Prepublished Copy

NATURE ENVIRONMENT AND POLLUTION TECHNOLOGY

Article In Press

PEER-REVIEWED PREPUBLISHED MANUSCRIPT

This is a peer-reviewed prepublished version of the paper to be published in **Vol. 26, No. 1, March 2027.**

Applying LightGBM to Enhance the Accuracy of Emission Inventory for the Hai An Port: A Premise for Developing Emission Mitigation Strategies

Ngo Nhu Tai, Le Cong Doanh, Doan Le Nam Khanh, Trinh Quoc Anh, Phan Van Hung and Nguyen Thanh Son

RECEIVED
25 May 2026

CITE THIS ARTICLE
Tai, N.N., Doanh, L.C., Khanh, D.L.N., Anh, T.Q., Van Hung, P. and Son, N.T., 2027. Applying LightGBM to Enhance the Accuracy of Emission Inventory for the Hai An Port: A Premise for Developing Emission Mitigation Strategies. *Nature Environment and Pollution Technology*, 26(1), D1942.

https://doi.org/10.46488/NEPT.2027.v26i01.D1942

EARLY-ACCESS STATEMENT

This manuscript, after peer review, is being released online in its original, unedited form to provide early access to its findings. Prior to final publication, it will undergo additional editing and formatting. Readers may note that, as an unedited version, the manuscript may have linguistic and other minor errors. The DOI will be active after the final publication of the paper.

PUBLICATION STATUS

PEER REVIEW	ONLINE VERSION	FINAL ISSUE
Completed	**Prepublished**	**March 2027**

Original Research

Applying LightGBM to Enhance the Accuracy of Emission Inventory for the Hai An Port: A Premise for Developing Emission Mitigation Strategies

Ngo Nhu Tai , Le Cong Doanh, Doan Le Nam Khanh, Trinh Quoc Anh, Phan Van Hung, Nguyen Thanh Son†
Faculty of Navigation, Vietnam Maritime University, Hai Phong, Vietnam
†Corresponding author: Nguyen Thanh Son; nguyenthanhson@vimaru.edu.vn
ORCID IDs of Authors: 0000-0002-4114-5295

Abstract: Sustainable port management is becoming a challenge due to the fact that maritime transportation is a significant GHG emitter and source of atmospheric pollutants. This study aims to offer an integrated framework which integrates conventional emission inventory approach with the Light Gradient Boosting Machine (LightGBM) algorithm and SHapley Additive exPlanations (SHAP) to enhance the accuracy and interpretability of ship emission estimation. The framework has been built based on operational data from 33 vessels (2197 vessel calls) in Hai An Port (Vietnam). Greenhouse gas (CO_2, CH_4, N_2O) and atmospheric pollutant (NO_x, SO_2, CO, VOC, PM_{10}, $PM_{2.5}$) emission inventories were created, and the predictive performance of Linear Regression, Random Forest, XGBoost, and LightGBM was compared. The results demonstrate that the predictive performance of LightGBM was the best with coefficients of determination (R^2) of 0.957 to 0.986 and mean absolute percentage error (MAPE) of 3.85% to 6.78%, which were obtained for all the pollutants investigated. SHAP analysis also found that the most important features that drive the emission estimates are engine power, engine load, vessel speed, operating mode and engine operating time. According to the annual emission inventories, CO_2 and NO_x were the major GHG and AP, respectively, and it is important to improve fuel efficiency and the operation management to reduce emission. The proposed framework will be accurate and interpretable for the ship emission inventory development, and will facilitate the data-driven environmental management, emission reduction strategies and sustainable green port development for port authorities and maritime stakeholders.

Keywords: Machine Learning, SHAP, Ship Emission Inventory, Hai An Port, Enhanced Management.

1. INTRODUCTION

Maritime carbon emissions have become an important aspect of ecological civilization and international climate regulations. Maritime transport constitutes a foundational element of the global shipping industry and represents a primary anthropogenic source of carbon emissions. Approximately 90% of global trade volume is transported by sea. According to the IPCC (2007), if the average global temperature increase exceeds 1.5–2.5°C (relative to 1980–1999), 20–30% of assessed species could face an elevated risk of extinction. It is estimated that nearly 70% of ship emissions occur within 400 km of land (Endresen et al., 2003; Eyring et al., 2005; Eyring et al., 2010). These nearshore emissions exert substantial impacts on the environment and coastal communities (Saxe and Larsen, 2004; Corbett et al., 2007). Evidence indi-cates that ship emissions in ports specifically affect the climate (GHGs), regional air quality (NOx and SOx causing acidification; NOx causing eutrophication and tropospheric ozone formation), as well as public health and ecosystems (NOx, SOx, PM, CO, and HC causing respiratory decline, lung cancer, allergies, and asthma), particularly for coastal populations (Corbett and Fischbeck, 1997; WHO, 2000; Bailey and Solomon, 2004; Eyring et al., 2010; Jing et al., 2014). As Vietnam progressively transitions into a global maritime trans-shipment hub, the impacts of ship emissions on the ecosystem are substantial. Since large-scale direct emission monitoring via sensors and equipment is economically infeasible, researchers and experts are actively seeking alternative methodologies to accurately predict maritime emissions (Chen et al., 2021; Tran et al., 2022). In general, two primary approaches are commonly employed to estimate ship emissions: the top-down method and the bottom-up method (Nunes et al., 2017).

Top-down methods are typically based on total global fuel consumption or fuel sales records (Peng et al., 2020). Conversely, bottom-up methods, also known as activity-based approaches, derive emission estimates by leveraging the individual activities of specific vessels (IMO, 2020). Bottom-up methods are considered superior in accuracy compared to top-down approaches for regional emission estimation and have been more widely adopted by researchers and authorities in recent years (Alver et al., 2018; Moreno-Gutiérrez et al., 2015).

In the Fourth IMO GHG Study, marine emission factors are categorized into fuel-based emission factors and energy-based emission factors , depending on the estimation methodology for the respective emission types. For instance, fuel-based emissions are estimated by applying emission factors to fuel consumption, while energy-based emissions are calculated by multiplying emission factors by the work output generated by the ship's engines (IMO, 2020). The energy-based emission group comprises nitrogen oxides (NO_x), methane (CH_4), nitrous oxide (N_2O), carbon monoxide (CO), non-methane volatile organic compounds (NMVOCs), and particulate matter ($PM_{2.5}$ and PM_{10}). Conversely, carbon dioxide (CO_2), sulfur oxides (SO_x), and black carbon (BC) fall under the fuel-based emission category (IMO, 2020; Nam T.T., 2024).

Emission factors have also been updated in accordance with the latest research and publications to refine computational results and reflect the actual state of maritime emissions. For example, the Fourth IMO GHG Study endeavored to make emission factors universally applicable worldwide. However, a significant limitation of previous studies is that models were restricted to the application of secondary emission factors provided by international organizations such as the EPA, EEA, and GHG Conversion Factor. This is not really the best solution because it does not consider the specific vessels and does not consider the complex non-linear relationship between the engine performance, cruising speed and idling/waiting time. Thus, it results in insufficient accuracy and is not suitable for assessment of the individual effects of operation and engine properties on the emissions.

These restrictions impair managers' capacity to make strategic decisions for emission mitigation. Machine learning models could be a possible approach to tackle these challenges. Machine learning models are trained with input and output data to capture complex non-linear relationships and can be used for the estimation of emission factors without the need to run time consuming traditional non-linear models. A thorough study has been conducted that merges more than 20 maritime emission estimation models and optimization algorithms (Hala Salem Al Nuaimi et al., 2025). The purpose of this current research is to find a suitable model that can be used to improve the accuracy of maritime emission inventories. This paper is structured in 4 sections:

Section 1 presents a systematic overview and analysis of research pertaining to maritime Greenhouse Gas (GHG) emission inventories and modelling, as well as air pollutant emission inventories and modelling, including emissions of NOx, SOx, CO, PM, and VOCs, highlighting the limitations of the current approaches and trends towards the use of Machine Learning in the study of maritime emissions.

Section 2 outlines the proposed research methodology, including the estimation of emissions using vessel specific operational data collected from ships calling at Hai An Port, the data processing workflow, and the LightGBM machine learning model to characterize non-linear relationships between the opera-tional factors and GHG and air pollutant emissions. Moreover, SHAP (SHapley Additive exPlanations) in-terpretation method is embedded to analyse and interpret the contribution levels of specific technical pa-rameters and operational factors to emission outcomes for both GHGs and air pollutants.

In section 3 the results of the simulations are presented and analyzed, with the main objective being to investigate the impact of different factors such as the characteristics of the main and auxiliary engines on the amount of GHGs and air pollutants emitted by the ships during their actual port operations. The results from SHAP analysis are used to provide insights on the relative importance of every group of factors, thus justifying the validity of the proposed machine learning model in capturing the real trends of the operation as well as interpreting its results for both GHG and air pollutant emission predictions.

Section 4 presents the main findings of the study and considers these findings in the context of practical fleet environmental management. This section highlights the potential of the integrated approach of emission inventories and machine learning to support ship owners, the port authorities and stakeholders to develop effective GHG and air pollutant mitigation strategies, with an improved understanding of ship operational behaviour and the technical factors influencing emissions.

2. LITERATURE REVIEW

2.1. Evaluation of Emission Inventory Methodological Trends

In response to IMO's Greenhouse Gas (GHG) strategy, ambitious targets are set to reach net zero emissions by 2050 (Mari-Liis Tombak et al., 2025; Bojić, F et al., 2025). To achieve this, IMO requires ships to meet technical (EEXI) and operational (CII) requirements. Moreover, Annex VI to MARPOL Convention contains provisions on Sulfur Content Limits (IMO 2020) and Emission Control Areas (ECAs) and in particular monitoring of maritime activities. It is a practical research with practical applications to help the Government's efforts to reach the goal of achieving net zero emissions in 2050 as stated at COP26. To participate in this common objective, the study details and high resolution emission inventory for the Hai An fleet is undertaken along the selected shipping routes. This improves the accuracy of the emission assessment process especially in a well-traveled sea area like the East Sea. The use of Artificial Intelligence for enhancing inventory accuracy aligns with Sustainable Development Goal 9 (Industry, Innovation and Infrastructure) and Sustainable Development Goal 13 (Climate Action) of the United Nations (Chen et al., 2024).

According to the existing research methods, it can be divided into two categories (Ning, Y. et al., 2025). The "Top-down" approach (Juan moreno-Gutiérrez et al., 2021; Cheng Cheng et al., 2024) adopts estimation models which rely on macro-level fuel sales statistics and are multiplied by emission factors; these are often used to estimate maritime emissions at global and regional levels (Kramel, D et al., 2022; Miola, A et al., 2011; Endresen, O et al., 2005; Deniz, C et al., 2008; Hulskotte, J et al., 2010). This is a straightforward approach, but several studies have found that it isn't very accurate. The main reasons for this inaccuracy stem from: Lack of spatiotemporal resolution. This approach does not identify where emissions are produced (berths, fairways) and when they are produced by considering the type, trajectory and routes of ships (Cheng Cheng et al., 2024; Dang, Q. V. et al., 2022). Statistical Discrepancies: Corbett, J.J. et al.'s (2003) research revealed that there were big discrepancies between the fuel sales statistics and the fuel consumption in the world which caused the inventories to have a big gap of error (Wang, C et al., 2008).

