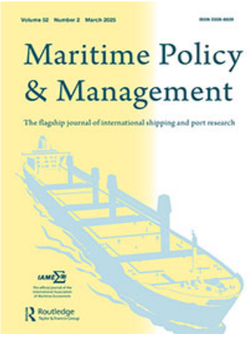

Routledge
Taylor & Francis Group

Maritime Policy & Management

The flagship journal of international shipping and port research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tmpm20

Impact of imputation methods for ship technical parameters on emission estimations in ports

Ruikai Sun, Wessam Abouarghoub, Emrah Demir & Andrew Potter

To cite this article: Ruikai Sun, Wessam Abouarghoub, Emrah Demir & Andrew Potter (12 Feb 2025): Impact of imputation methods for ship technical parameters on emission estimations in ports, Maritime Policy & Management, DOI: 10.1080/03088839.2025.2463635

To link to this article: https://doi.org/10.1080/03088839.2025.2463635

MARITIME POLICY & MANAGEMENT
https://doi.org/10.1080/03088839.2025.2463635

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS

Impact of imputation methods for ship technical parameters on emission estimations in ports

Ruikai Sun [a], Wessam Abouarghoub [a,b], Emrah Demir [a] and Andrew Potter [a]

[a]Logistics and Operations Management, Cardiff Business School, Cardiff University, Cardiff, UK; [b]Department of Operations and Project Management, College of Business, Alfaisal University, Riyadh, Saudi Arabia

ABSTRACT

Greenhouse gas emissions from ships have emerged as a pressing concern. Nevertheless, the quality of data in existing databases remains inadequate, with numerous instances of missing information. This presents significant challenges for accurately estimating emissions associated with ship activities in port. This paper uses three imputation methods and applies them to three ports as a case study to evaluate their performance in emission estimation. The mixed-method demonstrates high accuracy while covering nearly all cases of missing data, resulting in the smallest error in estimating daily emissions. The results indicate that if the data quality is not improved, at least 12% of CO_2 emissions may be underestimated. The cases of missing data that the imputation model can address also have a significant impact. For example, the multiple linear regression method, which only covers partial cases of missing data, leads to an underestimation of emissions by 2% to 6%. The findings highlight that an appropriate imputation method can significantly improve the accuracy of emission estimation. They also highlight the importance of data quality, which not only reduces estimation errors but also helps prevent the substantial underestimation of emissions.

ARTICLE HISTORY
Received 3 May 2024
Accepted 26 January 2025

KEYWORDS
Green port; missing data imputation; big data; ship emission; ship technical variable; data preprocess

1. Introduction

The value of the maritime supply chain for the global economy is indisputable, where more than 80% of global merchandise trade passes through ports (UNCTAD 2024). Shipping freight is notably a more carbon-efficient mode of transport compared to road and air freight. Recent technological advancements in engine efficiency and ship design have led to reductions in emissions and improvements in fuel efficiency. Due to the massive scale of maritime trade and its anticipated growth, the shipping industry continues to have a significant impact on global greenhouse gas emissions. The industry's carbon dioxide emissions account for approximately 3% of the global total, with an increase of 5.9% from 2012 to 2018 (IMO 2020). The increase in carbon dioxide (CO_2) emissions from shipping is a mounting concern. In response, the International Maritime Organization (IMO), the regulatory authority for the shipping industry, has formulated and enacted policies aimed at decarbonization. For instance, the IMO's initial GHG Strategy of 2023 focuses particularly on reducing the carbon intensity of shipping. The strategy sets a target to decrease CO_2 emissions per voyage by at least 40% by 2030, aiming for a 70% reduction by 2050, relative to 2008 levels. Ports, as critical nodes connecting water and land transport, concentrate a large number of

CONTACT Ruikai Sun sunr10@cardiff.ac.uk Logistics and Operations Management, Cardiff Business School, Cardiff University, Cardiff CF10 3EU, UK

ships within a confined area, leading to high emissions density (Yu, Sun, Sun, Wu, et al. 2022). It is estimated that 70% of ship CO_2 emissions occur within 400 km of coastlines (S. Chen et al. 2021). These excessive carbon emissions will increase the effect of greenhouse gases from shipping and accelerate global warming. Therefore, accurately estimating emissions from ships in ports is crucial for addressing these challenges effectively.

Recent developments in the field of emission estimation have led to a renewed interest in data quality. Furthermore, missing data is a critical issue affecting data quality (Ezzine and Benhlima 2018) and a common problem in maritime quantitative research (Ribeiro, Paes, and de Oliveira 2023; Sun et al. 2025). Smaller databases faced quality issues, but their manageable size allowed for easier data cleansing. With advances in computing power, efficiency and storage capacity making it possible to compile and analyze database of incredible size and complexity, the era of so-called 'big data' has been created. However, the advent of big data, enabled by leaps in computing power, efficiency, and storage capacity, has not resolved the issue of missing data. Despite the surge in available information for analysis, larger datasets do not inherently mean more complete datasets (Lall and Robinson 2022). In the shipping field, obtaining complete ship technical parameters may be difficult due to the existence of different organizations and data collection standards (H. Wang et al. 2016). Although this paper uses different databases (LSGE and Clarkson) for validation, only a few shipping industry researchers and employees have access to multiple comprehensive databases. These databases are usually only accessible to institutions and companies. Therefore, although some of the missing data can be searched in different databases, they can be treated as missing data due to workload and access issues. However, these data play a key role in estimating emissions from ships in port, and too much missing data may result in estimates that are significantly lower than the actual values (Doundoulakis and Papaefthimiou 2022; Peng et al. 2020; L. Wang and Li 2023). There has been little quantitative analysis of imputing missing data of ship technical parameters and no previous study has formally studied the impact of data quality and imputation methods on estimating port CO_2 emissions from ships.

This paper uses the data from Port of Busan, Los Angeles and Felixstowe to analyze the impact of ship data quality on port emission estimation. Three methods (multiple linear regression, curve fitting and mixed-method) are utilized to impute the missing ship technical parameters. A fuel-based bottom-up emission estimation method is employed in this study to calculate the ship emissions in the port based on the imputed database as well as the original database. This research makes the following three contributions to previous research. Firstly, this paper is the first to quantify the impact of different imputation methods and data quality on the estimation of port CO_2 emissions from ships. Secondly, this paper proposes a mixed approach to address the shortage of the IMO-multiple linear regression method to impute missing values when both ship service speed and main engine power are missing. Thirdly, the performance of each missing data imputation method for different ship technical parameters is evaluated under two types of port using cross-validation with databases from two sources.

The remaining part of the paper is structured as follows: Section 2 provides a literature review of port ship emission estimation methods and data imputation in the maritime field and identifies research gaps. Section 3 presents the introduction of methods and preparation of the case study. Section 4 reports the empirical results and discussion. Section 5 concludes the research and presents future research directions.

2. Literature review

Ports are recognized as focal points for the emission of air pollutants and GHG (IMO 2014). As ship emissions are the largest source of port emissions, there has been a growing interest in research aimed at quantifying the emissions from ships within a port.

Research in this area has demonstrated the use of two traditional approaches for estimating ship emissions: top-down methods and bottom-up methods (Miola and Ciuffo 2011). Typically, top-down estimation methods are often applied to the measurement of ship emissions on a national or regional scale, which is then geographically reduced by using proxy variables on a smaller scale (region or city) (Toscano and Murena 2019; Yu, Sun, Sun, and Shu 2022). The method is usually based on the average technical parameters and operating conditions of different types of ships combined with the number of ships, and then the data are aggregated in time and space to obtain an estimation. The bottom-up approach is based on the ship's activity, which is based on the ship's mode of operation (e.g. berthing, maneuvering and sailing), real-time speed, engine load factor, environmental conditions and sometimes even wind speed and wave height (Jalkanen et al. 2009; Tran et al. 2022; Tzannatos 2010). Ship emissions during this period of activity are then calculated based on the type of ship, main engine type, auxiliary engine type, service speed and various other technical parameters. Finally, the results of all activities are added together to obtain a total emission amount. It is generally agreed that a top-down approach is useful to obtain a preliminary estimate of local emissions, while a bottom-up approach allows for more precise and detailed results (Eyring et al. 2010; Ng et al. 2013).

In the bottom-up approach, there are two types of methods: the fuel-based method and the energy-based method (IMO 2014), the main difference between these two methods is the calculation of the emission factor. The fuel-based method's emission factor is based on the chemical composition of the fuel to estimate the emissions per unit of fuel, so the fuel-based method is more focused on calculating the fuel consumption of the ship (Zis et al. 2014). The emission factor of the energy-based method is an estimation of the number of emissions per kW based on the engine power. These two methods are each suitable for calculating different types of emissions (IMO 2020). Carbon dioxide and sulphur oxides are suitable for fuel-based emission factors, while NOx, Particulate Matter, Methane, Nitrous oxide, and non-methane volatile organic compounds (NMVOC) are more suitable for energy-based emission factors. This paper focuses on estimating CO_2 emissions, employing the fuel-based bottom-up method to calculate emissions from ships in port.

In addition to traditional methods, machine learning (ML) has significantly advanced the maritime emissions field. Yang et al. (2024) found that existing fuel consumption prediction models lack generalizability and proposed that developing a unified model based on ML is suitable for various vessel types could address this limitation. Furthermore, the application of ML models in unsupervised route planning for maritime autonomous surface ships can effectively reduce carbon emissions (Li and Yang 2023). In their study, Liu, Rong, and Guedes Soares (2023) and Q. Chen et al. (2022) enhanced emission accuracy and reduced model complexity by clustering AIS data and integrating vessel traffic density models with emission models.

