Transportation Research Part D 80 (2020) 102277

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Dynamic calculation of ship exhaust emissions based on real-time AIS data

Liang Huang[a,b], Yuanqiao Wen[a,b], Yimeng Zhang[c,*], Chunhui Zhou[c], Fan Zhang[c], Tiantian Yang[c]

[a] Intelligent Transportation System Research Center, Wuhan University of Technology, Wuhan, China
[b] National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan, China
[c] School of Navigation, Wuhan University of Technology, Wuhan, China

ARTICLE INFO

Keywords:
Ship exhaust emission
Real-time calculation
Dynamic inventory
AIS data stream

ABSTRACT

The activity-based methodology is becoming an increasing way to calculate exhaust emissions from ships in a port. Existing studies make great effort to build and analyze ship emission inventory in a variety of ports by applying this method to historical ship trajectory data. This kind of static emission inventory however, cannot meet the needs of real-time ship emission monitoring. This article proposes a method of dynamic calculation of ship exhaust emissions based on real-time ship trajectory data. Firstly, real-time ship AIS messages are partitioned into continuous data blocks and go through a series of pre-processing operations, including trajectory extraction, association and interpolation. Ship activity parameters are then determined by database querying and regression analysis based on ship attributes. Subsequently, an improved activity-based methodology is employed to estimate exhaust emissions from ships in a distributed way. Based on the grid model, regional ship exhaust emissions can be statistically and dynamically calculated by the spatial allocation of all ship emissions. In a case study, a real-time monitoring platform for ship exhaust emissions in Shenzhen port is developed to demonstrate the effectiveness of the proposed method.

1. Introduction

With the explosive increase of marine traffic volume, exhaust emissions from ships are rising at a rapid pace in recent years. ship exhaust emissions have become an important pollutant source in ports, channels and sea areas of congest shipping lanes and have already imposed a negative effect on human health and ocean environment (Geng et al., 2017, Wen et al., 2017, Wan et al., 2018). To reduce ship exhaust emissions, IMO (International Maritime Organization) has set strict regulations on emissions of SOx (sulfur oxides) and NOx (nitrogen oxides). Ministry of Transportation of China also establishes Emission Control Areas (ECAs) for vessels in the waters of the Pearl River Delta, Yangtze River Delta and Bohai Bay (Beijing-Tianjin-Hebei) by the end of 2015 to control the emissions of SOx, NOx and particulate matters (Moldanová et al., 2013, Cullinane and Bergqvist, 2014). For the implementation of these regulations, it is necessary to assess and supervise real-time ship exhaust emissions.

In existing researches, many scholars (Huang et al., 2017, Huang et al., 2018, Goldsworthy and Goldsworthy, 2019, Simonsen et al., 2019) have applied activity-based method (i.e. Ship Traffic Emission Assessment Model, STEAM) to historical datasets of Automatic Identification System (AIS) for estimating ship exhaust emissions in a region or port. For each ship, the activity-based

* Corresponding author.
E-mail address: Y.Zhang-30@tudelft.nl (Y. Zhang).

https://doi.org/10.1016/j.trd.2020.102277
Received 15 August 2019; Received in revised form 13 February 2020; Accepted 13 February 2020
Available online 19 February 2020
1361-9209/

method distinguishes different ship work conditions and corresponding emission factors of ship engines by identifying ship activities, and then yields the amount of exhaust emissions from the ship at each activity. A sum-up of all ships exhaust emissions will give birth to ship emission inventory. Based on grid model, all ships exhaust emissions can be allocated into different grid cells. Aggregation of all emissions in each cell can produce a density map of pollutants to indicate spatial distribution of ship emissions (Chen et al., 2017, Johansson et al., 2017). This kind of ship emission inventory called static inventory can provide a static view of general characteristics and distribution of ship exhaust emissions in a long period and can yield deep insight into source features of ship exhaust emissions by statistical analysis. However, such static inventory cannot provide real-time ship emissions that is essential and important for ship air pollution monitoring, especially in emission control area. To solve this problem, real-time ship AIS data should be received and processed for ship emissions assessment.

Real-time AIS data stream consists of static and dynamic ship AIS data. Static AIS data contains a group of ship fundamental information like ship name, ship type, ship length, etc. Dynamic AIS data records a series of ship movement information, including spatial coordinates, speed of ground, course of ground, timestamp, etc. Compared to historical data, real-time AIS data stream possesses the properties of mixing (static & dynamic messages), compound (multiple ships) and disorder (a range of timestamps). These characteristics bring some new problems to ship emission assessment based on STEAM model. Many efforts are needed to improve data quality and extract available ship movements for emissions assessment. Moreover, ship movement may be not continually updated at each sampling period. There is a big effort needed to identify various ship activities from limited ship trajectory points for ship emissions assessment. Besides, there will be higher requirements for the efficiency of real-time AIS data processing.

This paper introduces the method of how to acquire, reorganize and process real-time AIS data for ship exhaust emissions assessment with spark streaming technology. Processing methods of online AIS data stream, i.e. trajectory extraction, association and interpolation, have been detailed and utilized to obtain continuous ship trajectory segments for emission assessment. An improved STEAM model was then applied to estimate real-time exhaust emissions from ships in a distributed way. The dynamically estimated emissions on the order of minutes cover all in-and-out port ships. A grid-based activity allocation method was adopted to obtain regional ship exhaust emissions by allocating and aggregating all individual ship exhaust emissions to the grid. An experimental system in the case of Shenzhen port was developed to demonstrate the effectiveness of the proposed method. The method can provide detailed exhaust emissions from ships on the order of minutes, hours, or days. These results are of great value for maritime manager to aware the situation and dynamic change of ship air pollution. One of the most important is that dynamic emissions data of ships can be combined with real-time observation data from air pollution sensors in the same area to explore potential target ship with illegal exhaust emissions. Air pollution sensors are responsible for measuring background concentration values of air pollutant like SOx. Sudden concentration value change of pollutants indicates that there may be high emissions from ships passing by sensors. Dynamic emission data of ships can be used to detect the suspicious object.

Fig. 1 illustrates the flowchart of how to calculate exhaust emissions from ships based on the online AIS data stream. The entire workflow is divided into three stages, including AIS trajectory online process, individual ship emission calculation and regional ship emission statistics. At the first stage, received AIS data will be classified into ship static and dynamic information. Two distinguished processing strategies are set to get corresponding ship power parameters based on static information and extract individual ship trajectories from dynamic information. After an examination of trajectory continuity, individual ship exhaust emissions will be assessed in a distribute way. Finally, regional exhaust emissions from ships will be estimated by statistical analysis.

The rest of the paper is structured as follows. Section 2 introduces related work on the calculation of ship exhaust emission based on activity-based method. The following section details the methods of processing of real-time AIS data. Section 4 estimates exhaust emissions from the ship in different activities with the improved STEAM model. Section 5 validates the effectiveness of the proposed method by the experiment in Shenzhen port. The final section summarizes and discusses the work.

2. Related work

Recently, activity-based method has gradually become the main method of ship exhaust emissions assessment while ship trajectory data is increasingly available (Liu et al., 2016, Nunes et al., 2017). The idea of the method is ship activities and durations are identified firstly from trajectory data (i.e. AIS data) and then combined with ship power parameters (i.e. emission factors and load factors) to assess ship exhaust emissions. The STEAM (Ship Traffic Emissions Assessment Model) method, one of most popular activity-based method, is firstly proposed to calculate exhaust emissions from ships in the Baltic Sea using AIS information (Jalkanen et al., 2009). In the model, the effect of waves on ship speed and ship engine power has been considered. Moreover, Jalkanen et al. (2012) further proposed STEAM2 method. It not only added the calculation of PM and CO, but also considered the effect of shipload, ship fuel and emission reduction technology. For a long period, STEAM model has been increasingly utilized to estimate ship emission

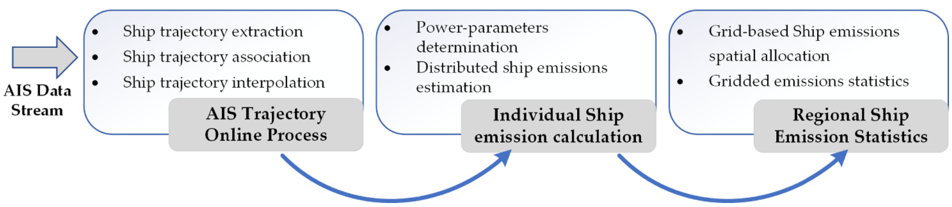


Fig. 1. Flowchart of ship emissions calculation based on online AIS data stream.

inventory in global scale (Corbett et al., 2008, Paxian et al., 2010), regional scale (Fan et al., 2016, Li et al., 2016, Chen et al., 2017), port or shipping lanes scale (Huang et al., 2017, Ng et al., 2013). The fineness and accuracy of ship emissions inventory have been continuously improved while more factors like ship handling behaviors are took account into the model (Wang et al., 2019, Wu et al., 2019).