To overcome these drawbacks, the bottom-up approach came into existence, and it has proven to be very successful in improving inventory accuracy. The Automatic Identification System (AIS) has spurred the "bottom-up" approach. This is done by using granular trajectory data (position, velocities, time, etc.) to estimate energy consumption for individual vessels in different operating modes, including cruising, maneuvering and anchoring (Ning et al., 2025; Mari-Liis Tombak et al., 2025; Xubiao Xu et al., 2025). However, the use of standardized models such as STEAM (Ship Traffic Emission Assessment Model) have become the main tools, allowing emission assessment to be performed at much higher resolution (Mei Sha et al., 2024; Mari-Liis Tombak et al., 2025; Hyangsook Lee et al., 2021).But challenges in the generation of accurate emission inventories remain. First, the use of default EFs, those published by bodies like US EPA or EMEP/EEA (Europe) or others and without any fine tuning for the specific vessel type and geographic region remains a challenge. Since there is variation in the operational speeds of vessels and they are subject to external environmental influences like waves and currents, emission factors are not static but dynamic (Juan Moreno-Gutiérrez et al., 2021; Zhong Shuo Chen et al., 2024). A study done in Singapore showed that there can be tremendous difference in applying default EFs without considering these meteorological factors. For example, the bottom-up method with default EFs underestimated emissions by up to 9% and overestimated CO emissions by up to 10% compared to the empirical measurements in a comparative analysis in Singapore (Zhong Shuo Chen et al., 2024; Pham N.D.K et al., 2023; Long, N. X., & Van Hung, P., 2025). Secondly, large deviations at low load are present, typical of ship activities in the port area. Marine diesel engines waste fuel at low loadings (less than 20% of the rated power). The baseline emission factor sets tend to be unable to reflect the non-linear rise of emissions (Su Song, 2014). The emission factors have been shown to be much higher at low loads than at high loads (Mojibul Sajjad et al., 2025). While the IMO gives LLADF factors, empirical analyses indicate that emissions are much higher than theoretical calculations, especially with regard to the emission of Particulate Matter (PM) and Particulate Number (PN) (Philipp Eger et al., 2023).

It can be found that traditional emission inventory techniques, like activity-based (bottom-up) approaches, show better accuracy than fuel-based (top-down) approaches but also have some difficulties in dealing with the complex non-linear relationship between operational factors and actual emission quantities. Physical models require large amounts of calculation, and their static emission factors can result in large errors.

2.2. Machine Learning

The use of Machine Learning (ML) models and Deep Learning (DL) models has been proposed as a solution for quantifying emissions caused by ship and engine use without sacrificing excessive computational time to process the non-linear relationships. Such models can be used to learn hidden patterns or relations in Big Data (Farhana Rahman Anonna et al., 2023) and research has shown that the use of machine learning can also help in decision making for green port and sustainable development goals (Alper Seyhan, 2025; Periklis Prousaloglou et al., 2025; Irmina Durlik et al., 2024). Many studies have been conducted in different countries around the world that shows the use of AI results in high efficiency regarding the emission inventory process. Hala Salem Al Nuaimi et al. (2025) reviewed and compiled more than 20 state-of-the-art Machine Learning (ML) models and Deep Learning (DL) models for maritime emissions assessment; the physical nature and suitability of each model were analyzed using the different datasets. In 2024, Guangnian Xiao et al. studied 476 research papers between the years of 2022–2022, which confirmed that the use of AI in the maritime transport sector is a trend that will grow at an exponential rate. Farzadmehr et al. (2025) sorted and assessed the benefits and costs of 30 AI initiatives now going on at organizations at seaports, as well as a series of models tested at global ports, including Explainable Artificial Intelligence (XAI) (Juhyang Lee et al., 2024) and the PrE-PARE model (Bojić Filip et al., 2025).Machine learning models show high performance in the emission calculation process. Existing popular studies can be categorized into two classes: most of them are based on Machine Learning algorithms and field datasets, the other one based on PEMS device which directly measures emission in the field. While PEMS data is very reliable at the instantaneous level, the aggregation of results of previous studies shows that the predictive performance of models developed from the PEMS data may be unstable when extended to longer time horizons or larger scale operations. Mądziel, M (2023) employed a Gradient Boosting model, with a coefficient of determination of $R^2 = 0.61$, following a comparison of the results with the literature on studies that used standardized datasets and long-term operational data that achieved R^2 above 0.93 (Al-Nefaie et al., 2023; Ullah et al., 2021; Tena-Gago et al., 2023; Hien et al., 2022; Xia et al., 2021). On the other hand, the PEMS studies related to NO_x, CO_2, etc., from diesel vehicles using Gradient Boosting Regression machine learning model gave the higher value for R^2 reaching the value of 99% (Wen et al., 2021); however, there was the limitation of input variables which resulted in restricted accuracy of the model.Therefore, it is clear that the use of Machine Learning models can be as well effective by utilizing direct measurement data such as PEMS, as well as the data standardization quality, the diversity of the input data and the algorithm's ability to learn nonlinear relationships. The proposed method of using standardized data and machine-learning models provides a better compromise between reliability, scalability and cost of implementation than the purely empirical measurement methods for long-term emission assessment for vessels. Therefore, in this research, a machine learning model integrated with the vessel's field-reported dataset will be employed.Overall, the application of AI in emission inventories has become highly prevalent; however, it is observable that developing countries still lack AI research in the maritime transport sector to develop vessel emission management tools utilizing artificial intelligence or advanced data analysis algorithms. Specifically, few studies have focused on constructing emission management support systems at the individual vessel level within specific seaports (Nguyen, P. N, 2025).

From this perspective, the authors of this study select an appropriate AI model for application at Vietnamese seaports, establishing a foundation for the implementation of Machine Learning and Deep Learning in subsequent research to deliver practical solutions for the sustainable development of seaports in developing countries. Based on existing research on machine learning models, it is evident that the selection of an appropriate model is contingent upon the dataset collected at the seaport. Following a comprehensive review and synthesis of the literature, the research team resolved to apply an integrated approach utilizing the LightGBM model and SHAP to construct the maritime emission inventory model.

The need for this research stems from the urgent need to tackle the environmental effects of container vessel emissions and to gain a better understanding of the processes that cause emissions to vary. Accurate forecasting models can help policymakers make informed assessments of emission mitigation, sustainable transportation and infrastructure investments. In addition, accurate emissions predictions can help consumers choose their cars, highlight sustainable driving habits, and help with sustainable transport.

In this work, we try to overcome the limitations of the current methods for estimating maritime emissions by building a light-weight machine learning model based on empirical data, which has several major benefits:

Developing and testing a low resource-consuming machine learning model using state-of-the-art methods to precisely estimate maritime emissions from extensive datasets.

Incorporation of empirical data to train the machine learning model along with Explainable Artificial Intelligence (XAI) to make the vessel emission prediction more practical and accurate.

Improving computational efficiency and scalability by creating a machine learning model that minimally processes data, allowing for its practical implementation and broader application in maritime emission estimation.

The LightGBM model is optimized to handle large amounts of data at much higher speeds and lower memory usage than other models. Gouline et al. showed that LightGBM can speed up the training process by up to 20 times, while keeping the same accuracy as the traditional Gradient Boosting models (Ke, G et al., 2017). Besides, LightGBM has a training speed of only 0.5 seconds, while LSTM and GRU are more than 400 seconds, and the error margin (RMSE, MAE) is comparable (Nguyễn, T. A et al., 2024). The LightGBM model is very well suited for tabular data, which is common in emission inventories, such as speed, size of the vessel, anchoring time, load, power etc. LightGBM's leaf-wise tree growth algorithm, not level-wise, ensures better loss minimization and the accurate modeling of complex non-linear relationships between operational parameters, engine characteristics, and emission volumes. Furthermore, LightGBM has been shown to have good generalization performance and is not prone to overfitting because of the tree depth optimization mechanisms (Ullah et al., 2021; Nguyen, T. A et al., 2024).

Although very precise, LightGBM is often seen as a "black-box" and makes it difficult for port managers to make a full evaluation of the emission volumes. To address the ‘black-box’ nature of machine learning models in the context of maritime emission inventories, this study combines the SHAP method, which sits within the Explainable Artificial Intelligence (XAI) paradigm, to explicitly interpret the impact mechanisms of the individual input variables to the output variables (Ullah et al., 2021). SHAP allows to convert complex technical forecasts into understandable results, which can help to build trust among stakeholders, to apply research results in practice to manage them, and to increase the overall scientific credibility (Ke, G et al., 2017).

SHAP quantifies the contribution of technical parameters such as main engine power, auxiliary engine load, engine rotational speed and anchoring time, rather than just their relative influence level, which is the case for traditional feature importance metrics. This capability is able to effectively explain the physical-operational nature of the emission process in the context of a container port environment. At the same time, its ability to analyse at both the global and local level means that the model can not only show the overall trend of the dataset, but also highlight the operational anomalies, which can help to carry out comprehensive risk assessment and operation optimisation.

Given this context, LightGBM is chosen as the base algorithm because of its ability to handle large-scale tabular data and model complex non-linear relationships between vessel engine characteristics and exogenous factors. The combination of LightGBM and SHAP represents a methodological framework that balances predictive power with interpretability, moving the machine learning model beyond a predictive tool to a more transparent decision-support system for seaport environmental management.

The proposed model is flexible with respect to data structure and highly scalable. In addition to evaluating its performance using vessel operation data collected from ships calling at Hai An Port, the model has the potential to be applied to other fleets operating under similar navigational and operational conditions. The model is built using LightGBM to handle large-scale, multi-source tabular data efficiently, and the SHAP method is used to enhance the interpretability of the model and generate clear insights into various operating scenarios for vessels.

This forms the foundation for the development of a national emission inventory system and to compare the environmental efficiency of ports. Meanwhile, it provides a scientific basis for the policy making process and scales up emission mitigation solutions in the Vietnamese seaport system.

3. METHODOLOGY

3.1. Research Framework

In this study, a bottom-up method in combination with machine learning models is used to estimate and ana-lyse the emissions from the Hai An fleet. This process is clearly broken down into three main stages:

(1) Vessel trajectory information collection and preprocessing;

(2) Preparation of a baseline inventory based on physical equations.