With the development of shipping technology, the quantity of data collected on ships is constantly improving. Based on these data, the methods for estimating emissions from ships are also improving. However, the extensive data volume comes with challenges related to data quality. Peng et al. (2020) found that the quality of data was poor since the technical parameters of ships were provided by multiple maritime administrations. The lack of this necessary information made it difficult to calculate ship emissions using an activity-based approach, the author adopted an estimation method based on a sampling method for calculating ship exhaust emission inventories to improve the data quality. Khan et al. (2018) calculated ship emission data for the port of Incheon in October 2014 with a total of 602 ships. Despite the assistance of the Korean maritime department and Incheon port authorities, 263 vessels were directly excluded from the emission calculations due to missing technical parameters data. This would obviously result in underestimating the actual values. The same issue has been extensively investigated in studies on emissions from passenger ships (Q. Chen et al. 2021, 2023). The study combined seven databases (Marine traffic, CCS, KR, RS, NK, Clarkson, and BLM-Shipping) and found that the technical parameters of the 952 ships used as subjects were seriously missing. The missing rates of flag, IMO number, year of ship built, Gross Tonnage (GT), ship service speed, and rated power comprised 8.59%, 53.78%, 60.89%, 61.19%, 67.70%, and 67.11%, respectively. In particular, the key indicators for calculating emissions, GT, ship service speed and rated power, are missing at a very high level. Furthermore, the use of high-resolution data for analysis is also prone to missing data. In D. Chen et al. (2017) study, the authors estimated ship emissions in China in 2014, which included 166,546 ships. The coverage of key ship technical parameters ranged from 26.5% to 87.8%, with missing data reaching 70.4% especially for ship service speed. The situation of missing data in the database has not improved significantly over time. In a study of Nepali ports in 2021, the percentage of missing data was 10.55% for power installed onboard and 5.72% for maximum speed (Toscano et al. 2021). These studies show that missing ship data is a common and critical issue in current emission estimation studies in the shipping industry.

In the maritime field, there are few studies on missing data regarding ship technical parameters. For instance, basic design parameters like Dead Weight Tonnage (DWT) and GT are often used to estimate key characteristics such as a ship's ship service speed and main engine power (Charchalis 2013, 2014). There is also imputing of the data by ML and regression models forming a combined model (Kim, Steen, and Muri 2022). Based on the limitations of the database, all of these methods have restrictions in their application. Due to the lack of a harmonized database for these technical parameters. The studies mentioned above could only rely on a single database or on a limited number of samples (Cepowski 2019; Cepowski and Chorab 2021). Therefore, it is important to mention again the IMO. Each member state uploads their maritime registry information, their database is very well developed and the proposed methods for imputing the technical parameters of ships are more relevant to the shipping industry. However, no studies have yet been conducted to test the effectiveness of these data imputation methods in practice.

This paper uses the missing data imputation methods used by IMO in the Fourth Greenhouse Gas Study 2020 (IMO 2020) and the Calculation of the attained energy efficiency existing ship index (EEXI) Resolution (IMO 2021) as test methods. The Felixstowe port and Busan port are analyzed for ship emission estimations. A mix-method based on these two methods was also added to the case study as the third method. The complete data from the Clarkson database was also used as a control group to test the performance of each method for different technical parameters.

3. Experiments

In this section, the detailed process of experiment is introduced, including methodology and description of case study process. We first impute the ship's missing technical parameters of ship service speed and main engine power, using IMO multiple linear regression, IMO curve fitting and mixed methods. Then, we apply a fuel-based bottom-up emission estimation model to calculate carbon emissions in ports. Finally, the Empirical Cumulative Distribution Function and Friedman–Nemenyi Test were used to evaluate their performance and impact on estimation of port CO_2 emissions from ships based on LSGE and Clarkson databases.

3.1. Methods for imputing missing ship technical parameters

3.1.1. Imo-curve fitting

Previous research has demonstrated through simple correlation analyses that there exists a power function relationship between ship technical indicators and DWT (Charchalis 2013, 2014; IMO 2021). In 2021, IMO released a guideline (IMO 2021) that includes estimating missing technical parameters through curve fitting with power functions. This approach is considered to have a stronger advantage in estimating missing maritime technical parameters, such as ship service speed and main engine power. The function to compute the ship service speed (knots) is as follows:

Ship service speed (knots):

$$V = A * DWT^{B} \tag{1}$$

where,
V is the ship's service speed (knots)

Table 1. Factors reated to service speed and main engine power calculation.

Ship type	A	B	C	D
Bulk carrier	10.658	0.027	23.751	0.541
Gas carrier	7.446	0.076	21.470	0.595
Tanker	8.136	0.054	22.842	0.558
Containership	3.240	0.183	0.504	1.030
General cargo ship	2.454	0.188	0.882	0.921
Refrigerated cargo carrier	1.060	0.3152	0.027	1.386
Combination carrier	8.139	0.0538	22.854	0.558
LNG carrier	11.054	0.050	20.710	0.635
Ro-ro cargo ship (vehicle carrier)	16.677	0.018	262.769	0.400
Ro-ro cargo ship	8.079	0.091	37.771	0.635
Ro-ro passenger ship	4.114	0.199	9.134	0.911
Cruise passenger ship having non-conventional propulsion	5.124	0.127	1.355	0.887

(Source: IMO 2021)

A and B are constants and their values depend on the ship type as shown in Table 1.
The function to compute the main engine power (kW) is as follows:

$$P_{ME} = C * DWT^{D} \tag{2}$$

where,
P_{ME} is the power of main engines (kW)
C and D are constants and their values depend on the ship type as shown in Table 1.

3.1.2. Imo-multiple linear regression (IMO-MLR)

IMO's Fourth GHG study (IMO 2020) introduced a revised methodology for imputing missing technical parameters of ships. This methodology is founded on an algorithm developed by University Maritime Advisory Services (UMAS) International Ltd. The algorithm employs multiple linear regression for each ship type, utilizing design parameters such as beam, draft, Length Overall (LOA), and DWT as variables (Cepowski 2019; IMO 2020). Johansson, Jalkanen, and Kukkonen (2017) demonstrated that this formula offers greater accuracy in estimating ship speed and main engine power, and it has been recommended for use by shipping researchers. The following formulas are the multiple regressions used to compute the missing ship service speed and main engine power.

Ship service speed (knots):

$$V = a_1 + a_2 \cdot LOA + a_3 \cdot P_{ME} + a_4 \cdot DWT \tag{3}$$

Main engine power (kW):

$$P_{ME} = b_1 + b_2 \cdot LOA + b_3 \cdot V + b_4 \cdot DWT \tag{4}$$

where,
V: the ship's service speed, measured in nautical miles per hour (knots), in deep water and assuming the weather is calm with no wind and no waves.
P_{ME} is the power of main engines (kW)
LOA is the length overall, measures ship's length (m)
$a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4$ are regression coefficients

3.1.3. The mixed-method

However, as we can see from IMO-multiple linear regression's formula, it cannot be used when both ship service speed and main engine power are missing. IMO-curve fitting requires only DWT for its input variables, so it is not very accurate. However, considering that missing DWT is rare (Kim, Steen, and Muri 2022; Sun, Abouarghoub, and Demir 2025), taking the port of Felixstowe in the case study as an example, none of the ships coming to Felixstowe from 2019 to 2020 are missing

DWT data. Therefore, when encountering a scenario where both ship service speed and main engine power are missing, IMO-curve fitting can be applied to impute the main engine power and then IMO-multiple linear regression can be applied to estimate the ship service speed based on the results of IMO-curve fitting. Thus, a mixed method is created. The formula for this method is presented below. It requires only DWT and LOA as input variables to estimate the missing technical parameters, effectively addressing the issue of incomplete technical parameters. Furthermore, the non-linear regression formula has been linearized, simplifying parameter estimation while preserving the model's capacity to capture non-linear relationships (Asghari et al. 2022; Molnar and Orosz 2024; Yan, Liu, and Wang 2024).

Ship service speed (knots):

$$V = a_1 + a_2 \cdot LOA + a_3 \cdot \left(C \cdot DWT^D\right) + a_4 \cdot DWT \tag{5}$$

Main engine installed power (kW):

$$P_{ME} = b_1 + b_2 \cdot LOA + b_3 \cdot \left(A \cdot DWT^B\right) + b_4 \cdot DWT \tag{6}$$

3.2. Methods for emissions calculation

A fuel-based bottom-up approach is used in this study to estimate the carbon emissions from ships in ports. The method is an advanced approach based on automatic identification systems (AIS) activity data, which has been gradually improved in recent years through extensive use and validation (Schwarzkopf et al. 2021; Spengler and Tovar 2022; Woo and Im 2021). Ship emissions are calculated for each time interval between two consecutive AIS reports using the ship AIS activity data and ship technical parameters. The AIS activity data includes variables such as the IMO number, AIS date and time, status, speed, and so forth. This data is dynamic and changes over time (Ribeiro, Paes, and de Oliveira 2023; Yang et al. 2024). In contrast, the ship technical parameters—such as ship type, main engine power, dead weight tonnage, and gross tonnage—remain relatively fixed after the ships are produced (Kim et al. 2020). The detailed relationship between these variables is illustrated in Figure 1. Based on these relationships, the selected input variables are listed in Table 2, which includes four ship activity variables and seven ship technical parameters. These variables can be used to estimate ship emissions, which primarily stem from fuel consumption by the main engine, auxiliary engines, and boilers (Zhou et al. 2024). Using the four selected ship activity variables, the ship's operation mode and operational time between two AIS signals can be determined, facilitating the estimation of auxiliary engine and boiler power (Chen et al. 2021). The seven selected ship technical parameters provide the main engine power, while emission factors and specific fuel consumption are derived from engine specifications. By integrating engine power, engine load, and their specific fuel consumption, the amount of fuel consumed per unit time during each engine's activity can be obtained. The total emissions can then be calculated based on the emission factor and the activity time. Additionally, if real-time engine load data is not available, it can be estimated using ship service speed, AIS speed, ship design draft, and AIS draft (USEPA 2020). The relationship between the variables shows that for ship emission calculations, ship service speed and main engine power are the two most important quantitative variables for ship technical parameters. The research subjects of the missing parameters in this paper will therefore be limited to ship service speed and main engine power.