As the establishment of global ECAs, strict regulations on fuel oil for ships are put forward to control ship exhaust emissions in the area. It is important and essential to assess real-time exhaust emissions of ships for monitoring. There are some difficult problems in applying the STEAM model directly to real-time AIS data. On one hand, assessing method in the STEAM model needs to be improved to identify ship activities from real-time AIS data. On another hand, efficient computation framework should be integrated to compatible with AIS data stream. In the traffic field, Apache Spark technology is actively used to realize steam-oriented computation and analysis on various big data. It is a fast and general engine designed for large-scale data processing and analysis (Zhu et al., 2019). The goal of Apache Spark is to make data analysis program run faster by offering a general execution model that optimizes arbitrary operator graphs and supports in-memory computation. Spark Streaming is the extension of Apache Spark, which increases the ability of online processing with interface functions like Spark, such as map, filter, reduce, and so on. Wang et al. (2016) identified traveling companions from a large-scale streaming traffic data with dynamic characteristics using the Spark Streaming framework. Gandewar and Phakatkar (2017) succeed to predict the traffic flow by applying Apache Spark to process a huge amount of traffic flow data. Peixoto et al. (2018) proposed a distributed parallel framework for efficient and scalable offline map-matching on top of the Spark framework. Gong et al. (2018) designed a real-time parallel clustering of spatio-temporal data with Apache Spark running over large-scale cloud resources. Based on the applications above, it can be concluded that the Apache Spark Streaming technology can be utilized to effectively acquire and process real-time ship AIS data stream in a distributed way.

3. Processing methods of real-time AIS data stream

3.1. Ship trajectory extraction

In existing researches, historical AIS records documenting full trajectories of ships are the main source of data for ship emissions assessing. It is relatively convenient to extract various ship activities like cruising and anchoring from historical data. In comparison, real-time AIS data stream can only provide limited trajectory sampling points for each ship at a time. Identifying ship activities become difficult. The received real-time AIS messages in sequence, with characteristics of discontinuity, mass and batching, may be static or dynamic parameters and even noisy data related to different ships. It is impossible to pre-organize and sort all received data in advance, and all processes should be completed online. Thus, distinguishing continuous trajectories for each ship online is an essential pre-process step before ship emissions estimation.

As shown in Fig. 2, the sampling period is specified as one second, and AIS data stream will be acquired in the form of continuous data blocks. In each data block, trajectory extraction will be conducted firstly for various ships. Since MMSI (Maritime Mobile Service Identify) plays an indispensable role in solely identifying a ship, AIS messages with the same MMSI in a data block can be aggregated

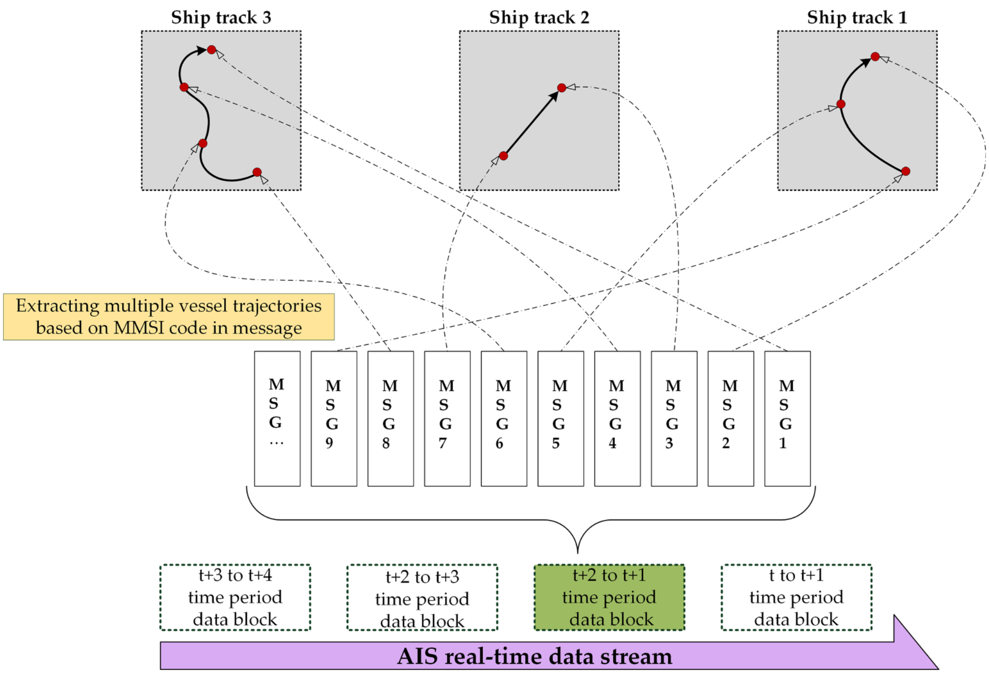


Fig. 2. Partition of AIS real-time data stream.

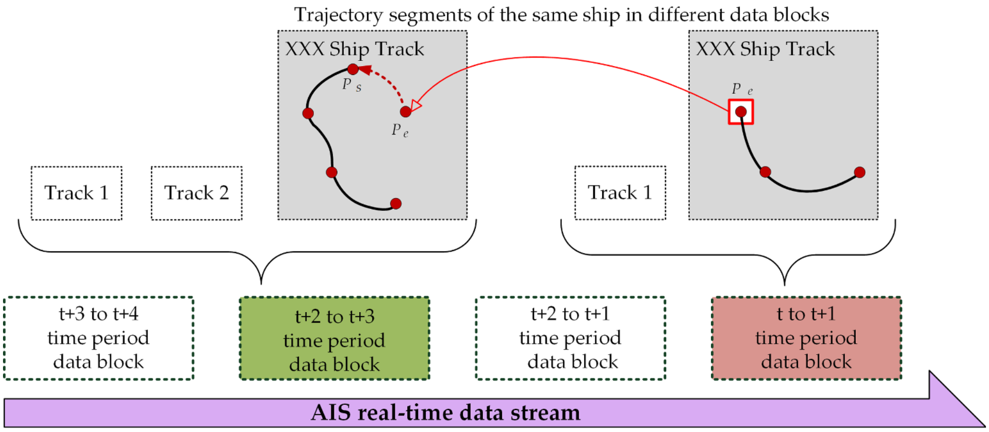


Fig. 3. Association of ship trajectories in different data blocks.

and parsed into a series of discrete trajectory points of the specific ship. All trajectory points of the same ship are sorted by their receiving timestamps, and per two time-adjacent points form a track segment of ship trajectory.

3.2. Ship trajectory association

Besides, trajectory continuity should be checked in different data blocks. In one data block, it is easy to obtain continuous ship trajectory segments. Meanwhile trajectory segments of the same ship extracted from different data blocks should also be associated. As shown in Fig. 3, two trajectory segments of XXX ship have been extracted from two discontinuous data blocks. Point $\boldsymbol{P_e}$ is the last track point of the ship identified in the previous data block and point $\boldsymbol{P_s}$ is the first sampling point of the ship received in the current data block. If only considering independent emission processing of individual data block, underestimation of exhaust emissions from XXX ship may result from the ignorance of ship movement from $\boldsymbol{P_s}$ to $\boldsymbol{P_e}$. Therefore, the latest position of each ship will be held in memory and used for trajectory association between data blocks in this paper. For current data block, the latest position of the individual ship will be inserted in front of the extracted and sorted track segment for subsequent calculation, and be updated with the last sampling point of each ship. In Fig. 3, point $\boldsymbol{P_e}$ will be denoted as the latest position of XXX ship when the calculation of data block in the previous time period has finished. Then it will be inserted into the location before point $\boldsymbol{P_e}$, which is at the top of the ship track obtained from the data block in the current time period. In this case, ship movement can be continuous for emission calculation in any one data block.

3.3. Ship trajectory segment interpolation

While completing trajectory extraction and association for an individual ship, the followed step is trajectory segment interpolation. Due to transmission errors or drop out of signal, the interval of two continuous AIS messages may vary from seconds to tens of minutes. Based on the work of Wang et al. (2015), this article applies the algorithm taking account of ship dynamic to acquire more accurate ship movement. Except for special cases, the main idea of the algorithm is to obtain interpolation point by weighted calculation of two points.

$\boldsymbol{P}$ and $\boldsymbol{Q}$ are two adjacent trajectory points of the same ship. Firstly, we can obtain two predicted values of position of interpolation point based on AIS information of endpoint $\boldsymbol{P}$ and $\boldsymbol{Q}$ with Eqs. (1) and (2).

$$\begin{cases} x_{ip} = x_p + v_p \sin\theta_p (t_i - t_p) \\ y_{ip} = y_p + v_p \cos\theta_p (t_i - t_p) \end{cases} \tag{1}$$

$$\begin{cases} x_{iq} = x_q + v_q \sin\theta_q (t_i - t_q) \\ y_{iq} = y_q + v_q \cos\theta_q (t_i - t_q) \end{cases} \tag{2}$$

where x and y are Gaussian coordinate of trajectory point of a ship at time t. v *and* θ represent speed of ground and course of ground recorded in AIS data when ship locates at the trajectory point. Then Gaussian coordinate of interpolation point position can be derived from the weighted average value of two predicted values, as shown Eq. (3).

$$\begin{cases} x_i = W_p x_{ip} + W_q x_{iq} \\ y_i = W_p y_{ip} + W_q y_{iq} \end{cases} \tag{3}$$

where W_p and W_p denote weight factors of two points respectively, which can be allocated in accordance with the time interval between interpolation point and two endpoints. The smaller the time difference, the bigger the weight factor, as shown in Eq. (4).

$$\begin{cases} W_p = 1 - (t_i - t_p)/(t_q - t_p) \\ W_q = 1 - (t_q - t_i)/(t_q - t_p) \end{cases} \tag{4}$$

In most cases, desirable interpolation results can be got directly from the algorithm with equators above. However, there are special conditions of given trajectory points should be considered in interpolation. The cases are list as follows.

