collision with other vessels during the navigation. In this state, the vessel is not located in a port or a dock, but the navigation mileage does not increase.

This study considers the following three vessel navigational states: berthing, temporary stopping and sailing, and proposes an algorithm for identification of navigational state using vessel monitoring data. We divide the service area and navigable area to some subsets of segments. For each sub-segment in either the service area or the navigable area, we calculate the average SOG and the average engine speed in a time interval Δt. The average values are then compared with thresholds of SOG and engine speed, SOG_0 and ES_0, to decide the navigational state. According to some preliminary analysis to the SOG and engine speed, we set $\Delta t = 10Min,\ SOG_0 = 1Km/h,\ ES_0 = 300$. The pseudo code of the identification algorithm is presented in Algorithm 1.

Algorithm 1 Identification of vessel navigational state

Input: $Segments\_Set$ (Segments data)
Output: $State\_B$ (Berthing state), $State\_T$ (Temporary stopping state), $State\_S$ (Sailing state)
Variable: $Segments\_Set$ (Segments data), t_first (first time), t_last (last time)
t_start (start time), t_end (end time), $State\_Temp$ (Temporary state)

[1] Initialise: $State\_B \leftarrow \emptyset$, $State\_T \leftarrow \emptyset$, $State\_S \leftarrow \emptyset$, $State\_Temp \leftarrow \emptyset$
[2] Extract the segments data from clean dataset, and save as $Segments\_Set$
$Segments\_Set = \{sub\_segment_1, \cdots, sub\_segment_n\}$
[3] ***for each*** $sub\_segment$ in $Segments\_Set$ ***do***
[4] Record t_first and t_last
$t\_first \leftarrow$ the time of the first record in $sub\_segment$
$t\_last \leftarrow$ the time of the last record in $sub\_segment$

[5] ***for*** $t = t\_first$ ***to*** t_last ***do***

[6] ***if*** $mean\_SOG(t, t + \Delta t) < SOG_0$ ***and*** $mean\_ES(t, t + \Delta t) < ES_0$ ***then***

[7] Record current time $t\_start = t$

[8] Find the end time t_end

[9] Append the data of between t_first and t_start to $State\_Temp$

[10] ***if*** $sub\_segment$ is within service area ***then***

[11] $State\_B = State\_B \cup State\_Temp$

[12] ***else***

[13] $State\_T = State\_T \cup State\_Temp$

[14] ***end if***

[15] Update the first time $t\_first = t\_end$

[16] ***else***

[17] $State\_S = State\_S \cup sub\_segment$

[18] ***end if***

[19] ***end for***

[20] ***end for***

3.4 Modelling using LSTM

The real-time vessel monitoring data provide a lot of valuable information for identification of vessel navigational status, trajectory analysis, risk assessment, traffic management and optimisation of vessel operations. However, if parts of the data are missing, the overall value of the data will be greatly affected and some useful information may not be extracted. This may further lead to erroneous conclusions and wrong decisions in operation. For instance, the absence of trajectory data can affect the aquatic accident investigations and the hot track mining. Engine speed and fuel consumption are also important attributes that help the optimisation of vessel speed and aid the decision-making for vessel operations. However, in some situations, the real-time data relating to these variables are not available. In these cases, it is necessary to build appropriate models to predict or recover the missing information. In this study, advanced ANN models based on the LSTM (Gers et al., 2000) structure were developed to tackle three tasks, i.e. trajectory reconstruction, engine speed modelling and fuel consumption prediction.

In this work, the original monitoring data were obtained by real-time sampling at a fixed sampling time. We may regard the samples as a time series of multi-dimensional data, in which case LSTM is good at modelling. LSTM is a special Recurrent Neural Network (RNN) model, as shown in Fig. 6 (Goodfellow et al., 2016). It was proposed to solve the problem of gradient dispersion in the RNN model.

It is a highly efficient Cyclic Neural Network (CNN) that is dedicated to process the sample data of time series. Due to the strong ability of self-learning (Alahi et al., 2016; Zhao et al., 2017), the LSTM network is suitable for constructing predictive models for vessel monitoring tasks.

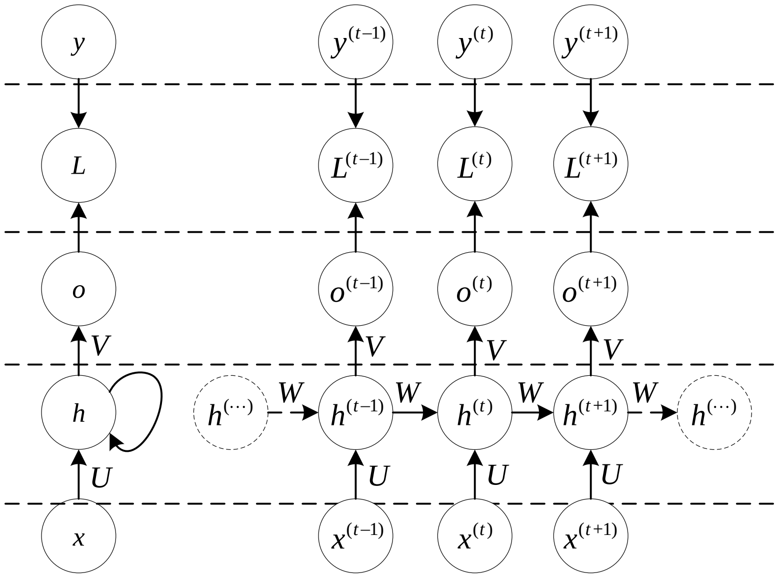


Fig. 6. The structure of RNN (Goodfellow et al., 2016)

In Fig. 6, x is the sequence of input with length T, h is the sequence of the hidden layer, O is the output sequence, L is the total loss and y is the sequence of target. U is the parameter matrix from the input layer to the hidden layer, W is a self-looping parameter matrix in the hidden layer and V is the parameter matrix from the hidden layer to the output layer. As can be seen from Fig. 6, the input nodes, the hidden nodes and the output nodes are all represented by small circles and they are fully connected. A self-loop feedback is added between the hidden nodes through weight sharing, which enables the network to process the data of indefinite length.

The main difference between the traditional RNN and LSTM lies in an information conveyor belt named "cell state" on the top of LSTM, which memorises information. The LSTM network has three control gates: forget gate, input gate and output gate. Forget gate is a control gate between the previous long state information and the current long state information. Input gate is a control gate between the short state information and the long state information. Output gate is a control gate between the current information and the output state information. The current information is the summation of the long state information and the short state information.