(3) Building the LightGBM model with SHAP to assess and analyse the variables that affect maritime emis-sions and the development of the LightGBM model

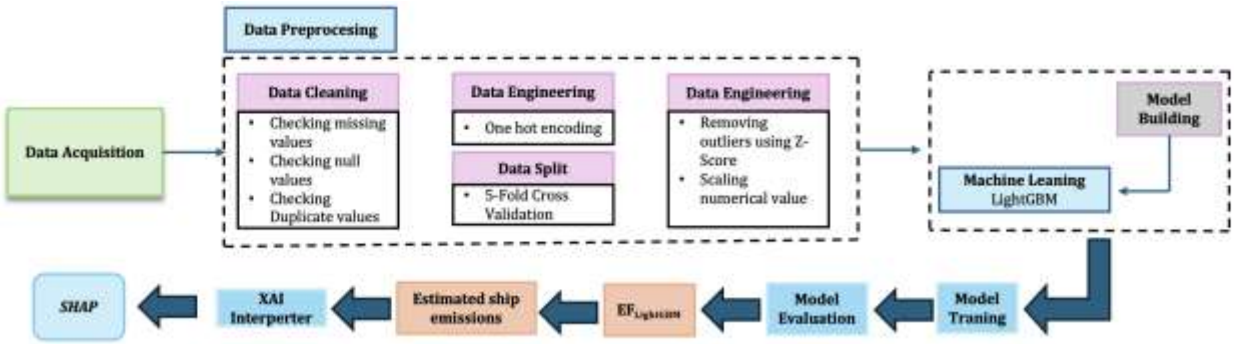


Fig 1: Methodological flowchart illustrating the entire emission inventory process

3.2. Data Collection and Processing

3.2.1. Data Collection

The data set included all the vessel traffic records of ships calling at Hai An Port from 28 February 2018 to 8 February 2026. There were 33 container vessels and 2,197 vessel calls in the data set. Table 1 presents the main technical parameters of each vessel and the number of port calls in the week according to the Hai An port operational information system.

Table 1: Principal Particulars of the 33 Ships Calling at Hai An Port

Vessel Name	Max. Speed	Deadweight	Year Built	Main Engine	RPM	Main Engine Power	Auxiliary Engine Power
Haian Beta	19.1	18852	2024	Man B & W 6S60ME-C10.5, Tier II	93	13200	2904
Haian Time	21.1	13267	2001	1DE: 2 SA 8 CY	100	11440	2516.8
Haian Alfa	18.5	18852	2023	Man B & W 6S60ME-	93	13200	2904

				C10.5, Tier II			
Haian Link	18	12559	2010	(1) 2T - 7 CYL - 500 MM X 1910 MM AT 127 RPM	127	9988	2197.36
Haian Park	17.3	9413	2000	1DE: 2 SA 6 CY	104	7980	1755.6
Haian Bell	19	14308	2003	7 S 50 MC-C MAN B&W	127	11060	2433.2
Haian Rose	21.6	17515	2008	1DE: 2 SA 7 CY	125	15820	3480.4
Haian Iris	20.1	9963	2011	1DE: 2 SA 6 CY	101	9480	2085.6
Haian East	20	18102	2008	7UEC60LSE	104	15785	3472.7
Haian Dell	21.6	17280	2008	1DE: 2 SA 7 CY	127	15820	3480.4
Haian View	21.5	17280	2009	1DE: 2 SA 7 CY	127	15820	3480.4
Pegasus Dream	19.6	9924	2024	MAN B&W6S46ME-B8.3	134	6.290	1383.8
Sm Tokyo	21.1	9928	2007	MAN-B&W 7L58/64	400	9730	2140.6
Haian City	21.5	20887	2008	1DE: 2 SA 7 CY	127	15820	3480.4
Anbien Bay	21.2	17515	2008	MAN B&W 7S60MC-C	105	15500	3410
Contship Fox	18.9	9909	2009	MAN-B&W 7L58/64	428	9730	2140.6
Dong Ho	16.5	6543	1998	MEDIUM SPEEDDIESEL	145	6250	1375
Haian Gate	17.1	6454	2013	2T - 6 CYL FEEDER	500	5040	1108.8
Haian Mind	21.8	18517	2012	MAN B&W 6S60MC	105	12930	2844.6
Haian Song	19.2	18517	2008	MAN B&W	127	10010	2202.2
Haian West	21.4	18517	2007	MAN B&W 7S60MC-C	105	13560	2983.2
Pegasus Peta	18.7	9988	2014	MAN B&W 6S46MC-C	134	7330	1612.6
Pegasus Tera	18.7	9988	2014	MAN B&W 6G70ME-C9.2	134	8650	1903
Pegasus Unix	19.1	9522	2007	MAN ELECTRONIC DIESEL	136	8950	1969
Pegasus Yotta	18.3	9522	2005	2T FEEDER DIESEL	115	9250	2035
Pegasus Zetta	189	9522	2005	MAN B&W	129	7860	1729.2
Josco Real	20.1	18885	2018	CONTAINER DIESEL	100	10857	2388.54
Pacific Grace	20.8	9352	1997	B&W model 6L50MC	133	7987	1757.14
Star Challenger	22	9955	2016	MAN B&W 6S46ME-B8	114	6700	1474
Tc Symphony	19	10384	2005	MEDIUM SPEED MARINE DIESEL	99	15600	3432
Voyager Elite	21.6	6195	1995	MAKITA MAN B&W 5L35MC.	210	3236	711.92
Wisdom Grace	22.5	13199	1998	B&W 7S50MC	91	9989	2197.58
Pelican	17	16770	2013	MAN-B&W 6S60MC	142	14280	3141.6

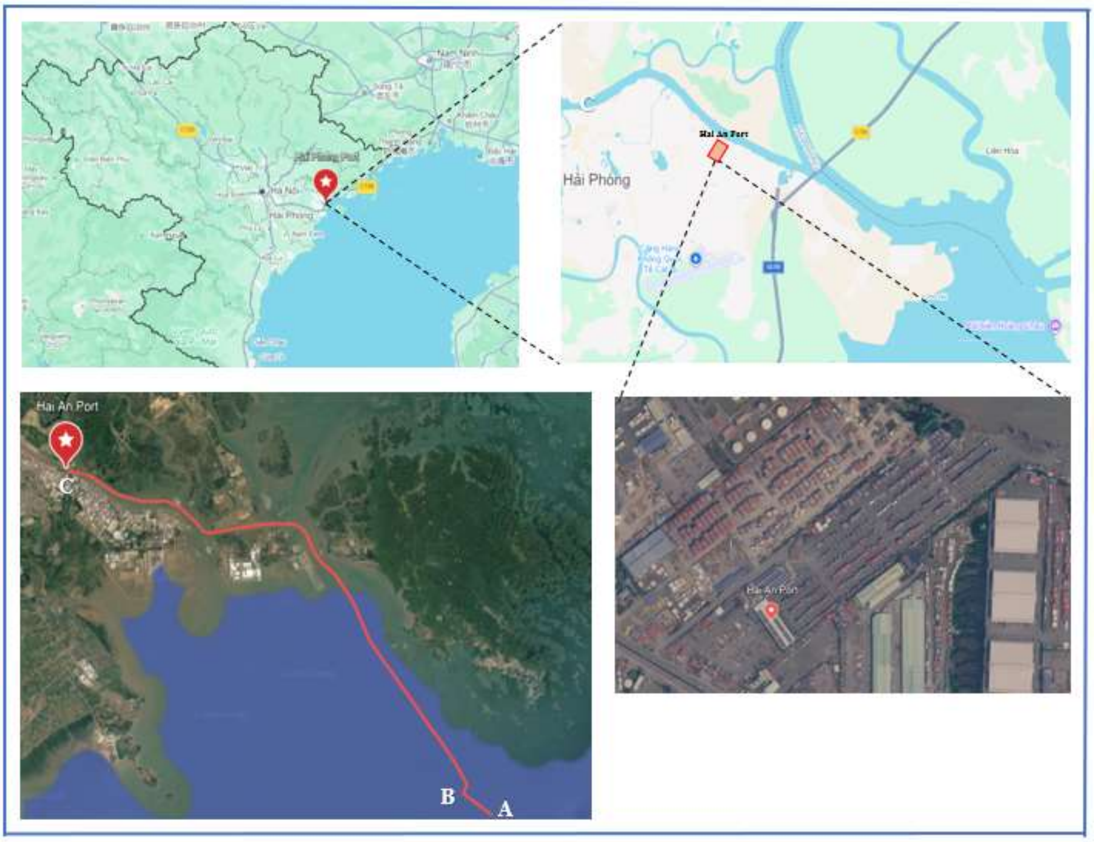


Fig 2: Geographical Location and Voyage Routes of Ships Calling at Hai An Port

The vessel route was determined based on the maritime traffic regulations of the Hai Phong port area and the operational procedures of Hai An Port, using three reference points A, B, and C, where A is the pilot boarding station, B is Fairway Buoy No. 0 that is used to separate the coastal navigation area and the port approach channel, and C is the assigned berth at Hai An Port. For this reason, the operation of the vessels was divided into three different operational phases: cruising (A-B), maneuvering within the navigation channel (B-C), and hoteling at berth (C). This kind of route segmentation allows the operating time of each phase to be determined accurately, thereby providing a solid foundation to estimate the engine load and quantify the greenhouse gas emissions during the port call.

Table 2: Definition of Reference Points and Operational Characteristics of Vessel Navigation to Hai An Port

Reference point	Location	Example timestamp	Operational phase	Typical speed (knots)	Purpose in Emission inventory
A	Pilot boarding station	18:00, 25/11/2025	Cruising	8 - 11	Estimation of emissions during the cruising phase.
B	Fairway buoy no. 0	20:15, 25/11/2025	Maneuvering	3–6	Estimation of emissions during maneuvering within the navigation channel.
C	Hai an port berth	01:00, 26/11/2025	Hoteling	0	Estimation of emissions during berthing and cargo-handling operations.

3.2.2. Data Processing

Data Cleaning

The main goal of this step is to remove inconsistencies and anomalies that may hinder or affect the model training process (Alam et al., 2025). Missing values and null values: Missing or null values (NaN/null) in the data set can lead to bias or loss of accuracy during training, especially if important features are missing from the data (Alam et al., 2025). For this purpose, the dataset needs to be given a thorough review. The missing rate is small when compared to the total number of records, standard processing techniques usually entail discarding incomplete records altogether, or using imputation methods like mean imputation or median imputation to improve the integrity and reliability of the dataset in question (Farhana Rahman Anonna et al., 2023). Record with "Missing value rate" > 20% for primary variables are not included in the data.

Data Engineering and Transformation

This is done to map raw data into formats that could be understood and acted upon optimally by Machine Learning algorithms. One-hot Encoding: The beauty and grace of machine learning models is that they can't directly work with text or categorical data. These categorical variables are converted to numerical using the One-hot encoding technique. Binary columns are created for each category which has a value of either 0 or 1 which allows the model to learn well from these columns since there is no hidden hierarchy created between these categories (which happens with Label Encoding). In addition, interestingly, for applying explainable AI (XAI), it has been shown that One-hot encoding can also improve the interpretability of the SHAP values.