1) If the speed values of both $\boldsymbol{P}$ and $\boldsymbol{Q}$ are less than 0.5 *knot*, which means the ship is at anchor or berthing, no interpolation is needed in this case.
2) If not, a ratio denoted as $\boldsymbol{R_{pq}}$ between the minimum speed and the maximum speed of two endpoints will be calculated with Eq. (5). While $\boldsymbol{R_{pq}}$ gets below a given threshold value $\boldsymbol{\delta}$, the smaller of the two speeds will be adjusted to $\boldsymbol{k}$ times of the bigger one, as shown in Eq. (6). Both $\boldsymbol{k}$ and $\boldsymbol{\delta}$ are two adjustable parameters that are mainly used to suitably balance speed values of ship at two endpoints. This is to prevent the situation that the ship is too fast at one endpoint but too slow at another one. To be specific, parameter $\boldsymbol{\delta}$ and parameter $\boldsymbol{k}$ are responsible for timing control and range control of speed regulation respectively. In the manuscript, $\boldsymbol{\delta}$ and $\boldsymbol{k}$ are set to 0.1 and 0.5. After that, two speed-adjusted points instead of raw endpoints are applied for weighted interpolation above.

$$\begin{cases} v_{min} = min(v_p, v_q) \\ v_{max} = max(v_p, v_q) \\ R_{pq} = v_{min}/v_{max} \end{cases} \tag{5}$$

$$v_{min} = k \times v_{max}, \quad if \quad R_{pq} < \delta \quad where \quad \delta \in (0, 1) \quad and \quad k \in (\delta, 1) \tag{6}$$

4. Dynamic calculation of ship exhaust emissions

In this section, details of dynamic calculation of ship exhaust emissions will be introduced. Fig. 4 shows the complete workflow that can be divided into three modules. The first module is responsible for processing static AIS message containing many ship attributes like ship MMSI and ship type. It is designed to determine ship activity parameters based on ship type. The second module is a task set for distributed calculation of ship exhaust emissions. When dynamic AIS message is received, online trajectory processing will be performed in the given calculation period. After that, real-time task of exhaust emission calculation can be created for each ship labelled by ship MMSI. All the results will then be updated to ship emission database. Based on this database, regional ship emission results within specified time can be derived by spatial allocation and statistical analysis. More details will be demonstrated in following subsections.

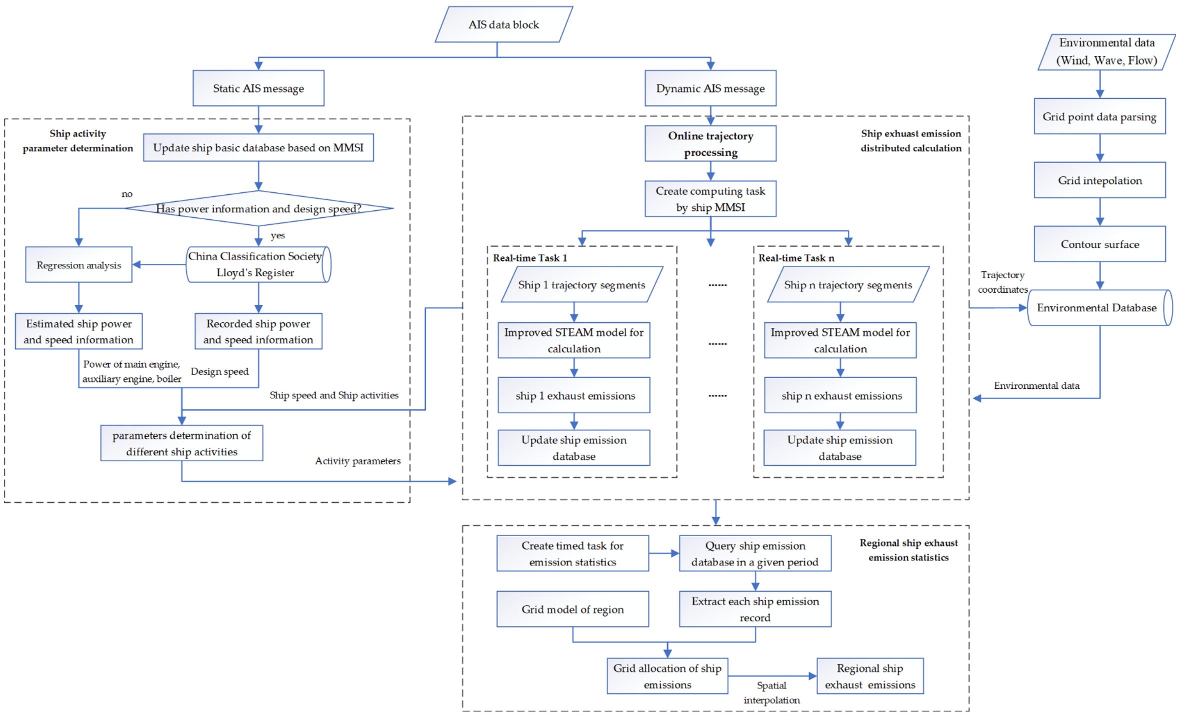


Fig. 4. Diagram of dynamic calculation of ship exhaust emissions.

In the diagram, an improved STEAM model proposed by Huang et al. (2018) is applied to estimate exhaust emissions from individual ship. Eq. (1) shows formulas of calculating exhaust emissions from ship in sailing and mooring.

$$\begin{cases} E_i = P_m \times \left(\frac{\vec{V}_{oss}}{\vec{V}_{max}}\right)^3 \times T_m \times EF_{i,m} + P_a \times LF_a \times T_a \times EF_{i,a}, & sailing \\ E_i = P_a \times LF_a \times T_a \times EF_{i,a} + P_b \times T_b \times EF_{i,b}, & mooring \end{cases} \tag{1}$$

In the Equator, subscript little i denotes the type of pollutant, i.e. *SOx*. Subscripts little *m, a* and *b* mean main engine, auxiliary engine and ship boiler respectively. Variable E represents the emission amount of pollutant. P means the power of the engine. Variables P and T are the power and working time of ship engine. Parameters *EF and LF* are the emission factor and load factor of ship engine. $\vec{V}_{max}$ is the design speed or the maximum service speed can be acquired from maritime database like Lloyd's Maritime Database. $\vec{V}_{oss}$ is the revised ship speed.

4.1. Ship activity parameters determination

Ship activity parameters are essential elements for ship exhaust emission assessment in the improve STREAM model. These parameters include the number, the category, the rated power and the designed speed of ship engines, etc. This study collects and stores fundamental data of more than 90,000 ships from China Classification Society and Lloyd's Register in the form of database. About 50,000 ships can provide valid activity parameters by ship MMSI (Maritime Mobile Service Identify) and ship type. However, the rated power and the designed speed of ship main engine, two of the key parameters however, are usually inaccessible due to transmission error or attributes missing.

There are mainly two methods to estimate ship main engine power (Zhou et al., 2019). One method obtains approximate value by referencing to main engine power of ship with similar ship type, ship size, gross tonnage (Chen et al., 2017), etc. But if there are no basic attributes of target ship, this method fails to provide the reference. Another one calculates the estimated value based on ship resistance (Sun, 2018). In this study, authors have conducted correlation coefficient analysis between main engine power and ship attributes. The result shows that the correlation between main engine power and the product of ship length and ship width is the highest, as shown in Fig. 5. As a result, regression analysis is adopted to dynamically calculate the value of unknown main engine power based on ship size.

The polynomial regression method is used to construct the regression models of main engine power for different ships. Regression model training was performed using the *sklearn* library of machine learning in Python, with 80% of the sample set as the training set and the rest as the test set. The advantage of polynomial regression is that the measured point can be approximated by adjusting the freedom degree d and increasing the high order term of the independent variable x until it is satisfied. In regression analysis, determination coefficient R^2 reflects the percentage of interpretable variation in the total variation. The greater the proportion of variation explained by the regression curve, the more reliable the regression function fits the actual situation. The range of R^2 varies from 0 to 1. The closer that this value is to 1, the better the performance of the model is. This article selects the most suitable fitting model depending on the coefficient.