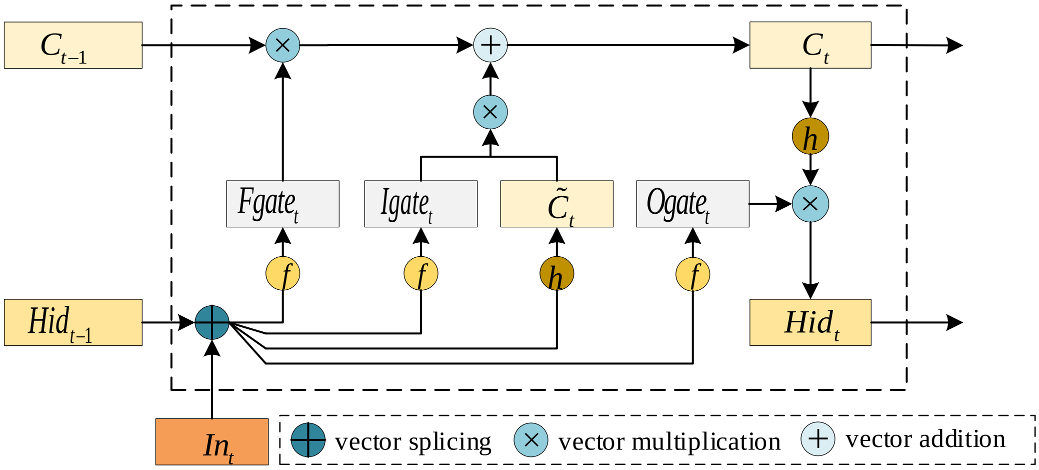


Fig. 7. The structure of the LSTM module (Gers et al., 2000)

Fig. 7 demonstrates the structure of a LSTM module, where In_t is the input sequence at time t, which can be a one-dimensional or multi-dimensional feature vector. Hid_t is the output of the hidden node at time t and Hid_{t-1} is the output of the hidden node at time $t-1$. C_t is the cell memory at time t and C_{t-1} denotes the cell memory at time $t-1$. $Fgate_t$, $Igate_t$ and $Ogate_t$ are the forget gate, input gate and output gate at time t, respectively. f is a sigmoid activation function and h is a hyperbolic tangent activation function $tanh$.

The calculation of information through the forget gate F_t, input gate I_t and output gate O_t is shown in Equations (1)-(3), respectively. Equation (4) shows the update of the cell memory, where F_t is used to control how much history information is forgotten and I_t is used to control how much new information is saved. Equation (5) calculates the output of the module, controlled by O_t. $W_{\{F,I,O,C\}}$ are the parameter matrices from the input layer to the hidden layer. $U_{\{F,I,O,C\}}$ are the self-looping parameter matrices of the hidden layer. $b_{\{F,I,O,C\}}$ are offset parameter matrices.

$$F_t = f(W_F In_t + U_F Hid_{t-1} + b_F) \quad (1)$$

$$I_t = f(W_I In_t + U_I Hid_{t-1} + b_I) \quad (2)$$

$$O_t = f(W_O In_t + U_O Hid_{t-1} + b_O) \quad (3)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot h(W_C In_t + U_C Hid_{t-1} + b_C) \quad (4)$$

$$Hid_t = O_t \odot h(C_t) \quad (5)$$

The modelling process with LSTM can be described as follows:

Step 1: Setting the input and output variables.

Step 2: Normalising the input and output data to the range of 0 to 1, and dividing the data set into a training set and a testing set.

Step 3: Reshaping the input and output vectors into the format of [samples, time steps, features], which the LSTM network adapts to.

Step 4: Creating and training the LSTM network. First, randomly initialising weights, and then training the network until termination criterion are satisfied.

Step 5: Testing the network with a separate data set and evaluating the network using evaluation functions, such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and coefficient of determination (R^2).

In Step 2, the method of min-max normalisation can be used to normalise the input data of the LSTM network, as shown in Equation (6).

$$x(t)^* = \left(x(t) - Min(x(t))\right) / \left(Max(x(t)) - Min(x(t))\right) \tag{6}$$

where $x(t)$ and $x(t)^*$ represent the initial data and the normalised data, respectively. In Step 5, *RMSE*, *MAE* and R^2 can be adopted to evaluate the performance of the developed models, as shown in Equations (7)-(9).

$$RMSE = \left(\frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y_t})^2\right)^{1/2} \tag{7}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \widehat{y_t}| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(y_t - \widehat{y_t})^2}{\sum_{t=1}^{T}(y_t - \overline{y_t})^2} \tag{9}$$

where t represents the index of a datum and T represents the number of data; y_t and $\widehat{y_t}$ are the real values and the predicted values of the t^{th} datum, respectively; $\overline{y_t}$ is the mean of y_t, where $t = 1,2,3 \ldots T$.

4. Case study

In this study, the host platform was a desktop PC, of which the CPU (Central Processing Unit) was Inter (R) Core (TM) i5-8500, the main memory was 16GB RAM (Random Access Memory) and the operating system was 64-bit Windows 10. The programming language was Python 3.7, where a Python IDE (integrated development environment) Spyder and an open-source ANN library Keras were employed.

4.1 Data source

The data used in this study came from the inland vessel monitoring system of Changjiang National Vesselping Group Co. Ltd. The data set includes 128 voyages and 543,224 data records of a vessel (IMO: CN20112508309, see Table 2 for its basic information), collected between August 12, 2018 and March 28, 2019. The normal sampling time is around 1 minute. An example about the data has been shown in Fig. 4, which relates to the voyage no. 1902 and were collected between 18:51 on February 21, 2019 and 16:27 on March 11, 2019. The original monitoring data include 10,828 data records. After the data pre-processing, we got a clean data set with 9,682 records. There were about 10.6% of the original data that were identified as abnormal and removed. In the following case study, the data of the segment Maoping to Fengjie (segment no. 12) were employed, which was in the reservoir area of the Three Gorges Dam that had very stable and relatively small current. To further reduce the effect of environmental factors to engine speed and fuel consumption, the relevant data were collected at the time with very little wind.

Table 2. The basic information of the vessel

Parameters	**Value**
Designed length (m)	110.0
Designed width (m)	19.2
Designed depth (m)	5.6
Deadweight (Mt)	7028
Main engine rated power (kW)	735 × 2
Main engine rated speed (rpm)	830

The method proposed in Section 3.2 was adopted to clear the abnormal data in the original data set. The algorithm of identification of vessel navigational state was then used to derive the navigational state, and three relevant data sets $State\_B$, $State\_T$ and $State\_S$ were obtained. Following the method and steps introduced in Section 3.4, three modelling tasks were accomplished and the details are provided in the following sections.

4.2 Vessel trajectory repair

Trajectory repair is an important research work in vessel trajectory data mining. The vessel trajectory data include longitude and latitude, obtained by periodic real-time sampling. Such data can be regarded as two-dimensional time series after sorting, and thus we set the input feature vector of trajectory as $Tra = \{longitude, latitude\}$ in the trajectory modelling. The principle of trajectory repair using the LSTM network is to estimate and fill in the missing trajectory values by learning the inter-relationship among the front and back trajectories data.

After a number of preliminary experiments as shown in Table 3, the following parameter settings were found to be appropriate and were used: the number of neurons was set to 72, the batch size was set to 100, the activation function of the dense layer (Zhao et al., 2017) was "*linear*", and the optimiser function was "*rmsprop*", which is root mean square propagation optimiser. We created two LSTM networks to perform bi-directional repair from forward and backward, respectively. For each LSTM model, we set the time step to 4, which means repairing one missing trajectory point needs previous four trajectory points.

