

Insights into Maritime Shipping's Carbon Footprint

Assessing the State of Emissions in Maritime Shipping Through Big Data Analysis

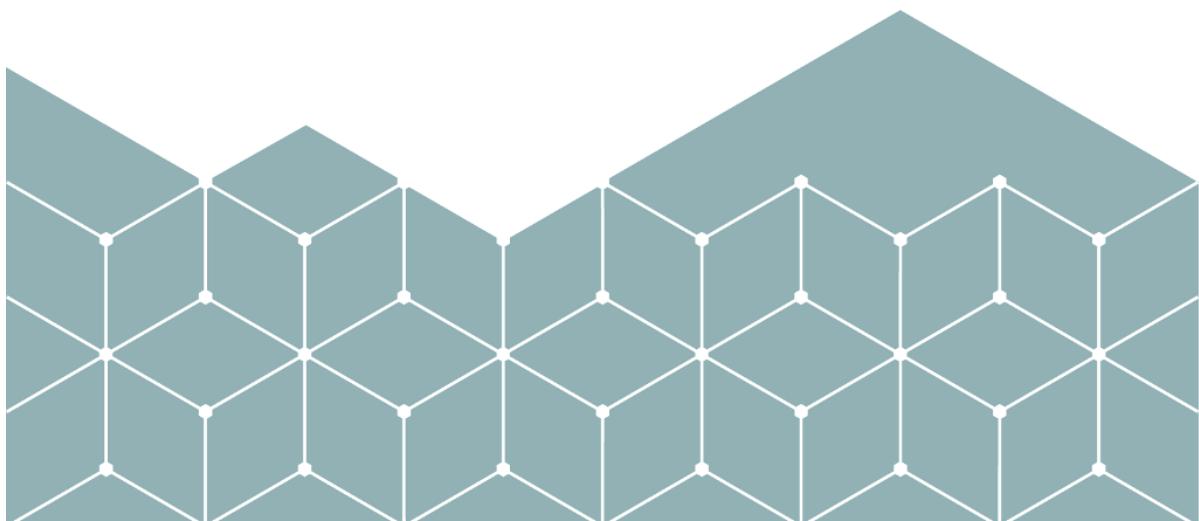
Freddy Rameshchandra Thobhani

Masters in Green Energy Technology (Smart Energy)

Department of Engineering

Lucian Mihet-Popa

August 14, 2023



Insights into Maritime Shipping's Carbon Footprint

*Assessing the State of Emissions in Maritime Shipping
Through Big Data Analysis*

Freddy Rameshchandra Thobhani

Abstract

This thesis offers an intricate exploration of maritime shipping's carbon emissions, unraveling the complexity through a systematic analysis. By categorizing vessels into distinct segments, the research provided a nuanced examination of emissions behavior and identified multifaceted factors affecting carbon emissions. Practical and cost-effective measures for emissions reduction were explored, contributing valuable insights to environmental sustainability in the maritime industry. The study also sets the stage for future research, emphasizing the inclusion of port emissions and analysis across more vessel segments. The findings signify a substantial advancement in understanding maritime shipping's carbon footprint, offering practical solutions, and laying a foundation for ongoing exploration and responsible industry practices.

Acknowledgements

I am deeply grateful to the numerous individuals who have played a crucial role in the successful completion of this thesis. First and foremost, I extend my sincerest appreciation to Professor Nicolae Lucian Mihet at Østfold University College, whose guidance, insights, and unwavering support have been invaluable throughout this research journey. I am also grateful to my supervisor, Kristian Amundsen Ruud, at Astrup Fearnley, for his expertise, thought-provoking ideas, and continuous support, which have significantly enriched the outcomes of this project.

I am equally thankful for the collaborative spirit and shared wisdom of the exceptionally talented team at Astrup Fearnley, with whom I had the privilege of working. Their support and contributions have undeniably elevated the quality of this work. I am also grateful for the unwavering support and encouragement of my cherished friends and family, who have been a constant source of motivation during this challenging academic pursuit.

While it is impossible to individually name all contributors, I wish to extend my heartfelt thanks to everyone who has contributed to this thesis in diverse ways. Your collective influence has been instrumental in shaping this endeavor, and your impact resonates profoundly in the culmination of this achievement.