The fitting results are shown in Fig. 6. It can be seen the regression is the linear regression when $d = 1$, the fitting coefficient of

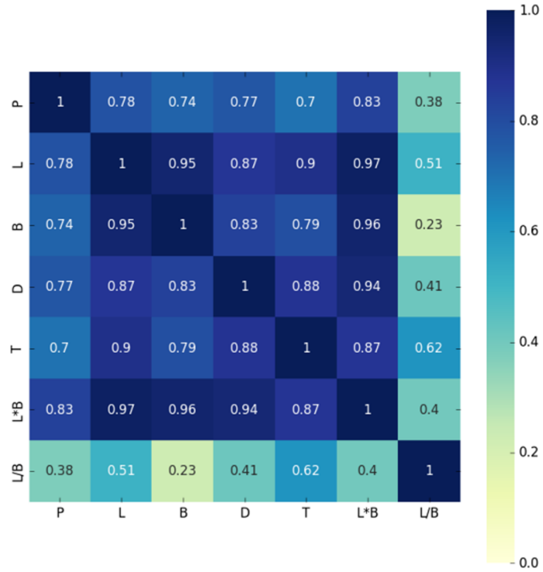
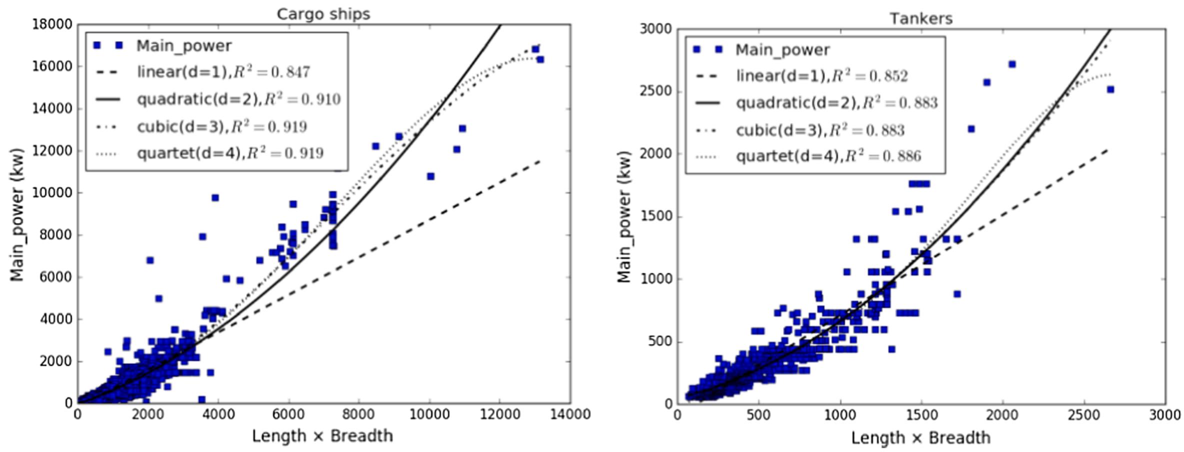


Fig. 5. Heat map of correlation coefficient for main engine power and ship attributes. (Note: P-main engine power, L-ship length, B-ship breadth, D-ship depth, T-ship tonnage).



(a) cargo ships (b) tankers

Fig. 6. Power scatter distribution and fitting curve of cargo ships and tankers.

the cargo ship is 0.847, and the fitting coefficient of the tanker is 0.836. The fitting effect is not good, indicating that the relationship between the two variables is not a simple linear relationship. When $d = 2$, the fitting coefficient of the cargo ship is 0.910, the fitting coefficient of the tanker is 0.873, and the fitting curve passed most of the measured points, so the fitting effect was better. When d increases, the coefficient of determination increases, but the growth is not obvious. For ease of application, the fitting equation with $d = 2$ is the best. The power estimation models of the cargo ship and the tanker are shown in Table 1.

When the main engine power can be determined, the unknown designed speed of ship can be also calculated. Cepowski (2019) has explored the regression relation between the designed speed and engine total power for tankers, container ships and bulk carriers. For tanker and cargo, main engine total power can be calculated by Eq. (2).

$$\begin{cases} CR = \alpha * DWT^{\beta} * V^{\gamma}, & Mtanker \\ MCR = \alpha * TEU^{\beta} * V^{\gamma}, & cargo \end{cases} \tag{2}$$

where MCR is the main engine total power [kW], DWT is deadweight capacity of tanker [t], TEU is the number of containers [–], V is the design speed [knots] and α, β, γ are the regression model coefficients.

Based on Eq. (2), the unknown designed speed of ship can be calculated backward based on determined ship engine power. Collected cargos and tankers with complete attributes are used to train and determine the regression model coefficients in Eq. (2) the trained predict model of the designed speed for tanker and cargo is shown in Eq. (3).

$$\begin{cases} V_T = \sqrt[0.6]{MCR_T/2.66DWT^{0.6}}, & tanker \\ V_C = \sqrt[0.4]{MCR_C/4.297DWT^{0.6}}, & cargo \end{cases} \tag{3}$$

4.2. Ship exhaust emissions distributed calculation

After online trajectory processing, raw AIS data block has been translated into trajectory segments of various ships. Then real-time calculation task will be assigned to each ship for dynamic emissions calculation in a distributed way. In this study, Apache Spark Streaming technology is used to implement dynamic ship emissions calculation. Spark Streaming achieves a near-real-time calculation by sequentially and quickly processing batched data blocks at specified intervals.

The processing mechanism of Spark Streaming for ship emissions is shown in Fig. 7. Firstly, trajectory segments of each ship will be split into different data blocks. Each trajectory segment stored in a data block acts as a basic data unit for dynamic emissions calculation. After all data blocks are submitted, Spark Engine will allocate individual calculation task for each data block. In the task, improved STEAM model is applied to estimate exhaust emissions of the given trajectory segment. All the tasks follow in a FIFO (First In, First Out) queue. Finally, Spark Engine can get a series of emission result data that is written directly into database.

During the calculation, some fundamental data like ship attributes and emission factors are essential for ship emission estimation

Table 1
Fitted models of cargo ships and tankers.

Type	x (Length × Breadth, m^2), y (main engine power, kW)	R^2
cargo ships	$y = 7.52 \times 10^{-5}x^2 + 0.59x - 41.48$	0.910
tankers	$y = 3.32 \times 10^{-4}x^2 + 0.27x + 57.20$	0.873

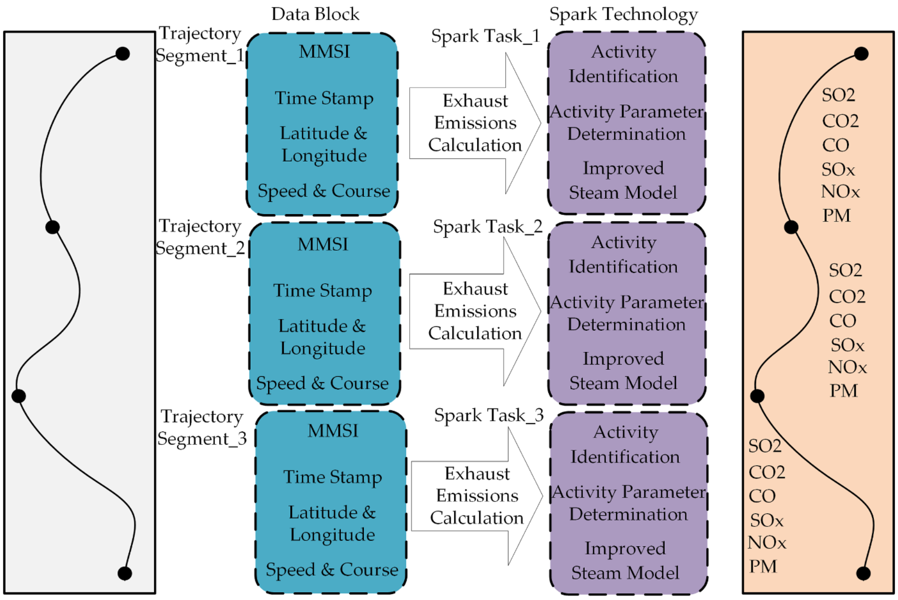


Fig. 7. Distributed calculation workflow of ship exhaust emissions.

based on STEAM model. It is time-consuming and difficult to fetch these data dynamically from the database. To meet the need of distributed calculation, memory cache like Redis memory database is integrated to provide essential information for calculation. Before the launch of real-time calculation task, these parameters will be queried and stored into memory cache in advanced. With the Redis connection pool, each computing node can be easy access to needed information.

4.3. Regional ship exhaust emissions statistics

Regional ship exhaust emissions statistics can provide a view of general characteristics and spatial distribution of ship emissions. Moreover, it is beneficial to explore hotspot regions or channels where pollutions from large number of ships are concentrated by allocating ship emissions into different grid cells or ship routes. In general, there are two popular ways to aggregating ship emissions over regions. One is dot density weighted algorithm that uses density of data dots to calculate spatial weights and allocate ship emission into grid cells (Liu et al., 2016, Li et al., 2016). Another is activity allocation method that the emissions assigned to a cell depend on the proportion of the trajectory length that lies within the cell (Goldsworthy and Goldsworthy, 2015). The emission of the cell is the weighted sum of emissions for trajectory segments that intersect the cell. Where a trajectory segments lies partially within a cell, the weighting is according to the proportion of the trajectory segment length that lies within the cell.

