Ship Carbon Dioxide Emission Estimation in Coastal Domestic Emission Control Areas using High Spatial-temporal Resolution Data: A China Case

Haijiang Li [a], Peng Jia [a, b, *], Xinjian Wang [c], Zaili Yang [d, e], Jin Wang [d], Haibo Kuang [a]

[a] School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, PR China
[b] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, PR China
[c] Navigation College, Dalian Maritime University, Dalian 116026, PR China
[d] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, L3 3AF, UK
[e] Transport Engineering College, Dalian Maritime University, Dalian 116026, PR China

Abstract: It is necessary to accurately calculate ship carbon emissions for shipping suitability. The state-of-the-art approaches could arguably not be able to estimate ship carbon emissions accurately due to the uncertainties of Ship Technical Specification Database (STSD) and the geographical and temporal breakpoints in Automatic Identification System (AIS) data, hence requiring a new methodology to be developed to address such defects and further improve the accuracy of emission estimation. Firstly, a novel STSD iterative repair model is proposed based on the random forest algorithm by the incorporation of13 ship technical parameters. The repair model is scalable and can substantially improve the quality of STSD. Secondly, a new ship AIS trajectory segmentation algorithm based on ST-DBSCAN is developed, which effectively eliminates the impact of geographical and temporal AIS breakpoints on emission estimation. It can accurately identify the ships' berthing and anchoring trajectories and reasonably segment the trajectories. Finally, based on this proposed framework, the ship carbon dioxide emissions within the scope of domestic emission control areas (DECA) along the coast of China are estimated. The experiment results indicate that the proposed STSD repair model is highly credible due to the significant connections between ship technical parameters. In addition, the emission analysis shows that, within the scope of China's DECA, the berthing period of ships is longer owing to the joint effects of coastal operation features and the strict quarantine measures under the COVID-19 pandemic, which highlights the emissions produced by ship auxiliary engines and boilers. The carbon intensity of most coastal provinces in China is relatively high, reflecting the urgent demand for the transformation and updates of the economic development models. Based on the theoretical models and results, this study recommends a five-stage decarbonization scheme for China’s DECA to advance its decarbonization process.

Keywords: Ship emissions, Ship decarbonization, Random Forest, ST-DBSCAN, AIS

1 Introduction

More than 80% of the volume of international trade in goods is carried by sea, and the percentage is even higher for most developing countries (UNCTAD, 2021). Maritime transport serves as the backbone of international trade and the global economy, supporting the smooth flow of international logistics supply chains and fostering global economic and trade development. However, as seaborne

* Corresponding author.
E-mail addresses: hj.li.dlmu@gmail.com (H. Li), jiapeng@dlmu.edu.cn (P. Jia)

trade expands, the issue of carbon dioxide emissions caused by maritime transport becomes increasingly concerned (Gossling et al., 2021). According to the International Maritime Organization (IMO), the current global greenhouse gas (GHG) emissions of total shipping have reached 1.076 billion tons (IMO, 2021a), which ranks the fifth in the list of all economies by carbon dioxide emissions (BP Group, 2021). Without any change, they are expected to rise by 50 to 250 percent by 2050. Therefore, it is urgent to control ship carbon dioxide emissions. In 2018, the IMO adopted a long-term emissions reduction objective for maritime transport, requiring that all maritime greenhouse gas (GHG) emissions be cut to 50% by 2050, independent of the expansion of maritime commerce. Although this carbon reduction objective is set quite high, there is still fear that it will fall short of the Paris Agreement's 1.5 °C climate change target (Bullock et al., 2022). Therefore, the International Chamber of Shipping (ICS) recently called on the IMO to urge governments to take a greater immediate action to assist the global maritime transport in reaching net zero carbon emissions by 2050 (Başhan & Kökkülünk, 2020; Department for Transport, 2019; McKinlay et al., 2021).

Since 2009, China has been the largest exporter of goods in the world. Approximately 95% of China's international trade goods are carried by sea, and its seaborne imports and exports account for 30% of the global seaborne trade volume. Therefore, China is crucial to achieving the IMO 2050 emission reduction targets. In 2020, China proposed the “dual carbon” goal, that is, striving to peak CO_2 emissions before 2030 and achieving carbon neutrality before 2060 (Chen et al., 2021), to demonstrate a strong commitment to decarbonization. This policy has a profound effect on shipping since it applies to all industries in China. As a developing economy, China aims to put great efforts to achieve the “dual-carbon goal”, which motivates the studies on decarbonization of transportation in the country. To solve the problem of shipping decarbonization, the estimation of ship carbon emissions is the most important and basic work. Although the IMO releases a new version of GHG report every few years (IMO, 2000, 2009, 2014, 2021a), these reports only cover the whole global sea area and numerous technological concerns are simplified to relieve the tremendous workload of IMO report. For example, the AIS trajectories are sampled in a long interval, i.e., hourly, to expedite computation. This is feasible in emission estimation studies on a global scale. However, in regional-level carbon emission studies, higher data quality is required to ensure the accuracy of emission estimates. Therefore, effective supplements to the IMO method are needed to meet the requirements of regional-level carbon emission estimation methods. The key point limiting the accuracy of carbon emission calculation is the quality of Ship Technical Specifications Database (STSD) and Automatic Identification System (AIS) data. There will inevitably be varying degrees of outliers and noise information when collecting STSD and receiving AIS data (Li et al., 2022; Silveira et al., 2022), owing to a variety of factors such as human errors and signal transmission. Therefore, some empirical formulas and regression models are proposed to repair STSD or replace some parameters that cannot be repaired, which improves the data quality to a certain extent. Unfortunately, there is no detailed processing strategy for the compensation of AIS data shortcomings in most research on ship emissions. The volume of AIS data is massive, and the impact on carbon emissions may be disastrous if AIS data are not handled properly. In addition, the COVID-19 pandemic has raged around the world in recent years, significantly impacting on the shipping industry, but the specific impact of the pandemic on ship

emissions is still unknown. To address the aforementioned issues, this paper aims to develop a bottom-up ship carbon dioxide emission estimation framework by incorporating STSD and AIS data with high spatial-temporal resolution into machine learning models. Based on this framework, the ship carbon dioxide emissions within the Domestic Emission Control Areas (DECA) along China's coast in 2019 (without the pandemic) and 2020 (with the pandemic) are estimated and compared. This study makes new contributions to building a regional-level data-driven ship carbon dioxide emission estimation framework, mining the spatiotemporal characteristics of China's ship carbon dioxide emissions in DECA, providing data support for the formulation of ship decarbonization policies, and analyzing the impact of the COVID-19 pandemic on China's ship carbon dioxide emissions.

The rest of the paper is organized as follows. The related literature on ship carbon dioxide emission estimating models, STSD data processing and AIS trajectory processing methods is reviewed in Section 2. The methodology of this study, including the STSD repair model, the ship AIS trajectory segmentation algorithm, and the carbon dioxide emission estimation model, is presented in Section 3. The performance of the STSD repair model and the carbon dioxide emission measurement results are analyzed and discussed in depth in Section 4. Finally, conclusions and limitations are addressed in Section 5.

2 Literature Review

In the current studies, there are mainly two methods to estimate ship carbon dioxide emissions: “top-down” and “bottom-up”. The top-down methods are primarily based on macro-statistical data to estimate carbon dioxide emissions, and the most widely used methods are fuel-based approach (Endresen et al., 2007) and trade-based method (Streets et al., 2000). The fuel-based approach estimates the total emissions by multiplying the ship’s fuel consumption using the corresponding fuel emission factors (Tzannatos, 2010). The data required by the fuel-based approach is reasonably straightforward to gather, and it is possible to swiftly estimate the emissions in a wide range of sea areas. However, since this approach ignores the specifications of ships, the uncertainty of its estimated results is therefore high. The trade-based method evaluates ship fuel consumption using data such as cargo turnover and cargo type (Yang et al., 2017), and then combines the emission factors to estimate carbon dioxide emissions. Although this method can reflect the relationship between ship emissions and economic development indicators, the vast number of assumptions used leads to low confidence in the estimation results. In general, due to the coarser granularity of the raw data, the top-down methods are typically suitable for estimating carbon dioxide emissions in a wide range of sea areas and analyzing the evolution trend of carbon dioxide emissions. However, most of the current estimation results cannot reflect the realistic carbon dioxide emissions and have a high degree of uncertainty (Peng et al., 2020).

In recent years, with the advances of ship equipment and the development of the big data technology, the bottom-up methods have progressively been dominant. The bottom-up methods estimate carbon dioxide emissions based on ship activity trajectories (Peng et al., 2020), such as Ship Traffic, Energy, and Environmental Model (STEEM) (Wang et al., 2007) and Ship Traffic Exhaust Assessment Model (STEAM) (Jalkanen et al., 2009). STEEM derives shipping lanes from the

Automated Mutual-Assistance Vessel Rescue System (AMVER) and the International Comprehensive Ocean Atmosphere Data Set (ICOADS) (US EPA, 2015), and calculates route-level emissions in conjunction with ship speed, main and auxiliary engine powers, load factors and other parameters. However, AMVER and ICOADS are mainly distributed in the Americas and have limited coverage in the other regions, which constrains the applications of STEEM. The STEAM model uses AIS data in conjunction with a series of parameters from the STSD, including ship build years, main engine powers, and fuel types and so forth to calculate carbon dioxide emissions. Moreover, AIS equipment is now mandatory on all cargo ships of 300 gross tons and over, as well as all passenger ships regardless of their sizes, which covers practically all operating ships in the world. There are no water restrictions when using the STEAM method. AIS data provides a very high spatial-temporal resolution. During ship operation, the AIS device will continuously broadcast the ship's dynamic and static information at extremely short time intervals (2 to 10 seconds in sailing conditions), including the Maritime Mobile Service Identity (MMSI) number, ship name, call sign, ship location, heading, speed and UTC time. The spatial resolution can approach 0.0001 arcminutes. As a result, the STEAM model provides emission calculations with high spatial-temporal resolution. The IMO has improved the STEAM model and used it to estimate global ship carbon dioxide emissions (IMO, 2021a). Jing et al. (2021) extended the STEAM model by adding such parameters as an ice correction factor, to evaluate carbon dioxide emissions along the Arctic shipping routes. Moreno-Gutiérrez et al. (2021) proposed a SENEM model (Ship's Energy Efficiency Model), that compared with STEAM, takes into account the influence of wind, waves, currents, and other factors on the real-time power of ships.

In conclusion, the STEAM model is a very appealing method for estimating ship carbon dioxide emissions. Although the amount of AIS data is vast, computing power is no longer a major issue to plague the estimation of ship carbon dioxide emissions, thanks to the constant iteration of big data technology and distributed computing architecture. At the moment, the biggest bottleneck influencing the accurate estimation of ship carbon dioxide emission is data quality. Although the STEAM model may generate emission estimation results with high spatial-temporal resolution, the accuracy of the results is still strongly influenced by the quality of the data. Many factors, including equipment stability and signal transmission, contribute to a substantial number of missing and incorrect data points in ship archives and AIS data. The massive volume of AIS data will make a small error infinitely accumulated and amplified during the calculation procedure, which considerably increased the uncertainty of the results. Therefore, high-quality data processing and repairing work are critical for accurate calculation of ship carbon dioxide emissions using the STEAM model.

STSD contains a variety of ship registration information, including the IMO number, MMSI number, ship name, call sign, ship type, length, width, deadweight, main and auxiliary engine parameters, and fuel type, all of which are critical for calculating emissions using the STEAM model. However, there are numerous flaws in STSD. Taking the STSD in this study as an example, only about 20% of the ships have fully canonical technical specifications. To overcome this limitation, Peng et al. (2020) adopted a stratified random sampling method to take the ship samples with complete technical specifications, and then extrapolated the total emissions using data such as population, skipping the procedure of STSD repair. Schwarzkopf et al. (2021) estimated the gross tonnage, main engine power,

and main engine RPM using a series of empirical formulas. Weng et al. (2020) calculated the main engine power of a ship using a nonlinear regression approach (D. Chen et al., 2017) and parameter matching of similar size ships. Huang et al. (2020) built a main engine power fitting function of oil tankers and cargo ships based on a polynomial regression method and calculated the main engine power according to DWT/TEU and the design speed. Johansson et al. (2017) proposed a concept of most similar vessel (MSV), and the missing information of the investigated ship can be replaced by the corresponding information of its MSV. Gan et al. (2022) established several fitting curves between ship technical specifications, such as fitting the gross tonnage according to its length, inferring the auxiliary engine power from the main engine power, and so on. Overall, these methods put all the STEAM-based carbon dioxide emissions estimation closer to reality. However, the factors considered in the data repair process are yet comprehensive no matter in the regression-based method or the MSV-based methods mentioned above, leaving many key influential factors unaddressed. In general, only ship type, design speed, length, and width are taken into account in the existing literature. It needs new experiments to test and improve the data repair effect when the model is supplemented with more factors such as flag state, draught and build year.

AIS data primarily records the real-time ship trajectory information, such as ship location, course, heading, speed, and UTC time. Similarly, there are abnormalities in AIS data, such as abnormal speed (Schwarzkopf et al., 2021), course, and ship location (Peng et al., 2020). These obvious abnormalities are generally simple to be corrected. For example, the latitude should be in the range [-90,90], and the heading and course should be in the range [0,360] (Peng et al., 2020). Beyond that, it is complex to detect and tackle the missing or sparse phenomenon of ship AIS trajectory points. Interpolation computation techniques are now routinely employed for the solutions to this sort of problems (Schwarzkopf et al., 2021; Weng et al., 2020). Since the trajectory interpolation computation will alter the distribution of the original trajectories, it is encouraged to investigate the uncertain impact on ship emission estimation first. To this end, Aulinger et al. (2016) proposed a method for extracting the routes from the AIS trajectories to correct the interpolated ship locations, which provided theoretical inspiration for future study.

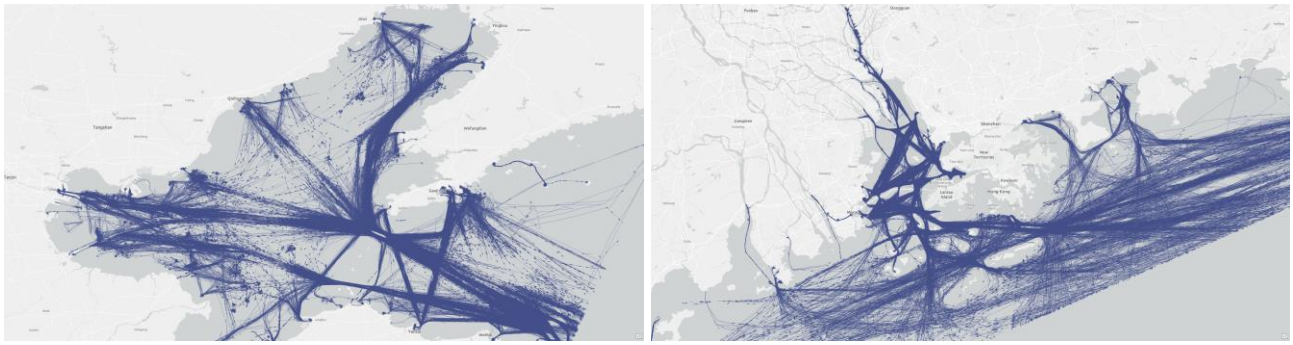
To conclude, estimating carbon dioxide emissions based on big AIS data has become attractive, and the STEAM model and its serial improved models provide a viable solution to estimating ship carbon dioxide emissions with high spatial-temporal resolution. However, the noise and defects of the raw data add much uncertainty to the carbon dioxide emission calculation. Prior research also reveals some problems in their applications when confronting the challenges such as various types of defective data filling problems in STSD and AIS trajectory processing scheme.

To address these outstanding challenges, this paper makes three new contributions, including 1) a pioneering iterative STSD repair model based on the random forest algorithm, which incorporates numerous ship technical specifications and features to predict and repair flaws to improve the STSD quality, 2) the application of a novel DBSCAN-based trajectory segmentation model to minimize the effect of geographical and temporal breakpoints on emission calculation results, and 3) An analysis of the case study on ship carbon dioxide emissions in China's DECA. The emission results reveal the distribution characteristics of ships' emissions (CO_2) along the coast of China against the main ship

types, engine types, size categories and age ranges that produce emissions, and the differences in the navigation status of ships caused by the COVID-19 pandemic. Finally, a five-stage decarbonization scheme for China's DECA is recommended.

3 Methodology

The bottom-up methods for ship carbon dioxide emission are big data driven. The bid data used in this paper is defined as follows. The time horizon for estimating ship carbon dioxide emissions in this study is 2019 and 2020. The STSD in this study comes from IHS Markit, the designated official agency by IMO, and was updated in July 2021. STSD provides the technical specifications of nearly 147,000 registered ships globally, which is consistent with the literature (Schwarzkopf et al., 2021). AIS data, including static and dynamic information, is retrieved via the MMSI number. The static information includes limited basic technical specifications such as the ship's MMSI number, IMO number, ship type, call sign, length and width. A total of 3.655 million pieces of static information were obtained in this study. The overall volume of dynamic data is approximately 10 TB. The trajectories of ships along the Chinese coast in January 2020 were plotted using ArcGIS Pro, as shown in Fig. 1. Due to space constraints, Fig. 1 only displays the ship trajectories of the Bohai Bay and the Pearl River Delta.