Contents

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 Background and Motivation	1
1.2 Big Data Analysis	2
1.2.1 Big Data	2
1.2.2 What is Big Data Analytics?	3
1.3 Indicators	4
1.3.1 Carbon Intensity Indicator (CII)	5
1.3.2 Energy Efficiency Operational Indicator (EEOI)	6
1.3.3 Energy Efficiency Design Index (EEDI)	7
1.3.4 Energy Efficiency eXisting ship Index (EEXI)	8
1.4 Segments	9
1.4.1 Bulk Carriers	9
1.4.2 Tank Carriers	10
1.5 Problem Statement	10
1.6 Research Question	11
1.7 Report Outline	11
2 Literature Review	13
3 Data Acquisition and Preprocessing	18
3.1 IHS Markit dataset	19
3.2 The Automatic Identification System (AIS) dataset	20
3.3 AIS Trade Flow System	21
3.3.1 Vessel Voyages	22
3.3.2 Port Stop and Port Area	22
4 Tools and Data Analysis Techniques	24
4.1 Tools	24
4.1.1 Databricks	24
4.1.2 Apache Spark	26
4.1.3 Astrup Fearnleys Code Emission Calculator API	27
4.2 Methodology	29
5 Results and Discussion	35
5.1 Grade Distribution	35
5.2 Projected Grade Trend	37

5.3	EEOI vs Grade	38
5.4	Average speed vs Grade	38
5.5	Grade by Build Country	39
5.6	Addressing Low Rating	40
5.7	Emissions	40
6	Conclusion	41
6.1	Future Work	41

List of Figures

1.1	Emission trajectories for different levels of ambition for emission reduction targets	2
1.2	Big Data: 3 V's [13]	3
1.3	Big Data: Beyond 3 V's - volume, velocity, variety, and complexity	3
1.4	Schematic diagram of the CII ratings and boundaries. [28]	6
1.5	Thesis Outline	11
2.1	Number of publications per year in energy efficiency and emission reduction in the maritime domain	13
2.2	Simulation framework	15
2.3	EEXI BULKER ESTIMATES VS. 2023 BASELINE	16
2.4	Application of Big data and AI in maritime industry	17
3.1	Vessel Specification Sheet [9]	19
3.2	Working of AIS	20
3.3	Trade flow from 150,000 AIS sampled signals.	21
3.4	AIS Raw Data Processing	22
3.5	Port Area	23
4.1	Databricks Notebook	25
4.2	Databricks Configuration	26
4.3	Apache Spark Configuration	27
4.4	Methodology Overview	29
4.5	a and C constants	32
4.6	dd vectors for determining the rating boundaries of ship types	33
4.7	CII grading based on dd vector	34
5.1	Grade Distribution	35
5.2	Combined Grade	36
5.3	Projected Grade Trend	37
5.4	Mean EEOI vs Grade	38
5.5	Average Speed vs Grade	39
5.6	Grade By Build Country	39
5.7	Percentage of Emissions by Segment	40

Chapter 1

Introduction

1.1 Background and Motivation

In the 21st century, climate change is the biggest challenge faced by humanity. It poses a substantial danger to the survival of the inhabitants of our planet. Human activities such as deforestation, extraction and burning of fossil fuels have led to a rise in global temperatures. The consequences of such activities are an increase in sea levels, extreme weather events, and loss of biodiversity. There is an urgent and undeniable need to reduce greenhouse gas emissions and transition to a sustainable and low-carbon future.

Maritime shipping is essential to the global economy. It accounts for transporting 90% of the world's goods by volume. It is also a major source of greenhouse gas emissions, with the International Maritime Organization (IMO) estimating that maritime shipping accounts for 3% of global carbon dioxide emissions. While 3% may seem small, it is important to note that this is a rapidly growing sector. Without action, maritime shipping contribution to carbon emissions can increase by up to 10-13% in the next few decades. Due to this fact, there is a growing global effort to reduce emissions from this sector. [12].

The European Green Deal is a significant initiative by the European Union to make Europe the world's first climate-neutral continent by 2050 . It aims to transform various sectors, including shipping, to reduce environmental impacts. Through meticulous analysis of total carbon emissions within the maritime sector using advanced data techniques, this research can significantly contributes to the goals of the European Green Deal. It can aid in formulating policies, monitoring progress, and promoting sustainable practices, aligning with the Green Deal's emphasis on innovation, global cooperation, and eco-conscious industrial transformation [22]. This study's alignment with the European Green Deal underscores its relevance and importance within the broader context of sustainability-focused endeavors.

In accordance with Sustainable Development Goal 13, in 2018, the initial strategy was adopted by IMO's Environmental Protection Committee (MEPC), during its 72nd session at IMO Headquarters in London, United Kingdom. According to this strategy, the IMO will work towards reducing the total annual greenhouse gas emissions from international shipping

by at least 50% by 2050 compared to 2008 [29]. In the 76th session of MEPC in 2021, several mandatory measures were adopted to reduce greenhouse gas emissions from international shipping, which will help in achieving the goal of reducing emissions by 50% by 2050 [30]. One of the important measures is the Carbon Intensity Indicator (CII).