In this research, activity allocation method is used to distribute emissions of all trajectory segments into a grid model. Namely, the more activity in a cell the ship has, the larger emission to the cell it emits. As shown in Fig. 8, there are four continuous trajectory segments (named $T_{p\text{-}1}$, T_p, T_{p+1} and T_{p+2}) and spatial grid cells around them. Firstly, spatial intersection operation will be conducted between trajectory segments and grid cells. Each raw trajectory segment will be divided into a group of sub-segments by the intersected grid cells. Sub-segments are the intersection results of the trajectory segment and grid cells. For instance, trajectory segment T_p intersecting with two grid cells $C_{i,j}$ and $C_{i\text{-}1,j}$, is divided into two sub-segments T_{p_sub1} and T_{p_sub2}. T_{p_sub1} is the common part of

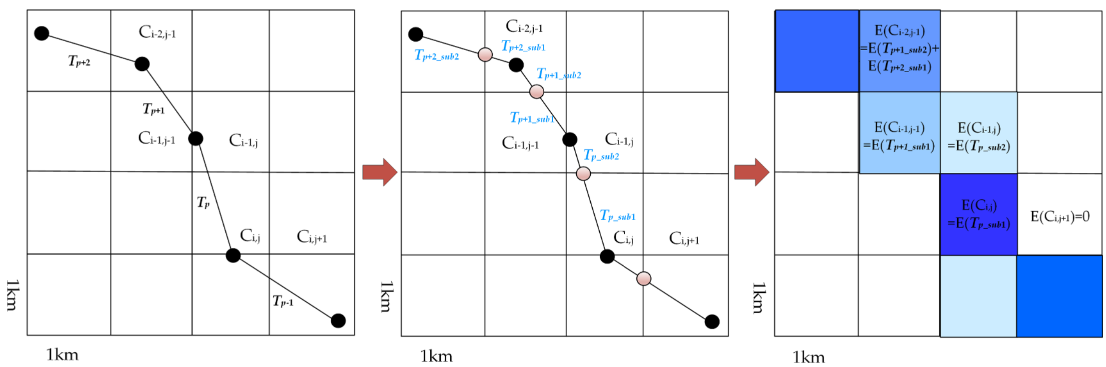


Fig. 8. Spatial allocation of trajectory segment emission to the grid.

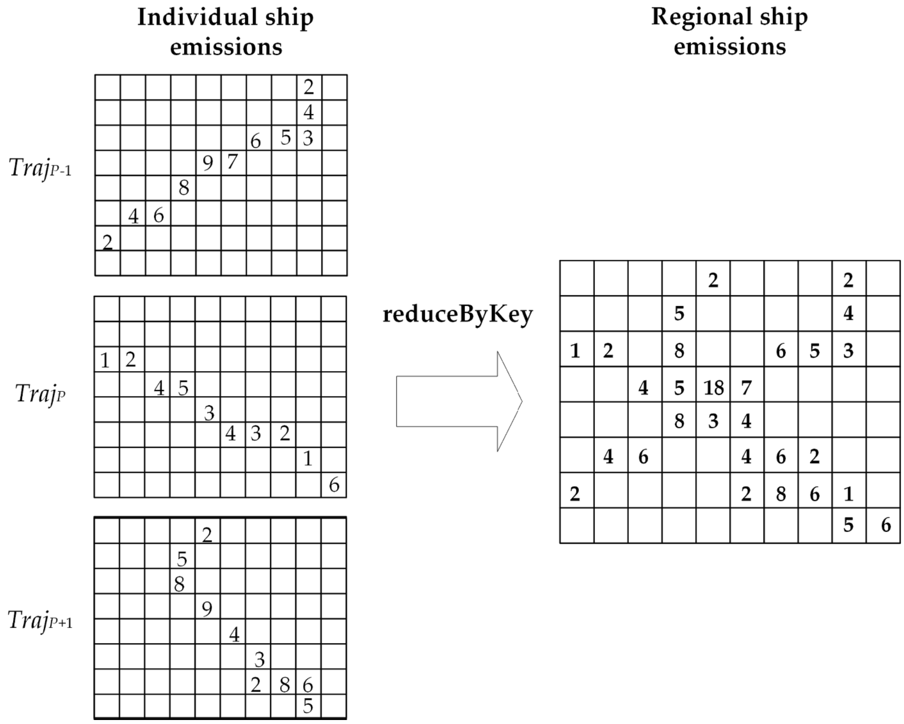


Fig. 9. Calculation of regional ship exhaust emissions.

segment T_p and cell $C_{i,j}$. T_{p_sub2} is the common part of segment T_p and cell $C_{i-1,j}$. Then, the length ratios of sub-segments to raw trajectory segment will be calculated and used as the weights of allocating ship emissions. Assuming that the total exhaust emissions of trajectory segment T_p is E(T_p), and the length ratios of T_{p_sub1} to T_p and T_{p_sub2} to T_p are α and β respectively ($\alpha + \beta = 1$), the emission of T_{p_sub1} and T_{p_sub2} can be represented as α*E(T_p) and β*E(T_p). Finally, the emission of a grid cell is calculated by the sum of emissions for sub-segments intersecting the cell. In the figure, most of cells contain only single sub-segment and the emission of the cell equals to the emission of the sub-segment it contains. But the cell $C_{i-2,j-1}$ contains two subsegments T_{p+1_sub2} and T_{p+2_sub1} that are the intersection results between it and two trajectory segment T_{p+1} and T_{p+2}. Its emission can then be assessed by the equation $E(C_{i-2,j-1}) = E(T_{p+1\_sub2}) + E(T_{p+2\_sub1})$.

By applying this method to all trajectory segments in each period, dynamic regional ship exhaust emissions can be aggregated. As shown in Fig. 9, each cell in the grid model is labelled with exclusive key consisting of row number and col number as the grid index. Exhaust emissions of individual ship will be firstly allocated into different cells based on the proposed method in distributed calculation task. The total amount of all kinds of emissions distributed to the same grid index is superimposed and merged by the *reduceByKey* method provided by Spark technology. As a result, regional ship emissions can be obtained in the form of grid.

5. Case study in Shenzhen port

5.1. Research area and data

In this section, the Shenzhen port in China is selected as a case study to demonstrate the effectiveness of the proposed method. As shown in Fig. 10(a), the experimental area is bounded to a rectangle range from 113°E to 115°E and 22°N to 23°N. The area is divided into grids at intervals of 5″ (about 140 m), as shown in Fig. 10(b). The data source is the online AIS data service provided a business company called Loongship Technology Co., Ltd. The service will push AIS static and dynamic data to the paid subscriber. In this experiment, the service will provide mixed AIS information of in-and-out ships equipped with AIS devices in the Shenzhen port. The data size of the service is about 100 records per second in the average level and can be up to 250 records per second. The response time of the service is 100 ms. Based on this service, AIS information related to nearly 5000 vessels have been collected for real-time ship exhaust emissions estimation from July 1st, 2018 to December 31st, 2018.

In addition, a prototype system has been developed to implement dynamic calculation of exhaust emissions from ships entering and leaving the Shenzhen port. The configuration of system hardware and software are shown in Table 2. In the system, Apache Kafka technology is used to access external real-time AIS messages. Apache Spark technology is integrated as a unified calculation and analytics engine for ship exhaust emissions. To support for the usage of these two technologies, a private small cluster is built with 6 high-performance computers as shown in Table 3.

Fig. 11 shows the performance of the system at the startup time and that of the system running for a period respectively. At the

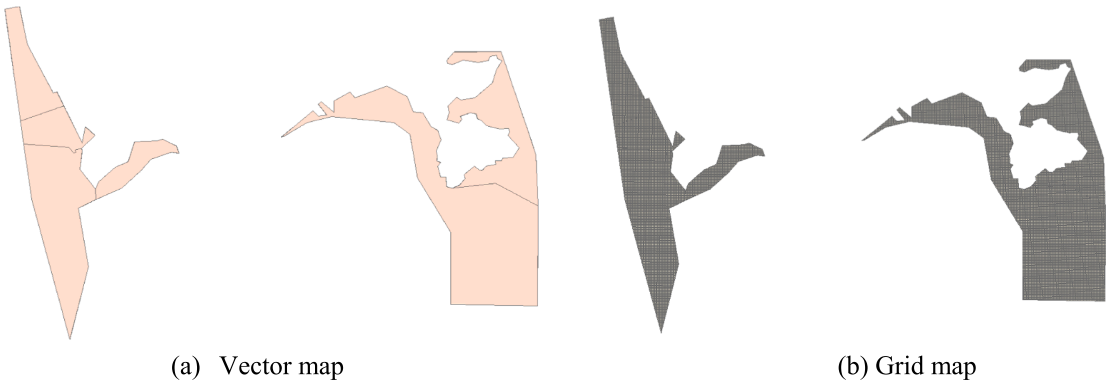


Fig. 10. Vector-based and Grid-based Shenzhen Port Sea Area.

Table 2
System configuration.