The amount of ship emission will be discrepant during different operation mode: cruising, maneuvering, and berthing/anchored (Toscano and Murena 2019). In cruising mode, the ship will travel at its service speed, main engines will operate at their highest load. Therefore, the auxiliary engines only hold lowest loads and boilers will be shut down. The maneuvering mode will be happened when ship close to its destination, it will require ship to travel at the slowest speed. Main engines will operate at low loads, auxiliary engines and boilers will start rising the loads to a high rate. In berthing and anchored mode ship will only use its auxiliary diesel engines maintain ship

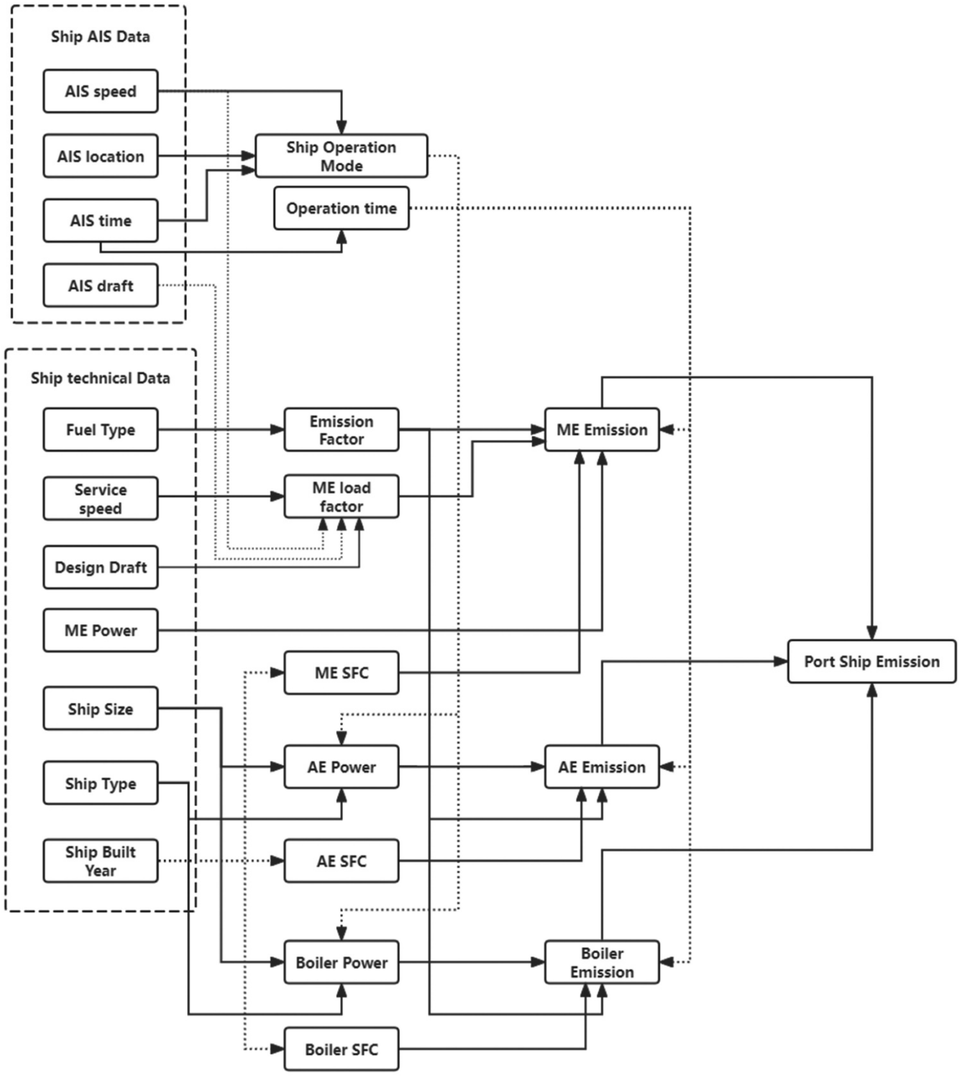

Figure 1. Relationship of ship emission estimation variables.

Table 2. Description of input data for emission calculation.

Variable	Description
AIS speed	Ship's real-time speed provided by AIS transponders.
AIS location	Ship's real-time location provided by AIS transponders, consists of latitude and longitude.
AIS time	Time information when upload AIS data.
AIS draft	Ship's real-time draft provided by AIS transponders.
Fuel type	Type of fuel used on ship.
Ship service speed	The average speed maintained by a ship under normal load and weather conditions.
Main engine power	Main engine maximum power output
Ship design draft	The distance between the sea level and the bottom of hull
Ship size	A number indicative of a ship's cargo carrying capacity.
Ship type	Type of ship.
Ship built year	Ship construction time

equipment's operation. And boilers are operated to keep main engines and fuel systems warm, in case the ship could leave port when there is an unexpected situation (IMO 2014). Therefore, according to previous studies (USEPA (2020); S. Chen et al. 2021, Q. Chen et al. 2021), different engine emissions need to be calculated for each operation mode with the following formulas:

Under cruising:

$$E_{cruising} = (P_{ME} \times LF \times SFC_{ME} + P_{AE} \times SFC_{AE}) \times time \times EF_f \tag{7}$$

$$LF = \left(\frac{V_{AIS}}{\mathrm{V}}\right)^3 \times \left(\frac{D_{AIS}}{D}\right)^{\frac{2}{3}} \tag{8}$$

Under maneuvering:

$$E_{manoeuvring} = (P_{ME} \times LF \times SFC_{ME} + P_{AE} \times SFC_{AE} + P_{boiler} \times SFC_{boiler}) \times time \times EF_f \tag{9}$$

$$LF = \left(\frac{V_{AIS}}{\mathrm{V}}\right)^3 \times \left(\frac{D_{AIS}}{D}\right)^{\frac{2}{3}} \tag{10}$$

Under berthing/anchored:

$$E_{hotelling} = (P_{AE} \times SFC_{AE} + P_{boiler} \times SFC_{boiler}) \times time \times EF_f \tag{11}$$

E_i are emissions by operating mode i (g)
EF_f is fuel-based emission factor, grams of pollutant per gram of fuel consumed (g/g)
LF is load factor of main engines (unitless)
P_{ME}: is the power of main engines (kW)
P_{AE} is the power of auxiliary engines (kW)
P_{boiler} is the power of boiler (kW)
V_{AIS} is AIS speed (knots)
V is ship service speed (knots)
D_{AIS} is AIS draft (m)
D is ship design draft (m)
SFC is certified specific fuel consumption (g/kWh)
time is duration time of this activity (hr)

3.3. Port background

This paper uses three ports of varying sizes as case study samples based on Lloyd's port throughput rankings (Lloyd, 2023). The selected ports are Busan, Los Angeles and Felixstowe, with respective throughputs in 2023 of 23,035,734 TEU, 9,911,155 TEU, and 3,297,000 TEU. These ports adequately represent the range of throughput among the world's top 100 ports. To facilitate a comparison of emissions across different ports, the emission estimation scope is defined as a 20 nautical-mile radius centered on each port, aligning with common practices in maritime research (Chang and Wang 2012; Merk 2014; USEPA 2020; Zhao, Chen, and Lee 2022). Due to the small number of emissions generated by miscellaneous ships in port, the complexity of calculating these emissions and the difficulty of accessing miscellaneous ship data (S. Chen et al. 2021; Murcia González 2021), this study assesses all types of cargo ships arriving in port, except for miscellaneous ships.

Table 3 summarizes the annual ship calls at the Ports of Busan, Los Angeles, and Felixstowe. The first column of the table identifies the case study, where the first two letters denote the port's name, and the last four digits indicate the year of the case. BS refers to Busan, LA to Los Angeles, and FX to Felixstowe. As shown in the table, container ships are the dominant ship type at all three ports, with Los Angeles having the highest proportion at 66%, followed by Felixstowe at 56% and Busan at 48%.

Table 3. Annual ship call type summary 2019–2020.

Case Study	Chemical Tankers	Containers	Dry Bulk	General Cargo	LPG Tankers	Oil Tankers	Other Dry	Other Tankers	Reefers	Ro-Ros
BS_2019	396	8,538	373	2,550	66	3,634	576	68	838	736
BS_2020	336	8,332	528	2,143	96	4,122	484	89	611	770
LA_2019	2	1,003	66	34	0	346	0	0	19	109
LA_2019	0	969	58	38	0	241	0	0	17	81
FX_2019	3	1,168	0	3	0	10	1	6	0	824
FX_2020	3	1,039	0	1	0	12	3	3	0	806

The second most common ship type at Busan and Los Angeles is oil tankers, accounting for approximately 20%. In contrast, the second largest ship type at Felixstowe is Ro–Ro ships, comprising 42%. This is because Felixstowe is the largest and busiest container port in the UK and one of the largest in Europe. The ship types at Busan and Los Angeles are more diverse than those at Felixstowe. Additionally, it is noteworthy that the combined share of General Cargo, Dry Bulk, and Other Dry ships at Busan Port reaches 19%. This distribution reflects Busan's dual role as South Korea's largest container port and its largest bulk port. The varied distributions of ship types across the three ports enhance the reliability of the experimental results.