Maritime shipping is a complex and highly volatile system, generating very large data sets. Big data analytics can be used to understand the complex system and make informed decisions. It can facilitate operations such as monitoring of emissions and predictive analysis of vessel performance. This can help in reducing emissions and improving the efficiency of the maritime sector [33].

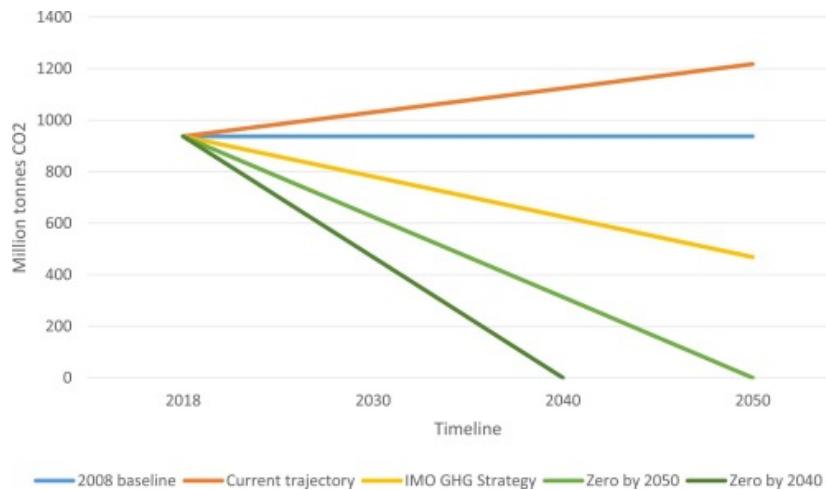


Figure 1.1: Emission trajectories for different levels of ambition for emission reduction targets

1.2 Big Data Analysis

Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things — *big data* and *analytics*.

1.2.1 Big Data

As the name suggests, big data is a large amount of data. There are other important attributes of big data. These are: data variety and data velocity.

Thus we can define big data using 3 V's: *volume*, *variety*, and *velocity* as shown in figure 1.2.

Beyond these three V's, Big Data is also about how complicated the computing problem is. Given the number of variables and number of data points for analysing the maritime shipping data. It is a very complicated problem. Thus, in addition to the three V's identified by IBM, it would also be necessary to take complexity into account as shown in figure 1.3 [15].

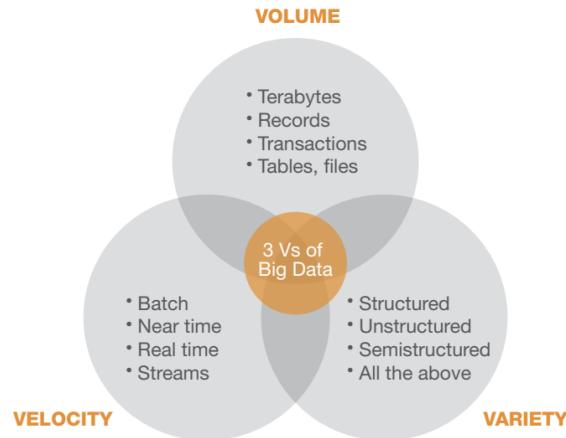


Figure 1.2: Big Data: 3 V's [13]

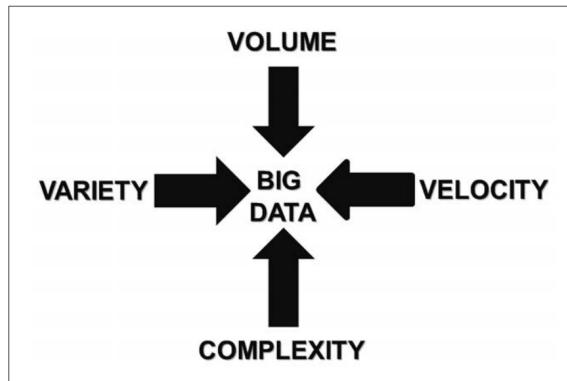


Figure 1.3: Big Data: Beyond 3 V's - volume, velocity, variety, and complexity

1.2.2 What is Big Data Analytics?

Big data analytics is the process of examining large and varied data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more-informed business decisions.