Category	Name	Description
Hardware	CPU	Intel(R) Xeon(R) CPU E5-2678 W v4, 3.00 GHz*4
	RAM	32 GB
	Disk Capacity	2.0 TB
Software	Operating System	Centos7
	Run Environment	JDK8
	Development Tool	Eclipse

Table 3
Machine configuration.

IP	System	Description	IP	System	Description
192.168.120.121	Centos7	Web server	192.168.120.124	Centos7	Spark cluster slave1
192.168.120.122	Centos7	Database server	192.168.120.125	Centos7	Spark cluster slave2
192.168.120.123	Centos7	Spark cluster master	192.168.120.126	Centos7	Spark cluster slave3

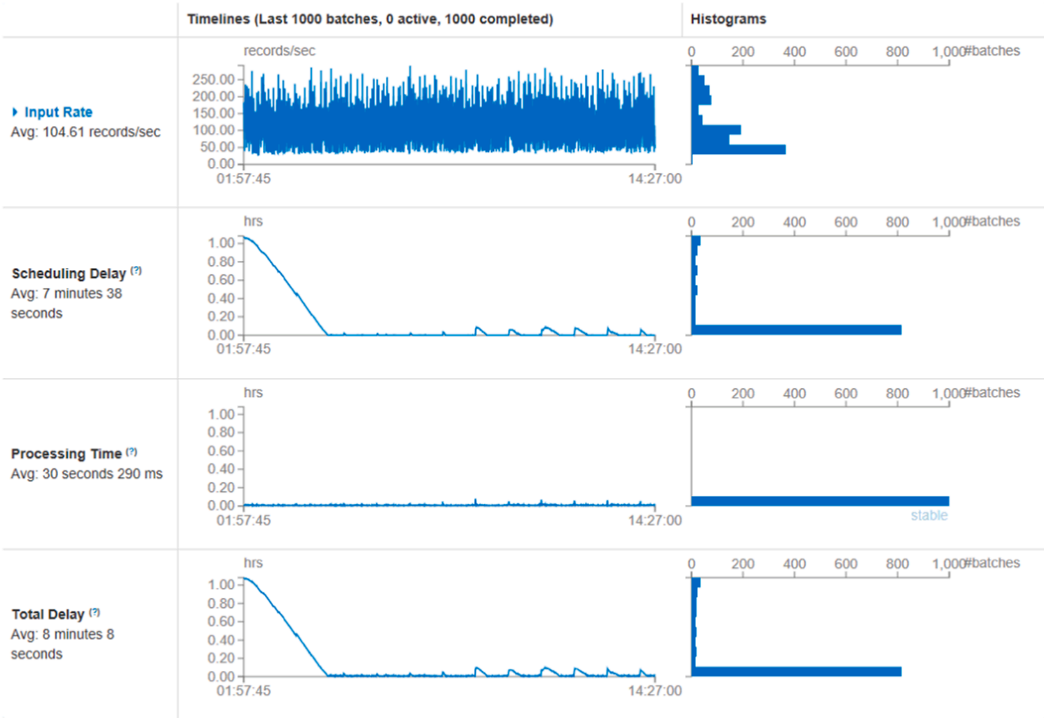
beginning, accumulated AIS data in the Apache Kafka reaches up to 300 records per second, which leads to a severe scheduling delay. While the data peak is slowly "digested" for a period, the calculation delay is gradually reduced to the second level. The inputted AIS data stream is then returned to about 100 records per second in the average level.

5.2. Individual ship exhaust emissions

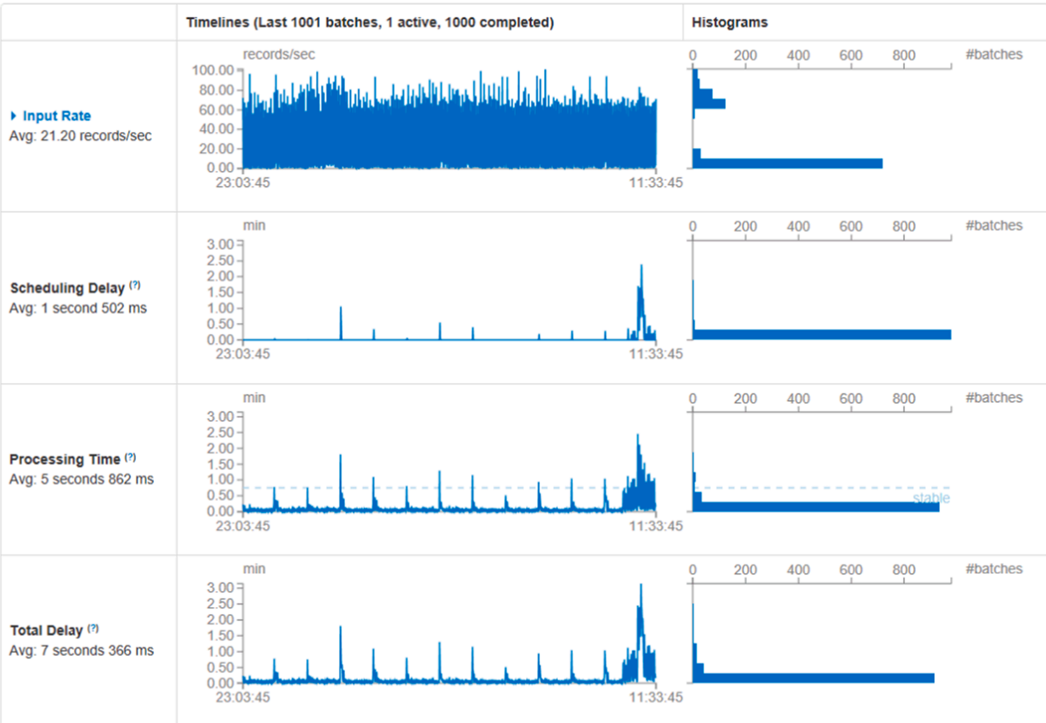
Before ship emissions assessment, real-time AIS data steam will be processed and translated into a series of trajectory segments. In this section, a cruise ship (MMSI: 477995428), a cargo ship (MMSI: 413468250) and a tanker (MMSI: 413900724) are selected to verify the effectiveness of the dynamic ship exhaust emissions measurement model proposed in this manuscript. Table 4 lists the extracted and sorted AIS data records of three ships received by Kafka from 12 pm (UNIX timestamp: 1541304000) to 12:10 pm (UNIX timestamp: 1541304600) on November 4th, 2018. We can see that cruise ship is decelerating and other two ship are sailing at Shenzhen sea area (China).

After a three-step processing, these data have been translated into a series of ship trajectory segments. In the interpolation stage, an interpolation point will be added every one minute if the time interval between two endpoints of a ship trajectory segment exceeds two minutes. Table 5 shows the interpolated ship trajectory segments. Based on these trajectory segments, several kinds of pollutants from ship exhaust were calculated in the given calculation period. Table 6 is the amounts of exhaust emissions from ships per minute. It seems that the proposed model can provide the supervision department with dynamic ship exhaust emissions and meet the requirement of on-site inspection.

In this experiment, the visualization of individual ship exhaust emissions is implemented to explore the hot-spot emission segment of a ship. Emission value has been added into raw trajectory segment as an attached attribute. Then ship trajectory is represented as spatial line vector with a grading rendering scheme from green to red (Fig. 12). The color of the line segment is associated with its emission value. The minimum and maximum values of ship emission in a period corresponding to green color and red color. When the color of a trajectory segment approaches to red color, it indicates that the segment has a large value of emissions and should be paid more attention by the manager.



(a) At the startup time



(b) At the running time

Fig. 11. System performance tests at the startup time and at the running time.

Table 4
Received AIS data stream of three ships in ten minutes.

MMSI	UNIX time	Longitude	Latitude	Course of ground	Speed of Ground	Classtype
477995428	1541304018	114.164965	22.291115	47.2	6.9	A
477995428	1541304087	114.166778	22.292568	45.8	6.5	A
477995428	1541304093	114.166778	22.292567	45.8	6.5	A
477995428	1541304124	114.167583	22.29336	38.9	5.5	A
477995428	1541304317	114.168063	22.293837	72.3	0.2	A
477995428	1541304397	114.168063	22.29382	72.3	0.1	A
477995428	1541304506	114.168067	22.293815	72.3	0.1	A
413468250	1541304019	113.710518	22.59523	142.3	8	B
413468250	1541304079	113.711947	22.593412	141.7	7.4	B
413468250	1541304379	113.719565	22.584775	124.2	7.5	B
413900724	1541304021	113.823192	22.518467	309.2	5.7	B
413900724	1541304561	113.811313	22.528825	308.1	6.5	B
413900724	1541304600	113.810053	22.529815	309.6	6.5	B

Table 5
Interpolated trajectory segments of three ships in ten minutes.