(a) Bohai Bay (b) Pearl River Delta

Fig. 1. The visualization of AIS dynamic data.

Relying on big maritime data resources, a new bottom-up ship emission estimation framework is established, as shown in Fig. 2.

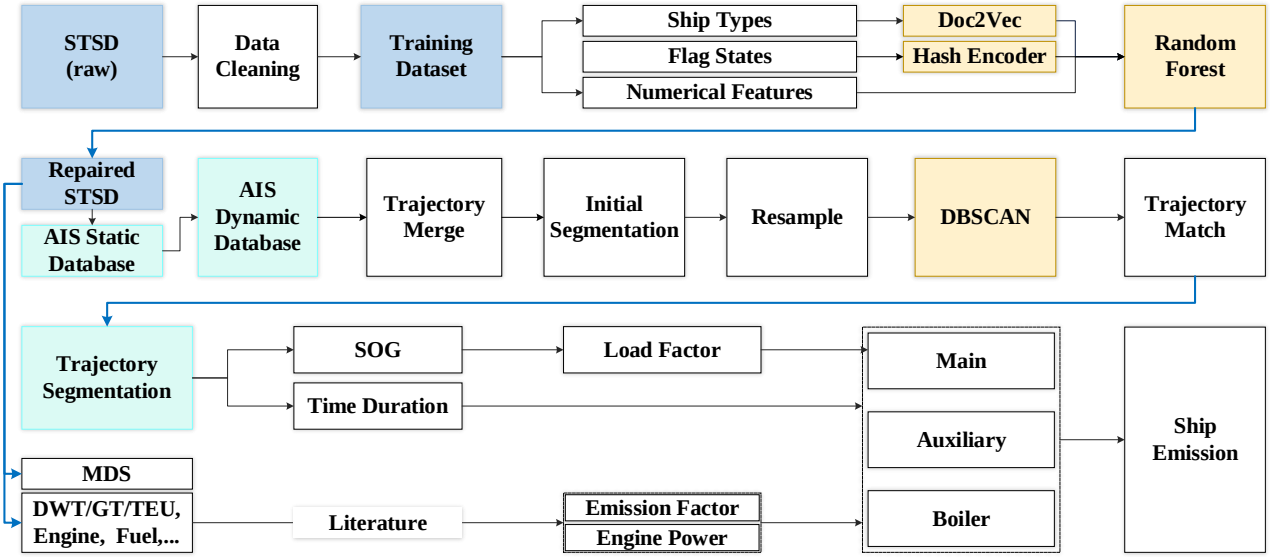


Fig. 2. Ship carbon dioxide emission estimation framework.

The framework consists of three major components: STSD repair, AIS trajectory processing, and carbon dioxide emission estimation. During the STSD repair stage, a random forest based technical specifications data repair model is developed, which can effectively improve the STSD quality. A DBSCAN-based ship trajectory segmentation algorithm is applied in the AIS trajectory processing to reduce the uncertainty of emission estimation. In the stage of emission estimation, the STEAM model is introduced to systematically calculate the emissions in China's coastal areas. Furthermore, the MMSI code of a ship may change for various reasons, and the ship trajectories of two different MMSI codes may belong to the same ship. Therefore, we merged the ship trajectories based on the IMO code using STSD and AIS static data.

3.1 Repair of STSD

There are defective data points and missing fields in the raw STSD (Johansson et al., 2017). In the STSD used in this study, the build year field is null for 3.0% of the ships, and another 3.1% of the ships in service were built after the date of the data update, which is obviously incorrect. The overall length of 7.4% of the ships is 0. Therefore, to address the poor quality in STSD data, a repair model based on a random forest algorithm is developed in this section.

3.1.1 Feature Processing

Based on the data cleaning techniques, the normative samples that can be used as training and validation sets according to the respective attributes of the fields are first selected in this study, and then a ship technical specification (STS) set is constructed. The STS set includes 13 features related to estimate carbon dioxide emissions: ship build year, ship type, length overall, length between perpendiculars, extreme breadth, moulded breadth, moulded draft, maximum speed, service speed, dead weight, TEU (Twenty Feet Equivalent Units) capacity, gross tonnage, and flag state. These features are grouped into two categories: numerical features and character features. Numerical features can be directly fed into tree-based machine learning models such as random forest for training. However, character features such as ship type and flag state cannot be directly fed into the models, and

they need to be processed as numerical vectors. One-hot encoding is a popular categorical feature encoding method (Rodríguez et al., 2018; Yu et al., 2019), which is widely used due to its simplicity. However, one-hot encoding has exposed some limitations in its previous applications. For example, when the feature is a very high dimensional data-type, it is easy to cause the curse of dimensionality. In addition, because the vectors obtained by one-hot encoding are orthogonal to each other, it is not sufficient enough to reveal the semantic relationship between different phrases (Wang & Chung, 2020).

To address these limitations, the word2vec model (Mikolov et al., 2013) developed by Google in 2013 was introduced to encode the ship type feature. Word2vec can consider the context of the current words during encoding, and it can avoid the curse of dimensionality when dealing with high-dimensional data types. A ship type is usually made up of several words, and its structure is more similar to a short sentence. Therefore, ship types were treated as short sentences and trained using the Doc2Vec module in gensim library. The Doc2Vec module is an extension of word2vec that can directly process short sentences and generate sentence vectors (Kim & Koo, 2017). Since only 266 ship types are involved in the STSD of this study, the number of categories is relatively small in comparison to text processing tasks in the areas of NLP (Natural Language Processing). Therefore, only the CBOW (Continuous Bag of Words) model (Mikolov et al., 2013) was employed to generate the vector of ship type. Table 1 shows the five ship types that have the greatest degree of cosine similarity (Jayakodi et al., 2016) to "Bulk Carrier". It is clear that the five results are substantially associated with "Bulk Carrier".

Table 1
Similar ship types to Bulk Carrier.

Ship Types	Similarity
Bulk Carrier Self-discharging	0.8734
Bulk Carrier Laker Only	0.8613
Bulk Sulphuric Acid Carrier	0.8371
Bulk Oil Carrier OBO	0.8273
Bulk Carrier Self-discharging Laker	0.8244

Although flag state and ship type are both categorical features, there is no semantic relationship between different flag states, hence the word2vec model is unnecessary for flag state encoding. The ships in the STSD are distributed among 220 flag states. Therefore, to avoid dimension explosion, the flag states were encoded using a Hash encoder (Weinberger et al., 2009). Feature hashing can compress high-dimensional feature vectors into lower-dimensional feature vectors while preserving the expressiveness of features to the maximum degree possible.

3.1.2 Random Forest

As the defects in technical specification data vary from ship to ship, the STSD data was first categorized based on the defects and then repaired iteratively for each category, as shown in Fig. 3. Using the feature of the current defects as the target, the STS set is separated into the feature set and the target set, and the training data set for the random forest algorithm is generated. The random forest

model for the current defect is trained using the training data set to repair the STSD. The loop then iterates until all defects are repaired.

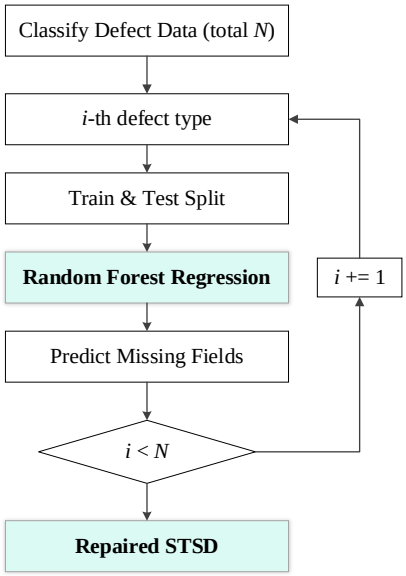


Fig. 3. Iterative repair flow of ship technical specification data based on random forest algorithm.

A Random Forest (RF) algorithm is a typical ensemble learning algorithm, which usually has better performance than traditional machine learning models. The CART (Classification and Regression Trees) (Lewis, 2000) is used as the base learner in RF, and the prediction outcomes of several decision trees are averaged using Bagging (Bootstrap aggregating) (Breiman, 1996) to obtain the final results, as illustrated in Fig. 4.

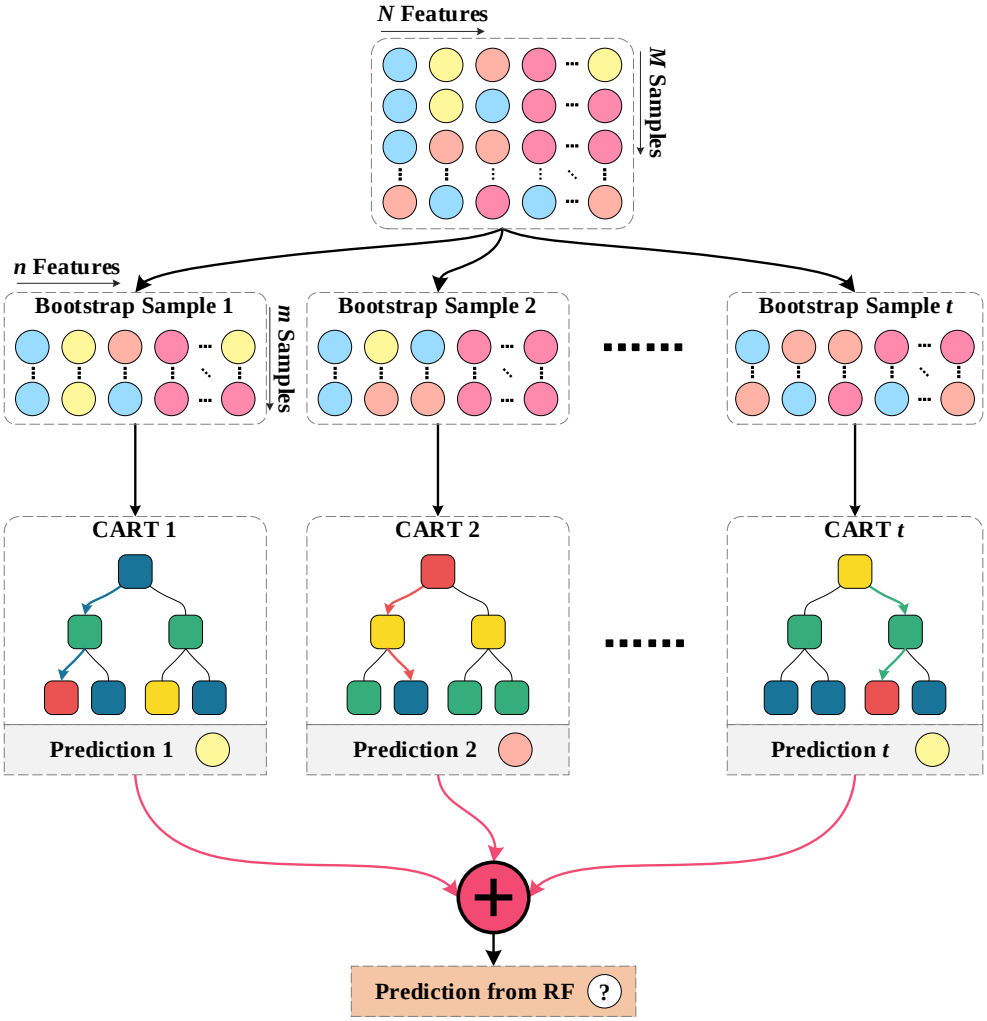


Fig. 4. Flow chart of random forest algorithm.

In Fig. 4, an $M \times N$ example dataset is illustrated, where each row represents a sample, each column denotes a feature. The various colors in the figure indicate different values and the use of them is to distinguish the nodes of different values. For a single CART t (the structure represented by the rounded rectangle in Fig. 4), the Bagging randomly takes a sample of sub-dataset $D_{m,t}$ (containing m ships, $m < M$) with replacement from the total training set D (containing M ships) as the CART training set using a bootstrap sampling method, such as the bootstrap sample t in Fig. 4. After the bootstrap sampling, about 63.2% of the ships in the training set D appear in the sample set, and the remaining 36.8% of "out-of-bag" ships in D may be utilized as the validation set of the current CART to improve the model's generalization ability. In terms of feature selection, for each node of CART, a subset containing n features is randomly chosen from the N input features, and then an optimal feature is selected from this subset for splitting. A random forest can be represented as an ensemble of multiple CARTs, as shown in Eq. (1).

$$\left\{\phi_{D_{m,t}} \middle| t = 1, 2, ..., T\right\} \tag{1}$$

where, t is the serial number of the CART, and the random forest is represented by set ϕ. As shown in Fig. 4, in the training phase, each CART produces a prediction result. Finally, the random forest will

average the results of each CART, as shown in Eq. (2).

$$\hat{\mathbf{Y}} = \frac{1}{T}\sum_{i=1}^{T}\phi_{D_{m,t}}(\mathbf{x}) \tag{2}$$

where, $\mathbf{x}$ is the input vector, $\hat{\mathbf{Y}}$ is the final predicted value.

3.2 Ship Trajectory Segmentation

AIS data will inevitably contain a significant amount of noise due to factors such as device stability and signal transmission, which will have a great impact on the accuracy of emission estimation. According to the requirements of international conventions, except under exceptional circumstances, the broadcast of ship AIS information should be continuous (Bole et al., 2013), yet there are many geographical and temporal breakpoints in the actually collected AIS data, as shown in Fig. 5.

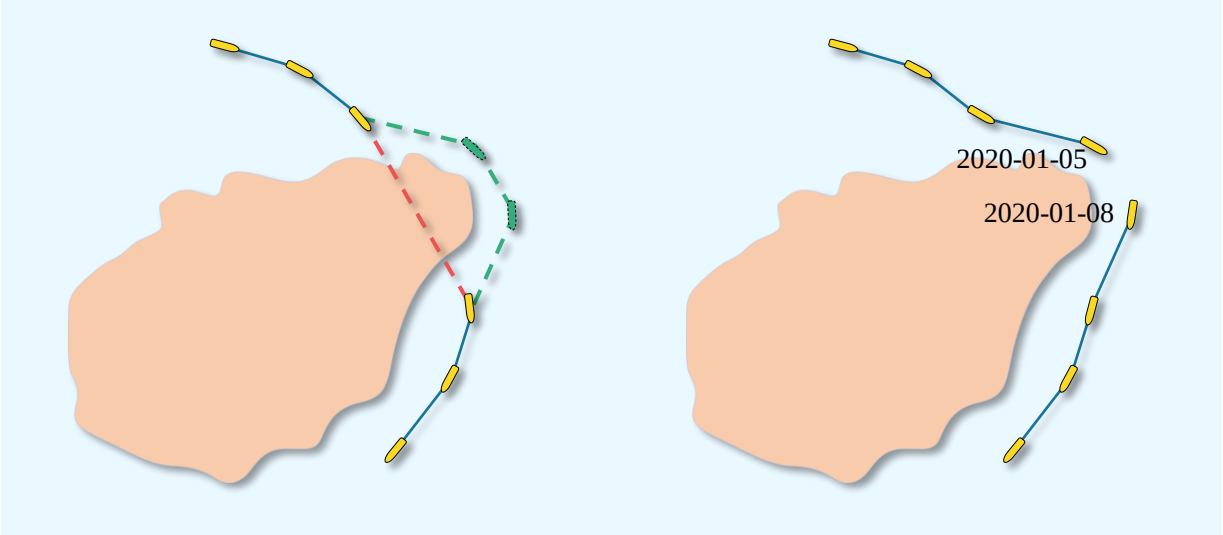


Fig. 5. Diagram of the geographical (left) and temporal (right) breakpoints of the AIS trajectory.

The left image in Fig. 5 shows the geographical breakpoints of the AIS trajectory. If the trajectory points are closely connected in time sequence, the trajectory will pass across the land (red dotted line), and the actual sailing trajectory may be the trajectory shown by the green dotted line, which will undoubtedly affect the emission calculation. The right image shows the temporal breakpoint of the AIS trajectory, that is, the time interval between two geographically close ship positions is unusually large. If it is not corrected, the emission estimation for this time period may be exaggerated. Due to the massive volume of AIS data, the ubiquitous existence of geographical and temporal breakpoints adds significant uncertainty to emission estimation.