3.4. Data description

The ship technical data and AIS activity data are mainly provided by LSGE. This database includes information on ship activity from all over the world. Between 2019 and 2020, 5,797,477 AIS activity data records and 42,151 ship calls technical information related to the Port of Busan, Los Angeles and Felixstowe, were collected. We also prepared a complete ship technical parameter database from Clarkson for validation. As shown in Figure 2, in 2019 and 2020, Busan recorded 17,775 and 17,511 ship calls, respectively, of which only 47% included complete ship technical information. Five percent of the data were missing only the service speed, while 32% were missing only ME power. Additionally, 12% of the data were missing both service speed and ME power, presenting challenges for data imputation. Less than 2% of ship calls lacked data DWT. In contrast, Los Angeles and Felixstowe had higher data completeness, with no cases of missing DWT. Los Angeles recorded 1,579 and 1,404 ship calls in 2019 and 2020, respectively, with 85% of port calls containing complete ship technical information. A 5% of the data were missing only the service speed, and 9% were missing only ME power, while fewer than 2% lacked both service speed and ME power. Felixstowe showed a similar pattern to Los Angeles, with 2,015 and 1,867 ship calls in 2019 and 2020, respectively. Among these ship calls, 88% contained complete ship technical information, 4% of the data were missing only the service speed, and 6% were missing only ME power. In 2020, only five cases at Felixstowe involved

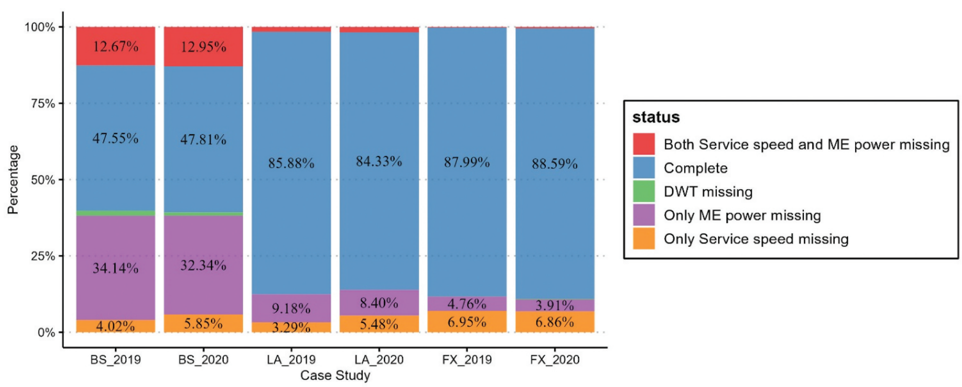

Figure 2. Ship call status summary.

missing both service speed and ME power. The varying patterns of data completeness at the three ports highlight differences in the performance of imputation methods.

3.5. Missing data imputation

After collecting the data, we start to impute the missing data. As mentioned in the Methodology, three different methods will be used to impute the technical parameters of the ships, and after substituting the data into the algorithms we will get the imputed database. At the same time, Clarkson database will be introduced as a control group to see if the distribution of the imputed data is consistent with the original data. Their database has completed technical parameters data for most ships. The data from Clarkson will be regard as actual database. When the data preparation part is finished, the next step will be emission estimation based on these five datasets, which includes three imputation methods' datasets, no impute dataset and actual dataset.

3.6. Emission estimation

In this section, a bottom-up fuel-based method is utilized to estimate port CO_2 emissions from ships. In the process of calculation, several parameters of emission need to be evaluated based on the real-time AIS data. The first parameter needs to be evaluated is the operating mode of the ship. According to the IMO (2020) study and the available AIS parameters, a ship in port operating mode decision matrix has been created. The ship's AIS speed and main engine status are used to determine the ship's operating mode in port to determine the ship's operating mode. The AIS ship status here refers to whether the ship's main engine is in a starting state or not, while the ship's AIS speed level is used to represent the engine load factor of the main engine. Finally, the operating mode of ship is determined based on the conditions of main engine, detailed information is shown in Table 4. The explanation has been mentioned in section 3. It can be seen from the data in Table 5, CO_2 emission factors of the ship obtained according to the fuel type. The auxiliary engine and boiler power of the ship is determined by the type of ship, DWT, and operating modes. The SFC value of the engine can be determined by the ship's build year, fuel type and engine type. Further information can be found in the tables in Appendices 1–3.

The imputed databases are input into the fuel-based bottom-up method combined with the corresponding AIS activity data to calculate CO_2 emissions for each activity of ship. The emissions

Table 4. Ship in port operating mode decision matrix.

AIS ship status	AIS speed	Ship operating mode
Underway Using Engine	<1	At berth
	1–5	Maneuvering
	>5	Cruising
Moored	<1	At berth
	1–3	Anchored
	>3	Maneuvering

(source: IMO 2020)

Table 5. Fuel-based CO_2 emission factor.

Fuel type	Carbon content	EF (g CO_2/g fuel)
HFO	0.8493	3.114
MDO	0.8744	3.206
LNG	0.7500	2.750
Methanol	0.3750	1.375
LSHFO 1.0%	0.8493	3.114

(source: IMO 2020)

are divided by the unit time to obtain the carbon intensity of the port, which allows a time series of emissions to be plotted as a line figure to further analyze the trends in port emissions.

4. Results and discussion

4.1. Results of ship technical parameter imputation

We compared imputed ship technical parameters using three different methods with the actual data. First, Empirical Cumulative Distribution Function (ECDF) plots were employed to observe differences in their distributions. ECDF is the distribution function associated with the empirical measure of a sample (Dekking et al. 2005). This cumulative distribution function is a step function that jumps up by 1/n at each of the n data points. It gives a clear indication of where the value is in the overall percentile. The y-axis of the graph is the cumulative probability, and the x-axis is the range of target data.

Figure 3 illustrates the performance of data imputation for ship service speed across different ports. For Busan, the mixed method produces a distribution that is closest to the actual data distribution. In Los Angeles and Felixstowe, where fewer ship calls lack service speed information, the mixed method and IMO-MLR exhibit comparable performance, both outperforming IMO-curve fitting. The range of missing ship service speeds varies across the three ports. For Busan and Los Angeles, missing speed values fall within the 8–26 knot range, while for Felixstowe, they range from 22 to 26 knots. This discrepancy arises because Busan and Los Angeles are general-purpose ports handling diverse vessel types, whereas Felixstowe is a container port, and this speed range aligns with the average service speed for container ships (Rodrigue 2020).

Figure 4 shows the performance of data imputation for main engine power across the ports. For Busan and Los Angeles, the mixed method performs better than the other methods. However, the results of Felixstowe are more complicated. IMO-curve fitting performs best when main engine power is less than 20,000 kW, whereas IMO-MLR is more effective when main engine power exceeds 20,000 kW. Further analysis is required to evaluate their performance. The range of missing values for main engine power is similar across the three ports, reflecting various vessel sizes.

Next, we use the Friedman–Nemenyi test to compare the performance of different methods on each estimation value. Friedman test is a nonparametric test, which is used to compare three or more paired groups (Friedman 1937). It gives a conclusion on whether there is a difference between the performance of the algorithms or models. If there is a difference, a post-hoc test is also required to find out which algorithms or models have statistical differences in performance between them. Therefore, the Nemenyi test is chosen as the post-hoc test in this paper. The Nemenyi test is used to compare the performance of algorithms against each other Nemenyi *(1963)*. The method compares the average ranking of each algorithm against the critical difference (CD). If the average ranking of two algorithms is less than the CD, it implies there is no statistically significant difference between

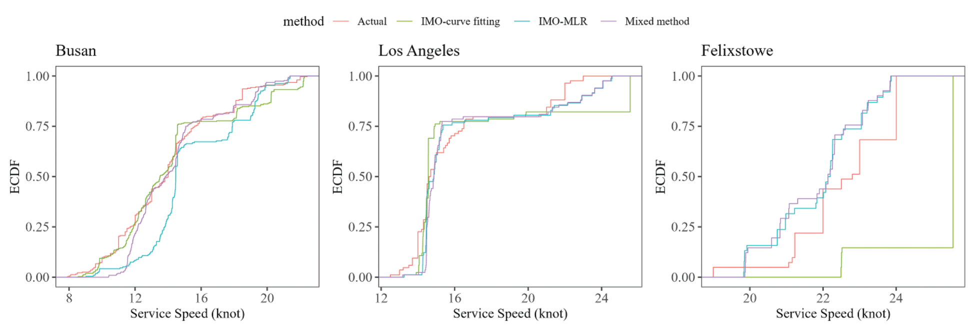

Figure 3. ECDF for imputed ship service speed.