MMSI	sUNIX time	sLongitude	sLatitude	eUNIX time	eLongitude	eLatitude
477995428	1541304018	114.164965	22.291115	1541304087	114.166778	22.292568
477995428	1541304087	114.166778	22.292568	1541304093	114.166778	22.292567
477995428	1541304093	114.166778	22.292567	1541304124	114.167583	22.29336
477995428	1541304124	114.167583	22.29336	1541304184	114.167801	22.294081
477995428	1541304184	114.167801	22.294081	1541304244	114.167957	22.294285
477995428	1541304244	114.167957	22.294285	1541304304	114.168051	22.293972
477995428	1541304304	114.168051	22.293972	1541304317	114.168063	22.293837
477995428	1541304317	114.168063	22.293837	1541304397	114.168063	22.293837
477995428	1541304397	114.168063	22.293837	1541304457	114.168063	22.293837
477995428	1541304457	114.168063	22.293837	1541304506	114.168067	22.293815
413468250	1541304019	113.710518	22.595233	1541304079	113.711947	22.593412
413468250	1541304079	113.711947	22.593412	1541304139	113.713542	22.591444
413468250	1541304139	113.713542	22.591444	1541304199	113.715099	22.589599
413468250	1541304199	113.715099	22.589599	1541304259	113.716618	22.587876
413468250	1541304259	113.716618	22.587876	1541304319	113.718099	22.586276
413468250	1541304319	113.718099	22.586276	1541304379	113.719565	22.584775
413900724	1541304021	113.823197	22.518467	1541304081	113.821773	22.519795
413900724	1541304081	113.821773	22.519795	1541304141	113.820380	22.521079
413900724	1541304141	113.820380	22.521079	1541304201	113.819011	22.522319
413900724	1541304201	113.819011	22.522319	1541304261	113.817666	22.523514
413900724	1541304261	113.817666	22.523514	1541304321	113.816347	22.524665
413900724	1541304321	113.816347	22.524665	1541304381	113.815051	22.525772
413900724	1541304381	113.815051	22.525772	1541304441	113.813781	22.526834
413900724	1541304441	113.813781	22.526834	1541304501	113.812535	22.527852
413900724	1541304501	113.812535	22.527852	1541304561	113.811313	22.528825

Table 6
Calculation results of ship exhaust emissions.

MMSI&Name	Type	Size L&W	Start Time	sUNIX Time	End Time	eUNIX Time	SO_x (g)	NO_x (g)
477995428&DAY STAR	Cruise Ship	34&12	12:01:27	1541304087	12:01:33	1541304093	0.219543	2.449603
			12:01:33	1541304093	12:02:04	1541304124	1.126291	12.57250
			12:02:04	1541304124	12:03:04	1541304184	2.157303	24.09745
			……	……	……	……	……	……
			12:07:37	1541304457	12:08:26	1541304506	0.70956	7.932736
413468250&HUI HONG 828	Cargo Ship	50&16	12:00:19	1541304019	12:01:19	1541304079	7.826275	82.03436
			12:01:19	1541304079	12:02:09	1541304129	7.014665	73.47229
			12:02:09	1541304129	12:02:19	1541304139	7.069901	74.04977
			……	……	……	……	……	……
			12:05:19	1541304319	12:06:19	1541304379	7.237398	75.800868
413900724&NANPU23	Tanker	56&12	12:00:21	1541304021	12:01:21	1541304081	23.81009	249.098993
			12:01:21	1541304081	12:02:21	1541304141	24.92122	260.715369
			12:02:21	1541304141	12:03:21	1541304201	26.066739	272.691218
			……	……	……	……	……	……
			12:08:21	1541304501	12:09:21	1541304561	33.691248	352.401989

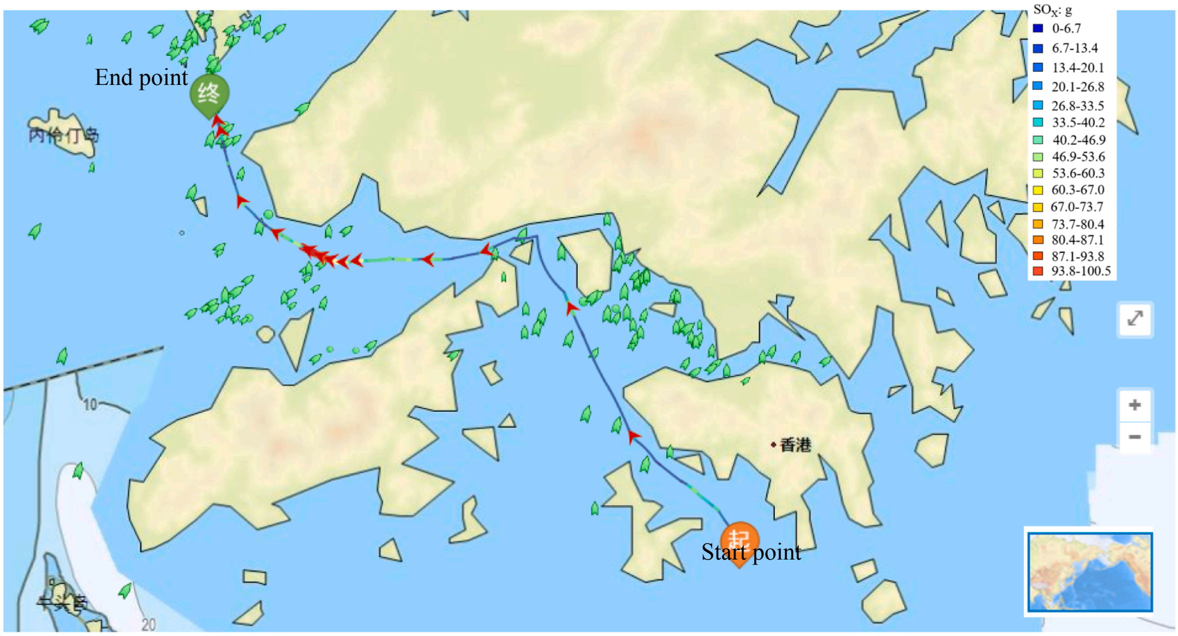


Fig. 12. Emission trajectory of individual ship.

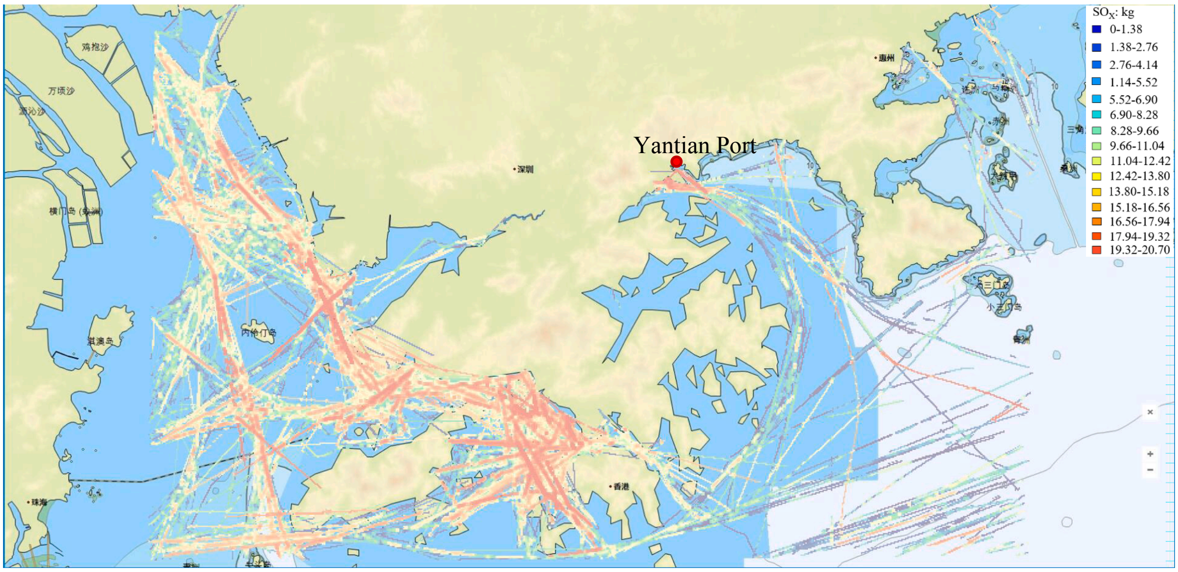


Fig. 13. Regional ship exhaust emissions.

5.3. Regional ship exhaust emissions

In the system, the heat map is utilized to represents regional ship exhaust emissions per hour, as shown in Fig. 13. Based on a 140 m * 140 m grid, exhaust emissions of all ships in the region are distributed into various cells. Moreover, a grading rendering scheme from green color to red color has also been applied to the heat map. The larger the emission of a cell is, the deeper its color is in the map. It can be viewed that most of the exhaust emissions from ships concentrated in the western waters of Shenzhen port. The emission amount of the major shipping route is explicitly larger than that of other places. These areas undertake massive ship traffic flows including international vessels and ships from the Pearl River and becomes obvious emission hotspots. In contrast, ship emissions in the ports and in-and-out ports are relatively low. The possible reason may be strict emission control regulations executed by the Chinese government. In addition, the Yantian Port seems to possess more ship emissions than other ports. It is an important and modern container port in South China, as well as an international transit deep-water port. A variety of container ships frequently enters in and leaves out the port for container loading and unloading. These ships of high ship activity parameters like main engine power contribute to relatively high exhaust emissions.