Therefore, in order to reduce the uncertainty, a ship trajectory segmentation algorithm was applied in this study based on the spatiotemporal DBSCAN (Density-Based Spatial Clustering of Applications with Noise) model (Birant & Kut, 2007). DBSCAN is a very classic spatial density clustering algorithm, which can accurately identify high-density clusters based on the concept of density reachability (Ester et al., 1996). Various improved DBSCAN models have been developed to cope with clustering issues in numerous scientific domains and to address the limitations of DBSCAN (Bushra

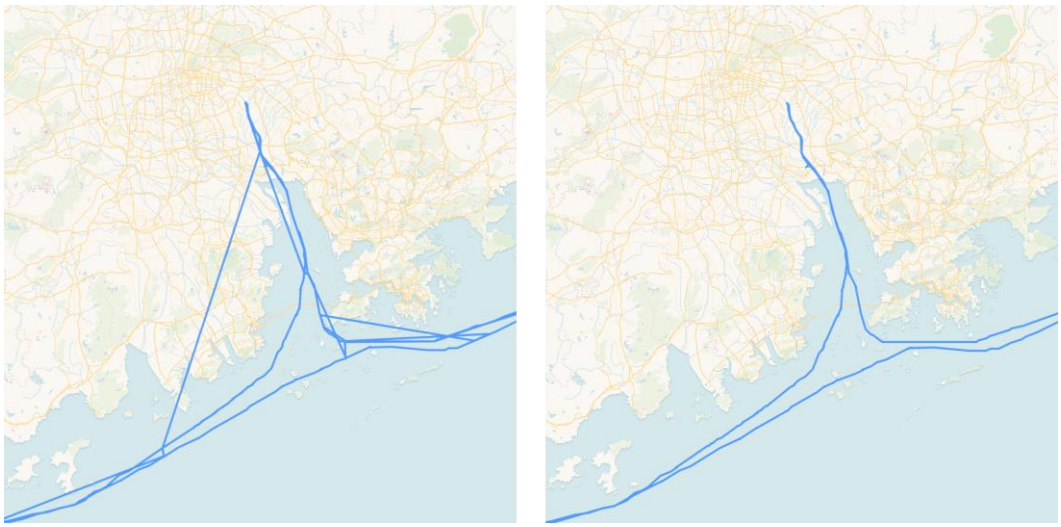
& Yi, 2021; Li et al., 2022), s uch as HDBSCAN (Hierarchical DBSCAN) (Campello et al., 2013), ST-DBSCAN (Spatial-temporal DBSCAN) (Birant & Kut, 2007) and ADBSCAN (Adaptive DBSCAN) (Khan et al., 2018). Considering the spatial-temporal characteristics and massive volume of AIS trajectory, ST-DBSCAN was adopted in this study. In the clustering process, ST-DBSCAN can mine the spatiotemporal features of ship trajectories by introducing the time thresholds, and its clustering effect is better than the DBSCAN based solely on the geographic distance. Moreover, the time threshold of ST-DBSCAN can limit the neighbor searching to a specified time range, which greatly reduces the search range and improves the efficiency. The ship trajectory segmentation algorithm has four steps:

Step 1: Initial segmentation of the trajectory. The ship trajectory is initially segmented based on the given temporal threshold η_t and geographical threshold η_d, as indicated in Table 2. The initial segmentation results are shown in Fig. 6. The left image is the original trajectory. It is evident that the trajectory crosses the land, which is inconsistent with the motion of ship and will also have an impact on emission estimation. The right image shows the trajectory after the initial segmentation, which has been significantly improved. The longest information broadcast time interval, according to the principle of AIS, is 6 minutes. Considering factors such as equipment acceptance delay, η_t is taken as 10 minutes in this study. Furthermore, based on the average speed of the world fleet (Martin, 2021), the mileage of the ship is about $2\ n\ mile$ within the time threshold of 10 minutes, hence, $\eta_d = 2\ n\ mile$.

Table 2
Initial segmentation of the trajectory.

Input: trajectory $\boldsymbol{T} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{v}, \boldsymbol{\theta}\}$
Output: trajectory $\boldsymbol{T} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{tid}\}$

Function $\boldsymbol{TrajectoryInitialSegmentation}(\boldsymbol{T})$:
    add feature $\boldsymbol{tid}$ to $\boldsymbol{T}$
    int $curr\_id = 0$, array $\Delta\boldsymbol{t}, \boldsymbol{d} =$ zeros_like($\boldsymbol{T}$.length, 1)
    for each p_i in $\boldsymbol{T}$: // $p_i = \{t_i, lng_i, lat_i, v_i, \theta_i, tid_i\}$
        $\Delta t_i = t_i - t_{i-1}$
        $d_i = GeodesicDistance((lng_i, lat_i), (lng_{i-1}, lat_{i-1}))$
        if $\Delta t_i \geq \eta_t \mid d_i \geq \eta_d$:
            $tid_i = curr\_id$
            $curr\_id \mathrel{+}= 1$



(a) Original trajectory (b) Segmented trajectory

Fig. 6. Initial segmentation of ship trajectory.

Step 2: Trajectory resampling. Because of the signal transmission, the distribution of ship locations on the initial trajectory is somewhat diffused. Ship locations in the same anchorage or harbor may be divided into different clusters due to longer geographic distances or time intervals. Considering the above problems, before clustering, a ship trajectory sampling algorithm was adopted based on a cubic spline function to sample the sub-trajectories after initial segmentation, as shown in Table 3. The sampled ship locations have equal time intervals, and there will be obvious aggregation phenomenon in the low-velocity area.

Table 3
Trajectory resampling.

Input: trajectory $\boldsymbol{T} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{tid}\}$	
Output: resampled trajectory $\boldsymbol{TS} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{tid}\}$	
01:	**Function** $\boldsymbol{ResampleTrajectory}(\boldsymbol{T})$
02:	define sampled trajectory: $\boldsymbol{TS} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{tid}\}$
03:	group trajectory $\boldsymbol{T}$ by feature $\boldsymbol{tid}$
04:	**for** each sub-trajectory group $\boldsymbol{Tg}$ in $\boldsymbol{T}$: // $\boldsymbol{T}_g = \{\boldsymbol{t}_g, \boldsymbol{lng}_g, \boldsymbol{lat}_g, \boldsymbol{tid}_g\}$
05:	$func = interpolate(\boldsymbol{t}_g, [\boldsymbol{lng}_g, \boldsymbol{lat}_g])$ // resample function
06:	$\boldsymbol{t}_g^{ts} = range(t_g[0], t_g[-1], \Delta t)$ // Δt sampling interval
07:	$\boldsymbol{lng}_g^{ts}, \boldsymbol{lat}_g^{ts} = func(\boldsymbol{t}_g^{ts})$
08:	$\boldsymbol{tid}_g^{ts} = \boldsymbol{tid}_g$

Step 3: DBSCAN-based ship location clustering. The resampled ship trajectories can be clustered using DBSCAN, as shown in Table 4. Since ship trajectories are typical time series data, a temporal threshold was introduced to constrain the scope of the neighborhood search. Compared with the geographically-based nearest neighbor search, the method additionally takes into account the speed features. The temporal threshold η_t^c and geographical threshold η_d^c of the neighborhood search in this study were defined depending on the sampling time interval in Step 2 and the ship's average speed.

According to the IMO (IMO, 2021a), anchoring or berthing usually occurs when the ship's speed is less than 3 knots. Therefore, Eqs. (3) and (4) were adopted to determine the value of thresholds:

$$\eta_t^c = \Delta t \cdot \left(N_{\text{min\_pts}} + \varepsilon_t \right) \tag{3}$$

$$\eta_d^c = v \cdot \Delta t \cdot \varepsilon_d \tag{4}$$

where, Δt is the sampling time interval in Step 2. $v = 3$ kn denotes the speed of the ship. $N_{\text{min\_pts}}$ is the minimum number of neighbors within "η_d^c" and "η_t^c" radius, which equals 5 here. ε_t is the temporal relaxation factor, usually taking 1 or 2. ε_d is the geographical scaling factor, which is used to fine tune the geographical neighborhood search scope, and it should usually not be greater than 1, and taken 0.5 in this study.

Table 4
DBSCAN-based ship location clustering.

Input: resampled trajectory $\boldsymbol{TS} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{tid}\}$
Output: clustered sampled trajectory $\boldsymbol{TS} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{c}, \boldsymbol{tid}\}$

```
Function TrainDBSCAN(TS)
    add feature c to TS and assign to -1
    define c_temp = 0
    for each p_i in TS:
        if p_i is unvisited:
            mark p_i as visited
            neighbors_i = FindNeighbors(p_i)
            if num_neighbors_i > min_pts:
                c_temp+= 1
                c_i = c_temp
                for each p_j in neighbors_i:
                    if p_j is unvisited:
                        mark p_j as visited
                        neighbors_j = FindNeighbors(p_j)
                        if num_neighbors_j > min_pts:
                            neighbors_i.append(neighbors_j)
                    if p_j is not a member of any cluster:
                        c_j = c_temp
```

As shown in Fig. 7, the actual trajectory of a ship in January 2020 (blue trajectory line in Fig. 7a) is plotted based on the Folium library. Each subfigure's labels are provided in the top-left corner. Google satellite tiles are used as the map tiles. Fig. 7 cannot display all trajectory points due to their large quantity. The colorful circular symbols with number labels are used to denote locations with densely scattered trajectory points. These number labels represent the number of trajectory points that

the symbol has covered. To further demonstrate the trajectory segmentation results, two typical colored symbols in Fig. 7a are zoomed in, as shown in Fig. 7b and Fig. 7c. The stay segments found by the DBSCAN method are highlighted with orange lines in Fig. 7b and Fig. 7c. It can be seen from Fig. 7 that the sub-trajectories of the ship at the two berths are accurately identified as the stay segment.

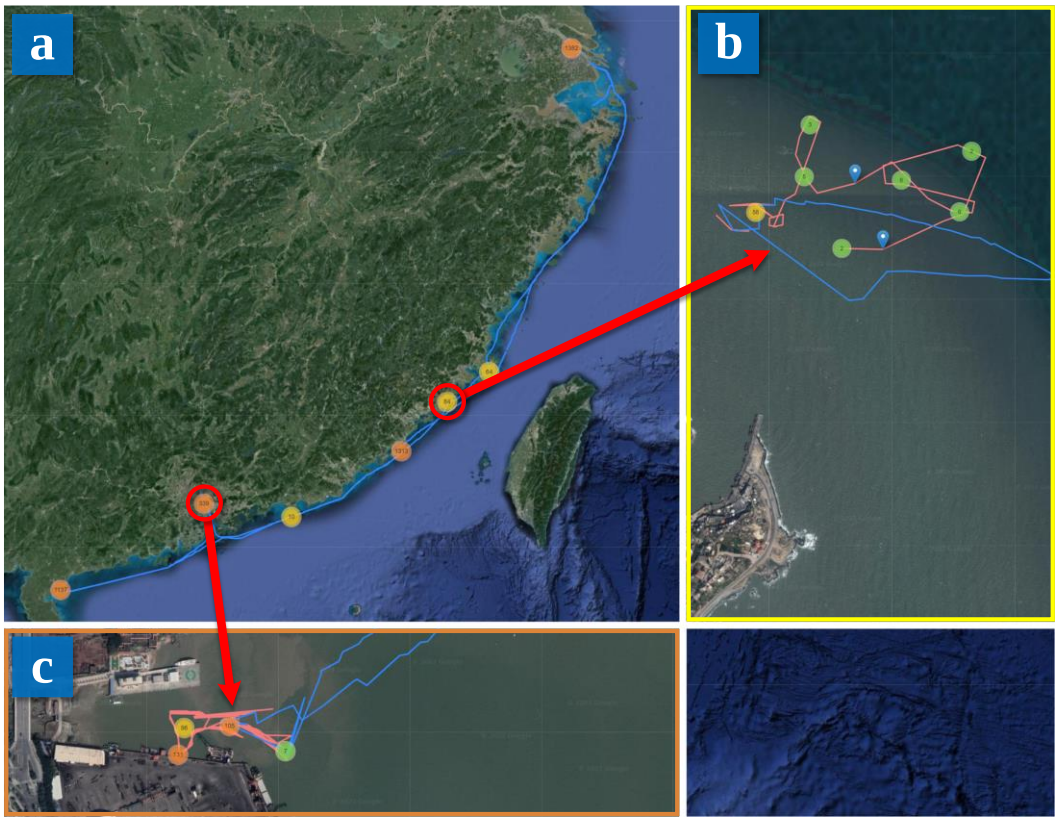


Fig. 7. Recognition of ship trajectory stay segments based on DBSCAN.

Step 4: Trajectory matching. Although the sampled trajectory maintains the characteristics of ship motion, interpolation changes the original trajectory of the ship, thus the influence on emissions is unknown and cannot be used to estimate emissions. Then the clusters of the sampled trajectories were matched to the original trajectories and continues to calculate emissions on the original trajectories. Table 5 shows the technique for a trajectory matching approach based on timestamps.

Table 5
Trajectory matching.

Input: clustered sampled trajectory $\boldsymbol{TS} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{c}, \boldsymbol{tid}\}$	
trajectory $\boldsymbol{T} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{tid}\}$	
Output: clustered trajectory $\boldsymbol{TC} = \{\boldsymbol{t}, \boldsymbol{lng}, \boldsymbol{lat}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{c}, \boldsymbol{tid}\}$	
01:	Function $\boldsymbol{MatchingTrajectories}(\boldsymbol{TS}, \boldsymbol{T})$
02:	add feature $\boldsymbol{c}$ to $\boldsymbol{T}$
03:	for each sub-trajectory group $\boldsymbol{Tg}$ in $\boldsymbol{T}$:
04:	define $tid = tid_g$
05:	$\boldsymbol{TS}_g = get\_group\_from\_ts(tid)$

06:	group $\boldsymbol{TS}_g$ by cluster feature $\boldsymbol{c}$
07:	for each cluster group $\boldsymbol{TS}_{g,c}$ in $\boldsymbol{TS}_g$:
08:	get time boundaries t_{max} and t_{min} of cluster group $\boldsymbol{TS}_{g,c}$
09:	assign the feature $\boldsymbol{c}$ of group $\boldsymbol{Tg}$ between t_{max} and t_{min} to $\boldsymbol{TS}_{g,c}$

Based on the technique given in Table 5, the stay segments in the ship's original trajectory can be accurately matched, and these stay segments include low-speed operations such as anchoring and berthing. The stay segment is utilized as the segmentation basis in this study. On the basis of the initial segmentation shown in Table 2, the trajectory is further segmented according to each stay segment, which can minimize the uncertainty in emission calculation caused by the geographical and temporal breakpoints in the trajectory shown in Fig. 5.

3.3 Carbon Dioxide Emission Estimation

A bottom-up approach based on ship activities was adopted in this study to estimate ship carbon dioxide emissions. The emissions in the studied waters are the sum of the emissions from all ships passing through the waters, as shown in Eq. (5).

$$E = \sum_{n=1}^{N} E_n \tag{5}$$

where, E is the total carbon dioxide emissions in grams from all ships sailing through the study waters. N is the number of ships sailing through the corresponding waters, and E_n is the emissions in grams of the n^{th} ship.

When a ship is sailing, the emissions mainly come from its power system, that is, the ship's main engine, auxiliary engine and boiler. Therefore, the emissions from each ship are the sum of the emissions from those systems, as shown in Eq. (6).

$$E_n = E_{n,\text{main}} + E_{n,\text{other}} \tag{6}$$

where, $E_{n,\text{main}}$ is the emissions in grams from the main engines of the n^{th} ship, and $E_{n,\text{other}}$ is the emissions in grams from its auxiliary engines and boilers.

3.3.1 Emissions from Main Engines

The emissions $E_{n,\text{main}}$ from the main engines are calculated by multiplying the instantaneous emissions by the duration, which indicated in Eq. (7).

$$E_{n,\text{main}} = P_{n,\text{main}} \times EF_{n,\text{main\_base}} \times \sum_{T} LF_t \times A_{LF,t} \times \Delta T_t \tag{7}$$

where, $P_{n,\text{main}}$ represents the maximum continuous rated power of the ship's main engine, in kWh. $EF_{n,\text{main\_base}}$ denotes the base emission factors of the main engine, in g/kWh. LF_t is the instantaneous load factor of the main engine, which mainly represents the ratio between the instantaneous power and the rated power of the main engine. $A_{LF,t}$ is emission adjustment factors when LF is lower than 20%, and the $A_{LF,t}$ value from the IMO Fourth GHG Study (IMO, 2021a) is

used in this study. ΔT_t represents the time interval between two adjacent AIS trajectory points, in seconds. Among them, the host instantaneous load factor is calculated as Eq. (8) (Goldsworthy & Goldsworthy, 2015):

$$LF_t = P_t / P_{n,\text{main}} = \left(v_t / MDS\right)^3 \tag{8}$$

where, P_t is the instantaneous power of the ship's main engine, in kWh. v_t denotes the instantaneous speed at time t and MDS is the maximum designed speed, in kn.

3.3.2 Emissions from Auxiliary Engines and Boilers

The carbon dioxide emissions of auxiliary engines and boilers are related to their power, and calculated by Eq. (9).

$$E_{n,\text{other}} = EF_{\text{other}} \times \sum_T (P_{\text{other},i,j} \times \Delta T_t) \tag{9}$$

where, EF_{other} is the emission factors for the auxiliary engines or boilers, in g/kWh. $P_{\text{other},i,j}$ is the power of the auxiliary engines or boilers, in kWh. The subscript i denotes the ship type and j denotes the navigation status. The navigation status is defined by speed and ME (main engine) load (Liu et al., 2016), as shown in Table 6. Due to a lack of data on the power of auxiliary engines and boilers, IMO estimated the power of auxiliary engines and boilers using parameters such as ship type, dead weight, gross tonnage, and navigation status. The power in this study is estimated using the method described in the IMO report (IMO, 2021a).