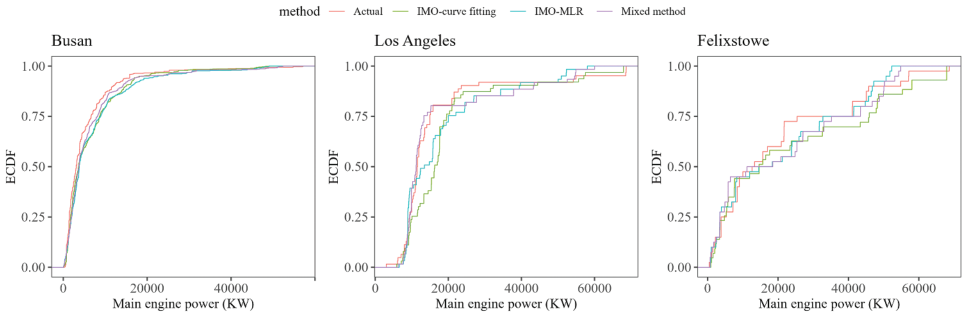

Figure 4. ECDF for imputed main engine power.

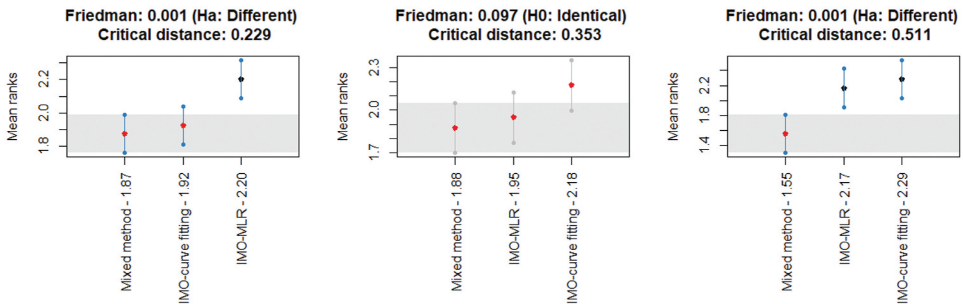

Figure 5. Friedman-Nemenyi test for ship service speed (left: Busan; mid: Los Angeles; right: Felixstowe).

them. In this paper, we will use the 'nemenyi ()' function of the 'tsutils' package in R, which produced by Kourentzes, to calculate and visualise the results of the Friedman–Nemenyi Test. We input the absolute error of the three methods with real values to the function to get the ranking of the models. A lower score means less error and better accuracy of the model.

The Figures 5 and 6 show the results of Friedman–Nemenyi Test. If one method's result is similar to other method, their CD area will overlap and the dot in the middle of each line will be red. The middle dot represents the mean rank value. The left and right dot of each line is the CD area boundary. The value on the x-axis is the ranking score of the model. Figure 5 presents the Friedman–Nemenyi test results for ship service speed, which corroborate the findings from the ECDF plots. The mixed method has the best performance across all ports, particularly at Felixstowe, where it is significantly more accurate than other methods. Figure 6 displays the Friedman–

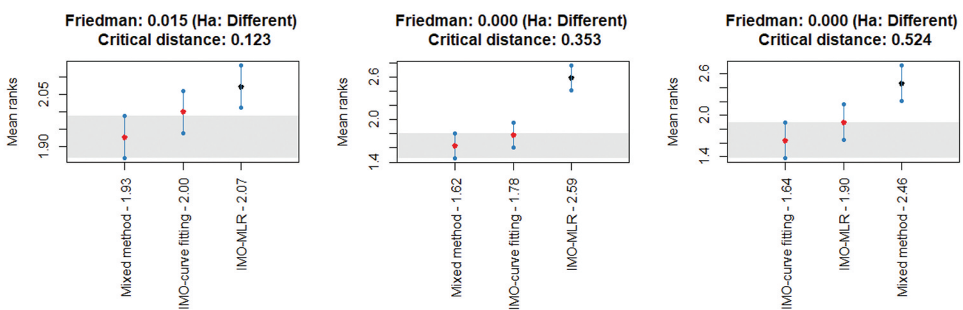

Figure 6. Friedman-Nemenyi test for main engine power (left: Busan; mid: Los Angeles; right: Felixstowe).

Nemenyi test results for main engine power. In Busan and Los Angeles, the mixed-method remains the most suitable method. However, at Felixstowe, the IMO-curve fitting method performs better. This is due to the port has limited ship types and smaller sample size. Overall, the mixed method is the most effective approach, but further validation is required by applying these three methods within the emission estimation model. This is because the coverage rate is also a critical performance indicator when implementing the model in practice (Sun, Abouarghoub, and Demir 2025).

4.2. Results of port emission estimation

After inputting the imputed data and no impute data of Felixstowe and Busan port into the fuel-based bottom-up method, we get the port CO_2 emissions from ships per day. Figure 7 illustrates the

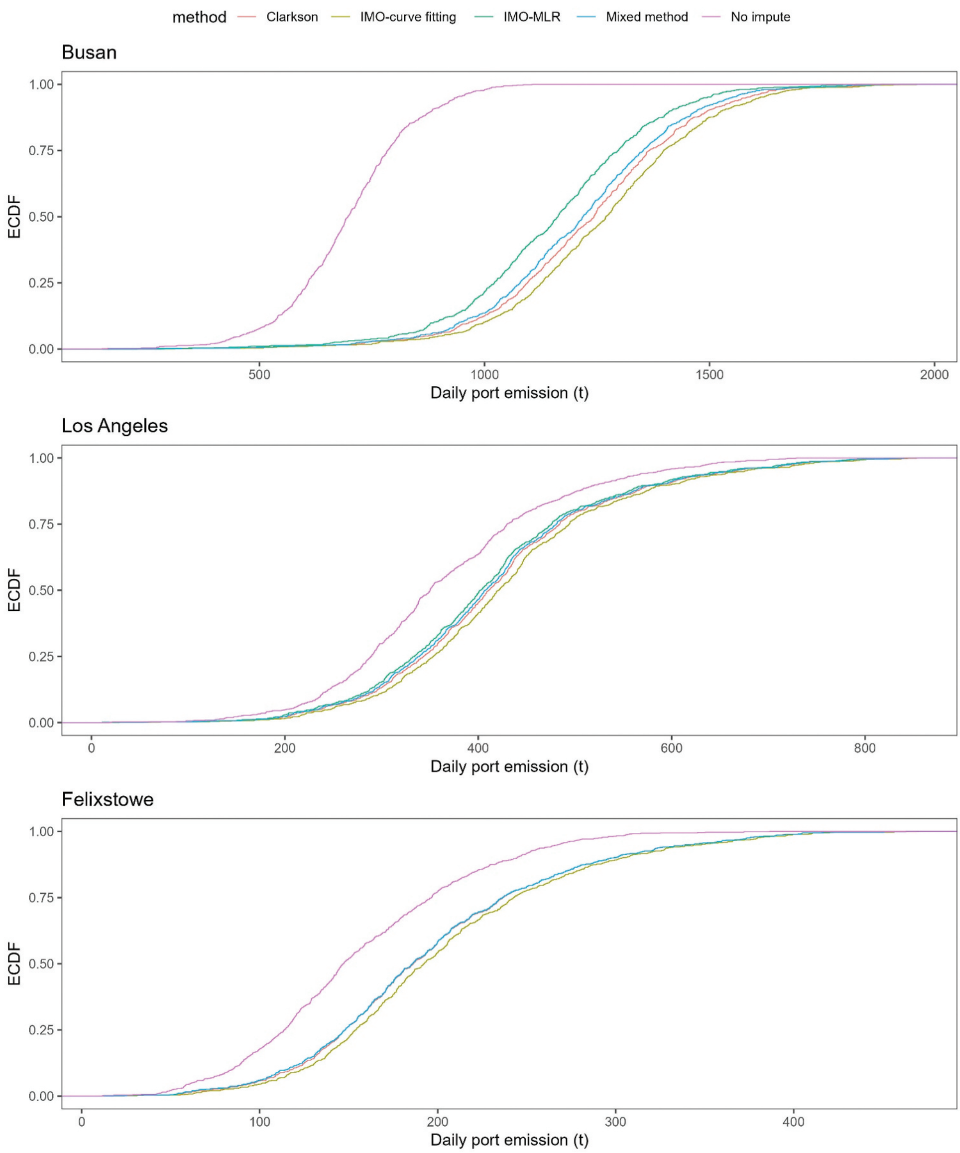

Figure 7. ECDF for port daily CO_2 emission.

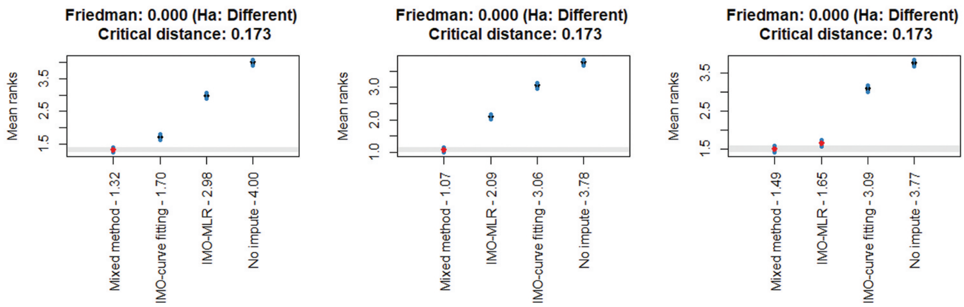

Figure 8. Friedman-Nemenyi test for daily CO_2 emission (left: Busan; mid: Los Angeles; right: Felixstowe).

Table 6. Port CO_2 emissions (t) from ships 2019–2020.