6. Conclusion and discussion

With the rapid development of marine transportation, air pollution from ships are also becoming increasingly serious. A dynamic calculation method of ship emissions based on real-time AIS data stream can be valuable and helpful to regulate ship emissions. This study proposes three-step processing methods of real-time AIS data stream. The methods are designed to solve the problems related to real-time AIS data stream like data loss and data error, and produce valid ship trajectory segments for emissions assessment. Moreover, A dynamic calculation method of ship exhaust emissions based on Spark Streaming technology is detailly illustrated. Firstly, regression analysis method is employed to estimate the main engine power of some ships, which is one of the most important activity parameters for emission assessment. Then, an improved STEAM model is applied to calculate individual ship exhaust emissions in distributed Spark tasks. Finally, regional ship exhaust emissions are calculated based on the aggregation of individual ship exhaust emissions with grid model. In the case study in Shenzhen port, system performance and machine configuration are stated. Individual and regional ship exhaust emissions based on real-time AIS data are calculated and verified. The experiment result shows that the proposed model provides technical methods for real-time ship emission monitoring.

However, there are some constraints in the proposed model. The model will be improved in the following two aspects. Firstly, the data structure in the Spark calculation engine can be further optimized. The data tables should be backed up, the main table are used for writing and the backup table for querying. Due to the continuous massive data storage, it is necessary to develop a data cleaning strategy for long-running. Secondly, real-time emission results should be certified with actual fuel consumption. In this study, authors have compared ship emissions based on online AIS data to that based on off-line data, and find that two results are the same. It however may be not a persuasive argument. The efficiency of the proposed model should be further certified. Finally, some parameters of the improved STEAM model are empirical values or reference values in other literatures. Completed and local ship activity parameters should be measured and verified.

CRediT authorship contribution statement

Liang Huang: Conceptualization, Methodology, Writing - review & editing. **Yuanqiao Wen:** Data curation, Supervision. **Yimeng Zhang:** Writing - original draft. **Chunhui Zhou:** Investigation. **Fan Zhang:** Validation. **Tiantian Yang:** Visualization.

Acknowledgments

This work was supported by the grant from the National Key R&D Program of China [Grant No. 2018YFC0213904]; the National Natural Science Foundation of China [Grant No. 41801375, Grant No. 51709218]; self-determined and innovative research funds of WUT [Grant No. 2019IVA069, Grant No. 2019III097CG]

References

Cepowski, T., 2019. Regression formulas for the estimation of engine total power for tankers, container ships and bulk carriers on the basis of cargo capacity and design speed. Polish Marit. Res. 26, 82–94.
Chen, D., Wang, X., Li, Y., Lang, J., Zhou, Y., Guo, X., Zhao, Y., 2017. High-spatiotemporal-resolution ship emission inventory of China based on AIS data in 2014. Sci. Total Environ. 609, 776–787.
Corbett, J.J., Winebrake, J.J., Green, E.H., Kasibhatla, P., Lauer, A., 2008. Mortality from ship emissions: a global assessment. Environ. Sci. Technol. 41, 8512–8518.
Cullinane, K., Bergqvist, R., 2014. Emission control areas and their impact on maritime transport. Transport. Res. Part D: Transp. Environ. 28, 1–5.
Fan, Q., Zhang, Y., Ma, W., Ma, H., Feng, J., Yu, Q., Yang, X., Ng, S.K.W., Fu, Q., Chen, L., 2016. Spatial and seasonal dynamics of ship emissions over the yangtze river delta and east china sea and their potential environmental influence. Environ. Sci. Technol. 50, 1322–1329.
Gandewar, R., Phakatkar, A., 2017. Classification approach for big data driven traffic flow prediction using apache spark. Int. Res. J. Eng. Technol. 4, 2395–12356.
Geng, X., Wen, Y., Zhou, C., Xiao, C., 2017. Establishment of the sustainable ecosystem for the regional shipping industry based on system dynamics. Sustainability 9, 742.
Goldsworthy, B., Goldsworthy, L., 2019. Assigning machinery power values for estimating ship exhaust emissions: Comparison of auxiliary power schemes. Sci. Total Environ. 657, 963–977.
Goldsworthy, L., Goldsworthy, B., 2015. Modelling of ship engine exhaust emissions in ports and extensive coastal waters based on terrestrial AIS data–An Australian case study. Environ. Modell. Software 63, 45–60.
Gong, Y., Sinnott, R.O., Rimba, P., 2018. Rt-dbscan: Real-time parallel clustering of spatio-temporal data using spark-streaming. In: International Conference on Computational Science, Wuxi, China. Springer, pp. 524–539.
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. Transport. Res. Part D: Transp. Environ. 59, 148–159.
Huang, L., Wen, Y., Geng, X., Zhou, C., Xiao, C., Zhang, F., 2017. Estimation and spatio-temporal analysis of ship exhaust emission in a port area. Ocean Eng. 140, 401–411.
Jalkanen, J.-P., Brink, A., Kalli, J., Pettersson, H., Kukkonen, J., Stipa, T., 2009. A modelling system for the exhaust emissions of marine traffic and its application in the Baltic Sea area. Atmos. Chem. Phys. 9, 9209–9223.
Jalkanen, J.-P., Johansson, L., Kukkonen, J., Brink, A., Kalli, J., Stipa, T., 2012. Extension of an assessment model of ship traffic exhaust emissions for particulate matter and carbon monoxide. Atmos. Chem. Phys. 12, 2641–2659.
Johansson, L., Jalkanen, J.P., Kukkonen, J., 2017. Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. Atmos. Environ. 167, 403–415.
Li, C., Yuan, Z., Ou, J., Fan, X., Ye, S., Xiao, T., Shi, Y., Huang, Z., Ng, S.K.W., Zhong, Z., 2016. An AIS-based high-resolution ship emission inventory and its uncertainty in Pearl River Delta region, China. Sci. Total Environ. 573, 1–10.
Liu, H., Fu, M., Jin, X., Shang, Y., Shindell, D., Faluvegi, G., Shindell, C., He, K., 2016. Health and climate impacts of ocean-going vessels in East Asia. Nat. Clim. Change 6, 1037–1041.
Moldanová, J., Fridell, E., Winnes, H., Holmin-Fridell, S., Boman, J., Jedynska, A., Tishkova, V., Demirdjian, B., Joulie, S., Bladt, H., 2013. Physical and chemical characterisation of PM emissions from two ships operating in European Emission Control Areas. Atmos. Meas. Tech. 6, 3931–3982.

Ng, S.K.W., Loh, C., Lin, C., Booth, V., Chan, J.W.M., Yip, A.C.K., Li, Y., Lau, A.K.H., 2013. Policy change driven by an AIS-assisted marine emission inventory in Hong Kong and the Pearl River Delta. Atmos. Environ. 76, 102–112.

Nunes, R.A.O., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2017. The activity-based methodology to assess ship emissions - A review. Environ. Pollut. 231, 87–103.

Paxian, A., Eyring, V., Beer, W., Sausen, R., Wright, C., 2010. Present-day and future global bottom-up ship emission inventories including polar routes. Environ. Sci. Technol. 44, 1333–1339.

Peixoto, D.A., Nguyen, H.Q.V., Zheng, B., Zhou, X., 2018. A framework for parallel map-matching at scale using Spark. Distribut. Parallel Databases 37, 1–24.

Simonsen, M., GöSSLING, S., WALNUM, H.J., 2019. Cruise ship emissions in Norwegian waters: A geographical analysis. J. Transp. Geogr. 78, 87–97.

Sun, F., 2018. Research on power forecasting method for ocean-going ship host master. Dalian Maritime University.

Wan, C., Yang, Z., Zhang, D., Yan, X., Fan, S., 2018. Resilience in transportation systems: a systematic review and future directions. Transp. Rev. 38, 479–498.

Wang, C., Ji, Y., Li, M., Chu, X., Wang, Y., 2015. An AIS track interpolation method considering the vessel's speed and course. Ship Sci. Technol. 37, 60–64.

Wang, X., Liu, C., Zhu, M., 2016. Instant traveling companion discovery based on traffic-monitoring streaming data. In: 13th Web Information Systems and Applications Conference (WISA), 23-25 Sept 2016 Wuhan, China. IEEE, pp. 89–94.

Wang, Y., Zio, E., Wei, X., Zhang, D., Wu, B., 2019. A resilience perspective on water transport systems: the case of Eastern Star. Int. J. Disaster Risk Reduct. 33, 343–354.

Wen, Y., Geng, X., Wu, L., Yip, T.L., Huang, L., Wu, D., 2017. Green routing design in short seas. Int. J. Shipping Transp. Logist. 9, 371–390.

Wu, B., Yip, T.L., Yan, X.P., Soares, C.G., 2019. Fuzzy logic based approach for ship-bridge collision alert system. Ocean Eng. 187, 12.

Zhou, C., Huang, H., Zhou, L., Sui, Z., Wen, Y., Xiao, C., 2019. Main engine power estimation method for the inland ship based on big data. J. Dalian Marit. Univ. 45, 106152.

Zhu, J.Y., Tang, B., Li, V.O., 2019. A five-layer architecture for big data processing and analytics. Int. J. Big Data Intell. 6, 38–49.