Table 6
Definition of navigation status.

Navigation Status	Speed (kn)	ME load
At berth	$v_t < 1$	-
At anchorage	$1 \le v_t \le 3$	-
Manoeuvring	$v_t \ge 3$	$< 20\%$
At sea	$v_t \ge 3$	$\ge 20\%$

3.3.3 Emission Factors

The emission factors for carbon dioxide can be calculated by Eq. (10) based on specific fuel consumption and the fuel-based emission factors.

$$EF_e = SFC \cdot EF_f \tag{10}$$

where, EF_e is energy-based emission factors (Xing, 2016), i.e. the emission factors in Eqs. (7) and (9). SFC stands for specific fuel consumption, in g/kWh (IMO, 2021a). EF_f is the fuel-based emission factors, in $g\ CO_2/g$. According to the Guidelines on the Method of Calculation of the Attained Energy Efficiency Design Index (EEDI) (IMO, 2021b), the value of fuel-based emission factors is shown in Table 7.

Table 7
Carbon dioxide fuel-based emission factors.

Fuel Type	EF_f
HFO	3.114
MDO	3.206
LNG	2.750

NB. HFO stands for marine Heavy Fuel Oil, MDO marine diesel oil, and LNG liquefied natural gas.

4 Results and Discussions

In this section, the performance of the STSD repair model is evaluated first using the prediction results of numerous competing models (Section 4.1), followed by the analysis of carbon dioxide emission calculation results in the coastal DECA of China (Section 4.2). This study analyzes the results from various perspectives, including the correlation analysis with the economic indicator cargo throughput (Section 4.2.2), and the emission difference analysis in different coastal provinces (Section 4.2.3). Also included is a comparison study of various ship sailing statistics such as ship types (Section 4.2.4), navigation statuses (Section 4.2.5), engine types (Section 4.2.6), ship size categories (Section 4.2.7) and ship ages (Section 4.2.8). Finally, an in-depth discussion of the above studies is given in Section 4.3. The objective of this section is to unravel the mechanism behind carbon dioxide emissions from ships along China's DECA and disclose the influence of the COVID-19 epidemic on emissions, which can further provide theoretical insights and administrative supports in the formulation of ship decarbonization policies.

4.1 Results Analysis of STSD Repair

4.1.1 Model Evaluation Indicators

In this study, mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) were used as the evaluation metrics for STSD repair models, as shown in Eqs. (11), (12) and (13).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{11}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{13}$$

where, y_i is the ground truth, $\hat{y}_i$ is the predicted values, and n is the number of samples in the test set.

4.1.2 Model Accuracy Evaluation

In STSD repair experiments, the performance of six different machine learning and deep learning

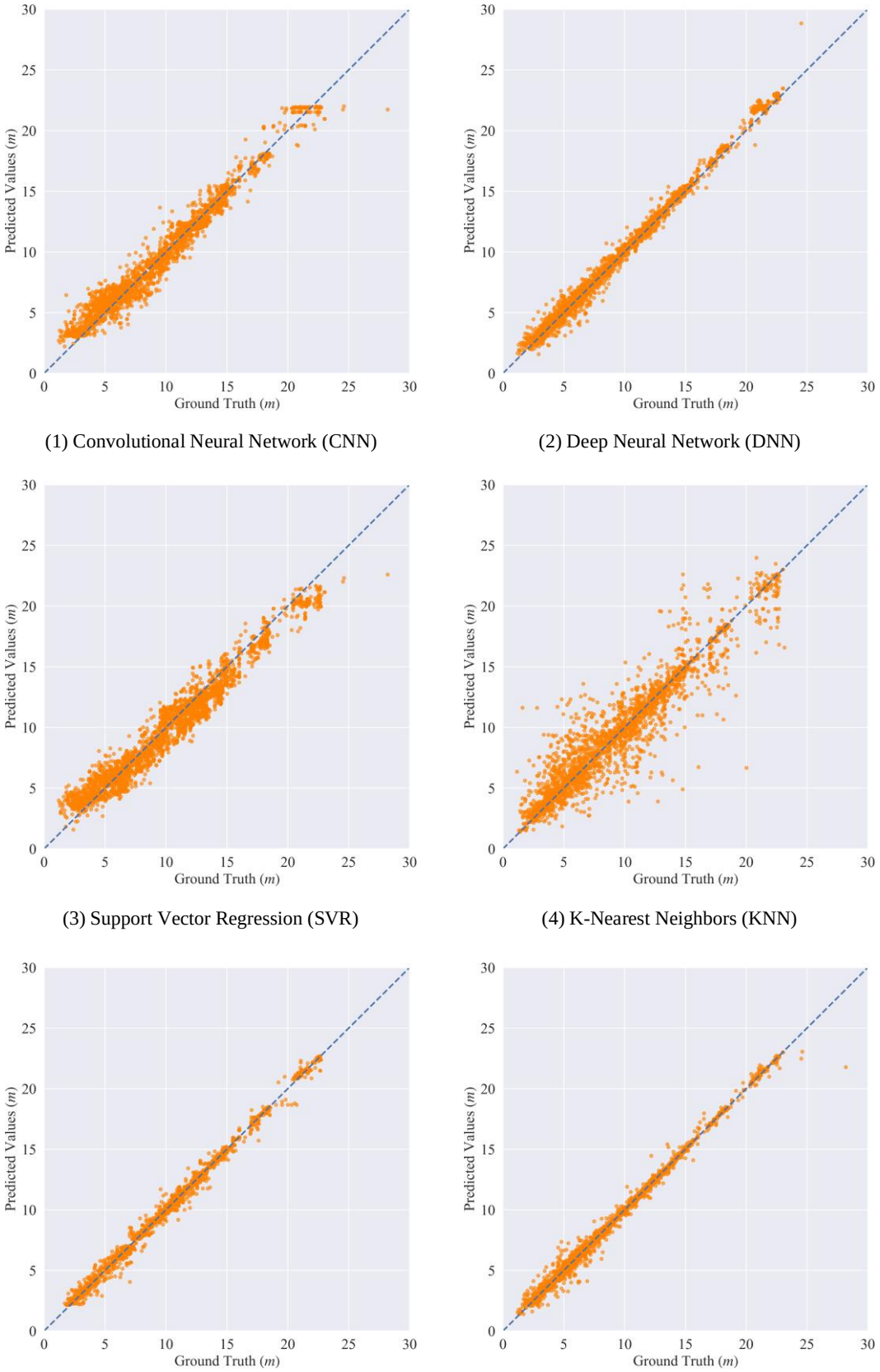
models was compared, i.e., Convolutional Neural Network (CNN), Deep Neural Network (DNN), Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Random Forest Algorithm (RF) and Extreme Gradient Boosting (XGBoost). There are 147,000 STSD samples in this study including ships with various service statuses. After data cleaning, 18,000 normative samples from STSD are selected as STS set, 75% of which is allocated for training with the remaining 25% of the data serving as the test data set. Taking the draft as an example, the draft is utilized as the label of the training data. Those six models mentioned above are employed to train the data. The grid search is used to optimize the hyperparameters of the machine learning models. Table 8 shows the final performance of evaluation results. Benefiting from the constraints of ship design dimensions specifications, there is a strong correlation between ship dimensions parameters. Therefore, compared with other regression or classification tasks, the prediction accuracy of the six models is higher. As shown in Table 8, RF performs better in MAPE (2.26%) and RMSE (0.3101) indicators and close to the best performance (0.1711 in comparison to 0.1496 of XGBoost) in MAE. Both RF and XGBoost are ensemble learning algorithms based on decision tree learners, and such algorithms perform better in the repair tasks of the STSD. Therefore, the random forest (RF) algorithm is selected to develop the STSD repair model in this study.

Table 8
Comparison of prediction accuracy on test set.

Models	MAE	MAPE	RMSE
CNN	0.6127	8.82%	0.6557
DNN	0.3228	4.52%	0.2485
SVR	0.9546	12.06%	1.1491
KNN	0.7643	10.93%	1.4299
RF	0.1711	**2.26%**	**0.3101**
XGBoost	**0.1496**	2.53%	0.3192

4.1.3 Prediction Results

Fig. 8 shows the correlation distribution between the predicted value and the ground truth of each model on the test set. As shown in Fig. 8, the data of DNN, RF and XGBoost are tightly surrounded around the reference line, and they perform better. The predicted data of the XGBoost model are dispersed in the range from 0 to10 m, but are tightly distributed in the interval between 10 to 30 m, which explains the lower MAE value of XGBoost. The predicted value of RF is evenly distributed over the intervals with stable performance.



(1) Convolutional Neural Network (CNN)

(2) Deep Neural Network (DNN)

(3) Support Vector Regression (SVR)

(4) K-Nearest Neighbors (KNN)

(5) Random Forest Algorithm (RF)

(6) Extreme Gradient Boosting (XGBoost)

Fig. 8. Comparison of prediction results of Convolutional Neural Network (CNN), Deep Neural Network (DNN), Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Random Forest Algorithm (RF) and Extreme Gradient Boosting (XGBoost) on the test set.

In addition, the performance of the new RF-based STSD repair model and the existing MSV (Johansson et al. 2017) is compared based on STS set. Referring to the research of Johansson et al. (2017), the gross tonnage is assigned as the label of the training dataset in this experiment. Moreover, this experiment relaxes the restriction of the MSV model on ship types and increases the number of candidate ships. This is because the STS set contains up to 266 ship types. Searching for candidates from ships of the same type may yield only a limited number of candidates, and even the most similar candidate ships may vary widely in size. Furthermore, in order to reveal the contribution of the two non-numerical features (ship type and flag state) to STSD repair, a dataset without flag-state and ship-type features is constructed. The performance of the models is shown in Table 9. It is discovered that, while MSV has achieved a good repair effect, the performance of RF remains the best. Moreover, after adding the two non-numerical parameters, both MAE and MAPE are improved, although the RMSE is slightly decreased. It therefore justifies the superiority of the addition of the two parameters in the RF model over the established MSV method.

Table 9
Performance of the RF and MSV models.

Models	MAE	MAPE	RMSE
RF	**648.91**	**3.21%**	1620.83
RF (without flag state and ship type)	655.03	3.27%	**1613.11**
MSV	915.02	5.32%	2603.66

4.2 Carbon dioxide Emission Estimation Results and Discussions

4.2.1 Overview of Carbon Dioxide Emissions

Based on the proposed carbon dioxide emission model, the carbon dioxide emissions from ships in China's coastal DECA in 2019 and 2020 were estimated, as shown in Table 10. The scope of the emission control area is the latest ship air pollutant emission control area issued by the Ministry of Transport of China in 2018 (MOT of China, 2018), as shown in Fig. 9 , which is designated to control and reduce emissions including SO_X, NO_X, particulate matters (PMs) and volatile organic compounds (VOCs) from vessels (Tan et al., 2021; Zhong et al., 2021). The precise data information in Fig. 9 can also be found in the literature (MOT of China, 2018).

Table 10
Carbon dioxide emissions from ships in China's coastal DECA (million tons).

Month	Emissions in 2019	Emissions in 2020
Jan.	2.50	1.95
Feb.	1.74	1.72
Mar.	2.22	2.01
Apr.	2.17	1.90
May	2.26	2.14
June	2.19	2.43

July	2.30	2.67
Aug.	2.51	2.69
Sept.	2.38	2.53
Oct.	2.33	2.41
Nov.	2.49	2.20
Dec.	2.57	2.20

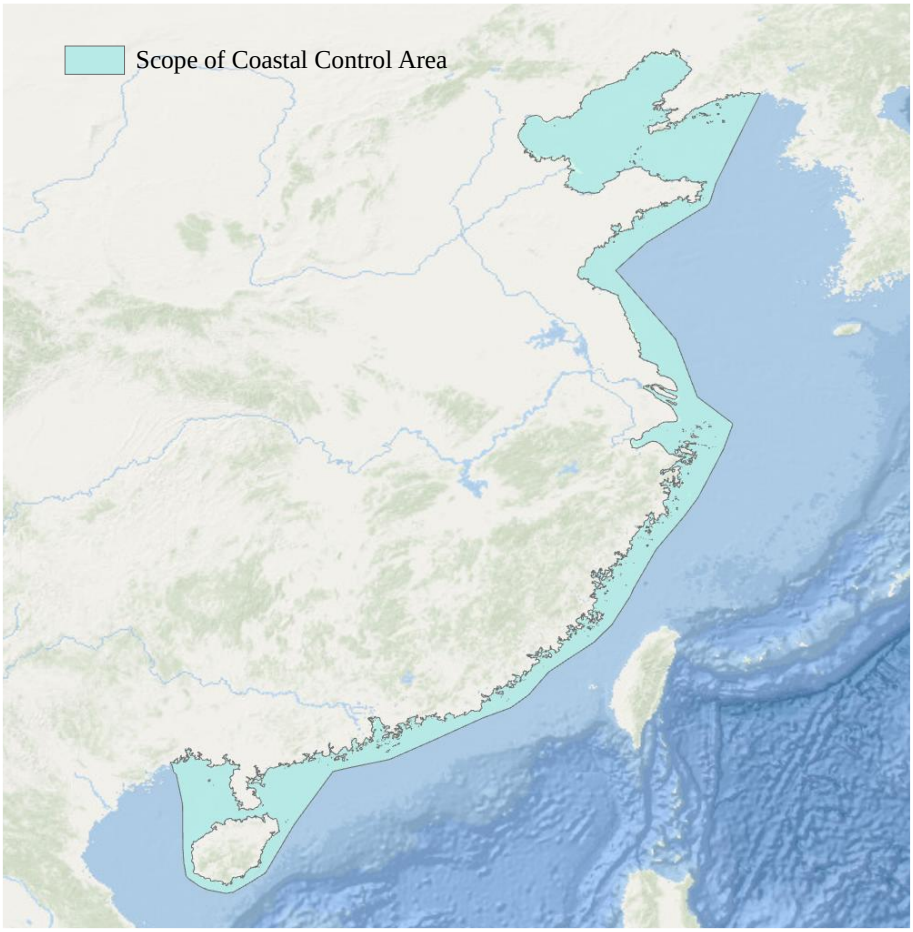


Fig. 9. Geographic scope of the emission control area in China.

IMO (2021a) classifies ships into four types based on the matching of AIS and technical specification data. Type 1 and type 2 cover most ocean-going ships, which matches to the IHS database. Type 3 is the domestic trade ships that match the Global Fishing Watch (GFW) database. Type 4 is listed in the IHS database as ships in service but miss their AIS data. This study primarily measures emissions of ships in type 1 and type 2. As shown in Table 10, the carbon dioxide emissions are 27.66 million tons in 2019 and 26.85 million tons in 2020. Ship activity and trade volume have dropped as a result of the COVID-19 pandemic (UNCTAD, 2020). Therefore, the emissions in 2020 are lower, but the total emissions in the two years are basically close. Since there is no officially released CO_2 emission data of ships within the scope of DECA in China, the research findings of the previous relevant studies in the literature are regarded as benchmarks in comparison with the result of this proposed framework in order to validate the effectiveness of this study. According to the research report released by the International Council on Clean Transportation (ICCT) (Mao & Meng, 2022), the

CO_2 emissions from ships along the coast of China in 2019 are 45 million tons. More specifically, the emission results shown in Table 10 are in a good harmony with those of the ICCT. It is therefore evident that the proposed emission estimation framework can deliver robust results. Compared to our findings, the ICCT however employs a relatively macro-level estimation framework and estimates emissions from ships along the coast of China based on only a limited number of ships, and the geographic scope is not specified.

4.2.2 Emissions in Coastal Provinces

The carbon dioxide emissions of coastal provinces in 2019 and 2020 are estimated and shown in Fig. 10 and Fig. 11. The emissions of each province fluctuate slightly between 2019 and 2020. Besides, Zhejiang, Shandong, Guangdong, Liaoning and Fujian are the top five coastal provinces in terms of carbon dioxide emissions owing to their longer coastlines and wider sea areas, and therefore ships that pass through their waters release more emissions.