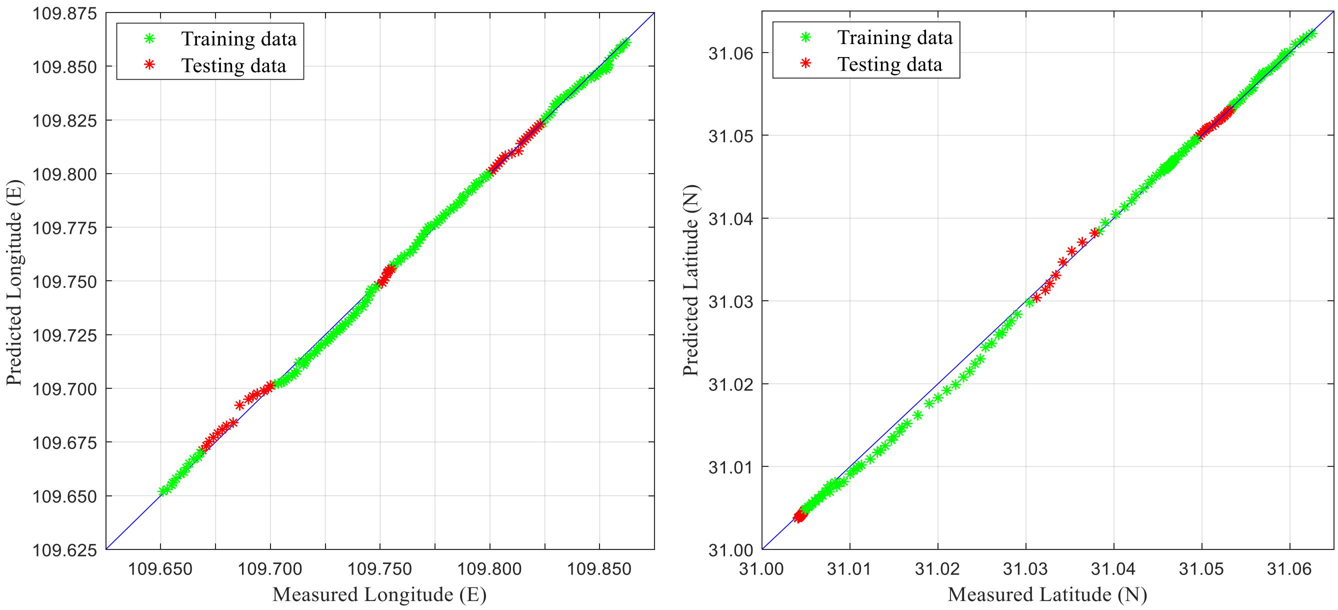
Table 3. Some experiments of vessel trajectory repair with different parameter settings

Group	Neurons number	Batch size	Time steps	Dense layer activation function	Optimiser function	*RMSE* ($\times 10^{-2}$)	
						Longitude	Latitude
1	72	50	3	*relu*	*rmsprop*	0.6203	0.0883
	72	50	3	***linear***	*rmsprop*	0.0183	0.0183
	72	50	3	*softsign*	*rmsprop*	0.2605	0.0223
	72	50	3	*sigmoid*	*rmsprop*	0.7115	0.0985
	72	50	3	*tanh*	*rmsprop*	0.4232	0.0475
2	72	50	2	*linear*	*rmsprop*	0.1311	0.0195
	72	50	**4**	***linear***	***rmsprop***	0.0028	0.0180
	72	50	5	*linear*	*rmsprop*	0.0928	0.0394
	72	50	6	*linear*	*rmsprop*	0.1009	0.0494
3	72	50	4	*linear*	*adamax*	0.1298	0.0140
	72	50	4	*linear*	*nadam*	0.0264	0.0177
	72	50	4	*linear*	*adadelta*	0.2273	0.0234
	72	50	4	*linear*	*adagrad*	0.1094	0.0183
	72	50	4	*linear*	*adam*	0.0967	0.0191
4	48	50	4	*linear*	*rmsprop*	0.0292	0.0186

	64	50	4	*linear*	*rmsprop*	0.0191	0.0189
	100	50	4	*linear*	*rmsprop*	0.0141	0.0239
	128	50	4	*linear*	*rmsprop*	0.0075	0.0273
5	72	40	4	*linear*	*rmsprop*	0.0028	0.0181
	72	80	4	*linear*	*rmsprop*	0.0027	0.0180
	72	**100**	**4**	***linear***	***rmsprop***	0.0027	0.0180
	72	120	4	*linear*	*rmsprop*	0.0028	0.0180

Activation functions: *relu*: a rectified linear unit function; *linear*: a linear activation function; *tanh*: a hyperbolic tangent function; *softsign*: similar to *tanh* but smoother; *sigmoid*: a common s-type function. **Optimisation functions:** *rmsprop*: root mean square propagation optimiser; *adam*: adaptive moment estimation; *adamax*: a variant of *adam* with infinity norm; *nadam*: Nesterov-accelerated adaptive moment estimation; *adagrad*: adaptive gradient algorithm; *adadelta*: extension of *adagrad* with smaller learning rate.

The modelling performance of the developed LSTM networks against the training data and the testing data is shown in Fig. 8. In order to verify the performance of the constructed model, we repaired a straight-line trajectory section and a curved-line trajectory section. The results are shown in Fig. 9 and Table 4. Fig. 9(a) is a complete vessel trajectory after repair, where two missing sections are marked by a black box (1) and a black ellipse (2). The straight-line part includes twenty repaired trajectory points, which consist of ten points of forward repair and ten points of backward repair. The curved-line part includes fifteen repaired trajectory points, which consist of seven points of forward repair and eight points of backward repair. The comparison between the real trajectory and the repaired trajectory is shown in Fig. 9(b) and Fig. 9(c).



(a) (b)

Fig. 8. The measured data vs. predicted data using bi-directional LSTM networks for (a) longitude and (b) latitude