Thus, Data analytics revolves around deriving valuable knowledge and meaningful insights from extensive sets of data. This process involves crafting hypotheses, often rooted in gathered experiences and uncovering correlations between variables, sometimes even through serendipitous discoveries. Data analytics can be classified into four distinct types [18]:

1. Descriptive Analytics

Descriptive analytics focuses on explaining past events and presenting them in a comprehensible manner. The collected data is structured into visual aids like bar charts, graphs, pie charts, maps, and scatter diagrams, facilitating easy interpretation that offers insights into the data's implications. This mode of data representation is often termed a dashboard, reminiscent of

a car's dashboard that provides details such as speed, engine status, fuel levels, and distance traveled. A classic instance of descriptive analytics involves displaying population census data, which categorizes a nation's population by gender, age brackets, education, income, population density, and similar criteria [18].

2. Predictive Analytics

Predictive analytics extends beyond existing data to forecast forthcoming events. It anticipates what is likely to occur in the immediate future. Techniques like time series analysis utilizing statistical methods, neural networks, and machine learning algorithms are employed for this extrapolation. A significant application of predictive analytics is seen in marketing, where it understands customer preferences and needs. For instance, when purchasing shoes online, an advertisement for socks may appear. Another prevalent application is in orchestrating election campaigns. This involves gathering diverse data, such as the demographics of voters in different areas and their perceived needs like infrastructure and local concerns [18].

3. Prescriptive Analytics

This process detects opportunities for enhancing existing solutions by analyzing collected data. Essentially, it guides us on the actions to undertake in order to accomplish a particular objective. An illustrative instance is observed in the aviation industry where airlines determine seat pricing through analysis of historical travel patterns, popular travel origins and destinations, significant events, holidays, and more. This approach is employed to optimize profit generation [18].

4. Exploratory or Discovery Analytics

This process uncovers unforeseen connections among variables within extensive datasets. The collection and analysis of data from diverse sources opens up new avenues for gaining insights and making serendipitous discoveries. One of its major applications involves the identification of patterns in customers' behavior by companies through sources like feedback, tweets, blogs, Facebook data, emails, and sales trends. By deciphering customer behavior, companies can potentially predict actions like renewing a magazine subscription, switching mobile phone service providers, or canceling a hotel reservation. Armed with this information, companies can devise appealing offers aimed at altering the anticipated course of action by the customer [18].

1.3 Indicators

Indicators play a crucial role in assessing and monitoring carbon emissions in the maritime shipping industry. They provide valuable insights into the environmental performance of ves-

sels, facilitate comparisons between different ships or fleets, and help track progress towards emission reduction targets. By measuring various aspects of emissions and energy efficiency, these indicators enable stakeholders to identify opportunities for improvement and implement effective strategies to mitigate the environmental impact of shipping operations.

In this section, we will discuss several key indicators commonly used in the monitoring and evaluation of carbon emissions in maritime shipping. These indicators cover a range of factors, including carbon intensity, energy efficiency, fuel consumption, and cargo transport work. Each indicator offers a unique perspective on emissions, providing researchers, policymakers, and industry stakeholders with valuable information to support decision-making and foster sustainable practices.

It is important to note that the selection and use of indicators may vary depending on the specific research objectives, data availability, and regulatory frameworks in place. The combination of different indicators allows for a comprehensive assessment of emissions and enables a deeper understanding of the efficiency and environmental performance of shipping activities.

Below is the list of indicators discussed in this section:

1. Carbon Intensity Indicator (CII)
2. Energy Efficiency Operational Indicator (EEOI)
3. Energy Efficiency Design Index (EEDI)
4. Energy Efficiency eXisting ship Index (EEXI)

1.3.1 Carbon Intensity Indicator (CII)

The International Maritime Organization (IMO) has introduced a new carbon intensity (CII) measure for ships, which is a more accurate way to evaluate a vessel's environmental impact than total carbon emissions. CII is calculated using the Annual Efficiency Ratio (AER) formula, taking into account a ship's fuel consumption, CO₂ emission factor, annual distance sailed, and design deadweight.

To calculate CII in the most basic form:

$$\text{CII} = \frac{\text{Carbon Emission}}{\text{Distance Travelled} \times \text{Cargo Capacity}} \quad (1.1)$$

Vessels are rated A to E based on their CII results, and those with a D or E rating for three consecutive years or an E rating in one year must submit a corrective action plan. The IMO will enforce CII regulations for all ships over 5,000 GT and require an enhanced Ship Energy Efficiency Management Plan (SEEMP) with CII-related content from January 2023. The

SEEMP must include the ship's required annual operational CII target and an implementation plan to achieve it over the next three years. [2].