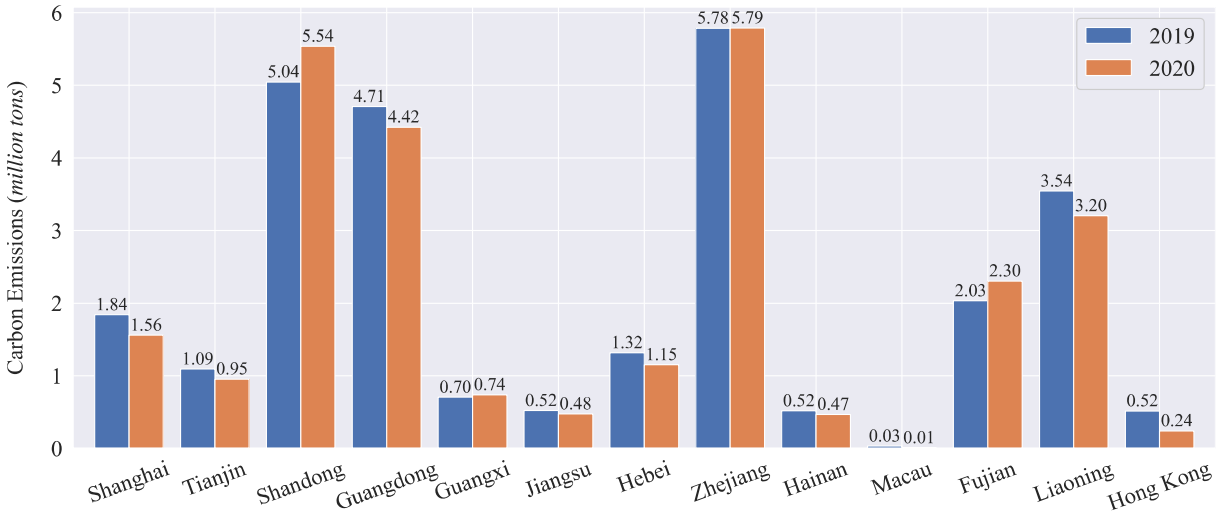
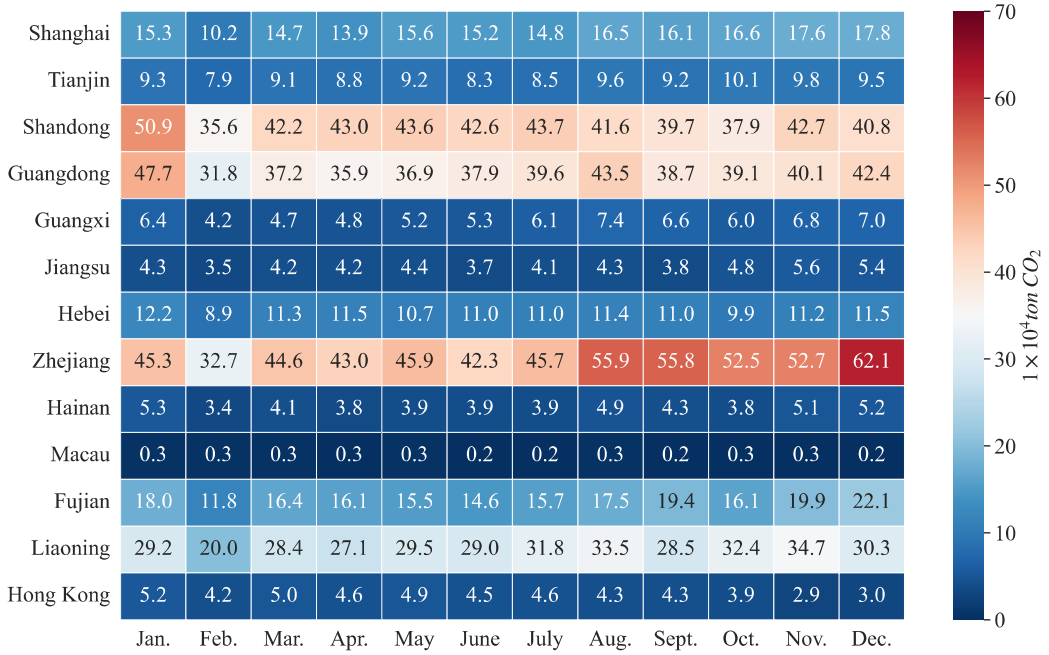
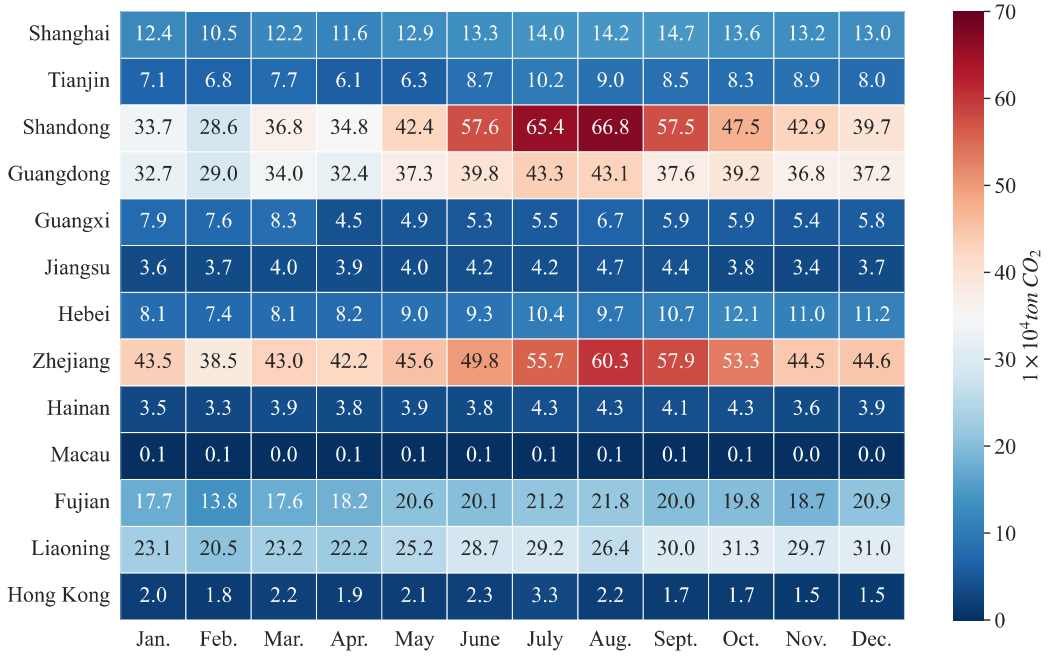


Fig. 10. Carbon dioxide emissions from coastal provinces.



(a) Monthly emissions in 2019.



(b) Monthly emissions in 2020.

Fig. 11. Monthly carbon dioxide emissions from coastal provinces.

Fig. 12 shows the average emissions from ships (calculated by per square kilometer of water) in each province. Compared with total emissions in Fig. 10, Shanghai, Hong Kong, and Macau have larger emissions per unit sea area. However, the top provinces in terms of overall emissions (i.e. Shandong, Guangdong, Liaoning and Fujian) have lower emissions per unit area. Furthermore, the statistics demonstrate that the emissions per unit area of Hong Kong and Macau are notably different before and after the COVID-19 pandemic, indicating that the pandemic has a greater impact on the maritime trade in Hong Kong and Macau compared with other provinces.

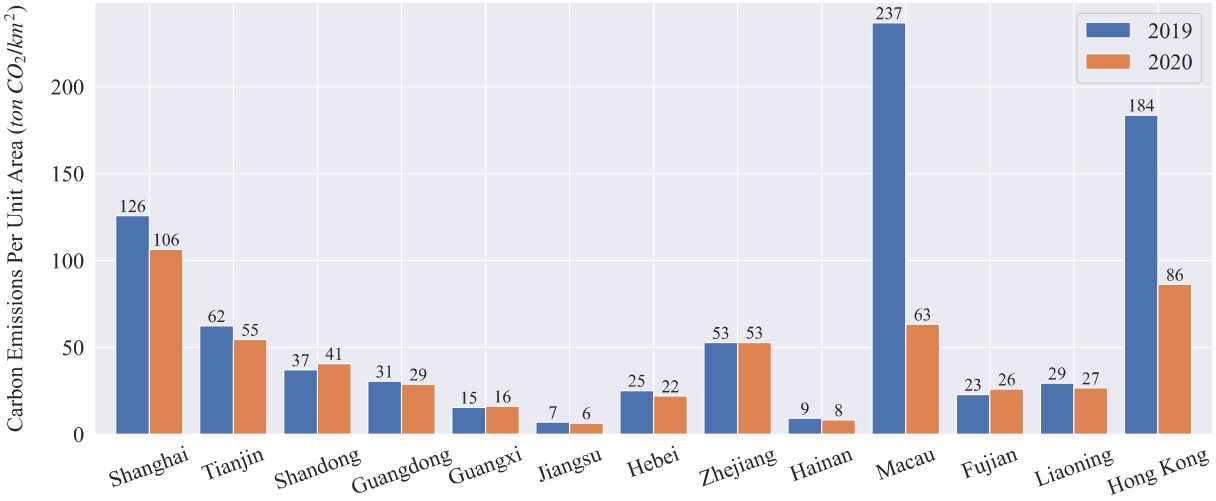


Fig. 12. Carbon dioxide emission per unit area of coastal provinces.

Fig. 13 shows the carbon intensity of coastal provinces, which is the volume of carbon dioxide emissions per unit of GDP. This metric is primarily used to assess the relationship between the

economy and carbon dioxide emissions. If the carbon intensity declines while the economy grows, it means that the region has better low-carbon development. Liaoning has the highest carbon intensity, which is closely related to the economic development model and industrial structure of the province. It is critical to optimize the industrial structure in order to achieve low-carbon development. Although Jiangsu also displays high total emissions, its carbon intensity is the smallest among all provinces. Carbon dioxide emissions are also high in Hainan, Zhejiang, Tianjin, and Shandong. As a result, China's ship decarbonization process still has a long way to go, and its economic development model and industrial structure must be shifted towards higher-quality green development. Furthermore, with the exception of Hong Kong and Macau, the carbon intensity of the remaining provinces remained essentially stable before and after the COVID-19 epidemic.

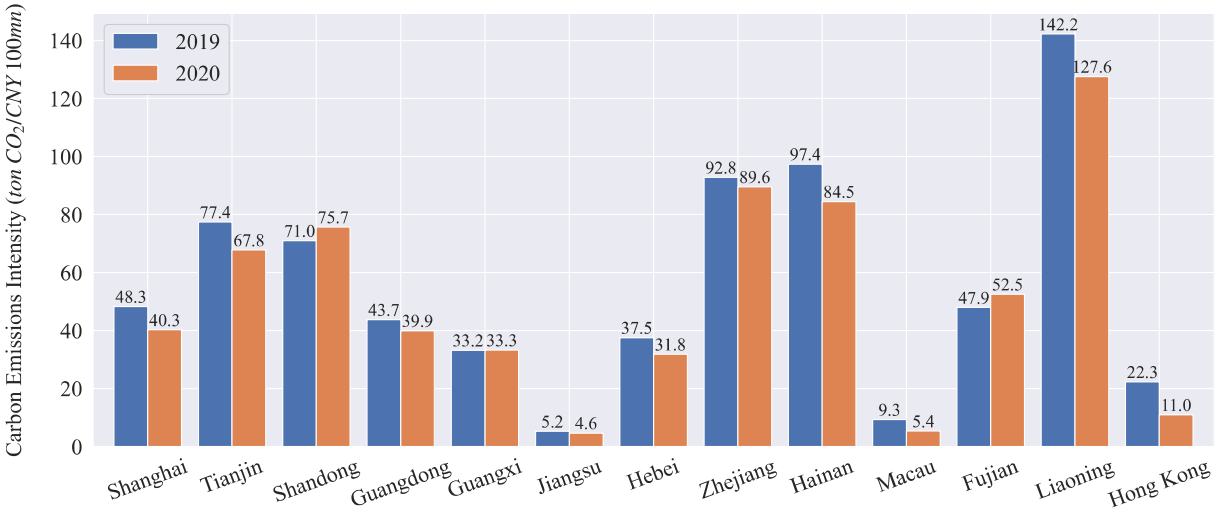


Fig. 13. Carbon intensity of coastal provinces.

4.2.3 Emission from Different Ship Types

The bulk carriers, containerships, oil tankers, chemical tankers, general cargo ships, liquefied gas tankers, and ro-ro cargo ships are the main types of ships that are reported in classical maritime statistical reports (EQUASIS, 2020; IMO, 2021a) and literatures (Ju & Hargreaves, 2021). To make this research compatible and comparable with the existing literature, in this study, ships are classified into eight types, as indicated in Fig. 14. The miscellaneous-other ships in Fig. 14 include tugs, fishing ships, yachts, service ships, and other small boats cruising close to shore. As shown in Fig. 14 and Fig. 15, ship carbon dioxide emissions were classified by ship types in 2019 and 2020. The three types of ships with the highest carbon dioxide emissions from coastal ships are containerships, oil tankers, and bulk carriers. This finding is in line with Ju and Hargreaves (2021) at Singapore port. In 2019, the three types accounted for 74% of the total emissions (31% for containerships, 24% for oil tankers, and 19% for bulk carriers). In 2020, the share is 75% (including 28% for containerships, 29% for oil tankers, and 18% for bulk carriers). According to Fig. 14, these three types of ships carry the majority of the freight volume. As a result, in the control of carbon dioxide emissions, adopting corresponding emission reduction strategies for these three ship types should be given priority to achieve better emission reduction effects compared to other ship types. Emissions from containerships, bulk carriers, general cargo ships, liquefied gas tankers, and ro-ro cargo ships decrease to varying degrees in 2020,

while the counterparts from oil tankers and chemical tankers increase.

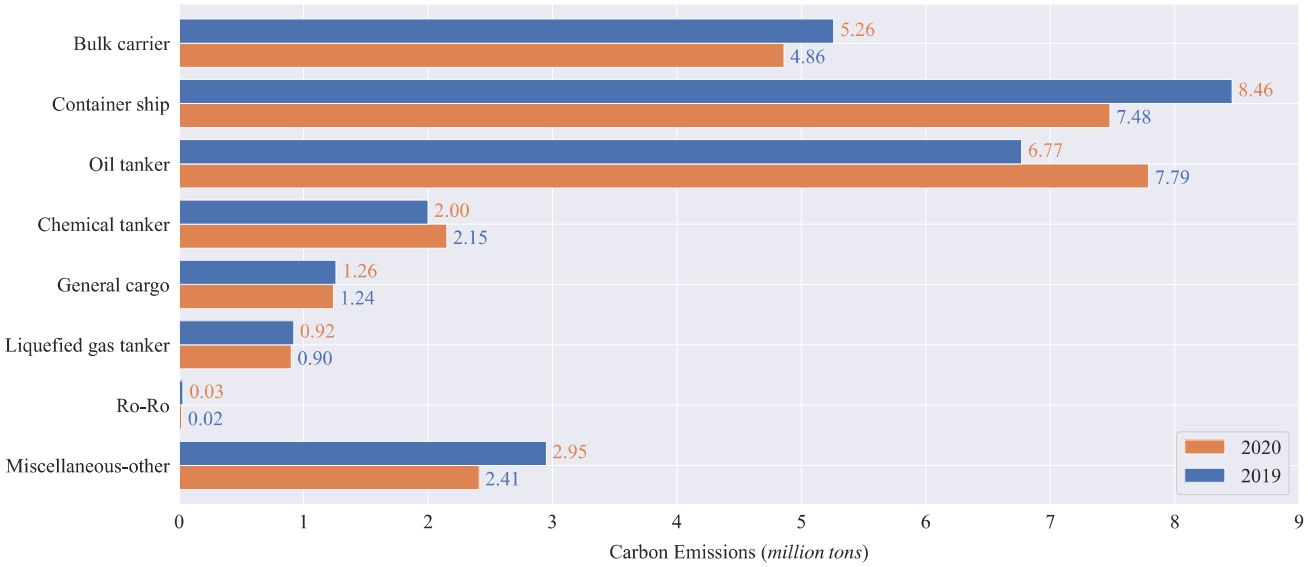
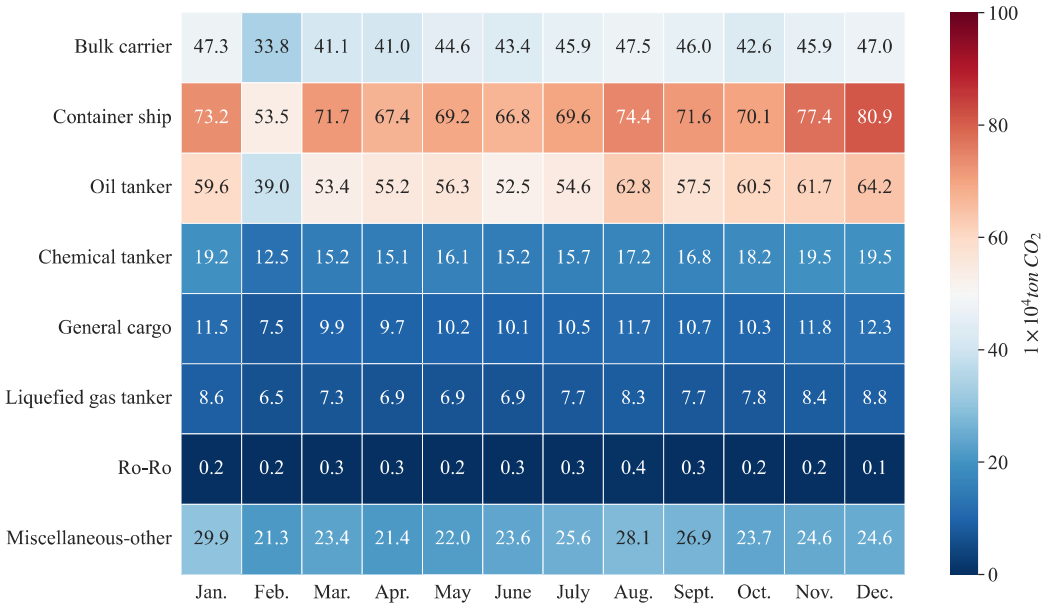
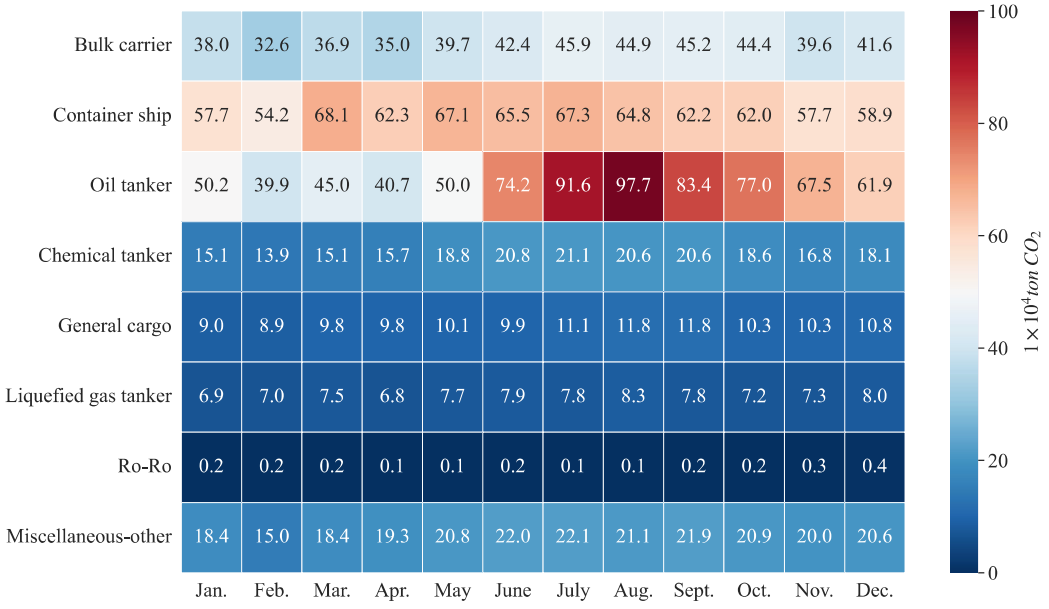