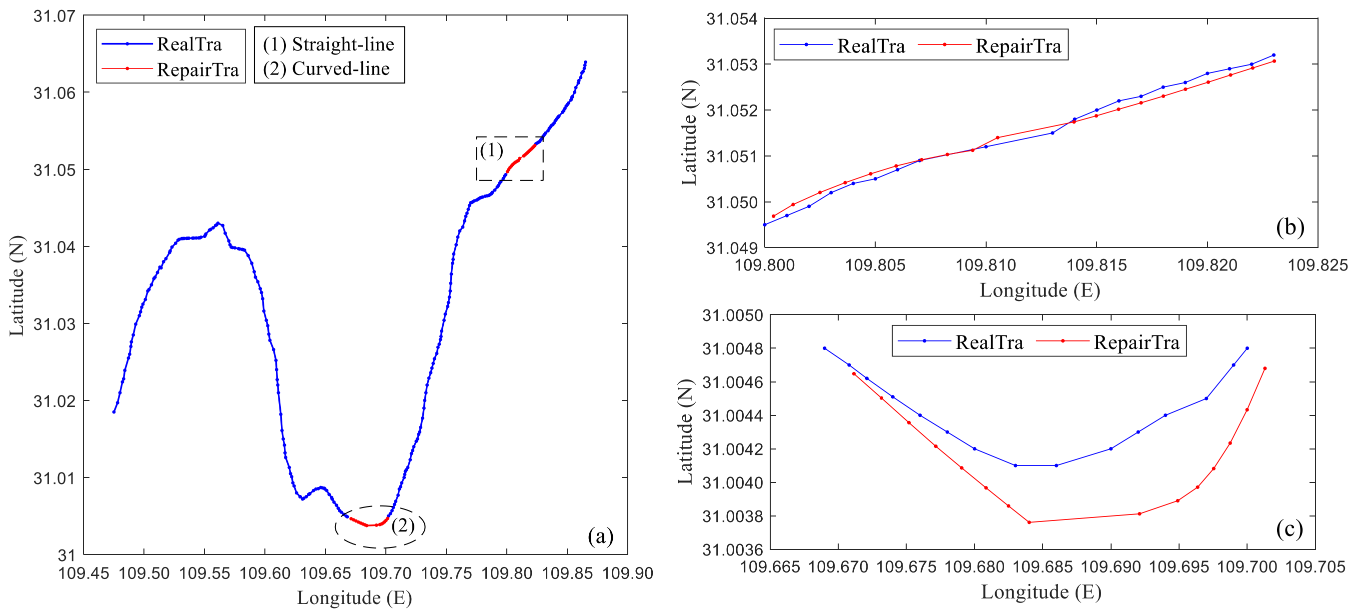


Fig. 9. The results of vessel trajectory repair using bi-directional LSTM networks: (a) a complete vessel trajectory after repair, (b) details of the straight-line section and (c) details of curved-line section

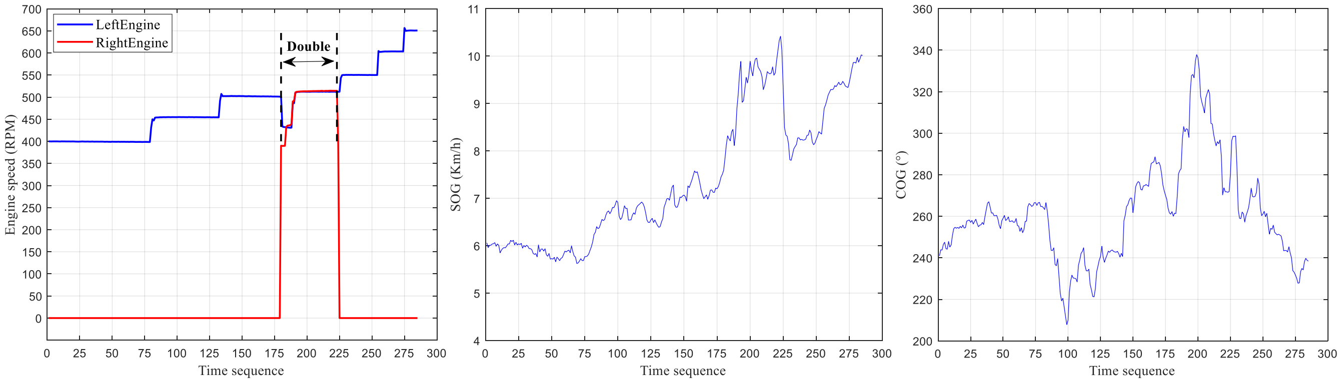
The developed network model was compared with widely used Back-Propagation Neural Networks (BP-NNs) to verify the merits of the LSTM network. We compared it with a single-layer BP-NN and a double-layer BP-NN using the same sample data. The results are shown in Table 4, and it indicates that the LSTM network we created outperforms conventional ANNs in accuracy in vessel trajectory repair.

Table 4. Comparison between different methods in vessel trajectory repair

Network	**Neurons**	***RMSE* (Straight-line)**		***RMSE* (Curved-line)**	
		Longitude	**Latitude**	**Longitude**	**Latitude**
BP-NN (Single-layer)	128	0.0072±0.0029	0.0016±0.0011	0.0086±0.0059	0.0019±0.0010
BP-NN (Double-layer)	[128 5]	0.0037±0.0020	0.0008±0.0005	0.0062±0.0038	0.0012±0.0009
Proposed LSTM	128	0.0008±0.0004	0.0003±0.0002	0.0032±0.0006	0.0003±0.0002

4.3 Engine speed modelling

Fig. 10 shows some vessel dynamic data extracted from the on-line monitoring samples, including the speeds of a left engine and a right engine. As shown in Fig. 10(a), the blue line indicates the speed of the left engine, which includes seven different stages between 400 and 650. The red line indicates the speed of the right engine, which is zero in most of the time except for when the vessel is in the double-engine operating state. By comparing Fig. 10(a) with Fig. 10(b) and Fig. 10(c), it can be seen that the engine speed has close relationship with SOG and COG. For example, the SOG increases with the increase of engine speed and reaches the maximum when both engines are running. In the cases that some engine speed data are missing or the engine speed reading is not available in some places, such as in the remote waterway traffic management centre, the SOG and COG data may help repair or generate the engine speed information.

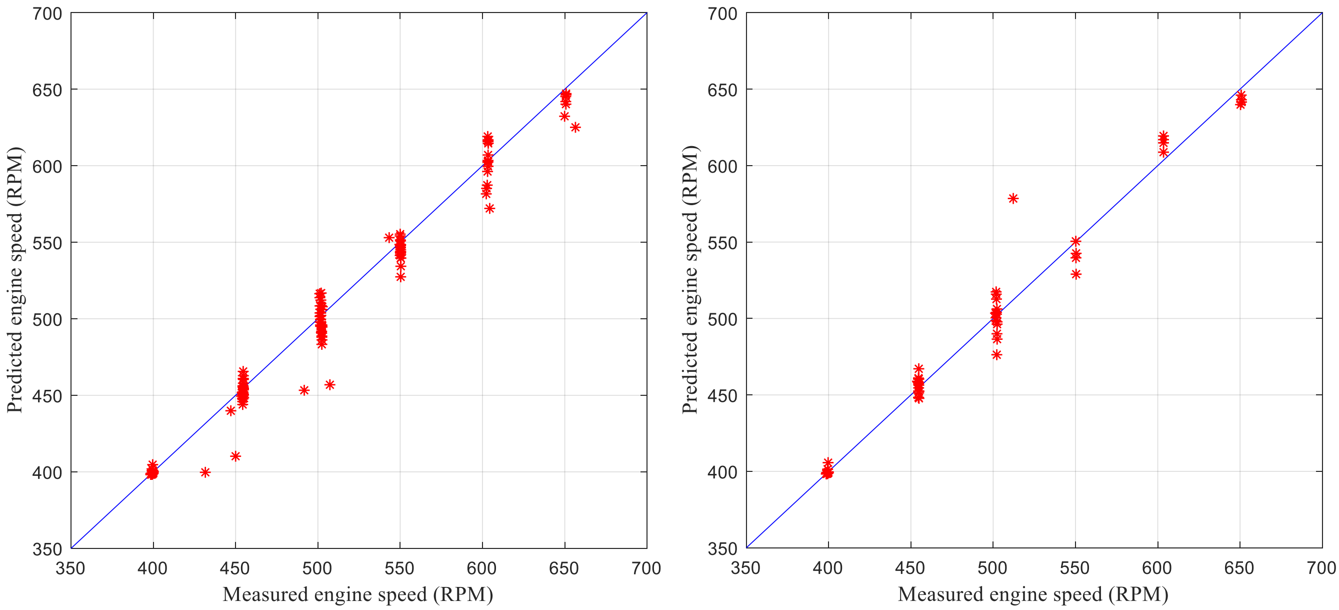


(a) (b) (c)