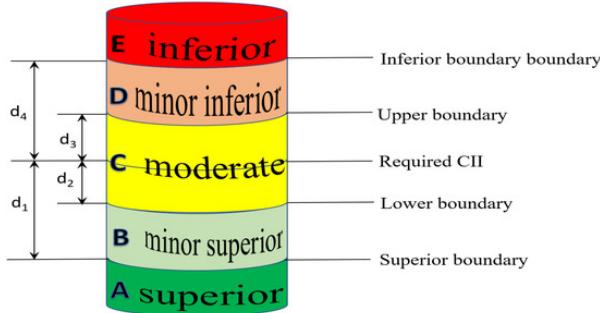


Figure 1.4: Schematic diagram of the CII ratings and boundaries. [28]

The article [20], discusses the increasing popularity of Energy and Carbon Intensity indicators among policy makers, which are calculated as units of energy or mass of emissions per unit of Gross Domestic Product (GDP). These indicators are gaining momentum due to support from think tanks, public organizations, consulting groups, and academics focused on energy and climate change policy development. The ways in which intensity indicators are framed and perceived in public debates generated by these intermediaries are important as they can influence policy making and the development of better sets of indicators to assess how well countries are addressing climate change and resource efficiency policies. Intensity indicators are appealing for emerging economies as they are not incompatible with high rates of economic growth and do not imply the imposition of absolute emission/energy caps.

1.3.2 Energy Efficiency Operational Indicator (EEOI)

The Energy Efficiency Operational Indicator (EEOI) is a tool used to measure the CO₂ gas emissions per unit of transport work, indicating the operational efficiency of a ship. The EEOI is calculated annually and is subject to changes after each voyage due to various external factors, such as navigation conditions, sea area, weather, temperature, and cargo weight.

The EEOI provides an accurate measure for each voyage, and its unit depends on the type of cargo or transport work, such as tons CO₂/(tons/nautical miles), tons CO₂/(TEU/nautical miles), or tons CO₂/(person/nautical miles). The formula for calculating the EEOI is represented by formula (1.2), where a lower value indicates a more energy-efficient ship [17].

$$\text{EEOI} = \frac{\text{Carbon Emission}}{\text{Performed Transport Work}} \quad (1.2)$$

For the calculation of EEOI for a specific voyage, formula (1.3) is used.

$$\text{EEOI} = \frac{\sum_j F_{Cj} \cdot C_{Fj}}{m_{\text{cargo}} \cdot D_j} \quad (1.3)$$

However, when dealing with a large number of ships, formula (1.3) is expressed as equation (1.4), taking into account parameters such as fuel type, voyage number, fuel consumption, fuel-to-CO₂ conversion factor, cargo weight, and distance traveled [27].

$$\text{Average}_{\text{EEOI}} = \frac{\sum_i \sum_j (F_{C_i} \cdot C_{F_j})}{\sum_i (m_{\text{cargo},i} \cdot D_i)} \quad (1.4)$$

where:

j : Fuel type used

i : Navigation voyage number

FC_{ij} : Mass of consumed fuel *j* at voyage *i*

CF_j : Fuel mass to CO₂ mass conversion factor with fuel *j*

m_{cargo} : Weight of cargo carried (tons) on ship

D_i : Distance of voyage *i* (nautical miles)

The fuel-to-CO₂ conversion factor (CF) is a non-dimensional factor that converts fuel consumption, measured in grams, to CO₂ gas emissions, also measured in grams, based on the carbon content. The below table

1.1 is showed the certain value of CF follows the type of fuel.

No.	Type of fuel	Reference	Carbon content	CF (t-CO ₂ /t-Fuel)
1	Diesel/gas oil	ISO 8217 Grades DMX through DMC	0.875	3.206000
2	Light fuel oil (LFO)	ISO 8217 Grades RMA through RMD	0.86	3.151040
3	Heavy fuel oil (HFO)	ISO 8217 Grades RME through RMK	0.85	3.114400
4	Liquefied petroleum gas (LPG)	Propane, butane	0.819, 0.827	3.000000, 3.030000
5	Liquefied natural gas (LNG)		0.75	2.750000

Table 1.1: The value of CF (t-CO₂/t-Fuel) [27].

1.3.3 Energy Efficiency Design Index (EEDI)

Energy Efficiency Design Index (EEDI) is a legislation proposed by the International Maritime Organization (IMO) to estimate the energy efficiency of ships and calculate their CO₂ emissions per unit of transport work done during the ship design phase. EEDI is based on a complex formula that takes into account the ship's emissions, capacity, and speed, and the lower the ship's EEDI index, the less CO₂ emissions it produces.

EEDI is a non-prescriptive mechanism that allows the shipping industry to use the latest technologies for designing commercial vessels as long as they meet the required energy efficiency levels and parameters. It lays down a minimum energy efficiency level, per capacity mile, for different ship types and sizes, including tankers, bulk carriers, gas carriers, general cargo ships, container ships, refrigerated cargo carriers, and combination carriers.