Removing Outliers using Z-Score

It is particularly important to remove the outliers, otherwise the model could predict values that are pulled by the outliers in the extreme. The Z-score method of analysis is used to find the extent of the deviation of the individual data point from the mean. The formula to calculate Z scores is:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Where xi is a particular piece of data, μ is the average of the data, and σ is the standard deviation. Outliers with Z values outside of certain threshold are identified and removed; the model is only trained on clean, high-representative data.

In addition, a domain expert rule and the statistical method Interquartile Range (IQR) are used to identify and remove outliers. In particular, any observation lying below the lower limit, [Q1 – 1,5× IQR] or above the upper limit, [Q1 + 1,5× IQR] is considered to be an outlier and is excluded. In the maritime domain however, mechanically applying statistical criteria can result in discarding of practically relevant observations: extremely large vessels (e.g. LNG carriers) may be statistically out of the data sample, but may be relevant operationally. So that only outliers as defined by the IQR method and also those that do not fulfill the basic technical requirements of the vessel system (e.g. negative speed or power values above the allowable limits) are eliminated.

Scaling Numerical Values

In many cases, numerical variables have heterogeneous units and large/small ranges. The data should be scaled so that all features make a fair contribution to the performance of the model, and to allow the algorithm to converge more quickly. Min-Max Scaling is a commonly used technique to scale each feature to a range of 0 to 1. The formula is the following:

$$\mathrm{x_{scaled}} = \frac{\mathrm{x} - \mathrm{x_{min}}}{\mathrm{x_{max}} - \mathrm{x_{min}}} \tag{2}$$

Where x_{max} and x_{min} are the maximum and minimum values that feature can get, respectively. This is especially important to minimize the effect of variables with large orders on the loss function.

3.3. LightGBM

The emission factor calibration process is shown in Fig 3. The base emission factor (Base EF) was mainly used in previous studies because of its simplicity. In practice, in marine engines, however, this base emission factor is not comprehensive and precise enough to truly represent the emission quantities of the individual pollutants, under different operating conditions for each engine.

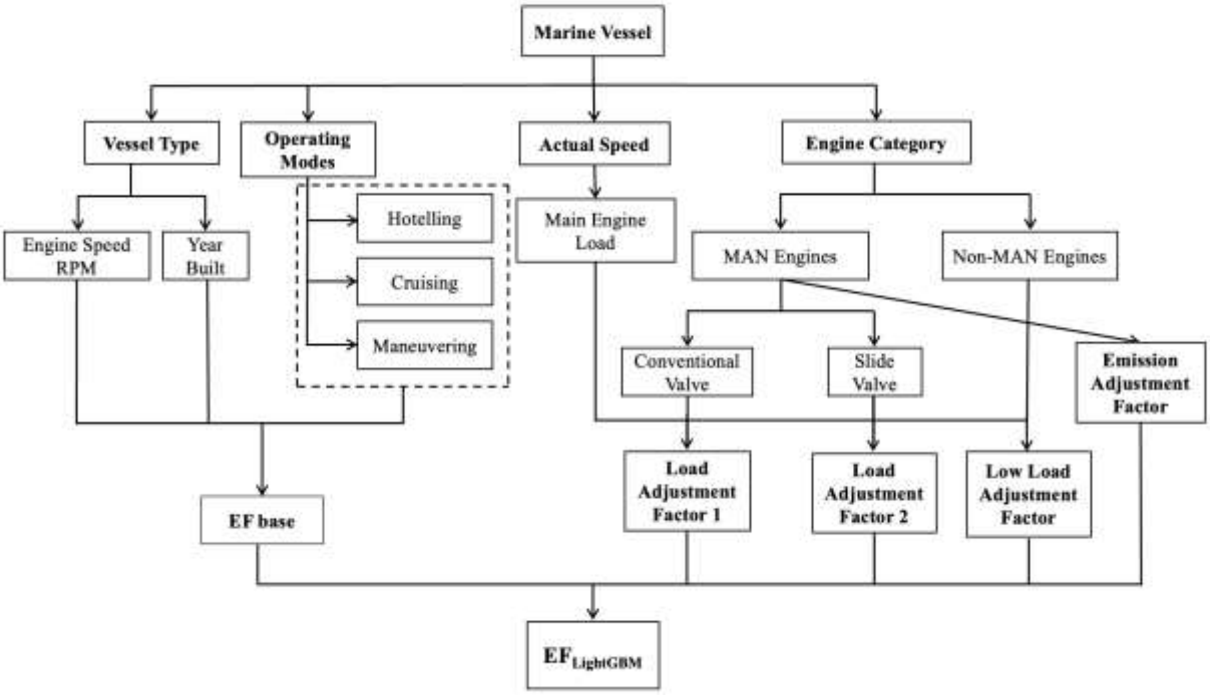


Fig 3: Emission factor calibration process

Therefore, calculations should be based on load adjustment factor suite created by the San Pedro Bay Ports (Starcrest Consulting Group, 2024). If a vessel is running in different modes of operation, the emission trend for each pollutant from the respective engine type will be different and the adjustment factor will not follow any deterministic rule, but rather will show a varying non-linear trend of increase and decrease.As such, this research acknowledges that there are some inherent margins of error with conventional methods of calculation. In addition, the process to manually interpolate and calibrate emission factors is very time consuming. This study utilizes LightGBM algorithm which gives a balance between accuracy and faster processing speed (Ullah et al., 2021) in order to evaluate the continuous effects of engine load as per the maritime emission evaluation requirements.

In this study, the LightGBM (Light Gradient Boosting Machine) advanced machine learning algorithm is chosen to build and test the research dataset to make a forecasting model. The simultaneous use of two models is not just a way to make use of the different model's advantages but also to provide objective and reliable results after comparison of their performance. LightGBM is used so as to take the best of its lightening speed and ability to process large data sets. The techniques used in this model include the Gradient-based One-Side Sampling (GOSS) method to select samples with large gradients, and the Exclusive Feature Bundling (EFB) method to reduce the dimensionality of the features, to achieve more efficient computation and memory conservation. Moreover, the histogram-based technique allows to describe the continuous data in discrete form, thus gaining great advantage for the reduction of computational cost. In particular, the leaf-wise tree growth strategy optimizes the reduction of the loss function for the sake of obtaining better predictive performance while repressing the occurrence of overfitting through constraining such as depth, etc.Traditional GBDT algorithm is to pre-sort all input features very carefully to identify the optimal split point for each node of the decision tree. While this helps to improve accuracy in data-splitting, it also comes with some significant added memory and computational cost. To alleviate this disadvantage, LightGBM adopts histogram-based algorithm along with a leaf-wise growth strategy (Ke, G et al., 2017).By combining and comparing these two models, the study is able to comprehensively evaluate the forecasting performance in the same data set by using three metrics: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and the Coefficient of Determination (R^2). This way helps to choose the best model that is accurate, general and efficient.

In addition, LightGBM can set the maximum depth of the tree and control the complexity of the model, which can prevent overfitting when training the model. The histogram-based learning and leaf-wise tree growth strategy boost efficiency without compromising prediction accuracy. In this study, the model was used to approximate the relationship between ship emission factor and selected input variables as:

$$EF\ pred = f\ (P, LF, RPM, V, Age, Type, Engine) \quad (3)$$

Where:

EF is the predicted ship emission factor (g/kW),

P is the engine power (kW),

LF is the engine load factor (%)

RPM is the engine speed (revolutions per minute),

V is the vessel speed (knots),

Age is the vessel age (years),

Type denotes the vessel type (e.g., container ship or bulk carrier),

Engine denotes the engine type (MAN engine, conventional engine, or engines with and without slide valves).

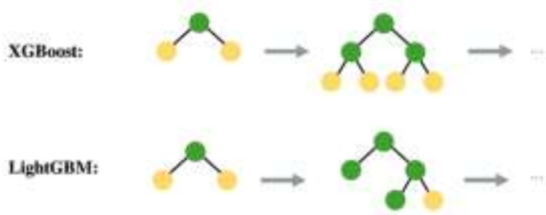


Fig 4: LightGBM and XGBoost models structure

For assessing the generalization ability of the model, the data is regularly divided into five equal-sized partitions, labeled D_1, D_2, D_3, D_4 and D_5. respectively. To guarantee the objectivity and strength of the model evaluation, the K-fold cross validation technique with K = 5 is implemented. In every iteration of the validation process, four subsets are used to train the model and the other subset to validate the model. This can be expressed mathematically as shown below:

Training Set: D_i với $i \in \{1, 2, 3, 4, 5\}$

Validation Set: D_j với $j \in \{1, 2, 3, 4, 5\}, j \neq i$

This process is repeated 5 times so that each subset is used to validate the model once and the other four sets are used to train the model. In particular, the iterations are done in the following manner:

Iteration 1: Training on D_2, D_3, D_4, D_5 – Validation on D_1

Iteration 2: Training on D_1, D_3, D_4, D_5 – Validation on D_2

Iteration 3: Training on D_1, D_2, D_4, D_5 – Validation on D_3

Iteration 4: Training on D_1, D_2, D_3, D_5 – Validation on D_4

Iteration 5: Training on D_1, D_2, D_3, D_5 – Validation on D_5

At the end of the 5 iterations, the model's performance evaluation metrics are combined together by averaging the results of the 5 different validation iterations. This will give a more reliable estimation of the capability of the model to predict on unseen data.The K-fold cross validation technique on the other hand distributes the data splitting task evenly and helps to avoid bias due to random data partitioning, since every observation in the data set is used for both training and validation. This approach, therefore, provides a better and consistent measure of the generalization power of the forecasting model.

3.4. Model Evaluation

Common metrics for accuracy evaluation are used, such as R^2, MAE, and MAPE, in the model to check the accuracy of the forecasting model compared to traditional models.

R^2

R^2, or Coefficient of Determination, is a metric that represents the percentage of the variation in the dependent variable y that is explained by the model using x. The closer the R2 value is to 1, the better the model can describe the relationship between the explanatory variables and the Emission Factor (EF). The formulas for calculating R^2 and Adjusted R^2 are as follows:

$$R^2 = 1 - \frac{ESS}{TSS} \tag{4}$$

$$R^2_{hc} = 1 - \left(\frac{\frac{ESS}{\frac{n}{k}}}{\frac{TSS}{n-1}} \right) = 1 - \frac{n-1}{n-k} \times (1 - R^2) \tag{5}$$

Where:

R^2 is the explained variance from the regression,

ESS is the Error Sum of Squares (residual sum of squares),

TSS is the Total Sum of Squares,n is the number of observational samples,

k is the number of model parameters, which equals the number of independent variables plus 1.