Port	year	Actual	No impute	IMO- MLR	IMO-curve fitting	Mixed method
Busan	2019	470,574	259,644	442,069	481,328	463,237
Busan	2020	503,269	289,628	469,740	515,304	490,251
Felixstowe	2019	70,171	56,984	69,810	72,337	70,358
Felixstowe	2020	81,619	63,134	80,751	83,773	81,300
Los Angeles	2019	168,147	146,584	165,207	171,778	167,031
Los Angeles	2020	158,821	136,281	155,573	162,456	157,399

ECDF of port daily CO_2 emissions in 2019 and 2020, which covers a total of 731 days. It indicates a significant gap between the results without imputation and those with imputation. The distributions produced by the three imputation methods also differ. The daily emissions estimated using the mixed-method are closest to the actual data. The distribution for IMO-MLR is noticeably lower compared to the other two methods and the actual data, while the IMO-curve fitting distribution is slightly higher. We conducted a Friedman–Nemenyi test for daily CO_2 emissions, and Figure 8 provides the detailed results. The results of the Friedman–Nemenyi test are consistent with the ECDF plot, with the mixed-method having the smallest estimation error for daily emissions after imputation.

Table 6 gives a clearer view of the impact of imputation methods on total emissions. First, we need to validate the accuracy of the data. For Los Angeles, where official emissions estimates are available, the emissions calculated using actual ship technical parameters are consistent with the official figures (LA 2024). This demonstrates that our emissions calculations are aligned with realistic values. The table underscores the importance of selecting an appropriate imputation method. Using IMO-MLR results in underestimations of annual emissions by 2% to 6%. This is because IMO-multiple linear regression is not effective at handling cases where both ship service speed and main engine power are missing simultaneously. This limitation becomes more significant in cases with more missing data, such as Busan. There is no significant difference in accuracy between mixed method and IMO-curve fitting, but IMO-curve fitting overestimates annual emissions by 2%, while mixed-method keeps the error within 0.6%. Hence, the mixed method also performs best in annual emissions calculations. Furthermore, data quality has a significant impact on emission estimates. Without using imputation methods to improve data quality, emissions would be underestimated by at least 42% in Busan, 12.82% in Los Angeles, and 18.79% in Felixstowe. This highlights the critical role of imputation in ensuring reliable emission estimation.

5. Conclusions

The purpose of the current study was to quantify the impact of data quality on port emission. This study compares the performance of three methods used for imputing missing ship technical parameters in port emissions calculations. The Port of Busan, Los Angeles, and Felixstowe were selected as case studies for this paper. Based on 42,151 ship calls with 5,797,477 AIS activity records, the fuel-based bottom-up method was used to calculate ship emissions in the port, after imputing missing technical parameters. The ECDF plots and the Friedman–Nemenyi Test were used to quantify the impact of imputing results. The main findings of this study are as follows. First, the calculations show that if the missing data is not imputed to improve the quality of the data, it will have a significant impact on the calculation of port ship emissions. In the Port of Los Angeles and Felixstowe, if raw data is not processed, annual CO_2 emissions are underestimated by 13% and 14%, respectively, compared to actual values. The situation would be even worse in larger ports. In the Port of Busan, using unprocessed data would result in an underestimation of 42% of CO_2 emissions. Secondly, for annual port CO_2 emissions from ships, the mixed method performs best across all case studies. The IMO-curve fitting method tends to overestimate CO_2 emissions by 2%, while the IMO-MLR method underestimates them by 2% to 6%. As the error rate increases, the underestimation with the IMO-MLR method becomes more significant. The ability of the imputation method to handle missing data is also a crucial indicator for evaluating the model's performance. Thirdly, for daily port CO_2 emissions from ships, the mixed-method consistently provides the smallest estimation error after imputation. Finally, the three imputation methods vary in performance when applied to the technical parameters of ships. The mixed-method generally performs best in estimating ship service speed and main engine power, while the IMO-curve fitting method is more accurate for estimating main engine power in container ships. However, several limitations need to be noted regarding the present study. Although the ports of Busan, Los Angeles and Felixstowe are different types of ports, the number of ports is only three. It is necessary to include more ports in the case study for more generalize findings. Furthermore, the parameters derived in this paper only focus on the service speed and main engine power of the ships, and more parameters can be added for analysis later. The current study employs traditional methods for data imputation, which may not capture the complexities associated with missing data in this context.

Future research should address the identified limitations to advance the field. First, a systematic study can be conducted to examine the impact of data quality on ports of different sizes and types. Second, more accurate imputation models for missing data can be developed using machine learning models. Third, a unified evaluation standard can be created to assess maritime data quality levels. Finally, apart from port emission estimation, data quality issues in other maritime traffic research can also be further investigated.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The support provided by China Scholarship Council (No. 2021108060143) and Cardiff University is acknowledged.

ORCID

Ruikai Sun http://orcid.org/0000-0002-9015-3515
Wessam Abouarghoub http://orcid.org/0000-0002-1647-1291
Emrah Demir http://orcid.org/0000-0002-4726-2556
Andrew Potter http://orcid.org/0000-0002-3157-9735

References

Asghari, M., A. M. Fathollahi-Fard, S. M. J. Mirzapour Al-E-Hashem, and M. A. Dulebenets. 2022. "Transformation and Linearization Techniques in Optimization: A State-Of-The-Art Survey." *Mathematics* 10 (2): 283.

Cepowski, T. 2019. "Determination of Regression Formulas for Main Tanker Dimensions at the Preliminary Design Stage." *Ships and Offshore Structures* 14 (3): 320–330.

Cepowski, T., and P. Chorab. 2021. "Determination of Design Formulas for Container Ships at the Preliminary Design Stage Using Artificial Neural Network and Multiple Nonlinear Regression." *Ocean Engineering* 238 (1): 109727. https://doi.org/10.1016/j.oceaneng.2021.109727.

Chang, C.-C., and C.-M. Wang. 2012. "Evaluating the Effects of Green Port Policy: Case Study of Kaohsiung Harbor in Taiwan." *Transportation Research Part D: Transport & Environment* 17 (3): 185–189.

Charchalis, A. 2013. "Dimensional Constraints in Ship Design." *Journal of KONES* 20 (1): 29–34.

Charchalis, A. 2014. "Determination of Main Dimensions and Estimation of Propulsion Power of a Ship." *Journal of KONES* 21 (2): 39–44.

Chen, D., X. Wang, Y. Li, J. Lang, Y. Zhou, X. Guo, and Y. Zhao. 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, Q., Y.-E. Ge, Y.-Y. Lau, M. A. Dulebenets, A. K. Ng, and Z. Wang. 2022. "Capturing Effects of COVID-19 on Global Cruise Vessel Density and Emissions." *Available at SSRN 4211079*.

Chen, Q., Y.-E. Ge, Y.-Y. Lau, M. A. Dulebenets, X. Sun, T. Kawasaki, A. Mellalou, and X. Tao. 2023. "Effects of COVID-19 on Passenger Shipping Activities and Emissions: Empirical Analysis of Passenger Ships in Danish Waters." *Maritime Policy & Management* 50 (6): 776–796.

Chen, Q., Y.-Y. Lau, Y.-E. Ge, M. A. Dulebenets, T. Kawasaki, and A. K. Ng. 2021. "Interactions Between Arctic Passenger Ship Activities and Emissions." *Transportation Research Part D: Transport & Environment* 97 (19): 102925. https://doi.org/10.1016/j.trd.2021.102925.

Chen, S., Q. Meng, P. Jia, and H. Kuang. 2021. "An Operational-Mode-Based Method for Estimating Ship Emissions in Port Waters." *Transportation Research Part D: Transport & Environment* 101 (1): 103080. https://doi.org/10.1016/j.trd.2021.103080.

Dekking, F. M., C. Kraaikamp, H. Paul Lopuhaä, and L. Erwin Meester. 2005. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Vol. 488. London: Springer.

Doundoulakis, E., and S. Papaefthimiou. 2022. "A Comparative Methodological Approach for the Calculation of Ships Air Emissions and Fuel-Energy Consumption in Two Major Greek Ports." *Maritime Policy & Management* 49 (8): 1135–1154.

Eyring, V., I. S. Isaksen, T. Berntsen, W. J. Collins, J. J. Corbett, O. Endresen, R. G. Grainger, J. Moldanova, H. Schlager, and D. S. Stevenson. 2010. "Transport Impacts on Atmosphere and Climate: Shipping." *Atmospheric Environment* 44 (37): 4735–4771.

Ezzine, I., and L. Benhlima. 2018. "A Study of Handling Missing Data Methods for Big Data." Paper presented at the 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), Marrakech, Morocco.

Friedman, M. 1937. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Journal of the American Statistical Association* 32 (200): 675–701.

IMO. 2014. *Port Emissions ToolkitGuide No.1: Assessment of Port Emissions*. International Maritime Organization.

IMO. 2020. *Fourth IMO Greenhouse Gas Study 2020*. International Maritime Organization.

IMO. 2021. *Calculation of the Attained Energy Efficiency Existing Ship Index (EEXI) Resolution MEPC.333(76)*. International Maritime Organization.

Jalkanen, J.-P., A. Brink, J. Kalli, H. Pettersson, J. Kukkonen, and T. Stipa. 2009. "A Modelling System for the Exhaust Emissions of Marine Traffic and Its Application in the Baltic Sea Area." *Atmospheric Chemistry & Physics* 9 (23): 9209–9223.

Johansson, L., J.-P. Jalkanen, and J. Kukkonen. 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.

Khan, S., Y.-T. Chang, S. Lee, and K.-S. Choi. 2018. "Assessment of Greenhouse Gas Emissions from Ships Operation at the Port of Incheon Using AIS." *Journal of Korea Port Economic Association* 34 (1): 65–80.