The EEDI formula has two components: attained EEDI and required EEDI. The attained EEDI is calculated using a complex formulation based on the vessel's emissions, capacity, and

speed, while the required EEDI is the minimum level of energy efficiency that a ship must meet as per its ship type and size. The attained EEDI is verified based on the ship's design and construction, and the required EEDI is the target that the ship must achieve during its operation [19].

EEDI calculation module as part of Marpol Annex VI, following the directive MEPC.1/Circ.681 at the MEPC meeting conducted by the IMO in 2011. This regulation came into effect on January 1, 2013. The EEDI formula (Equation 1) specified by IMO (2011) is represented by the equation (1.5) [26].

$$EEDI = \frac{P \cdot SFC \cdot Cf}{DWT \cdot V_{ref}} \quad (1.5)$$

where:

P : 70% of the power of the engine (main and auxiliary) in kW

SFC : Amount of fuel burned by the engines in kW (specific fuel consumption)

Cf : Emission rate of fuel used by the ship (presented in Table 1)

DWT : Ship's capacity (in tons)

V_{ref} : Speed of the ship (in knots)

1.3.4 Energy Efficiency eXisting ship Index (EEXI)

The Energy Efficiency Existing Ship Index (EEXI) is a regulation introduced by the International Maritime Organization (IMO) aimed at improving the energy efficiency of existing ships. It is part of the broader effort to reduce greenhouse gas emissions from the shipping industry and combat climate change. The EEXI is designed to complement the Energy Efficiency Design Index (EEDI), which focuses on new ship designs [4].

The EEXI is part of a comprehensive framework that includes short-term, mid-term, and long-term measures. The short-term measures focus on technical and operational improvements, such as retrofitting ships with energy-efficient technologies. The mid-term measures involve market-based mechanisms to incentivize emission reductions, while the long-term measures explore alternative fuels and propulsion systems [3].

The calculation of the Energy Efficiency Existing Ship Index (EEXI) can be optimized by considering the ship's maximum continuous rating (MCR) at 100% capacity. This approach ensures that improvements in technical efficiency closely align with the ship's actual operational fuel use. By accounting for the engine power limits (EPLs) within the engine margin, which have minimal impact on ship operations, a more accurate assessment of energy efficiency can be achieved. Currently, the proposed calculation methods for the EEXI involve using either 75% of the limited MCR (MCRlim), similar to the Energy Efficiency Design Index (EEDI), or a higher value of 87% MCRlim, which only considers the engine margin. However, utilizing ship characteristics data from IHS Markit and applying the appropriate calculation method, the attained EEXI score can still be estimated, providing valuable insights

into a ship's energy performance [21].

$$\text{Attained EEXI} = 3.1144 \times \frac{\text{MESFOC} \times \sum_{i=1}^{nME} P_{ME,i} + \text{AESFOC} \times P_{AE}}{\text{Capacity} \times V_{ref}} \quad (1.6)$$

The estimation of the attained EEXI score using ship characteristics data from IHS Markit, as outlined in Equation (1.6), contributes to a more comprehensive understanding of a ship's energy efficiency. This information supports decision-making processes related to optimizing operational fuel consumption, implementing retrofit measures, and promoting environmental sustainability in the maritime industry. By calculating the EEXI at 100% MCR, the assessment takes into account the EPLs within the engine margin, which are not expected to significantly affect ship operations. This approach ensures that technical efficiency improvements are properly aligned with the ship's actual operational fuel use, enabling informed decision-making for enhancing operational fuel consumption and reducing environmental impact. Through the utilization of ship characteristics data and the appropriate calculation method, the attained EEXI score serves as a valuable tool for assessing energy efficiency and driving advancements in the maritime sector [21].

1.4 Segments

The maritime shipping industry encompasses many more segments, this thesis will specifically focus on bulk carriers and tank carriers.

1.4.1 Bulk Carriers

Bulk carriers are specialized vessels designed to transport unpackaged bulk cargo such as grains, coal, ores, and cement. These vessels play a crucial role in global trade, ensuring the efficient and cost-effective movement of essential raw materials. Bulk carriers can be further categorized into several segments based on their size, design, and specific functionality:

- **Supramax:** A class of bulk carriers typically ranging from 50,000 to 60,000 deadweight tons (DWT), offering flexibility in port access and cargo handling.
- **Aframax:** Medium-sized vessels with a DWT of approximately 80,000, often used for short to medium-haul routes.
- **Handysize:** Smaller vessels ranging from 15,000 to 35,000 DWT, ideal for accessing ports with size restrictions.
- **Handymax:** Slightly larger than Handysize, with a DWT ranging from 40,000 to 50,000.
- **Panamax:** Specifically designed to pass through the Panama Canal, Panamax bulk carriers have a maximum DWT of around 80,000.