MAPE

Mean Absolute Percentage Error (MAPE) is used for the relative assessment of the difference between the emission factor predicted by LightGBM model and the reference emission factor obtained from EPA standards. MAPE expresses the prediction error as a percentage, allowing to compare the effectiveness of the models when large differences in the magnitude of the emission factors exist between the vessels. MAPE is calculated as:

$$MAPE = \frac{1}{n} \times \sum_{i=1}^{n} \left| \frac{y_i - \hat{y_i}}{y_i} \right| \tag{6}$$

Where:

Y_i is the standard emission factor according to the EPA for vessel i (g/kW),

$\widehat{Y_i}$ is the emission factor predicted by LightGBM for vessel i (g/kW),

n is the number of vessels in the dataset

MAE

$$MAE = \frac{\sum_{t=1}^{n} |\varepsilon_t|}{n} = \frac{\sum_{t=1}^{n} |Y_i - \hat{Y_i}|}{n} \tag{7}$$

Where:

Y_i is the standard emission factor according to the EPA for vessel i (g/kW),

$\widehat{Y_i}$ is the emission factor predicted by LightGBM for vessel i (g/kW),

n is the number of vessels in the dataset.

To evaluate the proposed approach, the performance of LightGBM was compared with three widely used regression models, namely Linear Regression, Random Forest, and XGBoost, using the same training and testing datasets and the same evaluation metrics (R^2, MAE, and MAPE). As shown in Table 3, LightGBM produced the lowest prediction errors for most pollutants. Compared with Linear Regression, the MAPE decreased from 12.43–60.27% to 3.85–6.78%, while the MAE was reduced by approximately 72–80% for major pollutants, including CO_2, NO_x, and SO_x.

These results indicate that the relationship between vessel operating conditions and emission factors is highly nonlinear and cannot be adequately represented by a linear model.

Compared with Random Forest, LightGBM reduced the MAE by approximately 9–17% and the MAPE by up to 1.8 percentage points for most pollutants. While both approaches use a decision tree, LightGBM learns the trees by minimizing the errors of the previous trees, which helps the model to learn more complex interactions between the input variables accurately.

Most pollutants had high coefficients of determination (R^2>0.97) in comparison with XGBOOST. However, LIGHTGBM always produced lower MAE and MAPE, with the MAE reduced by about 6-12%. This improvement is especially relevant to the creation of ship emission inventories, where small reductions in prediction error can result in more reliable emission estimates for the millions of vessel calls. Overall, the results suggest that LightGBM provides a suitable balance between prediction accuracy and model robustness for estimating ship emission factors under different operating conditions.

Table 3: Model error evaluation

Model	Metric	PM_{10}	$PM_{2.5}$	NOx	SO_2	CO	VOC	CO_2	N_2O	CH_4
Linear Regression	R^2	0.945	0.945	0.740	0.962	0.914	0.936	0.962	0.963	0.936
	MAE (kg)	3.883	3.477	465.593	8.481	23.837	8.579	14051.1	0.587	0.171
	MAPE (%)	15.36	15.36	60.27	12.44	17.41	15.19	12.44	12.43	15.19
Random Forest	R^2	0.965	0.965	0.973	0.979	0.946	0.96	0.979	0.978	0.967
	MAE (kg)	1.297	1.156	122.438	2.817	8.569	2.956	4615.49	0.194	0.058
	MAPE (%)	6.515	6.515	8.41	4.701	8.61	6.702	4.66	4.84	6.73
XGBoost	R^2	0.979	0.979	0.981	0.986	0.959	0.978	0.986	0.985	0.978
	MAE (kg)	1.221	1.095	101.336	2.645	7.992	2.715	4389.62	0.186	0.054
	MAPE (%)	5.107	5.083	6.619	4.112	7.372	5.037	4.143	4.12	5.037
LightGBM	R^2	0.972	0.972	0.986	0.982	0.957	0.973	0.982	0.983	0.973
	MAE (kg)	1.097	0.982	95.556	2.396	7.314	2.464	3992.92	0.166	0.049
	MAPE (%)	4.759	4.713	6.525	3.93	6.78	4.971	3.92	3.85	4.97

3.5. Emission Inventory Methodology

The overall data used for the vessel and the emission estimation methodologies as discussed below, cover the emissions from the main engine, auxiliary engines and auxiliary boilers.Any differences between the estimation procedures used in different operational modes (at anchor, at berth, or maneuvering) are explicitly explained for each section.This study adopts an inventory method to improve the emission calculation methodology, inspired by the San Pedro Bay Ports, which developed an emission inventory approach (Starcrest Consulting Group, 2024), while ensuring compliance with the technical guidelines for creating emission inventories from marine vessel sources released by the Vietnam Ministry of Agriculture and Environment (Ministry of Natural Resources and Environment, 2024).The following calculation formulas are used:

For MAN engines:

$$E = LF \times EF_{base} \times P \times A \times EFA \times LAF_i \times 10^{-6} \tag{8}$$

For non-MAN engines:

$$E = \begin{cases} LF \times EF_{base} \times P \times A \times 10^{-6} \text{ if } LF \geq 20\% \\ LF \times EF_{base} \times P \times A \times EFA \times LLA \times 10^{-6} \text{ if } < 20\% \end{cases} \tag{9}$$

Where:

E is the vessel's emission volume (tons);

P is the engine power (kW);

LF is the engine load factor (%);

A is the vessel's operational time (hours) (across anchoring, maneuvering, and sea passage phases);

EF base is the default emission factor provided by the EPA;

LAF_i is the load adjustment factor (slide valve and standard valve);

EFA is the emission adjustment factor by fuel type;

LLA is the low load adjustment factor, dimensionless.

Furthermore, to elucidate the emission volume across each operational phase for individual engines, the formula is as follows:

$$E = E_1 + E_2 + E_3 + E_4 + E_5 \quad (10)$$

Where:

E is the vessel's emission volume (tons);

E_1 is the main engine's emission volume during the sea passage phase (tons);

E_2 is the main engine's emission volume during the maneuvering phase in port (tons);

E_3 is the auxiliary engine's emission volume during the sea passage phase (tons);

E_4 is the auxiliary engine's emission volume during the maneuvering phase in port (tons);

E_5 is the auxiliary engine's emission volume during the anchoring phase in port (tons).

It is possible to disaggregate the engine's emission volume in various operational phases to accurately assess the impacts of each individual factor to the total emission output, which in turn aids in developing a better emission mitigation policy.

$$LF = (\text{Actual Speed/ Maximum Speed})^3 \quad (11)$$

For the auxiliary engine, the load factor is calculated based on the calculation standards of the container ships in different operational phases: sea passage: 0.13, maneuvering: 0.48, unloading: 0.19.

Table 4: Low load adjustment factors

Load	PM	NO_x	SO_2	CO	VOC	CO_2	N_2O	CH_4
2%	7.29	4.63	3.30	9.68	21.18	3.28	4.63	21.18
3%	4.33	2.92	2.45	6.46	11.68	2.44	2.92	11.68
4%	3.09	2.21	2.02	4.86	7.71	2.01	2.21	7.71
5%	2.44	1.83	1.77	3.89	5.61	1.76	1.83	5.61
6%	2.04	1.60	1.60	3.25	4.35	1.59	1.60	4.35
7%	1.79	1.45	1.47	2.79	3.52	1.47	1.45	3.52
8%	1.61	1.35	1.38	2.45	2.95	1.38	1.35	2.95
9%	1.48	1.27	1.31	2.18	2.52	1.31	1.27	2.52
10%	1.38	1.22	1.26	1.96	2.18	1.25	1.22	2.18
11%	1.30	1.17	1.21	1.79	1.96	1.21	1.17	1.96
12%	1.24	1.14	1.17	1.64	1.76	1.17	1.14	1.76
13%	1.19	1.11	1.14	1.52	1.60	1.14	1.11	1.60
14%	1.15	1.08	1.11	1.41	1.47	1.11	1.08	1.47
15%	1.11	1.06	1.09	1.32	1.36	1.08	1.06	1.36
16%	1.08	1.05	1.06	1.24	1.26	1.06	1.05	1.26
17%	1.06	1.03	1.05	1.17	1.18	1.04	1.03	1.18
18%	1.04	1.02	1.03	1.11	1.11	1.03	1.02	1.11
19%	1.02	1.01	1.01	1.05	1.05	1.01	1.01	1.05
20%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The application of the standard load adjustment factor suite, as recommended by the United States Environmental Protection Agency (EPA), improves the accuracy of the tool and increases its credibility in the seaport emission inventory process.At the same time, adherence to the internationally agreed-upon database system is strict, which helps to avoid static errors and strengthens the reliability, scientific value and the validity of the inventory results.

4. RESULTS

4.1. Analysis of the relationships among input variables in the LightGBM model

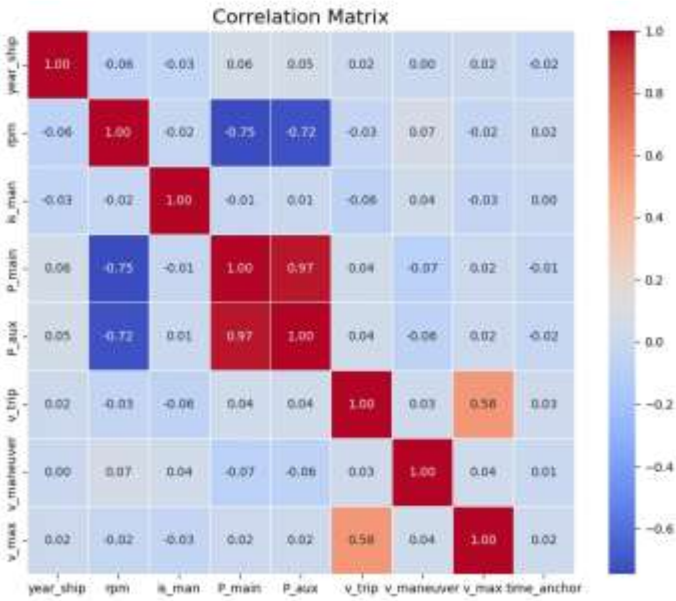


Fig 5: Correlation Matrix of Technical and Operational Characteristics of Container Ships at Hai An Port.

The Pearson correlation matrix of the input parameters used in LightGBM model training is shown in Figure 5, which includes the technical parameters of the vessel and operational conditions. The results obtained indicate that there is a high significant correlation between the main and the auxiliary engine power (r≈ 0,97). This is typical of the design features of the practical application of container ships, where a large main engine is likely to be accompanied by a large auxiliary engine system to meet energy requirements for anchoring and port operations. In addition, rpm has a negative correlation to the power of the main and auxiliary engines (r = -0.72 to -0.75). It is in accordance with the engineering principle of marine diesel engines, traditionally characterised by a small range of operating speeds, but high power for a better efficiency and mechanical durability.

4.2. Greenhouse gas emission inventory results

To assess the practical applicability of the proposed model, the emission factors were incorporated in the emission inventory framework to estimate the greenhouse gas emissions and air pollutants that would be produced by vessel operations during the years of 2019 to 2025 at Hai An Port.