Fig. 8. The collected data from real-life vessel experiments: (a) engine speed, (b) SOC and (c) COG

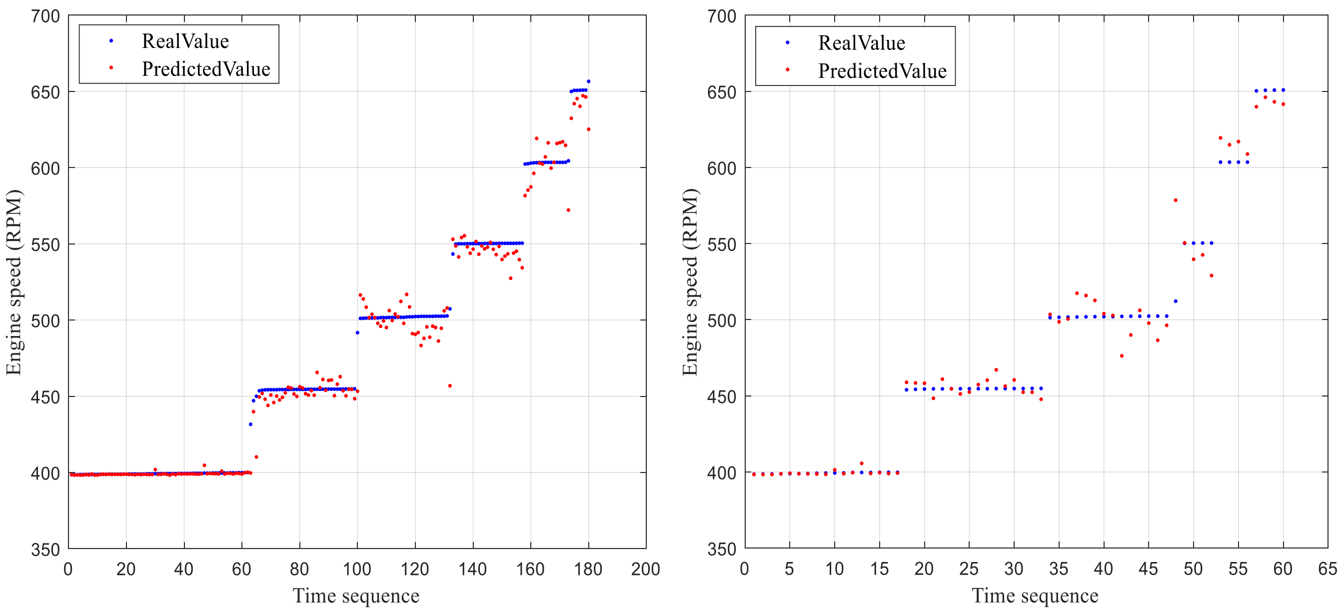
In order to estimate the relationship among engine speed, SOG and COG, we constructed a model using the LSTM structure. *SOG* and *COG* were set as the input variables and *engine speed* was set as the output variable, where the effect of environmental factors is not considered. 75% of the available data were randomly selected as the training set and the remaining 25% were the testing set. Two-dimensional series of input data $\{SOG, COG\}$ and one-dimensional of output data $\{engine\ speed\}$ were presented to the LSTM network with the time step being 1. After a number of preliminary experiments similar to Table 3, the following parameter settings were found to be appropriate and were used: the number of neurons was set to 128, the batch size was set to 72, the activation function of the dense layer was "*tanh*", and the optimiser function was "*rmsprop*", where "*tanh*" is the hyperbolic tangent function and "*rmsprop*" is the root mean square propagation optimiser.

Figs. 11 and 12 show the modelling performance of the developed model in both training and testing. In Fig. 12, the blue points represent the measured values of engine speed and the red points represent the predicted values. We also compared the LSTM method with some other widely used methods, including Stepwise Regression (SR) (Kolasa-Wiecek, 2015), Interaction Regression (IR) and Pure Quadratic Regression (PQR) (Kumar et al., 2019), and BP-NNs. We calculated the R^2, *RMSE* and *MAE* values of each model based on the same training and testing data sets. The results shown in Table 5 reveal that LSTM outperforms other methods in the engine speed modelling. The testing *RMSE* of the LSTM model is 11.84, while the testing *RMSE*s of the other models are between 14.31 and 21.19.



(a) (b)

Fig. 9. The measured data vs. predicted data using the LSTM network for engine speed modelling: (a) training data and (b) testing data



(a) (b)

Fig. 10. The performance of the developed engine speed model using the LSTM network in (a) training and (b) testing

Table 5. Comparison of different methods in engine speed modelling

Method	Training			Testing		
	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*
SR	0.941±0	18.154±0	13.245±0	0.920±0	21.195±0	14.185±0
IR	0.960±0	14.798±0	11.767±0	0.932±0	19.514±0	13.058±0
PQR	0.964±0	13.974±0	11.071±0	0.936±0	18.271±0	12.545±0
BP-NN (Single-layer)	0.976±0.0067	11.517±1.442	8.547±1.583	0.952±0.0126	16.140±2.038	10.149±1.575
BP-NN (Double-layer)	0.978±0.0085	10.894±1.955	7.974±1.873	0.962±0.0124	14.312±2.127	9.512±1.697
Proposed LSTM	0.989±0.0019	9.412±0.406	5.772±0.513	0.975±0.0023	11.841±0.622	6.663±0.522

4.4 Fuel consumption prediction

The fuel consumption in this paper refers to the amount of fuel consumed by a vessel in unit time. It has a big impact on ship manoeuvring and economic speed control (Coraddu et al., 2017). Unlike the vessel trajectory and engine speed data, the fuel consumption data were not involved in the collected data base, but can be obtained by calculating the change of the total amount of fuel in all fuel tanks. However, the derived fuel consumption data have large fluctuations, which can be observed from Fig. 13(a). We thus applied a median filtering method (Chen et al., 2017) and a moving average method (Paul et al., 2015) to obtain smoothed data, which facilitate further analysis and modelling. As shown in Fig. 13(b), the smoothed fuel consumption varies among seven distinct levels, which well reflects the situation shown in Fig. 13(a). If Fig. 13 is compared with Fig. 10, it is obvious to see the correlations among fuel consumption, engine speed and SOG, i.e., when the engine speed and/or SOG increases, the fuel consumption of the vessel will roughly increase.