- **Capesize:** The largest bulk carriers, often exceeding 100,000 DWT, are named for their inability to traverse the Cape of Good Hope or Cape Horn, requiring them to navigate around these capes.

1.4.2 Tank Carriers

Tank carriers, also known as tanker ships, are designed to transport liquids such as crude oil, petroleum products, and chemicals. These vessels play a vital role in the energy sector, connecting oil-producing regions with global markets. Tank carriers can be categorized into several segments, including:

- **VLCC (Very Large Crude Carriers):** Among the largest tank carriers, VLCCs typically have a DWT ranging from 200,000 to 320,000 tons.
- **VLGC (Very Large Gas Carriers):** Designed to transport liquefied petroleum gas (LPG) and other gas products, with a capacity ranging from 70,000 to 85,000 cubic meters.

1.5 Problem Statement

Carbon emissions from maritime shipping have been identified as a major contributor to global greenhouse gas emissions, with the International Maritime Organization estimating that shipping is responsible for around 3% of global CO₂ emissions [12]. To address this issue, the shipping industry has set targets to reduce its carbon footprint, and governments and international organizations have introduced policies and regulations to encourage emissions reduction.

However, measuring and monitoring carbon emissions from maritime shipping can be challenging due to the complexity of the industry and the lack of reliable data. The Energy Efficiency Operational Indicator (EEOI) and the Carbon Intensity Indicator (CII) have been proposed as two metrics to assess the carbon efficiency of ships and enable comparison between different vessels and fleets [34, 3]. However, there is a need to better understand the relationship between EEOI and carbon emissions, as well as to identify the factors that influence this metrics.

This thesis aims to address this problem by leveraging big data analysis and detailed ship characteristics data to provide a more precise and holistic view of maritime shipping's carbon footprint. By doing so, it seeks to enhance decision-making processes related to fuel optimization, retrofit measures, and environmental sustainability, contributing to a more responsible and efficient maritime industry.

1.6 Research Question

This thesis will focus on answering following research questions:

1. How to understand complexity of emissions with simple system?
2. How does the categorisation of vessels into different segments facilitate a more intricate examination of carbon emissions?
3. What are the factors affecting the carbon emissions of a vessel?
4. Can the carbon emissions be reduced with cost effective and realistic measures?

1.7 Report Outline

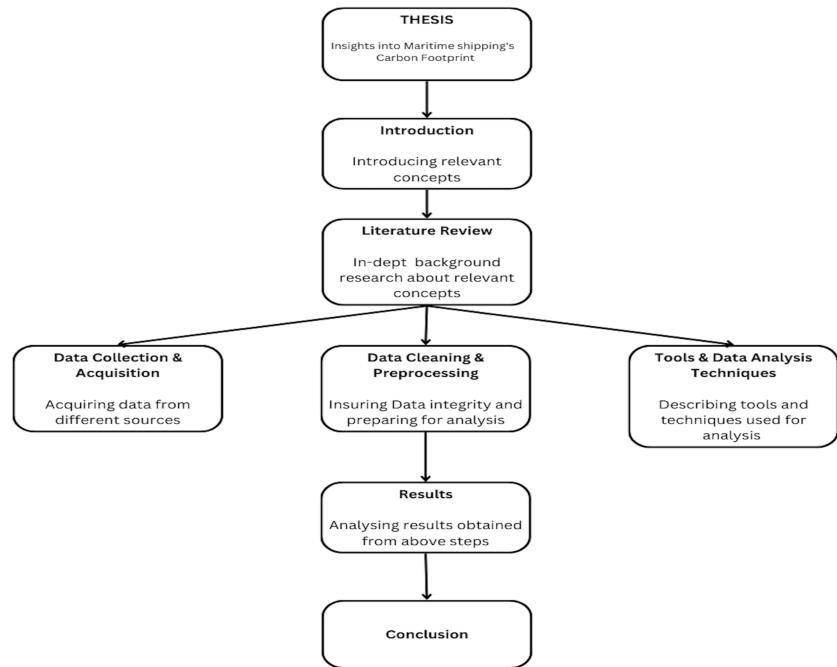


Figure 1.5: Thesis Outline

Chapter 2: Literature Review: This chapter covers the background information and literature review of the thesis. This section covers comprehensive review of existing papers and research related to carbon emissions in maritime shipping. The aim is to provide a comprehensive overview of the current state of research in this field and identify any gaps or opportunities for further exploration.