Kim, S. H., M. I. Roh, M. J. Oh, S. W. Park, and I. I. Kim. 2020. "Estimation of Ship Operational Efficiency from AIS Data Using Big Data Technology." *International Journal of Naval Architecture and Ocean Engineering* 12:440–454. https://doi.org/10.1016/j.ijnaoe.2020.03.007.

Kim, Y., S. Steen, and H. Muri. 2022. "A Novel Method for Estimating Missing Values in Ship Principal Data." *Ocean Engineering* 251 (3): 110979. https://doi.org/10.1016/j.oceaneng.2022.110979.

LA. 2024. "Annual Inventory of Air Emissions." *The Port of Los Angeles*.

Lall, R., and T. Robinson. 2022. "The MIDAS Touch: Accurate and Scalable Missing-Data Imputation with Deep Learning." *Political Analysis* 30 (2): 179–196.

Li, H., and Z. Yang. 2023. "Incorporation of AIS Data-Based Machine Learning into Unsupervised Route Planning for Maritime Autonomous Surface Ships." *Transportation Research Part E: Logistics & Transportation Review* 176:103171. https://doi.org/10.1016/j.tre.2023.103171.

Liu, D., H. Rong, and C. Guedes Soares. 2023. "Shipping Route Modelling of AIS Maritime Traffic Data at the Approach to Ports." *Ocean Engineering* 289 (1): 115868. https://doi.org/10.1016/j.oceaneng.2023.115868.

Lloyd. "One Hundred Container Ports 2023." *Lloyd's list*, Accessed 02/12/2024. https://www.lloydslist.com/one-hundred-container-ports-2023.

Merk, O. 2014. "Shipping Emissions in Ports."

Miola, A., and B. Ciuffo. 2011. "Estimating Air Emissions from Ships: Meta-Analysis of Modelling Approaches and Available Data Sources." *Atmospheric Environment* 45 (13): 2242–2251.

Molnar, T. G., and G. Orosz. 2024. "Destroying Phantom Jams with Connectivity and Automation: Nonlinear Dynamics and Control of Mixed Traffic." *Transportation Science* 58 (6): 1319–1334. https://doi.org/10.1287/trsc.2023.0498.

Murcia González, J. C. 2021. "Analysis and Measurement of SOx, CO2, PM and NOx Emissions in Port Auxiliary Vessels." *Environmental Monitoring and Assessment* 193 (6). https://doi.org/10.1007/s10661-021-09145-4.

Nemenyi, P. B. 1963. *Distribution-Free Multiple Comparisons*. ProQuest Dissertations & Theses Global: Princeton University.

Ng, S. K., C. Loh, C. Lin, V. Booth, J. W. Chan, A. C. Yip, Y. Li, and A. K. Lau. 2013. "Policy Change Driven by an AIS-Assisted Marine Emission Inventory in Hong Kong and the Pearl River Delta." *Atmospheric Environment* 76:102–112. https://doi.org/10.1016/j.atmosenv.2012.07.070.

Peng, X., Y. Wen, L. Wu, C. Xiao, C. Zhou, and D. Han. 2020. "A Sampling Method for Calculating Regional Ship Emission Inventories." *Transportation Research Part D: Transport & Environment* 89:102617. https://doi.org/10.1016/j.trd.2020.102617.

Ribeiro, C. V., A. Paes, and D. de Oliveira. 2023. "AIS-Based Maritime Anomaly Traffic Detection: A Review." *Expert Systems with Applications* 231:120561. https://doi.org/10.1016/j.eswa.2023.120561.

Rodrigue, J.-P. 2020. *The Geography of Transport Systems*. London: Routledge.

Schwarzkopf, D. A., R. Petrik, V. Matthias, M. Quante, E. Majamäki, and J. P. Jalkanen. 2021. "A Ship Emission Modeling System with Scenario Capabilities." *Atmospheric Environment X*: 12. https://doi.org/10.1016/j.aeaoa.2021.100132.

Spengler, T., and B. Tovar. 2022. "Environmental Valuation of In-Port Shipping Emissions per Shipping Sector on Four Spanish Ports." *Marine Pollution Bulletin* 178 (1): 113589. https://doi.org/10.1016/j.marpolbul.2022.113589 .

Sun, R., W. Abouarghoub, and E. Demir. 2025. "Enhancing Data Quality in Maritime Transportation: A Practical Method for Imputing Missing Ship Static Data." *Ocean Engineering* 315 (1): 119722. https://doi.org/10.1016/j.oceaneng.2024.119722.

Sun, R., W. Abouarghoub, E. Demir, and A. Potter. 2025. "A Comprehensive Analysis of Strategies for Reducing GHG Emissions in Maritime Ports." *Marine Policy* 171:106455. https://doi.org/10.1016/j.marpol.2024.106455.

Toscano, D., and F. Murena. 2019. "Atmospheric Ship Emissions in Ports: A Review. Correlation with Data of Ship Traffic." *Atmospheric Environment X* 4 (5): 100050. https://doi.org/10.1016/j.aeaoa.2019.100050.

Toscano, D., F. Murena, F. Quaranta, and L. Mocerino. 2021. "Assessment of the Impact of Ship Emissions on Air Quality Based on a Complete Annual Emission Inventory Using AIS Data for the Port of Naples." *Ocean Engineering* 232 (5): 109166. https://doi.org/10.1016/j.oceaneng.2021.109166.

Tran, N. K., J. Siu Lee Lam, H. Jia, and R. Adland. 2022. "Emissions from Container Vessels in the Port of Singapore." *Maritime Policy & Management* 49 (3): 306–322.

Tzannatos, E. 2010. "Cost Assessment of Ship Emission Reduction Methods at Berth: The Case of the Port of Piraeus, Greece." *Maritime Policy & Management* 37 (4): 427–445.

UNCTAD. 2024. "Review of Maritime Transport 2024." In *United Nations Conference on Trade and Development*, 4. New York and Geneva: United Nations.

USEPA. 2020. *Ports Emissions Inventory Guidance: Methodologies for Estimating Port-Related and Goods Movement Mobile Source Emissions*. Washington, DC, USA: United States for Environmental Protection Agency.

Wang, H., X. Zhuge, G. Strazdins, Z. Wei, G. Li, and H. Zhang. 2016. "Data Integration and Visualisation for Demanding Marine Operations." *Paper Presented at the OCEANS 2016-Shanghai.*

Wang, L., and Y. Li. 2023. "Estimation Methods and Reduction Strategies of Port Carbon Emissions-What Literatures Say?" *Marine Pollution Bulletin* 195 (7): 115451. https://doi.org/10.1016/j.marpolbul.2023.115451.

Woo, D., and N. Im. 2021. "Spatial Analysis of the Ship Gas Emission Inventory in the Port of Busan Using Bottom-Up Approach Based on AIS Data." *Journal of Marine Science and Engineering* 9 (12): 1457.

Yan, R., Y. Liu, and S. Wang. 2024. "A Data-Driven Optimization Approach to Improving Maritime Transport Efficiency." *Transportation Research Part B: Methodological* 180 (1): 102887. https://doi.org/10.1016/j.trb.2024.102887.

Yang, Y., Y. Liu, G. Li, Z. Zhang, and Y. Liu. 2024. "Harnessing the Power of Machine Learning for AIS Data-Driven Maritime Research: A Comprehensive Review." *Transportation Research Part E: Logistics & Transportation Review* 183 (4): 103426. https://doi.org/10.1016/j.tre.2024.103426.

Yu, Y., R. Sun, Y. Sun, and Y. Shu. 2022. "Integrated Carbon Emission Estimation Method and Energy Conservation Analysis: The Port of Los Angles Case Study." *Journal of Marine Science and Engineering* 10 (6): 717.
Yu, Y., R. Sun, Y. Sun, J. Wu, and W. Zhu. 2022. "China's Port Carbon Emission Reduction: A Study of Emission-Driven Factors." *Atmosphere* 13 (4): 550.
Zhao, T., M. Chen, and H. Lee. 2022. "A Study on the Framework for Estimating Ship Air Pollutant Emissions—Focusing on Ports of South Korea." *Atmosphere* 13 (7): 1141.
Zhou, S., A. Sun, C. Lou, P. Zhou, H. Xi, M. Shreka, H. Wang, Y. Zhu, and Y. Feng. 2024. "Gaseous and Particulate Pollutant Emissions from Ocean-Going Tankers in the Context of Carbon Reduction: Main Engine, Auxiliary Engine, and Auxiliary Boiler." *Energy (Oxford)* 313 (12): 133831. https://doi.org/10.1016/j.energy.2024.133831.
Zis, T., R. Jacob North, P. Angeloudis, W. Yotto Ochieng, and M. Geoffrey Harrison Bell. 2014. "Evaluation of Cold Ironing and Speed Reduction Policies to Reduce Ship Emissions Near and at Ports." *Maritime Economics & Logistics* 16 (4): 371–398. https://doi.org/10.1057/mel.2014.6.