To generate a predictive model for fuel consumption, *SOG*, *COG* and *engine speed* were set to be the input variables, and *fuel consumption* was set to be the output variable, where the effect of environmental factors is not considered. 185 data records were used as the training data and 40 data records were used as the testing data. Three-dimensional series of input data $\{SOG, COG, engine\_speed\}$ and one-dimensional of output data $\{fuel\_consumption\}$ were presented to the LSTM network with the time step being 1. With some preliminary experiments similar to Table 3, we initialised the network as follows: the number of neurons was 128, the batch size was 72, the activation function of the dense layer was "*tanh*", and the optimiser function was "*rmsprop*". The *RMSE* value for training and testing are 0.0123 and 0.0413, respectively.

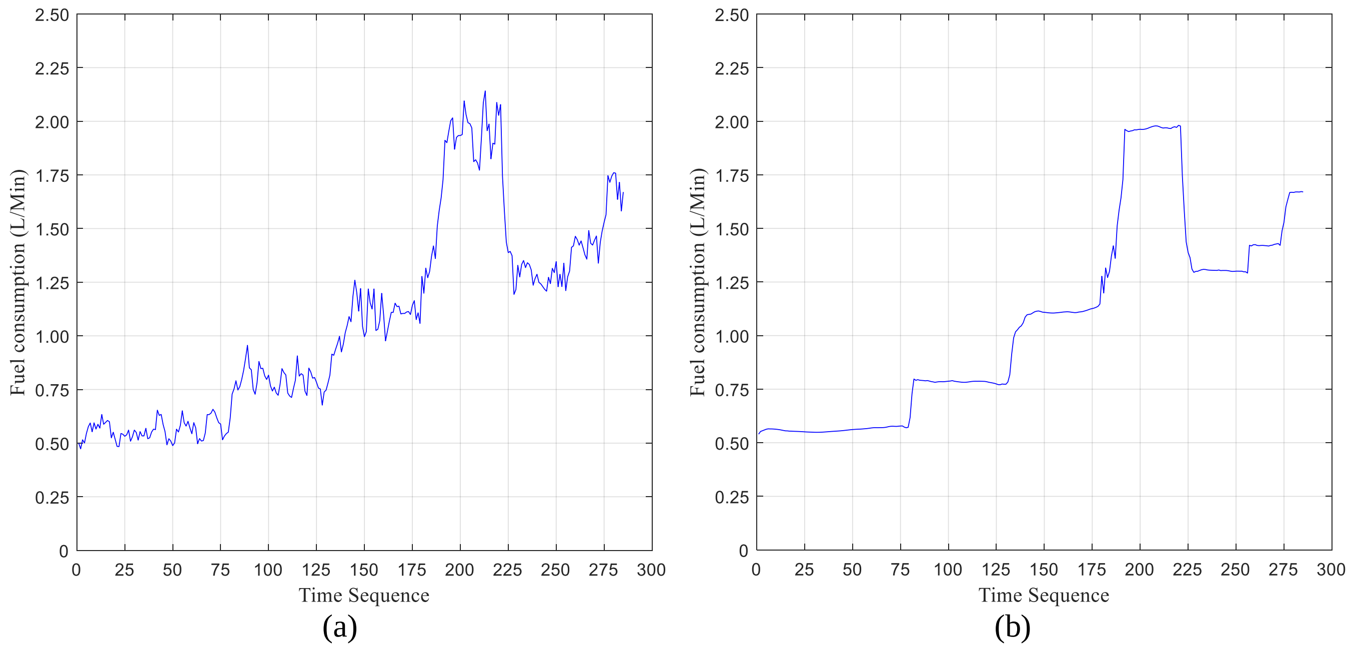


Fig. 11. Fuel consumption data: (a) Unsmoothed and (b) Smoothed

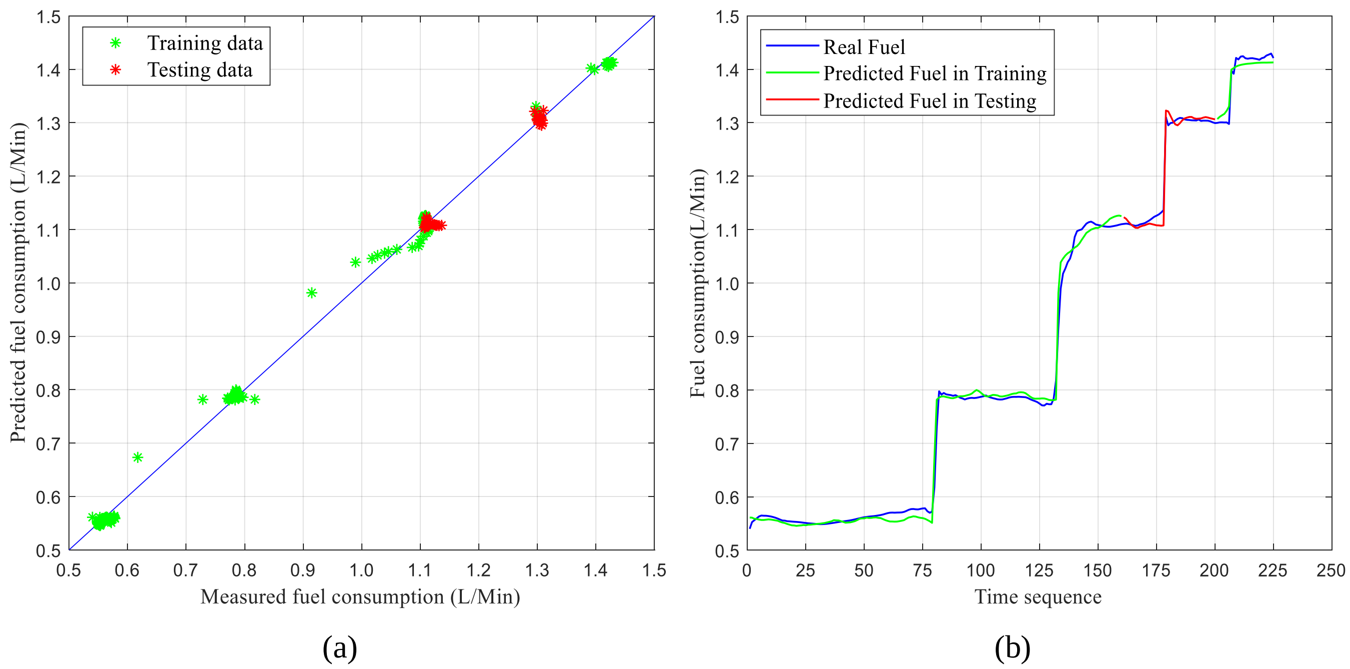
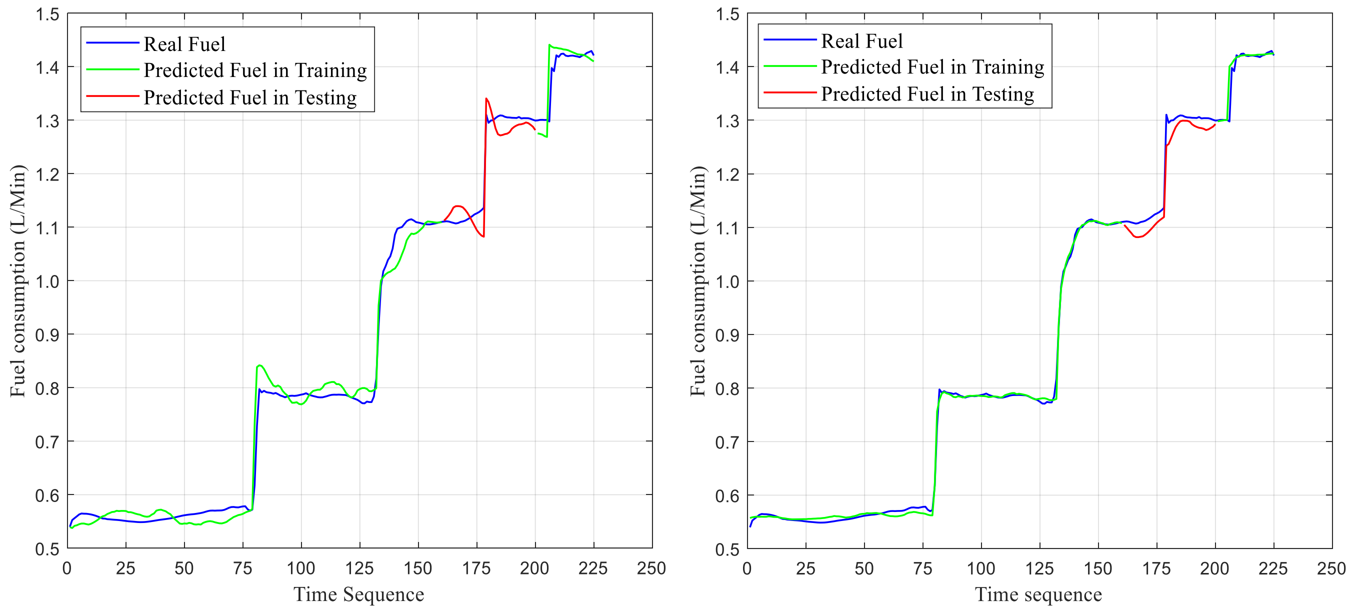