Fig. 14. Carbon dioxide emissions of different ship types.



(a) Monthly emissions in 2019.



(b) Monthly emissions in 2020.

Fig. 15. Monthly carbon dioxide emissions of different ship types.

Furthermore, the number of different types of ships being active in the research water is derived by matching AIS data with STSD, as shown in Table 11.

Table 11
Number of active archival ships within the DECA of China.

Ship Type	2019	2020
Bulk carrier	7958	6097
Container ship	3003	2206
Oil tanker	2131	1641
Chemical tanker	1896	1510
General cargo	2656	1972
Liquefied gas tanker	862	684
Ro-Ro	37	38
Miscellaneous-other	2596	1968

The numbers of all ship types operating within the DECA of China in 2020 is generally lower than in 2019, but the carbon dioxide emissions of oil tankers and chemical tankers have increased compared with 2019, as demonstrated in Fig. 14. The average single-ship emissions of various types of ships were also calculated in Fig. 16, indicating that the single-vessel emissions in 2020 were significantly higher than in 2019 due to the pandemic. Although the overall number of ships drops, the duration and frequency against the individual ships operating in coastal areas of China rise. In addition, oil tankers are the ship type with the largest increase in single-vessel emissions, increasing by nearly 50% in 2020 compared with 2019, followed by chemical tankers and general cargo ships, which climbed by 35% and 32%, respectively.

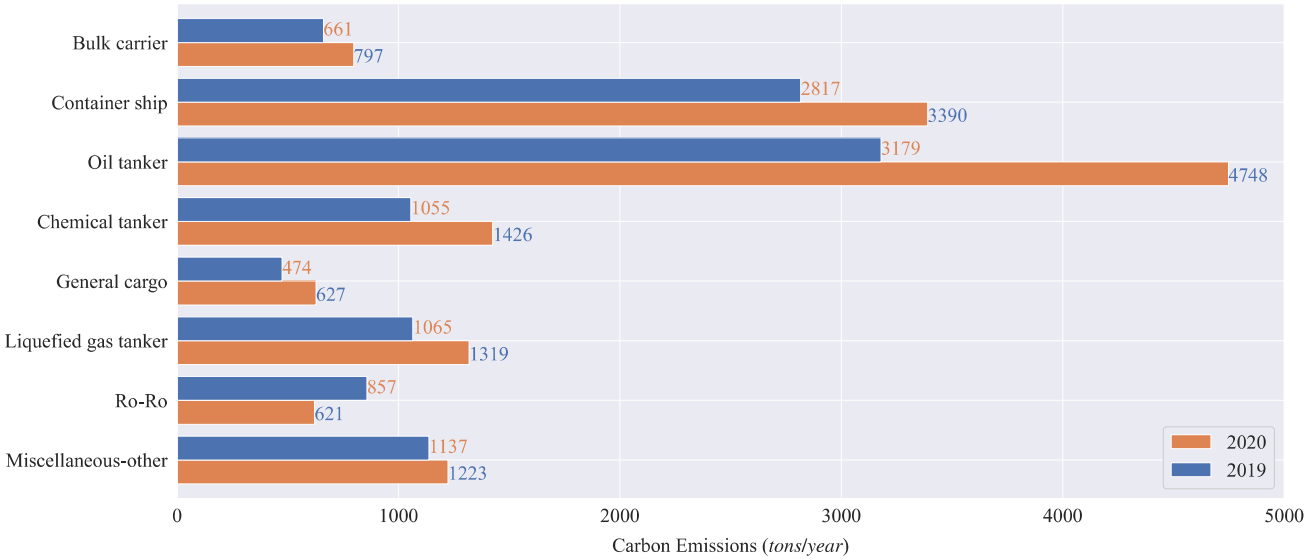
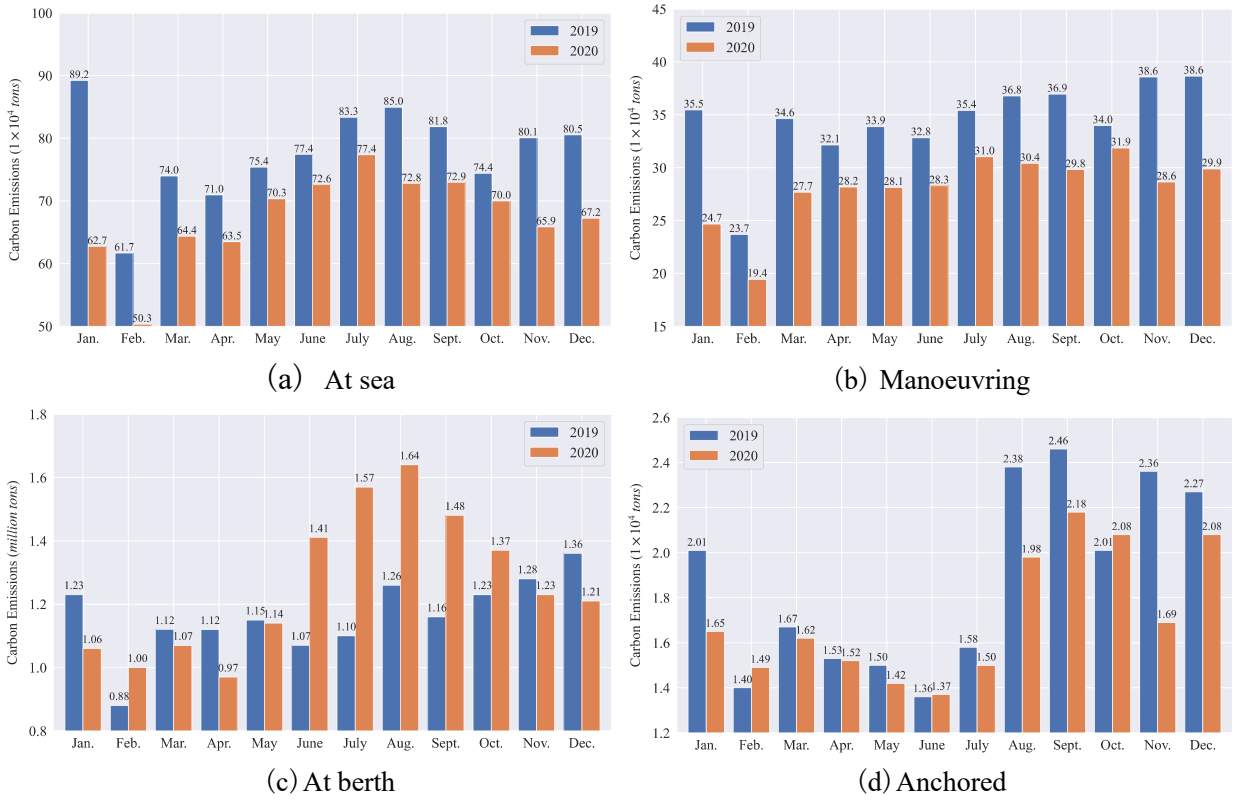


Fig. 16. Average annual carbon dioxide emissions per ship.

4.2.4 Emissions under Different Navigation Statuses

In Fig. 17, the carbon dioxide emissions of ships under different sailing states are calculated. In 2020, the emissions of ships in various sailing states were lower than in 2019 in most months, the emissions of ships at berth increased significantly from June to October 2020. In addition, at anchor status, emissions in February, June, and October 2020 also increased to varying degrees compared with

2019.



(a) At sea (b) Manoeuvring

(c) At berth (d) Anchored

Fig. 17. Ship emissions under different navigation status.

In order to further reveal the reasons, the average activity duration of different types of ships in the object area is analyzed in Fig. 18. It can be seen that, except for ro-ro cargo ships, the average activity duration of almost all types of ships in the waters within coastal DECA in 2020 has increased compared with 2019, which also explains the phenomenon that although the number of ships in 2020 is relatively small, there are still some months producing higher emissions compared with 2019. Due to the strict quarantine measures during the COVID-19 pandemic, most ships, particularly the ocean-going ships, have to berth in ports for extended periods of time. Furthermore, in 2019 and 2020, the ranking of the average activity duration of various types of ships is consistent. In Fig. 18, the majority of the "Miscellaneous-other" ships operates in coastal regions, so their activity duration in coastal areas is the longest, which also conforms to their operating characteristics. Every year, containerships, bulk carriers, oil tankers and chemical tankers are the most active ship types in China's DECA.

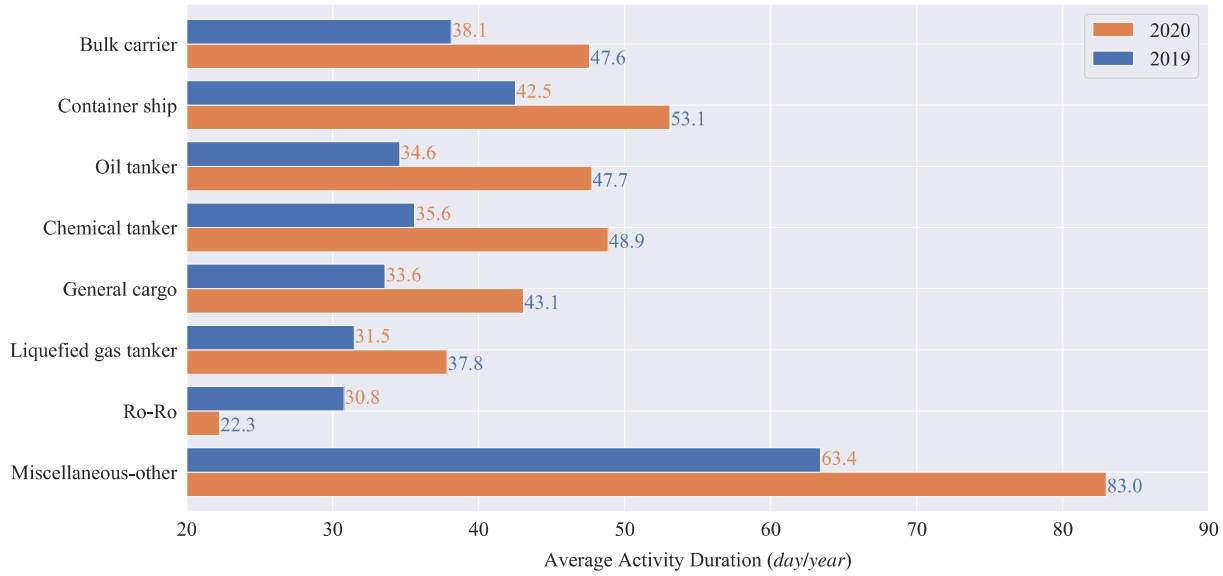
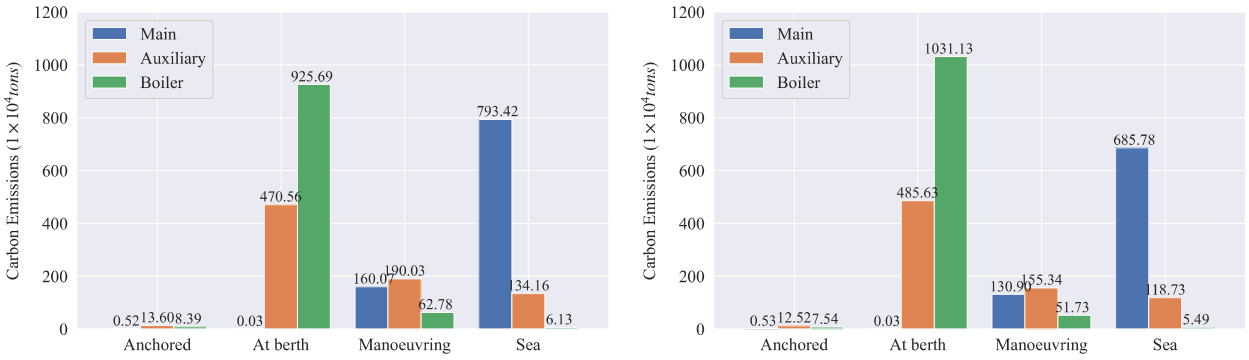


Fig. 18. Average activity duration of a ship within the China DECA (day/year).

4.2.5 Emissions from Different Engines

The emission sources of a ship during navigation are obviously variable depending on the navigation status. Fig. 19 shows the emissions of the ship main engines, auxiliary engines, and boilers under various navigation statuses. When the ship is at berth or anchor where the domestic electricity is required, the primary emission sources come from boilers and auxiliary engines. When the ship is at sea, the main engine is the main source of power rather than the boiler. From the perspective of emissions, there is a large number of ships at berth in the study area of coastal waters, so the emissions at berth are higher than the counterparts at sea. In contrast, the ships that are in manoeuvring and at anchor release fewer emissions.



(a) Emissions in 2019 (b) Emissions in 2020

Fig. 19. Emissions from main engines, auxiliary engines and boilers of ships under different navigation status.

The emission proportions of main engines, auxiliary engines and boilers of different types of ships in the research area are calculated and shown in Fig. 20 and Table 12. The detailed emission values can be calculated using the data in Fig. 14 and Table 12. The results demonstrate that the main and auxiliary engines are the primary generators of emissions for the majority of ship types. However, the boilers on oil tankers, chemical tankers and liquefied gas tankers emit a higher proportion of emissions, with boiler emissions on oil tankers accounting for more than 70% of total oil tanker emissions, and

boiler emissions on chemical tankers and liquefied gas tankers both exceeding 50%. Combined with Fig. 19 and Fig. 20, it can be revealed that these three types of ships have the longest turnaround times in ports. Shi and Weng (2021) calculated CO_2 emissions from ships in the Shanghai port and reached the same conclusion. However, on a global scale, the main engines of most ships produce the greatest emissions (IMO, 2021a), meaning in general a ship's emissions are significantly affected by its navigation statuses. Ships are usually kept in a sailing status longer than berthing, hence more studies and resources are attracted to deal with emissions of sailing ships. For coastal waters, monitoring and controlling CO_2 emissions from ships in berthing or anchored status should also be encouraged, and upgrading shore power facilities are deemed to be highly beneficial.