The Global Warming Potential (GWP) characterization factors recommended by the IPCC (Greenhouse Gas Protocol, 2024) were used to convert the amounts of CO_2, CH_4, and N_2O emissions into carbon dioxide equivalent (CO_2e) emissions (1 CO_2e = 1 tonne of CO_2). The emission inventory obtained (shown in Table 5) reveals that Hai An Port also had a large emission volume of both GHGs and atmospheric pollutants in the study period of 2019-2025. The greenhouse gas with the highest proportion of emissions was consistently carbon dioxide, ranging from 8,393.41 tonnes in 2021 to 17,353.73 tonnes in 2023. As a result, changes in CO_2e were closely associated with changes in CO_2 emissions, while the direct contribution from CH_4 and N_2O was relatively small. However, these gases have much higher GWPs than CO_2, so when emissions are added as CO_2e, their presence resulted in a higher climate impact.

The interannual variations in the inventory showed a well-defined temporal pattern. Total emissions have steadily reduced between 2019 and 2021, but rose in 2022, peaked in 2023, and then dropped in 2024, before rising slightly again in 2025. Similar patterns were found for greenhouse gases and air pollutants, indicating that the emission fluctuations from year to year were largely driven by the changes in the intensity of the vessel traffic, the duration of engine operations and fuel consumption. In contrast, the impact of technical characteristics and emission factors was quite consistent during the study period, suggesting that the main changes in annual emissions were due to activity changes.

Table 5: Emission inventory of greenhouse gases and air pollutants from vessel operations at Hai An Port (2019–2025) (unit: tons)

Year	PM_{10}	$PM_{2.5}$	NO_x	SO_x	CO	VOC	CO_2	N_2O	CH_4	CO_2e
2019	3.63	3.26	234.35	7.92	21.12	8.35	13119.94	0.55	0.17	13276.14
2020	2.97	2.67	186.45	6.35	17.34	7.54	10518.93	0.45	0.15	10646.43
2021	2.39	2.15	149.84	5.07	14.00	6.17	8393.41	0.36	0.12	8495.37
2022	2.67	2.40	199.50	5.80	17.35	8.74	9581.75	0.45	0.17	9711.12
2023	6.04	5.47	396.10	10.52	34.16	31.01	17353.73	0.86	0.62	17607.22
2024	3.65	3.28	211.67	7.99	22.08	10.10	13216.98	0.60	0.20	13385.45
2025	4.19	3.77	204.22	8.89	24.58	11.59	14706.68	0.64	0.23	14888.12

NO_x was always the largest share of the atmospheric pollutants emitted, between 149.84 and 396.10 tonnes per year. This outcome is due to the combustion behaviour of marine diesel engines, which increases the combustion temperature and thus the formation of thermal NO_x. Sulphur oxide emissions were significantly reduced and were mainly dependent on the sulphur content of the marine fuel used. Emission of carbon monoxide and volatile organic compounds, on the other hand, showed a strong correlation with incomplete combustion and were more sensitive to loading condition and efficiency of the engine. Even though the mass of $PM_{2.5}$ and PM_{10} represented a relatively small portion of the total emitted mass, their environmental and public health impacts are important. Fine particulate matter can penetrate deeply into the human respiratory system and lead to cardiovascular and respiratory diseases, especially in the port areas where the population is exposed to the highest levels of fine particulate matter.

Overall, the inventory shows that shipping activities at Hai An Port are a significant source of greenhouse gas emissions and air pollutants. The high concentrations of CO_2 and NO_x indicate that reducing fuel use and improving engine operating efficiency are likely to provide the biggest environmental improvements. Practical steps such as the more extensive use of shore power when vessels are at berth, optimizing vessel operating practices, and the shift to cleaner marine fuels and low emission propulsion technologies thus offer good emission reduction opportunities. The inventory is not only useful for operational decision making, but also serves as a quantitative basis for long term environmental management and sustainable emission reduction strategies for port authorities, shipping companies and other maritime stakeholders.

4.3. SHapley Additive exPlanations

The SHAP method (Lundberg, S.M.et al., 2017) is a global interpretation method which uses game theory to assess the influence of individual attributes on all the observations in the dataset.To understand the effects of different features on CO_2 emission predictions, robust eXplainable Artificial Intelligence (XAI) techniques, SHapley Additive exPlanations (SHAP), were used in this study.

The SHAP values offer an excellent tool to understand each individual prediction: they measure the importance of each feature to the model's prediction.The Shapley value function is as shown below (N.Q.Huy et al., 2023):

$$[\emptyset_i(v) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!}[v(S\cup\{i\}) - v(S)]] \tag{12}$$

Where:

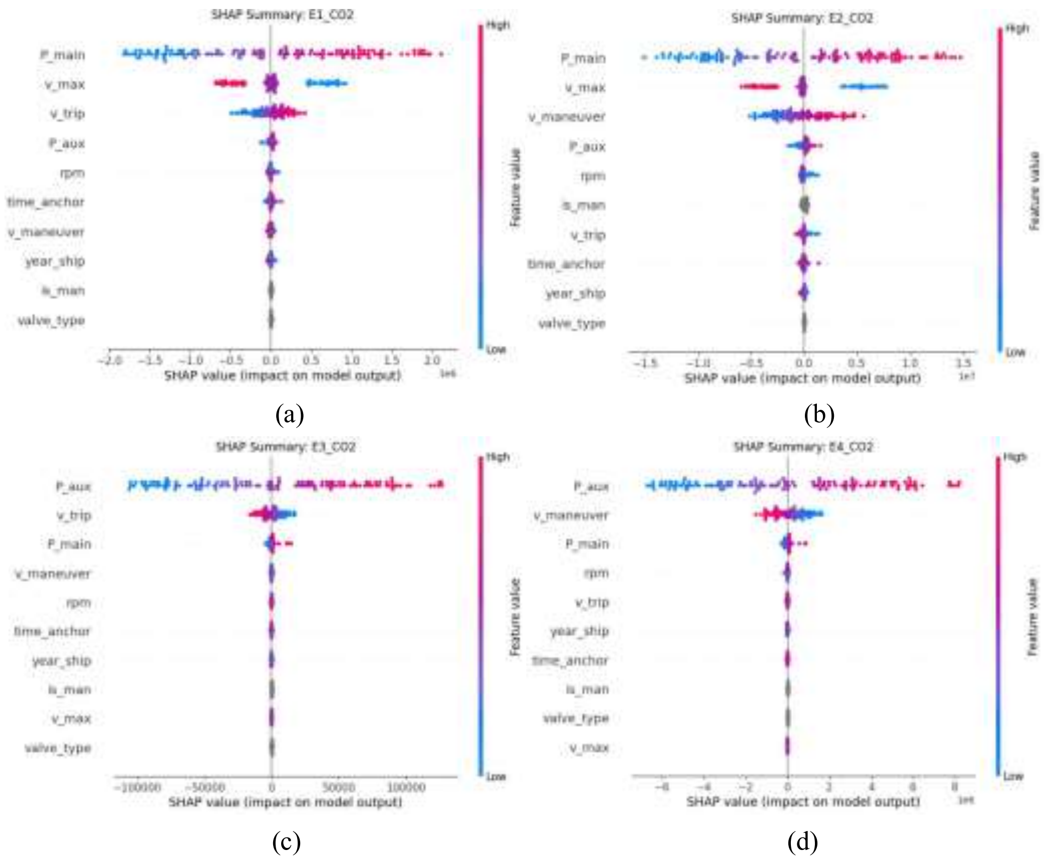
$\emptyset_i(v)$ is the contribution of the i-th feature to the value function;

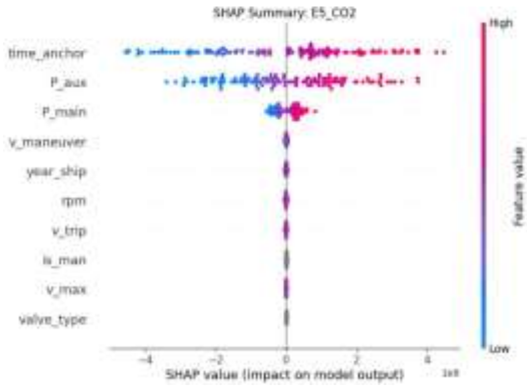
N is the set comprising n features (from 1 to n);

S is a subset of N that excludes feature i;

$|S|$ Is the number of elements in subset S.

SHAP method considers all possible combinations of the features and the features values, and the corresponding output, which ensures that the model's prediction is fairly distributed among the features.This means that the impact of each of the input variables is considered in the context of all other features.The features used in this study include ship age, main engine, deadweight, maximum speed, operational speed in different ship phases, and auxiliary engine.The research goal is to properly evaluate the influence of each of the input factors on the volume of the emission.





(e)

Fig 7: SHAP summary plots for CO_2 emissions under different engine types and operating phases: (a) main engine–cruising, (b) main engine–maneuvering, (c) auxiliary engine–cruising, (d) auxiliary engine–maneuvering, and (e) auxiliary engine–hoteling.

To check the contribution of the various input variables in the LightGBM model, SHAP analysis is carried out on the predicted CO_2 emission factors. CO_2 was chosen as the greenhouse gas of interest due to being the primary greenhouse gas emitted from marine vessels and being directly related to fuel usage. The importance of main engine power is high in E1 and E2, but this is largely due to the fact that it is a basic technical parameter that is used to estimate emissions and not a factor in operation. The variables varying during the operation of the vessel should therefore be given a greater attention. The clear influence of vessel speed on both E1 and E2 is associated with the fact that the more frequent speed changes during manoeuvre result in greater changes in the engine load and hence emissions. For E3 and E4, auxiliary engine emissions are highly correlated to onboard electrical demand, especially during maneuvering when auxiliary equipment and navigation systems are running continuously. In E5, hoteling time emerges as the major factor, suggesting that extended port stays have a significant impact on auxiliary engine emissions. This finding suggests the potential of operational measures such as shore power, just in time arrival and better berth scheduling to mitigate port generated emissions. Finally, the SHAP analysis shows that operational conditions offer more relevant information for emission mitigation than static vessel characteristics.

The impact of these emissions on the environment and health is significant.Even in the UK, with its sophisticated mitigation systems, more than 40,000 premature deaths each year can be blamed on the direct effects of emissions from maritime transport. There are no such in-depth studies in developing countries in South Asia and Southeast Asia, including Vietnam.Lack of such advanced emission reduction applications in these high traffic maritime zones has a significant effect on the health of the population living in coastal and port cities.