Fig. 12. The performance of the developed fuel consumption model using the LSTM network: (a) measured data vs. predicted data in modelling and (b) comparison in fuel consumption prediction

Fig. 14 demonstrates the modelling performance and the fuel consumption prediction result using the developed network, where in Fig. 14(b) the blue line represents the measured value, the green line represents the predicted value for training data, and the red line represents the predicted value for testing data. To verify the advantages of the LSTM network, we compared it with other five methods, three regression methods (including SR, IR and PQR) and two BP-NNs. The testing *RMSE* value of the LSTM model is 0.0143, while the testing *RMSE* values of other models are all above 0.0257, which are the mean results of 20 experiments. The prediction results of IR and double-layer BP-NN are shown in Fig. 15. The results shown in Table 6, Fig. 14 and Fig. 15 reveal that the proposed method outperforms other methods in fuel consumption prediction.



(a) (b)

Fig. 13. Fuel consumption predictions by (a) interaction regression and (b) double-layer back-propagation neural network

Table 6. Comparison of different methods in consumption prediction

Method	**Training**			**Testing**		
	R^2	***RMSE***	***MAE***	R^2	***RMSE***	***MAE***
SR	0.9879 ±0	0.0321 ±0	0.0233 ±0	0.7929 ±0	0.0495 ±0	0.0452 ±0
IR	0.9935 ±0	0.0235 ±0	0.0179 ±0	0.9166 ±0	0.0257 ±0	0.0221 ±0
PQR	0.9934 ±0	0.0237 ±0	0.0184 ±0	0.8995 ±0	0.0298 ±0	0.0232 ±0
BP-NN (Single-layer)	0.9969 ±0.0026	0.0151 ±0.0060	0.0124 ±0.0011	0.7904 ±0.2347	0.0496 ±0.0424	0.0461 ±0.0122
BP-NN (Double-layer)	0.9992 ±0.0002	0.0080 ±0.0011	0.0081+0.0005	0.8252 ±0.2100	0.0352 ±0.0266	0.0254 ±0.0110
Proposed LSTM	0.9983 ±0.0003	0.0123 ±0.0012	0.0078 ±0.0007	0.9954 ±0.0011	0.0143 ±0.0020	0.0124 ±0.0025

5 Conclusions

In this paper, a framework of data processing, analysing and modelling has been proposed for understanding and utilising multi-source monitoring data of inland vessels. The LSTM neural network has been tailored and employed in three important tasks: vessel trajectory repair, engine speed modelling and fuel consumption prediction. These works have been successfully validated using real-life monitoring data and LSTM has been compared with a number of well-known modelling methods. The LSTM network was first shown to be able to repair both straight-line and curved-line trajectories and it greatly outperformed BP-NNs in terms of repair accuracy. In engine speed modelling and fuel consumption prediction, LSTM also outperformed other modelling methods, such as stepwise regression, interaction regression, pure quadratic regression and BP-NNs, with at least 17% and 44% improvement in accuracy, respectively. Besides the modelling work, an algorithm has been designed to identify the vessel navigational state, and a data cleaning method has been proposed to detect and remove abnormal data in the original samples. To the best of our knowledge, for the first time, LSTM is employed into the applications relating to inland vessel monitoring, and for the first time, the engine speed model is developed in a data-driven manner.

In future, the structure of the neural network may be further improved to enhance its adaptability, where a developed model will work with some intermittent or partially missing input values. The case study will be further extended for the whole Yangtze River trunk and more data will be collected and considered in analyses of vessel dynamic status, including the data of wind, wave and current.

Acknowledgments

This work was supported by the National Key Research and Development Program of China [grant number 2018YFC1407400], the National Natural Science Foundation of China (NSFC) [grant numbers 51709219 and 51609195], and the China Scholarship Council (CSC).

References:

An, S. H., Lee, B. H., & Shin, D. R. (2011, July). A survey of intelligent transportation systems. In 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks (pp. 332-337). IEEE.

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

Arguedas, V. F., Pallotta, G., & Vespe, M. (2017). Maritime Traffic Networks: From historical positioning data to unsupervised maritime traffic monitoring. IEEE Transactions on Intelligent Transportation Systems, 19(3), 722-732.

Chen, W., Yuan, J., Chen, Y., & Gan, S. (2017). Preparing the initial model for iterative deblending by median filtering. Journal of Seismic Exploration, 26(1), 25-47.

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.

Filipiak, D., Stróżyna, M., Węcel, K., & Abramowicz, W. (2018). Big Data for Anomaly Detection in Maritime Surveillance: Spatial AIS Data Analysis for Tankers. Scientific Journal of Polish Naval Academy, 215(4), 5-28.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. Neural Computation. 12 (10), 2451-2471.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Sequence Modeling: Recurrent and Recursive Nets. Deep Learning, 367-415.