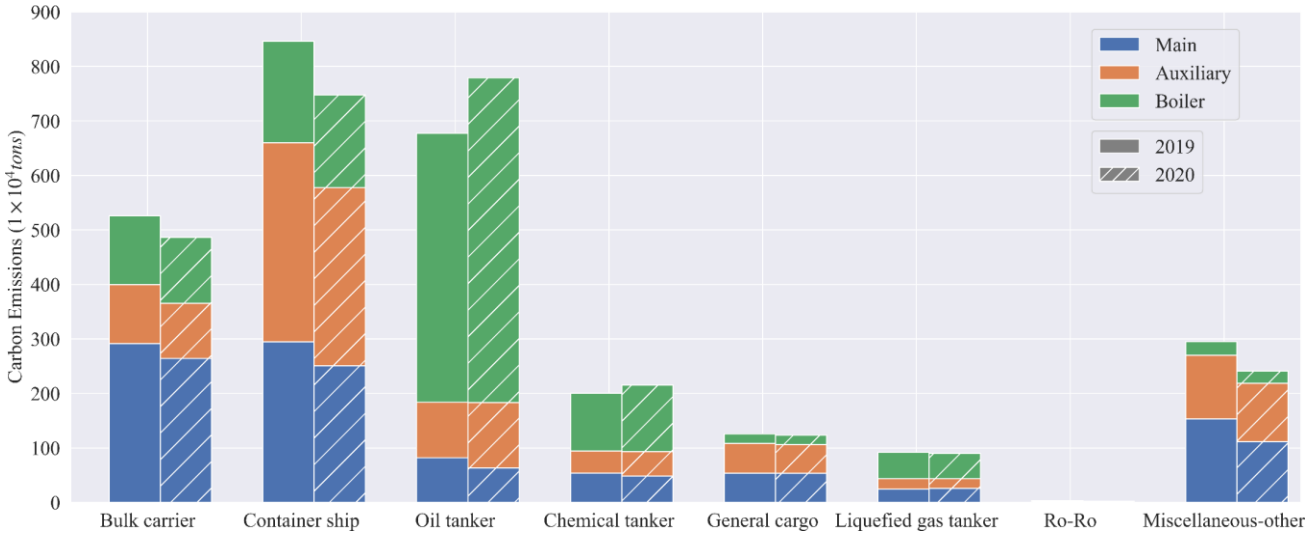


Fig. 20. Emissions from various engines of different types of ships.

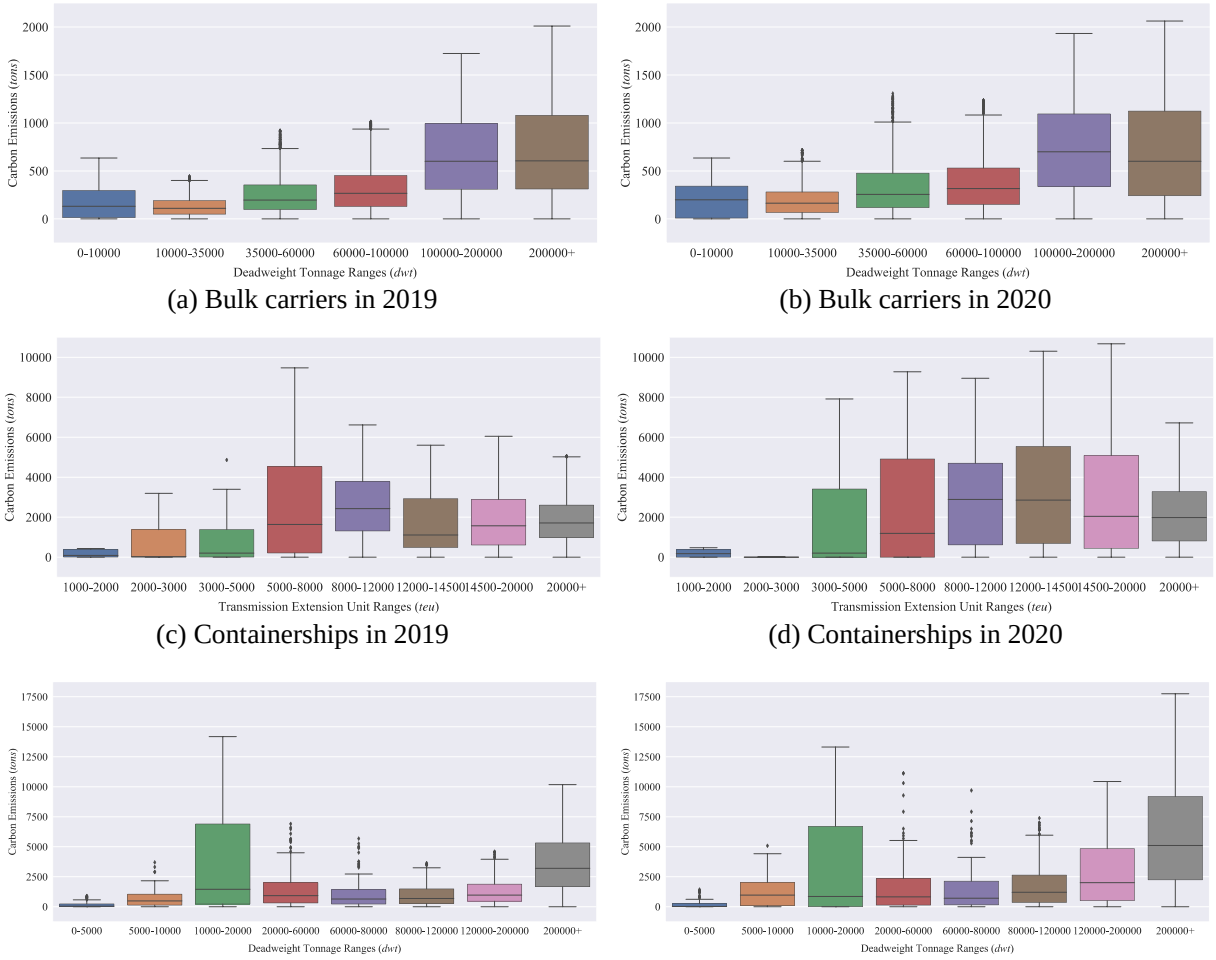
Table 12
The proportion of different types of engine emissions from various types of ships.

Ship Type	Main engine		Auxiliary engine		Boiler	
	2019	2020	2019	2020	2019	2020
Bulk carrier	55.4%	54.3%	20.6%	20.9%	24.0%	24.8%
Container ship	34.8%	33.5%	43.3%	43.8%	22.0%	22.7%
Oil tanker	12.1%	8.1%	15.0%	15.4%	72.9%	76.5%
Chemical tanker	27.0%	22.6%	20.0%	20.6%	52.9%	56.8%
General cargo	42.7%	43.0%	43.1%	42.8%	14.2%	14.2%
Liquefied gas tanker	26.4%	28.2%	21.0%	20.0%	52.6%	51.8%
Ro-Ro	29.3%	29.1%	50.8%	49.5%	19.9%	21.4%
Miscellaneous-other	51.9%	46.2%	39.6%	44.7%	8.4%	9.1%

4.2.6 Ship Size Categories and Emissions

Fig. 21 shows the relationship between ship sizes and carbon dioxide emissions, which does not follow a straightforward linear relationship. The ships were classified into different size categories based on the attributes of different ship types, referring the literature (IMO, 2021a) for details, and the

distribution of ship carbon dioxide emissions is displayed in each size category using boxplots. In the time domain, the distribution characteristics of emissions in 2019 and 2020 are similar from the perspective of the size categories of each ship type. However, the distribution statuses of the same ship type vary with different size categories. Within the deadweight range from 0 to 100,000 tons, the bulk carrier emissions under four deadweight ranges are comparatively low and almost even, as seen in the blue, orange, green, and red bars of subfigures (a) and (b) in Fig. 21. However, the bulk carriers with deadweights of above 100,000 tons emit extensive carbon dioxide, especially for the deadweight range of over 100,000 tons (see the dark brown bar). In subfigures (c) and (d) of Fig. 21, the emissions of containerships with fewer than 5,000 TEUs are relatively low, reporting an obvious skewed distribution. While containerships with more than 5,000 TEUs are balanced, their emissions basically present a normal distribution. In subfigures (e) and (f) of Fig. 21, most oil tankers emit low levels of emissions, and the median emissions of oil tankers in each size category are similar, except in the range of 10,000 to 20,000 tons (small ships) where the data shows an obvious skewed distribution. The emissions of large oil tankers above 200,000 tons grow dramatically. In addition, comparing the results in the IMO report (IMO, 2021a), the distribution of emissions across various ship types is significantly different from that shown in Fig. 21. According to the IMO, larger ships produce more emissions, which is consistent with reality, as larger sizes tend to consume more fuel. However, within China's DECA, smaller ships may produce more emissions, particularly for container ships and oil tankers in Fig. 21. Since there are numerous small-size ships for domestic trade within China's DECA, and large ships for international trade sometimes only stay within the DECA for a short period. This phenomenon is especially noticeable on containerships. Therefore, in the coastal areas of China, it is necessary to strengthen the emission control of domestic trade containerships.



(a) Bulk carriers in 2019 (b) Bulk carriers in 2020

(c) Containerships in 2019 (d) Containerships in 2020

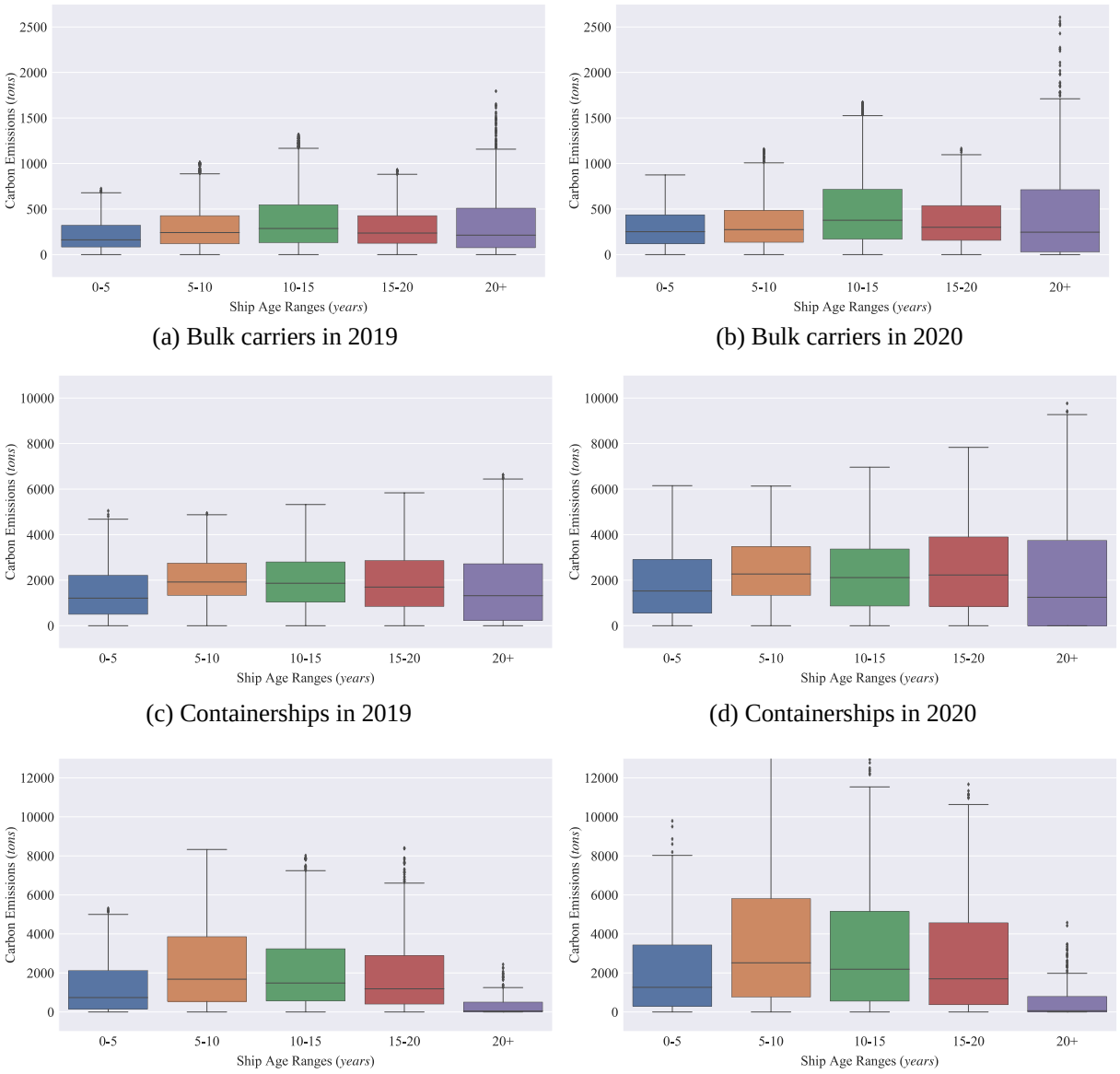
(e) Oil tankers in 2019 (f) Oil tankers in 2020

Fig. 21. Variability in emissions at different size categories across the three highest emitting ship types.

4.2.7 Ship Ages and Emissions

Fig. 22 shows the relationship between ship emissions and ship ages. The ship ages are divided into five groups with an interval of 5 years. The emission patterns of ships of the same type in 2019 and 2020 are largely identical in the time domain. Comparing the emissions of the same ship type at different ages, bulk carriers aged 10 to15 years emit more emissions, and most of the bulk carriers of beyond 20 years of age emit less emissions, but there are also certain ships with higher emissions. The distribution of containerships is balanced, which means the emission characteristics of containerships in all age groups are similar other than a higher emission of containerships over 20 years old. Among the oil tankers, the emissions from the 5 to 10-year oil tankers are relatively high, while the emissions from the oil tankers older than 20 years are comparatively low. The reason why the statistic shows older ships over 20 years emits lower CO_2 is that there are fewer ships over 20 years in service. Ships aged 5 to 15 years are the most common in service, and the government should develop a regulatory framework to promote the commissioning of new ships that satisfy stricter emission standards. The ships that already in service should be recommended to install emission control equipment, or upgrade and transform to clean-energy ships as much as possible. Although the abovementioned suggestions may bring some economic burdens in a short term, it is admirable to further consolidate the emission control of ships. Furthermore, the current production and supply infrastructures using clear energy

sources still require technological advances to facilitate their applications in the real world.



(a) Bulk carriers in 2019

(b) Bulk carriers in 2020

(c) Containerships in 2019

(d) Containerships in 2020

(e) Oil tankers in 2019

(f) Oil tankers in 2020

Fig. 22. Variability in emissions at different ages across the three highest emitting ship types.

4.3 Discussions and Implications

Based on all the aforementioned statistical results and discussions, referring to (IMO, 2021a; Nguyen et al., 2022; Shi & Weng, 2021), this study recommends that ship CO_2 emission control and management should be carried out gradually and in stages. **Stage 1**: Shore power facilities should be promoted in more ports, as the emissions from berthing ships within the scope of DECA account for a high proportion. **Stage 2**: Emission control and management should be concentrated on domestic ships because they operate within DECA for a longer period and produce more emissions. **Stage 3**: Corresponding emission reduction policies should be formulated for containerships, oil tankers, and bulk carriers, as these three types of ships emit the most CO_2. **Stage 4**: It is beneficial to encourage the commissioning of more new ships that meet the latest emission standards, and carry out carbon

reduction transformation and upgrades for those already in service. **Stage 5**: It is important to improve the clear energy supply system and promote the clean energy.

In general, this study improves the accuracy of carbon emission estimations, and its theoretical and practical implications are as follows: 1) In terms of theoretical significance, for STSD, an iterative data repair model based on the random forest algorithm is developed, which improves the quality of STSD and provides a novel solution for STSD repair. For geographical and temporal breakpoints in AIS data, a ST-DBSCAN-based ship trajectory segmentation algorithm is innovatively proposed, which minimizes the uncertainty of emission calculations. For the analysis of emission results, a set of classification analysis modes are applied, including analysis and visualization methods based on emission areas, ship types, sailing states, engine types, size categories, ship ages, etc., which are useful for in-depth exploration of ship emission characteristics and internal mechanisms. 2) In terms of practical implications, the emission estimating framework proposed in this study can significantly increase the efficiency and effectiveness of emission control measures, which is crucial for China to achieve the IMO 2050 decarbonization target and peak carbon emissions in 2030. In addition, through the multi-perspective analysis of the ship emission results in China's DECA, a series of emission control policies are recommended, such as reducing the emission from domestic trade ships and berthed ships, promoting the use of shore power facilities as soon as possible, and promoting use of clean energy, particularly in the maritime sector. This study can provide data support and a scientific basis for the competent authorities to formulate emission control policies and regulations, making the policies more targeted and improving the efficiency of emission control.

5 Conclusions

Relying on the ship AIS data and STSD, combined with machine learning models and the STEAM model, a bottom-up ship carbon dioxide emission estimation framework was built in this paper. The following findings are produced from the research on STSD data repair, AIS trajectory segmentation, and emission estimation:

(1) An iterative STSD repair model has been developed to address the low quality of ship technical specification data. The model, which is based on the Random Forest algorithm, takes into account 13 ship technical specification features. Furthermore, non-numerical features such as ship type and flag state can be processed using machine learning models. Since ships are constructed to meet certain technological specifications, there is a significant connection between ship parameters, therefore machine learning models have high credibility in handling such tasks. Therefore, the developed model can effectively repair the defects in STSD and improve the accuracy of emission estimation. Moreover, the model also has strong scalability, such as considering more ship technical parameters (e.g. main engine power, fuel type, etc.), or repairing multiple defect features of the same ship with only a few existing features. Finally, this study gives a feasible solution for STSD repair; nevertheless, there are certain limitations, such as one that the unequal distribution of the sample data will affect the accuracy of the repair of ships with fewer types, which is worth of further investigation.