Thus, this research is not only a technical engineering problem, but it can be used as an important public health risk analysis tool. Port areas are extreme emission hotspots, which are associated with higher levels of mortality risk and pose an alarming situation that cannot be resolved without action. This study quantifies emissions during different operating conditions and engine types, thus providing an empirical basis for port authorities and policy makers to develop micro level management strategies. These involve making best use of shipping schedules to reduce cargo handling and anchoring times near residential areas.Strategic port emission reduction systems can only be developed when taking into account multifaceted impact factors. Several studies have been published to propose optimal operational measures: Yu H.et al.(2021) proposed a “virtual arrival” measure, which is to share real-time operational information with the incoming vessels in the seaport.In periods of congestion, ships autonomously slow down over the water and arrive at ports at the correct time. This reduces unnecessary fuel use, emissions and unnecessary anchoring. Berth and Schedule Optimization: With the growing number of traffic and the need for short-term infrastructure expansion, the work of Jiang, X.et al., (2022) proved that operational optimization is more feasible than physical expansion. They suggested combining the vessel scheduling problem in restricted channels with berth allocation, considering the tides, traffic conflicts and vessel-berth compatibility to solve resource conflicts.Park, S.A.et al.(2025) found relative wind to be a direct governing factor, which is wind-assisted routing. In headwind conditions (>12 knots), engine loads rise by about 12 per cent, which leads to an SFOC increase of 8.4 g/kWh and emissions of NOx over 1500 ppm a non-linear outbreak threshold. On the other hand, the crosswind/tailwind operations (6 - 10 knots at 30 - 40% engine load) provide the best operating conditions, with up to 6.7 g/kWh saved by the SFOC reduction.Considering wind in operational strategies can therefore result in a 5 - 10% reduction in emissions whilst avoiding spikes of emissions.

5. CONCLUSION

This research successfully applied the LightGBM–SHAP framework to monitor vessels calling at Hai An Port and demonstrated that it provides a highly accurate and data-driven alternative to conventional top-down emission inventory approaches. The results reveal that substantial emissions are generated during port operations and anchoring activities, where emission characteristics exhibit pronounced nonlinear behavior, highlighting the need for more effective emission control policies. At the regional level, the maritime sector should speed up the shift to cleaner marine fuels, including LNG and Methanol, and increase bunkering infrastructure to enable low carbon shipping. Port authorities are also urged to implement emission control measures progressively in coastal zones to minimise public exposure to NO_x and $PM_{2.5}$ emissions, such as carbon-restricted zones and shore power facilities. Through this study, a practical foundation is laid to develop data-driven emission inventories for vessels operating at Hai An port. Future research will involve combining transfer learning and federated learning to allow for collaborative training of models across multiple fleets while maintaining data privacy. It is expected that the accuracy of prediction can be improved for various vessel types and the proposed framework can be further developed into an exhaustive ship port emission inventory system. Such a system would enable scientific support for intelligent management of port environments and contribute to the pathway of the maritime sector towards the IMO 2050 Net-Zero Emission goals.

Author Contributions: Conceptualization, P.H., N.T and L.D.; methodology, D.K. and P.H.; validation, N.T. and T.A.; formal analysis, P.H.; in-vestigation, N.T. and T.A.; data curation, D.K.; writing original draft preparation, L.D. and D.K.; writing review and editing, P.H. and D.K.; visualization, N.T. and T.A. All authors have read and agreed to the published version of the manuscript.
Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
consent has been obtained from the patient(s) to publish this paper" if applicable.
Acknowledgments: This research was supported by Vietnam Maritime University. The authors are grateful to Vietnam Maritime University and colleagues for their support during the research. Furthermore, the authors would like to thank viewers for providing reviews that enhanced the manuscript.
Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

1. Alam, G. M. I., Tanim, S. A., Sarker, S. K., Watanobe, Y., Islam, R., Mridha, M. F., & Nur, K., 2025. Deep learning model based prediction of vehicle CO_2 emissions with explainable AI integration for sustainable environment. *Scientific Reports*, *15*, Article 3655. https://doi.org/10.1038/s41598-025-87233-y.
2. Al Nuaimi, H. S., Acquaye, A. and Mayyas, A., 2025. Machine learning applications for carbon emission estimation. *Resources, Conservation & Recycling Advances*, 27, 200263. doi: 10.1016/j.rcradv.2024.200263.
3. Al-Nefaie, A. H. & Aldhyani, T. H. H., 2023. Predicting CO2 emissions from traffic vehicles for sustainable and smart environment using a deep learning model. *Sustainability* doi: 10.3390/su15097615.
4. Alver, F., Saraç, B.A., Alver S¸ ahin, Ü., 2018. Estimating of shipping emissions in the Samsun Port from 2010 to 2015. *Atmospheric Pollution Research*, 9 (5), 822–828. doi: 10.1016/j.apr.2018.02.003.
5. Anonna, F. R., Mohaimin, M. R., Ahmed, A., Nayeem, M. B., Akter, R., Alam, S., Nasiruddin, M. and Hossain, M. S.,2024. Machine Learning-Based Prediction of U.S. CO2 Emissions: Developing Models for Forecasting and Sustainable Policy Formulation. *Journal of Environmental and Agricultural Studies*, 5(1), pp. 10–22. doi: 10.32996/jeas.2023.4.3.12
6. Bailey, D., Solomon, G., 2004. Pollution prevention at ports: Clearing the air. *Environmental Impact Assessment Review*, 24, 749 - 774. doi: 10.1016/j.eiar.2004.06.005.
7. Bojić Filip, Gudelj Anita, Bošnjak Rino., 2025. A Comprehensive Model for Quantifying, Predicting, and Evaluating Ship Emissions in Port Areas Using Novel Metrics and Machine Learning Methods. *Journal of Marine Science and Engineering*, tr.1162. doi: 10.3390/jmse13061162.
8. Bojić, F., Gudelj, A. and Bošnjak, R., 2025. A comprehensive model for quantifying, predicting, and evaluating ship emissions in port areas using novel metrics and machine learning methods. *Journal of Marine Science and Engineering*, 13, 1162. doi: 10.3390/jmse13061162.
9. Ministry of Natural Resources and Environment (MONRE)., 2024. Official Letter No. 1074/BTNMT-KSONMT dated 21 February 2024 on Technical Guidelines for Emission Inventories of Particulate Matter and Air Pollutants from Point, Area, and Mobile Sources. [in Vietnamese].
10. Chen, S., Meng, Q., Jia, P., Kuang, H., 2021. An operational-mode-based method for estimating ship emissions in port waters. *Transportation Research Part D: Transport and Environment*, 101, 103080. doi: 10.1016/j.trd.2021.103080.

11. Chen, W., Men, Y., Fuster, N., Osorio, C. and Juan, A. A., 2024. Artificial Intelligence in Logistics Optimization with Sustainable Criteria: A Review. *Sustainability*, 16(21), 9145. doi: 10.3390/su16219145.
12. Chen, Z. S., Lam, J. S. L. and Xiao, Z., 2024. Prediction of harbour vessel emissions based on machine learning approach. *Transportation Research Part D: Transport and Environment*, 131, 104214. doi: 10.1016/j.trd.2024.104214.
13. Cheng, C., Li, Z., Yan, Y., Cui, Q., Zhang, Y. and Liu, L., 2024. Maritime Freight Carbon Emission in the U.S. using AIS data from 2018 to 2022. *Scientific Data*, 11, 542. doi: 10.1038/s41597-024-03391-0.
14. Corbett, J.J.; Fischbeck, P.S., 1997. Emissions from ships. *Science* 278 (5339), 823 - 824. doi: 10.1126/science.278.5339.823.
15. Corbett, J.J.; Koehler, H.W., 2003. Updated emissions from ocean shipping. *Journal of Geophysical Research: Atmospheres*, 8, 87–107. doi: 10.1029/2003JD003751.
16. Corbett, J.J., Winebrake, J.J., Green, E.H., Kasibhatla, P., Eyring, V., Lauer, A., 2007. Mortality from ship emissions: a global assessment. *Environmental Science & Technology*, 41, 8512-8518. doi: 10.1021/es071686z
17. Dang, Q. V., Nguyen, M. D., & Van Hung, P., 2022. Assessment of potentially cutting GHG emissions from shipping in relation to energy consumption trends using Fuzzy Analytic Hierarchy Process. *Australian Journal of Maritime & Ocean Affairs*, 14(3), 189–203. https://doi.org/10.1080/18366503.2021.1997172
18. Deniz, C.; Durmusoglu, Y., 2008. Estimating shipping emissions in the region of the Sea of Marmara, Turkey. *Science of the Total Environment*, 390, 255–261. doi: 10.1016/j.scitotenv.2007.09.033.
19. Durlik, I., Miller, T., Kostecka, E., Łobodzińska, A. and Kostecki, T., 2024. Harnessing AI for Sustainable Shipping and Green Ports: Challenges and Opportunities. *Applied Sciences*, 14(14), 5994. doi: 10.3390/app14145994.
20. Eger, P., Mathes, T., Zavarsky, A. and Duester, L., 2023. Measurement report: Inland ship emissions and their contribution to NOx and ultrafine particle concentrations at the Rhine. *Atmospheric Chemistry and Physics*, 23(14), pp. 8769–8788. doi: 10.5194/acp-23-8769-2023.
21. Endresen, O.; Bakke, J.; Sorgard, E. Improved modelling of ship SO2 emissions a fuel-based approach., 2005. *Atmospheric Environment*, 39, 3621–3628. doi: 10.1016/j.atmosenv.2005.02.041.
22. Endresen, O., Sorgard, E., Sundet, J.K., Dalsoren, S.B., Isaksen, I.S.A., Berglen, T.F., Gravir, G., 2003. Emission from international sea transportation and environmental impact. *Journal of Geophysical Research*, 108, 4560. doi: 10.1029/2002JD002898.
23. Eyring, V., Kohler, H.W., van Aardenne, J., Lauer, A., 2005. Emissions from international shipping: 1. The last 50 years. *Journal of Geophysical Research*, 110, D17305. doi: 10.1029/2004JD005619.
24. Eyring, V., Isaksen, I.S.A., Berntsen, T., Collins, W.J., Corbett, J.J., Endresen, O., Grainger, R.G., Moldanova, J., Schlager, H., Stevenson, D.S., 2010. Transport impacts on atmosphere and climate: shipping. *Atmospheric Environment*, 44, 4735-4771. doi: 10.1016/j.atmosenv.2009.04.059.
25. Farzadmehr, M., 2025. AI-powered solutions assessment in port and maritime sector. PhD thesis. University of Antwerp.
26. Greenhouse Gas Protocol, 2024. IPCC Global Warming Potential Values (Version 2.0). World Resources Institute and World Business Council for Sustainable Development.
27. Hien, N. L. H. & Kor, A.-L., 2022. Analysis and prediction model of fuel consumption and carbon dioxide emissions of light-duty vehicles. *Applied Sciences.* doi: 10.3390/app12020803.
28. Hulskotte, J.; van der Gon, H.D., 2010. Fuel consumption and associated emissions from seagoing ships at berth derived from an on-board Survey. *Atmospheric Environment*, 44, 1229–1236. doi: 10.1016/j.atmosenv.2009.10.018.
29. IMO. (2020). Reduction of GHG Emissions from Ships - Fourth IMO GHG Study 2020 - Final report (MEPC 75/7/15). I. M. Organization.
30. IPCC, 2007. Climate Change 2007: The Physical Science Basis. Cambridge University Press, Cambridge, United Kingdom.
31. Jing M. A., Shuping H., and Kasjewruy P., 2014. Model Selection for Emission Models Based on Emission Factors. *Nature Environment and Pollution Technology.* 13(4), 781-786.
32. Jiang, X., Zhong, M., Shi, J., Li, W., Sui, Y. and Dou, Y., 2022. Overall Scheduling Model for Vessels Scheduling and Berth Allocation for Ports with Restricted Channels That Considers Carbon Emissions. *Journal of Marine Science and Engineering*, 10(11), 1757. doi: 10.3390/jmse10111757.
33. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 30, pp. 3146–3154. https://www.researchgate.net/publication/378480234_LightGBM_A_Highly_Efficient_Gradient_Boosting_Decision_Tree.
34. Kramel, D.; Muri, H.; Kim, Y.R., 2022. Correction to "Global Shipping Emissions from a Well-to-Wake Perspective: The Maritime Model". *Environmental Science & Technology*, 56, 5999–6000. doi: 10.1021/acs.est.2c01301.