Jordan, R. J., Herndon, D. C., McMorrow, J. A., Harrington, J. E., Constantine, H. E., & Linzey, M. R. (2001). U.S. Patent No. 6,249,241. Washington, DC: U.S. Patent and Trademark Office.

Kazemi, S., Abghari, S., Lavesson, N., Johnson, H., & Ryman, P. (2013). Open data for anomaly detection in maritime surveillance. Expert Systems with Applications, 40(14), 5719–5729.

Kolasa-Wiecek, A. (2015). Stepwise multiple regression method of greenhouse gas emission modeling in the energy sector in Poland. Journal of Environmental Sciences, 30, 47-54.

Kowalska, K., & Peel, L. (2012, July). Maritime anomaly detection using Gaussian process active learning. In 2012 15th International Conference on Information Fusion (pp. 1164-1171). IEEE.

Kumar, A., Chinnam, R. B., & Tseng, F. (2019). An HMM and polynomial regression based approach for remaining useful life and health state estimation of cutting tools. Computers & Industrial Engineering, 128, 1008-1014.

Last, P., Bahlke, C., Hering-Bertram, M., & Linsen, L. (2014). Comprehensive analysis of automatic identification system (AIS) data in regard to vessel movement prediction. The Journal of Navigation, 67(5), 791-809.

Laxhammar, R. (2008, June). Anomaly detection for sea surveillance. In 2008 11th international conference on information fusion (pp. 1-8). IEEE.

Li, H., Liu, J., Liu, R., Xiong, N., Wu, K., & Kim, T. H. (2017). A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis. Sensors, 17(8), 1792.

Li, H., Liu, J., Wu, K., Yang, Z., Liu, R. W., & Xiong, N. (2018). Spatio-temporal vessel trajectory clustering based on data mapping and density. IEEE Access, 6, 58939-58954.

Li, Z., Li, J., Wang, Y., & Wang, K. (2019). A deep learning approach for anomaly detection based on sae and lstm in mechanical equipment. The International Journal of Advanced Manufacturing Technology (1), 1-12.

Li, Y., Liu, R. W., Liu, J., Huang, Y., Hu, B., & Wang, K. (2016, October). Trajectory compression-guided visualization of spatio-temporal AIS vessel density. In 2016 8th International Conference on Wireless Communications & Signal Processing (WCSP) (pp. 1-5). IEEE.

Lou, D. M., Bao, S. M., Hu, Z. Y., & Tan, P. Q. (2017). Cruise speed optimization of tugboat based on real fuel consumption and emission. Journal of Traffic and Transportation Engineering, 17(01), 93-100.

Mascaro, S., Nicholso, A. E., & Korb, K. B. (2014). Anomaly detection in vessel tracks using Bayesian networks. International Journal of Approximate Reasoning, 55(1), 84-98.

Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G. B. (2018). An automatic identification system (AIS) database for maritime trajectory prediction and data mining. In Proceedings of ELM-2016 (pp. 241-257). Springer, Cham.

Nguyen, D., Vadaine, R., Hajduch, G., Garello, R., & Fablet, R. (2018, October). A Multi-task Deep Learning Architecture for Maritime Surveillance using AIS Data Streams. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 331-340). IEEE.

Notteboom, T., Yang, D., & Xu, H. (2020). Container barge network development in inland rivers: A comparison between the Yangtze River and the Rhine River. Transportation Research Part A: Policy and Practice, 132, 587-605.

Pan, J., Jiang, Q., & Shao, Z. (2014). Trajectory clustering by sampling and density. Marine Technology Society Journal, 48(6), 74-85.

Paul, B., & Jayadas, N. H. (2015). A Demand Forecast Model based on Moving Average and Markov Method. Journal of Statistics and Management Systems, 18(1-2), 139-159.

Rhodes, B. J., Bomberger, N. A., Seibert, M., & Waxman, A. M. (2005, October). Maritime situation monitoring and awareness using learning mechanisms. In 2005 IEEE Military Communications Conference (pp. 646-652). IEEE.

Ristic, B., La Scala, B., Morelande, M., & Gordon, N. (2008, June). Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In 2008 11th International Conference on Information Fusion (pp. 1-7). IEEE.

Riveiro, M., Pallotta, G., & Vespe, M. (2018). Maritime anomaly detection: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(5), e1266.

Robards, M. D., Silber, G. K., Adams, J. D., Arroyo, J., Lorenzini, D., Schwehr, K., & Amos, J. (2016). Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)—a review. Bulletin of Marine Science, 92(1), 75-103.

Sang, L. Z., Wall, A., Mao, Z., Yan, X. P., & Wang, J. (2015). A novel method for restoring the trajectory of the inland waterway vessel by using AIS data. Ocean Engineering, 110, 183-194.

Shi, H., & Wang, X. (2018, December). Research on the Development Path of Blockchain in Shipping Industry. In 2018 the Asia-Pacific Conference on Intelligent Medical & International Conference on Transportation and Traffic Engineering (pp. 243-247). ACM.

Tian, H. B., Wu, B., & Yan, X. P. (2018, April). Challenges and developments of water transport safety under intelligent environment. In Progress in Maritime Technology and Engineering: Proceedings of the 4th International Conference on Maritime Technology and Engineering (MARTECH 2018), May 7-9, 2018, Lisbon, Portugal (p. 111). CRC Press.

Will, J., Peel, L., & Claxton, C. (2011, October). Fast maritime anomaly detection using kd-tree gaussian processes. In IMA Maths in Defence Conference.

Willems, N., Van De Wetering, H., & Van Wijk, J. J. (2009, June). Visualization of vessel movements. In Computer Graphics Forum (Vol. 28, No. 3, pp. 959-966). Oxford, UK: Blackwell Publishing Ltd.

Wu, B., Yip, T. L., Yan, X., & Soares, C. G. (2019). Fuzzy logic based approach for ship-bridge collision alert system. Ocean Engineering, 187, 106152.

Wu, B., Cheng, T., Yip, T. L., & Wang, Y. (2020). Fuzzy logic based dynamic decision-making system for intelligent navigation strategy within inland traffic separation schemes. Ocean Engineering, 197, 106909.

Xie, J., Yao, G., Sun, M., Ji, Z., Li, G., & Geng, J. (2017). Ocean surface wind direction inversion using vesselborne high-frequency surface wave radar. IEEE Geoscience and Remote Sensing Letters, 14(8), 1283-1287.

Xie, Z., & Liu, Q. (2018, January). LSTM networks for vessel traffic flow prediction in inland waterway. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 418-425). IEEE.

Yan, X., Zhang, H., & Wu, C. (2012, October). Research and development of intelligent transportation systems. In 2012 11th International Symposium on Distributed Computing and Applications to Business, Engineering & Science (pp. 321-327). IEEE.

Zhang, L., Meng, Q., Xiao, Z., & Fu, X. (2018). A novel vessel trajectory reconstruction approach using AIS data. Ocean Engineering, 159, 165-174.

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.

Zissis, D., Xidias, E. K., & Lekkas, D. (2016). Real-time vessel behavior prediction. Evolving Systems, 7(1), 29-40.