(2) An ST-DBSCAN-based AIS trajectory segmentation algorithm is proposed to address the geographical and temporal breakpoint issues in ship trajectory. The algorithm includes four stages:

initial segmentation, trajectory resampling, spatial-temporal clustering, and trajectory matching, which can be used effectively to identify the stay segments, and subsequently segment the AIS trajectories depending on the stay segments. This segmentation is required because the geographical and temporal breakpoints have a considerable impact on the accuracy of emission estimation. Since the motion characteristics of ships are quite complex, trajectories with breakpoints cannot be forcibly segmented according to a unified spatial-temporal threshold, and the DBSCAN method provides a good solution. This study does not optimize the efficiency of DBSCAN. Of course, this is very flexible. The neighborhood search can be optimized based on grid-based or tree -based search strategies, thereby improving the overall efficiency of the algorithm.

(3) The carbon dioxide emissions within China's coastal DECA in 2019 and 2020 were calculated using the proposed emission estimation framework to be 27.66 million tons and 26.85 million tons, respectively. Obviously, this result is not in line with the assumption that the COVID-19 pandemic reduced ship carbon dioxide emissions significantly. While emissions have dropped, the reduction is moderate. Furthermore, it reveals that the number of ships active in the DECA has decreased and ships had a longer berthing period. Therefore, we conclude that shipping has not been affected by the pandemic in a sense that ships had a significant berthing period which leads to the emission increase during berthing, decrease in sailing and hence moderate reduction overall. This finding is in line with that of Mujal-Colilles et al. (2022). When they analyzed the emissions of Barcelona port, they observed the same pattern. In addition, comparative data analysis reveals that the dependency of carbon emissions on the economic development of the coastal provinces varies. Jiangsu Province has the lowest carbon intensity, indicating that it has achieved a low-carbon development model. Most of the coastal provinces represented by Liaoning Province have high carbon intensity, indicating that their economic development still relies on high-emission industries, and the economic development model urgently needs to be transformed and upgraded.

(4) The multi-perspective analysis of ship emissions can reveal the emission characteristics of different ship attributes, so as to recommend more targeted emission control measures. Statistics found that within China's DECA, containerships, oil tankers, and bulk carriers are the three ship types producing the highest emissions. In contrast to the scenario with the highest proportion of main engines emissions on a global scale, the proportion of auxiliary engines and boilers emissions for oil tankers, chemical tankers and liquefied gas tankers is higher in DECA. Moreover, there is a lot of small-scale ships in DECA, and their impact on CO_2 emissions cannot be underestimated. Therefore, it is necessary to strengthen the monitoring and controlling of the emissions of the aforementioned ship groups. Then, the commissioning of new and upgraded ships that meet emission standards should be encouraged, and while this may bring some economic burdens in a short term, the initiative will contribute to the maritime environment in the long run.

Limited by the research conditions, the missing trajectory points are not filled while processing the geographical and temporal breakpoints of the trajectories in order to follow the original rules of the trajectories. Therefore, the emissions calculated in this study may have small difference with real ship emissions. It is helpful and promising to create a shipping route network in the future work to repair the trajectories more scientifically and increase the accuracy of the emissions estimation.

CRediT authorship contribution statement

Haijiang Li: Created the research concepts and designed the scientific experiments of this study, Construct the computational models and frameworks. Big data processing and analysis, Writing – review & editing. **Peng Jia:** Created the research concepts, Data acquisition and analysis, Revise this article. **Xinjian Wang:** Designed the scientific experiments, Writing – review & editing. **Zaili Yang:** Designed the scientific experiments, Writing – review & editing. **Jin Wang:** Designed the scientific experiments, Writing – review & editing. **Haibo Kuang:** Created the research concepts, Data acquisition, Revise this article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer

This paper is the opinion of the authors and does not represent the belief and policy of their employers.

Acknowledgements

This work was supported by National Key Research and Development Program of China [grant no. 2019YFB1600400], the Key Project of the Liaoning Province Social Science Planning Fund [grant no. L21ACL002], National Natural Science Foundation of China [grand no. 72174035, 71774018], and Liaoning Revitalization Talents Program [grand no. XLYC2008030].

References

Aulinger, A., Matthias, V., Zeretzke, M., Bieser, J., Quante, M., & Backes, A. (2016). The impact of shipping emissions on air pollution in the greater North Sea region – Part 1: Current emissions and concentrations. *Atmospheric Chemistry and Physics*, *16*(2), 739–758. https://doi.org/10.5194/acp-16-739-2016

Başhan, V., & Kökkülünk, G. (2020). Exergoeconomic and air emission analyses for marine refrigeration with waste heat recovery system: a case study. *Journal of Marine Engineering & Technology*, *19*(3), 147–160. https://doi.org/10.1080/20464177.2019.1656324

Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, *60*(1), 208–221. https://doi.org/10.1016/J.DATAK.2006.01.013

Bole, A. G., Wall, A., & Norris, A. (2013). *Radar and ARPA manual: radar, AIS and target tracking for marine radar users*. Butterworth-Heinemann.

BP Group. (2021). *Statistical Review of World Energy 2021*.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Bullock, S., Mason, J., & Larkin, A. (2022). The urgent case for stronger climate targets for international shipping. *CLIMATE POLICY*, *22*(3), 301–309. https://doi.org/10.1080/14693062.2021.1991876

Bushra, A. A., & Yi, G. (2021). Comparative Analysis Review of Pioneering DBSCAN and Successive Density-Based Clustering Algorithms. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3089036

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). *Density-Based Clustering Based on Hierarchical Density Estimates* (pp. 160–172). https://doi.org/10.1007/978-3-642-37456-2_14

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. *Science of The Total Environment*, *609*, 776–787. https://doi.org/10.1016/J.SCITOTENV.2017.07.051

Chen, M., Li, Y., Hu, M., & Diego, M. (2021). *What to Expect in China's Second Nationally Determined Contribution: Towards a Carbon Neutral Global Future 2021 Edition* (2021st ed.).

Department for Transport. (2019). *Clean maritime plan*.

Endresen, Ø., Sørgård, E., Behrens, H. L., Brett, P. O., & Isaksen, I. S. A. (2007). A historical reconstruction of ships' fuel consumption and emissions. *Journal of Geophysical Research*, *112*(D12), D12301. https://doi.org/10.1029/2006JD007630

EQUASIS. (2020). *The 2020 World Merchant Fleet Statistics from Equasis*.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*.

Gan, L., Che, W., Zhou, M., Zhou, C., Zheng, Y., Zhang, L., Rangel-Buitrago, N., & Song, L. (2022). Ship exhaust emission estimation and analysis using Automatic Identification System data: The west area of Shenzhen port, China, as a case study. *Ocean & Coastal Management*, *226*, 106245. https://doi.org/10.1016/J.OCECOAMAN.2022.106245

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. *Environmental Modelling & Software*, *63*, 45–60. https://doi.org/10.1016/J.ENVSOFT.2014.09.009

Gossling, S., Meyer-Habighorst, C., & Humpe, A. (2021). A global review of marine air pollution policies, their scope and effectiveness. *OCEAN & COASTAL MANAGEMENT*, *212*. https://doi.org/10.1016/j.ocecoaman.2021.105824

Huang, L., Wen, Y., Zhang, Y., Zhou, C., Zhang, F., & Yang, T. (2020). Dynamic calculation of ship exhaust emissions based on real-time AIS data. *Transportation Research Part D: Transport and Environment*, *80*, 102277. https://doi.org/10.1016/J.TRD.2020.102277

IMO. (2000). *Study of Greenhouse Gas Emissions from Ships*.

IMO. (2009). *Second IMO Greenhouse Gas Study 2009*.

IMO. (2014). *Third IMO Greenhouse Gas Study 2014*.

IMO. (2021a). *Fourth Greenhouse Gas Study 2020*.

IMO. (2021b). *MEPC.76/6/9 - Proposed Amendments to the 2018 Guidelines on the Method of Calculation of the Attained Energy Efficiency Design Index (EEDI) for New Ships*.

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. *Atmospheric Chemistry and Physics*, *9*(23), 9209–9223. https://doi.org/10.5194/acp-9-9209-2009

Jayakodi, K., Bandara, M., Perera, I., & Meedeniya, D. (2016). WordNet and Cosine Similarity based Classifier of Exam Questions using Bloom's Taxonomy. *International Journal of Emerging Technologies in Learning (IJET)*, *11*(04), 142. https://doi.org/10.3991/ijet.v11i04.5654

Jing, D., Dai, L., Hu, H., Ding, W., Wang, Y., & Zhou, X. (2021). CO2 emission projection for Arctic shipping: A system dynamics approach. *Ocean & Coastal Management*, *205*, 105531. https://doi.org/10.1016/J.OCECOAMAN.2021.105531

Johansson, L., Jalkanen, J. P., & Kukkonen, J. (2017). Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. *Atmospheric Environment*, *167*, 403–415. https://doi.org/10.1016/J.ATMOSENV.2017.08.042

Ju, Y., & Hargreaves, C. A. (2021). The impact of shipping CO2 emissions from marine traffic in Western Singapore Straits during COVID-19. *Science of The Total Environment*, *789*, 148063. https://doi.org/10.1016/j.scitotenv.2021.148063

Khan, M. M. R., Siddique, M. A. B., Arif, R. B., & Oishe, M. R. (2018). ADBSCAN: Adaptive density-based spatial clustering of applications with noise for identifying clusters with varying densities. *4th International Conference on Electrical Engineering and Information and Communication Technology, ICEEiCT 2018*. https://doi.org/10.1109/CEEICT.2018.8628138

Kim, D., & Koo, M.-W. (2017). Categorization of Korean News Articles Based on Convolutional Neural Network Using Doc2Vec and Word2Vec. *Journal of KIISE*, *44*(7), 742–747. https://doi.org/10.5626/JOK.2017.44.7.742

Lewis, R. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*.

Li, H., Lam, J. S. L., Yang, Z., Liu, J., Liu, R. W., Liang, M., & Li, Y. (2022). Unsupervised hierarchical methodology of maritime traffic pattern extraction for knowledge discovery. *Transportation Research Part C: Emerging Technologies*, *143*, 103856. https://doi.org/10.1016/j.trc.2022.103856

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. *Nature Climate Change*, *6*(11), 1037–1041. https://doi.org/10.1038/nclimate3083

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*. https://doi.org/10.1007/978-3-319-57421-9_20

Mao, X., & Meng, Z. (2022). *Decarbonizing China's coastal shipping: The role of fuel efficiency and low-carbon fuels*.

Martin, P. (2021). *Average speed of global merchant fleet by ship type 2018*.

McKinlay, C. J., Turnock, S. R., & Hudson, D. A. (2021). Route to zero emission shipping: Hydrogen, ammonia or methanol? *International Journal of Hydrogen Energy*, *46*(55), 28282–28297. https://doi.org/https://doi.org/10.1016/j.ijhydene.2021.06.066

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations*

in Vector Space.

Moreno-Gutiérrez, J., & Durán-Grados, V. (2021). Calculating ships' real emissions of pollutants and greenhouse gases: Towards zero uncertainties. *Science of The Total Environment*, *750*, 141471. https://doi.org/10.1016/J.SCITOTENV.2020.141471

MOT of China. (2018). *Implementation Scheme of the Domestic Emission Control Areas for Atmospheric Pollution from Vessels*.

Mujal-Colilles, A., Guarasa, J. N., Fonollosa, J., Llull, T., & Castells-Sanabra, M. (2022). COVID-19 impact on maritime traffic and corresponding pollutant emissions. The case of the Port of Barcelona. *Journal of Environmental Management*, *310*, 114787. https://doi.org/10.1016/J.JENVMAN.2022.114787

Nguyen, P.-N., Woo, S.-H., & Kim, H. (2022). Ship emissions in hotelling phase and loading/unloading in Southeast Asia ports. *Transportation Research Part D: Transport and Environment*, *105*, 103223. https://doi.org/10.1016/j.trd.2022.103223

Peng, X., Wen, Y., Wu, L., Xiao, C., Zhou, C., & Han, D. (2020). A sampling method for calculating regional ship emission inventories. *Transportation Research Part D: Transport and Environment*, *89*, 102617. https://doi.org/10.1016/J.TRD.2020.102617

Rodríguez, P., Bautista, M. A., Gonzàlez, J., & Escalera, S. (2018). Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing*, *75*, 21–31. https://doi.org/10.1016/J.IMAVIS.2018.04.004

Schwarzkopf, D. A., Petrik, R., Matthias, V., Quante, M., Majamäki, E., & Jalkanen, J. P. (2021). A ship emission modeling system with scenario capabilities. *Atmospheric Environment: X*, *12*, 100132. https://doi.org/10.1016/J.AEAOA.2021.100132

Shi, K., & Weng, J. (2021). Impacts of the COVID-19 epidemic on merchant ship activity and pollution emissions in Shanghai port waters. *Science of The Total Environment*, *790*, 148198. https://doi.org/10.1016/J.SCITOTENV.2021.148198

Silveira, P., Teixeira, A. P., & Guedes Soares, C. (2022). A method to extract the Quaternion Ship Domain parameters from AIS data. *Ocean Engineering*, *257*, 111568. https://doi.org/10.1016/J.OCEANENG.2022.111568

Streets, D. G., Guttikunda, S. K., & Carmichael, G. R. (2000). The growing contribution of sulfur emissions from ships in Asian waters, 1988–1995. *Atmospheric Environment*, *34*(26), 4425–4439. https://doi.org/10.1016/S1352-2310(00)00175-8

Tan, Z., Liu, H., Shao, S., Liu, J., & Chen, J. (2021). Efficiency of Chinese ECA policy on the coastal emission with evasion behavior of ships. *Ocean & Coastal Management*, *208*, 105635. https://doi.org/10.1016/J.OCECOAMAN.2021.105635

Tzannatos, E. (2010). Ship emissions and their externalities for Greece. *Atmospheric Environment*, *44*(18), 2194–2202. https://doi.org/10.1016/J.ATMOSENV.2010.03.018

UNCTAD. (2020). COVID-19 and Maritime Transport: Impact and Responses. *UNCTAD Policy Brief No. 75*.

UNCTAD. (2021). *Review of Maritime Transport 2021*.

US EPA. (2015). *Development of the First National Mexican Emission Inventories for Ships Using the*

Waterway Network Ship Traffic, Energy, and Environmental Model (STEEM).

Wang, C., Corbett, J. J., & Firestone, J. (2007). Modeling Energy Use and Emissions from North American Shipping: Application of the Ship Traffic, Energy, and Environment Model. *Environmental Science & Technology*, *41*(9), 3226–3232. https://doi.org/10.1021/es060752e

Wang, P., & Chung, A. C. S. (2020). *DoubleU-Net: Colorectal Cancer Diagnosis and Gland Instance Segmentation with Text-Guided Feature Control* (pp. 338–354). https://doi.org/10.1007/978-3-030-66415-2_22

Weinberger, K., Dasgupta, A., Langford, J., Smola, A., & Attenberg, J. (2009). Feature hashing for large scale multitask learning. *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, 1–8. https://doi.org/10.1145/1553374.1553516

Weng, J., Shi, K., Gan, X., Li, G., & Huang, Z. (2020). Ship emission estimation with high spatial-temporal resolution in the Yangtze River estuary using AIS data. *Journal of Cleaner Production*, *248*, 119297. https://doi.org/10.1016/J.JCLEPRO.2019.119297

Xing, H. (2016). *Study on Quantification of Exhaust Emissions from Ships* [Ph.D]. Dalian Maritime University.

Yang, D., Wu, L., Wang, S., Jia, H., & Li, K. X. (2019). How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*, *39*(6). https://doi.org/10.1080/01441647.2019.1649315

Yang, H., Ma, X., & Xing, Y. (2017). Trends in CO2 Emissions from China-Oriented International Marine Transportation Activities and Policy Implications. *Energies*, *10*(7), 980. https://doi.org/10.3390/en10070980

Yu, H., Chen, X., Li, Z., Zhang, G., Liu, P., Yang, J., & Yang, Y. (2019). Taxi-Based Mobility Demand Formulation and Prediction Using Conditional Generative Adversarial Network-Driven Learning Approaches. *IEEE Transactions on Intelligent Transportation Systems*, *20*(10), 3888–3899. https://doi.org/10.1109/TITS.2019.2923964

Zhong, H., Guo, C., Yip, T. L., & Gu, Y. (2021). Bi-perspective sulfur abatement options to mitigate coastal shipping ships emissions: A Case Study of Chinese coastal zone. *Ocean & Coastal Management*, *209*, 105658. https://doi.org/10.1016/J.OCECOAMAN.2021.105658