Appendix 1. Boiler power output

			Boiler Power Output (kW)			
Ship Type	Size Range	unit	At berth	Anchored	Manoeuvring	Sea
Bulk carrier	0–9999	dwt	70	70	60	0
Bulk carrier	10000–34999	dwt	70	70	60	0
Bulk carrier	35000–59999	dwt	130	130	120	0
Bulk carrier	60000–99999	dwt	260	260	240	0
Bulk carrier	100000–199999	dwt	260	260	240	0
Bulk carrier	200000+	dwt	260	260	240	0
Chemical tanker	0–4999	dwt	670	160	130	0
Chemical tanker	5000–9999	dwt	670	160	130	0
Chemical tanker	10000–19999	dwt	1000	240	200	0
Chemical tanker	20000–39999	dwt	1350	320	270	0
Chemical tanker	40000+	dwt	1350	320	270	0
Container	0–999	TEU	250	250	240	0
Container	1000–1999	TEU	340	340	310	0
Container	2000–2999	TEU	460	450	430	0
Container	3000–4999	TEU	480	480	430	0
Container	5000–7999	TEU	590	580	550	0
Container	8000–11999	TEU	620	620	540	0
Container	12000–14499	TEU	630	630	630	0
Container	14500–19999	TEU	630	630	630	0
Container	20000+	TEU	700	700	700	0
General cargo	0–4999	dwt	0	0	0	0
General cargo	5000–9999	dwt	110	110	100	0
General cargo	10000–19999	dwt	150	150	130	0
General cargo	20000+	dwt	150	150	130	0
Liquefied gas tanker	0–49999	cbm	1000	200	200	100
Liquefied gas tanker	50000–99999	cbm	1000	200	200	100
Liquefied gas tanker	100000–199999	cbm	1500	300	300	150
Liquefied gas tanker	200000+	cbm	3000	600	600	300
Oil tanker	0–4999	dwt	500	100	100	0
Oil tanker	5000–9999	dwt	750	150	150	0
Oil tanker	10000–19999	dwt	1250	250	250	0
Oil tanker	20000–59999	dwt	2700	270	270	270
Oil tanker	60000–79999	dwt	3250	360	360	280
Oil tanker	80000–119999	dwt	4000	400	400	280
Oil tanker	120000–199999	dwt	6500	500	500	300
Oil tanker	200000+	dwt	7000	600	600	300
Other liquids tankers	0–999	dwt	1000	200	200	100
Other liquids tankers	1000+	dwt	1000	200	200	100
Ferry-pax only	0–299	gt	0	0	0	0
Ferry-pax only	300–999	gt	0	0	0	0
Ferry-pax only	1000–1999	gt	0	0	0	0
Ferry-pax only	2000+	gt	0	0	0	0
Cruise	0–1999	gt	1100	950	980	0
Cruise	2000–9999	gt	1100	950	980	0
Cruise	10000–59999	gt	1100	950	980	0

(*Continued*)

(Continued).

			Boiler Power Output (kW)			
Ship Type	Size Range	unit	At berth	Anchored	Manoeuvring	Sea
Cruise	60000–99999	gt	1100	950	980	0
Cruise	100000–149999	gt	1100	950	980	0
Cruise	150000+	gt	1100	950	980	0
Ferry-RoPax	0–1999	gt	260	250	170	0
Ferry-RoPax	2000–4999	gt	260	250	170	0
Ferry-RoPax	5000–9999	gt	260	250	170	0
Ferry-RoPax	10000–19999	gt	390	380	260	0
Ferry-RoPax	20000+	gt	390	380	260	0
Refrigerated bulk	0–1999	dwt	270	270	270	0
Refrigerated bulk	2000–5999	dwt	270	270	270	0
Refrigerated bulk	6000–9999	dwt	270	270	270	0
Refrigerated bulk	10000+	dwt	270	270	270	0
Ro-Ro	0–4999	dwt	260	250	170	0
Ro-Ro	5000–9999	dwt	260	250	170	0
Ro-Ro	10000–14999	dwt	390	380	260	0
Ro-Ro	15000+	dwt	390	380	260	0
Vehicle	0–9999	gt	310	300	250	0
Vehicle	10000–19999	gt	310	300	250	0
Vehicle	20000+	gt	310	300	250	0
Yacht		gt	0	0	0	0
Service-tug		gt	0	0	0	0
Miscellaneous-fishing		gt	0	0	0	0
Offshore		gt	0	0	0	0
Service-other		gt	0	0	0	0
Miscellaneous-other		gt	110	110	90	0

(source: IMO 2020)

Appendix 2. Auxiliary engine power output

			Auxiliary Engine Power Output (kW)			
Ship Type	Size Range	unit	At berth	Anchored	Manoeuvring	Sea
Bulk carrier	0–9999	dwt	110	180	500	190
Bulk carrier	10000–34999	dwt	110	180	680	190
Bulk carrier	35000–59999	dwt	150	250	1100	260
Bulk carrier	60000–99999	dwt	240	400	1100	410
Bulk carrier	100000–199999	dwt	240	400	1100	410
Bulk carrier	200000+	dwt	240	400	1100	410
Chemical tanker	0–4999	dwt	110	170	190	200
Chemical tanker	5000–9999	dwt	330	490	560	580
Chemical tanker	10000–19999	dwt	330	490	560	580
Chemical tanker	20000–39999	dwt	790	550	900	660
Chemical tanker	40000+	dwt	790	550	900	660
Container	0–999	TEU	370	450	790	410
Container	1000–1999	TEU	820	910	1750	900
Container	2000–2999	TEU	610	910	1900	920
Container	3000–4999	TEU	1100	1350	2500	1400
Container	5000–7999	TEU	1100	1400	2800	1450
Container	8000–11999	TEU	1150	1600	2900	1800
Container	12000–14499	TEU	1300	1800	3250	2050
Container	14500–19999	TEU	1400	1950	3600	2300
Container	20000+	TEU	1400	1950	3600	2300
General cargo	0–4999	dwt	90	50	180	60
General cargo	5000–9999	dwt	240	130	490	180
General cargo	10000–19999	dwt	720	370	1450	520
General cargo	20000+	dwt	720	370	1450	520
Liquefied gas tanker	0–49999	cbm	240	240	360	240
Liquefied gas tanker	50000–99999	cbm	1700	1700	2600	1700
Liquefied gas tanker	100000–199999	cbm	2500	2000	2300	2650
Liquefied gas tanker	200000+	cbm	6750	7200	7200	6750
Oil tanker	0–4999	dwt	250	250	375	250
Oil tanker	5000–9999	dwt	375	375	560	375
Oil tanker	10000–19999	dwt	690	500	580	490
Oil tanker	20000–59999	dwt	720	520	600	510
Oil tanker	60000–79999	dwt	620	490	770	560
Oil tanker	80000–119999	dwt	800	640	910	690
Oil tanker	120000–199999	dwt	2500	770	1300	860
Oil tanker	200000+	dwt	2500	770	1300	860
Other liquids tankers	0–999	dwt	500	500	750	500
Other liquids tankers	1000+	dwt	500	500	750	500
Ferry-pax only	0–299	gt	190	190	190	190
Ferry-pax only	300–999	gt	190	190	190	190
Ferry-pax only	1000–1999	gt	190	190	190	190
Ferry-pax only	2000+	gt	520	520	520	520
Cruise	0–1999	gt	450	450	580	450
Cruise	2000–9999	gt	450	450	580	450
Cruise	10000–59999	gt	3500	3500	5500	3500
Cruise	60000–99999	gt	11500	11500	14900	11500

(*Continued*)

(Continued).

Ship Type	Size Range	unit	Auxiliary Engine Power Output (kW)			
			At berth	Anchored	Manoeuvring	Sea
Cruise	100000–149999	gt	11500	11500	14900	11500
Cruise	150000+	gt	11500	11500	14900	11500
Ferry-RoPax	0–1999	gt	105	105	105	105
Ferry-RoPax	2000–4999	gt	330	330	330	330
Ferry-RoPax	5000–9999	gt	670	670	670	670
Ferry-RoPax	10000–19999	gt	1100	1100	1100	1100
Ferry-RoPax	20000+	gt	1950	1950	1950	1950
Refrigerated bulk	0–1999	dwt	520	570	560	570
Refrigerated bulk	2000–5999	dwt	1100	1200	1150	1200
Refrigerated bulk	6000–9999	dwt	1500	1650	1600	1650
Refrigerated bulk	10000+	dwt	2850	3100	3000	3100
Ro-Ro	0–4999	dwt	750	430	1300	430
Ro-Ro	5000–9999	dwt	1100	680	2100	680
Ro-Ro	10000–14999	dwt	1200	950	2700	950
Ro-Ro	15000+	dwt	1200	950	2700	950
Vehicle	0–9999	gt	800	500	1100	500
Vehicle	10000–19999	gt	850	550	1400	510
Vehicle	20000+	gt	850	550	1400	510
Yacht		gt	130	130	130	130
Service-tug		gt	100	80	210	80
Miscellaneous-fishing		gt	200	200	200	200
Offshore		gt	320	320	320	320
Service-other		gt	220	220	220	220
Miscellaneous-other		gt	150	150	430	410

(source: IMO 2020)

Appendix 3. Engine specific fuel consumption

Engine Type	Fuel Type	before 1983	1984–2000	2001+
SSD	HFO	205	185	175
	MDO	190	175	165
	MeOH	–	–	350
MSD	HFO	215	195	185
	MDO	200	185	175
	MeOH	–	–	370
HSD	HFO	225	205	195
	MDO	210	190	185
LNG–Otto (dual–fuel, medium–speed)	LNG	–	173	156
LNG–Otto (dual–fuel, slow–speed)	LNG	–	–	148
LNG–Diesel (dual–fuel)	LNG	–	–	135
LBSI	LNG	–	156	156
Gas Turbines	HFO	305	305	305
	MDO	300	300	300
	LNG	–	–	203
Steam Turbines (and boilers)	HFO	340	340	340
	MDO	320	320	320
	LNG	285	285	285
Auxiliary engines	HFO	225	205	195
	MDO	210	190	185
	LNG	–	173	156

(source: IMO 2020)