35. Lee, H., Pham, H. T., Chen, M. and Choo, S., 2021. Bottom-Up Approach Ship Emission Inventory in Port of Incheon Based on VTS Data. *Journal of Advanced Transportation*, 5568777. doi: 10.1155/2021/5568777.
36. Lee, J., Eom, J., Park, J., Jo, J. and Kim, S., 2024. The Development of a Machine Learning-Based Carbon Emission Prediction Method for a Multi-Fuel-Propelled Smart Ship by Using Onboard Measurement Data. *Sustainability*, 16(6), 2381. doi: 10.3390/su16062381.
37. Long, N. X., & Van Hung, P., 2025. Energy efficiency and carbon emissions assessment of container terminal equipment at Tan Cang – Cai Mep Thi Vai. *International Journal of Sustainable Engineering*, 18(1). https://doi.org/10.1080/19397038.2025.2580726
38. Lundberg, S. M., & Lee, S. I., 2017. Explaining black box models with Shapley Values. *In Advances in Neural Information Processing Systems* (pp. 4765-4774). doi: 10.48550/arXiv.1911.11888
39. Mądziel, M., 2023. Liquified petroleum gas-fuelled vehicle CO_2 emission modelling based on portable emission measurement system, on-board diagnostics data, and gradient-boosting machine learning. *Energies*, 16, 2754, doi: 10.3390/en16062754.
40. Miola, A.; Ciuffo, B., 2011. Estimating air emissions from ships: Meta-analysis of modelling approaches and available data sources. *Atmospheric Environment*, 45, 2242–2251. doi: 10.1016/j.atmosenv.2011.01.046.
41. Moreno-Gutiérrez, J., Calderay, F., Saborido, N., Boile, M., Rodríguez Valero, R., Durán-Grados, V., 2015. Methodologies for estimating shipping emissions and energy consumption: a comparative analysis of current methods. *Energy*, 86, 603–616. doi: 10.1016/j.energy.2015.04.083.
42. Moreno-Gutiérrez, J. and Durán-Grados, V., 2021. Calculating ship's real emissions of pollutants and greenhouse gases: Towards zero uncertainties. *Science of the Total Environment*, 750, 141471. doi: 10.1016/j.scitotenv.2020.141471.
43. Nam, T.T., Hien, N.C., Long, N.X., Son, N.T., Dieu, N.T., Tai, N.N. and Hung, P.V., 2024. Control of Environmental Pollution from Maritime Activities. In: Maccarrone, V., Fadzil Akhir, M. (eds) *Coastal Sustainability. Coastal Research Library*, *vol 39. Springer, Cham*. https://doi.org/10.1007/978-3-031-75749-5_11
44. N. Q. Huy and T. L., 2023. Giải thích đặc trưng thẻ tín dụng theo phương pháp LIME và SHAP sau giai đoạn học sâu. *Ho Chi Minh City University of Education Journal of Science*, pp. 1746–1760, 2023. doi: 10.54607/hcmue.js.20.10.3906.
45. Nguyen, P. N., 2025. Ship emissions and social costs at Southeast Asian container ports. *Transportation Safety and Environment*, 7(4), tdaf067. doi: 10.1093/tse/tdaf067.
46. Nguyễn, T. A., Phạm, M. H., Lê, T. D., Nguyễn, N. T., Vũ, T. A. T., Vũ, M. P. và Trần, T. T., 2024. Đánh giá các mô hình dự báo bức xạ mặt trời: LightGBM, LSTM và GRU. *Tạp chí Khoa học & Công nghệ - Đại học Công nghiệp Hà Nội*, 60(5), tr. 147–152. doi: 10.57001/huih5804.2024.256.
47. Ning, Y., Li, T., Yang, L. and Chen, B., 2025. Data-Driven Analysis of Regional Ship Carbon Emission Reduction: The Bohai Bay Area Case Study. *Sustainability*, 17(3), 1159. doi: 10.3390/su17031159.
48. Nunes, R.A.O., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2017. The activity-based methodology to assess ship emissions - a review. *Environmental Pollution*, 231, 87–103. doi: 10.1016/j.envpol.2017.07.099.
49. Park, S.A., Je, M.-A., Jung, S.-H. and Park, D.-J., 2025. Data-Driven Optimization of Ship Propulsion Efficiency and Emissions Considering Relative Wind', *Journal of Marine Science and Engineering*, 13(11), 2120. doi: 10.3390/jmse13112120.
50. 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. doi: 10.1016/j.trd.2020.102617
51. Pham N. D. K., Dinh G.H., Pham H.T., Kozak J., Nguyen H.P., 2023. Role of Green Logistics in the Construction of Sustainable Supply Chains." *Polish Maritime Research*, 30 (3),191-211. https://doi.org/10.2478/pomr-2023-0052.
52. Prousaloglou, P., Kyriakopoulou-Roussou, M.-C., Stavroulakis, P. J., Tsioumas, V. and Papadimitriou, S., 2025. Artificial intelligence in the service of sustainable shipping. *Journal of Ocean Engineering and Marine Energy*, 11, pp. 621–653. doi: 10.1007/s40722-025-00390-0.
53. Sajjad, M., Chu-Van, T., Shahariar, G. M. H., Suara, K. A., Surawski, N., Bodisco, T. A., Ristovski, Z. D., Brown, R. J. and Zare, A., 2025. Ship emissions and fuel economy under transient conditions: Revisiting the propeller law. *Marine Pollution Bulletin*, 221, 118535. doi: 10.1016/j.marpolbul.2025.118535.
54. Saxe, H., Larsen, T., 2004. Air pollution from ships in three Danish ports. *Atmospheric Environment*, 38, 4057-4067. doi: 10.1016/j.atmosenv.2004.03.055.
55. Seyhan, A., 2025. A Novel Study for Machine-Learning-Based Ship Energy Demand Forecasting in Container Port. *Sustainability*, 17(13), 5612. doi: 10.3390/su17125612
56. Sha, M., Zhai, J., Wan, Z., Wu, Z., Jin, Y. and Zhu, E., 2024. Insights into the Global Characteristics of Shipping Exhaust Emissions at Berth. *Journal of Marine Science and Engineering*, 12(8), 1391. doi: 10.3390/jmse12091527.
57. Song, S., 2014. Ship emissions inventory, social cost and eco-efficiency in Shanghai Yangshan port. *Atmospheric Environment*, 82, pp. 288–297. doi: 10.1016/j.atmosenv.2013.10.006.

58. Starcrest Consulting Group, LLC., 2024. *San Pedro Bay Ports emissions inventory methodology report (Version 5)*. Prepared for the Port of Los Angeles and the Port of Long Beach. 2024.
59. Tena-Gago, D., Golkarnarenji, G., Martinez-Alpiste, I., Wang, Q. & Calero, J. M. A., 2023. Machine-learning-based carbon dioxide concentration prediction for hybrid vehicles. *Sensors* doi: 10.3390/s23031350.
60. Tombak, M.-L., Tapaninen, U. and Kotta, J., 2025. Methods for Calculating Greenhouse Gas Emissions in the Baltic Sea Ports: A Comparative Study. *Sustainability*, 17(2), 639. doi: 10.3390/su17020639.
61. Tran, N.K., Lam, J.S.L., Jia, H., Adland, R., 2022. Emissions from container vessels in the port of Singapore. *Maritime Policy & Management*, 49 (3), 306–322. doi: 10.1080/03088839.2021.1980236.
62. Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R.E. and Jamal, A., 2021. A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability. *Energy & Environment.* doi: 10.1177/0958305X211044998.
63. Wang, C.; Corbett, J.J.; Firestone, J,. 2008. Improving Spatial Representation of Global Ship Emissions Inventories. *Environmental Science & Technology*, 42, 193–199. doi: 10.1021/es0700799.
64. Wen, H.-T., Lu, J.-H., Lu, J. & Jhang, D.-S,. 2021.Features importance analysis of diesel vehicles NOx and CO2 emission predictions in real road driving based on gradient boosting regression model. *International Journal of Environmental Research and Public Health* doi: 10.3390/ijerph182413044.
65. World Health Organization,. 2000. *Air Quality Guidelines for Europe*, second ed. WHO Regional Publications, European Series No. 91, Copenhagen.
66. Xia, Y., Jiang, L., Wang, L., Chen, X., Ye, J., Hou, T., Wang, L., Zhang, Y., Li, M., Li, Z., Song, Z., Jiang, Y., Liu, W., Li, P., Rosenfeld, D., Seinfeld, J. H., & Yu, S,.2021. Rapid assessments of light-duty gasoline vehicle emissions using on-road remote sensing and machine learning. *Science of the Total Environment* doi: 10.1016/j.scitotenv.2021.152771.
67. Xiao, G., Yang, D., Xu, L., Li, J. and Jiang, Z,. 2024. The Application of Artificial Intelligence Technology in Shipping: A Bibliometric Review. *Journal of Marine Science and Engineering*, 12(4), 624. doi: 10.3390/jmse12040624.
68. Xu, X., Liu, X., Feng, L., Yap, W. Y. and Feng, H,. 2025. Emission Estimation and Spatiotemporal Distribution of Passenger Ships Using Multi-Source Data: A Case from Zhoushan (China). *Journal of Marine Science and Engineering*, 13(1), 168. doi: 10.3390/jmse13010168.
69. Yu, H., Fang, Z., Fu, X., Liu, J. and Chen, J,. 2021. Literature review on emission control-based ship voyage optimization. *Transportation Research Part D: Transport and Environment*, 93, 102768. doi: 10.1016/j.trd.2021.102768.