Chapter 3: Data Collection and Understanding: In this chapter, the focus will be on gathering and understanding the data required to perform analysis to understand emissions in maritime

shipping. Various data sources will be explored, including industry databases, research publications, and government reports. The aim is to acquire a comprehensive dataset that covers different aspects of carbon emissions in the maritime sector. Additionally, this chapter will delve into the intricacies of the collected data, understanding its structure, variables, and potential limitations.

Chapter 4: Data Cleaning and Preprocessing: Before conducting any data analysis, it is crucial to ensure the quality and integrity of the dataset. This chapter will discuss the steps involved in cleaning and preprocessing the data. This process may involve handling missing values, dealing with outliers, standardizing formats, and resolving inconsistencies. By performing these necessary data cleaning procedures, the dataset will be prepared for further analysis, ensuring reliable and accurate results.

Chapter 5: Data Analysis Techniques: In this chapter, various data analysis techniques specific to big data will be explored and applied to the cleaned dataset. These techniques may include statistical analysis, machine learning algorithms, and data visualization methods. The goal is to extract meaningful insights and patterns from the data to gain a comprehensive understanding of carbon emissions in maritime shipping. Additionally, this chapter will discuss the tools and technologies utilized for data analysis and highlight any specific challenges encountered during the process.

Chapter 6: Evaluation of Results: After performing the data analysis, this chapter will focus on evaluating and interpreting the obtained results. The findings will be compared against existing literature, industry benchmarks, and regulatory standards to assess the significance and implications of the analysis. The strengths and limitations of the analysis approach will be discussed, and recommendations for future research or practical applications will be provided. This chapter aims to provide a comprehensive evaluation of the insights gained from the data analysis and their potential impact on the maritime shipping industry.

Conclusion: The conclusion chapter will summarize the key findings and contributions of the thesis. It will highlight the significance of the conducted big data analysis on emissions in maritime shipping and its implications for sustainability and environmental initiatives. The conclusion will also discuss any potential limitations or challenges encountered during the research and suggest avenues for further exploration in this field.

Chapter 2

Literature Review

In response to the urgent need to reduce carbon emissions and combat climate change, researchers and industry stakeholders have focused on developing and implementing strategies to reduce carbon emissions in maritime shipping.

Figure 2.1 shows that the number of publications on energy efficiency and emission reduction in the maritime industry has grown exponentially since 2016. The number of publications from 2006 to 2015 was 76, while from 2016 to 2021, there were 260 publications, indicating a significant increase in interest in decarbonization in the maritime industry. [10]

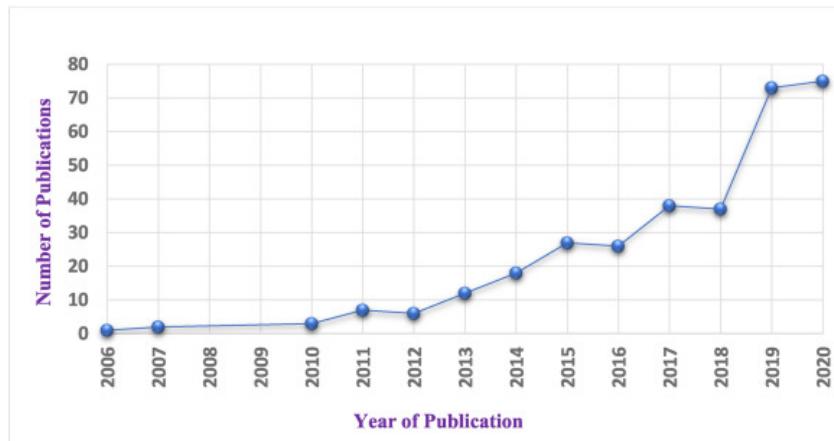


Figure 2.1: Number of publications per year in energy efficiency and emission reduction in the maritime domain

One promising area of research is the use of big data analysis to measure and improve carbon efficiency in maritime shipping. Big data analysis involves the collection and analysis of large and complex data sets to identify patterns, trends, and insights. In the context of maritime shipping, big data analysis can be used to measure carbon emissions and identify opportunities for improvement.

The purpose of this literature review is to examine the current state of research on carbon emissions in maritime shipping, with a focus on the Energy Efficiency Existing ship Index (EEXI) and Carbon Intensity Indicator (CII) as key metrics for measuring carbon

efficiency. The review will provide an overview of the current state of research on these metrics, their strengths and limitations, and their relevance for the maritime shipping industry.

The review will begin by exploring the importance of reducing carbon emissions in maritime shipping and the regulatory and policy frameworks that have been established to address this issue. It will then provide an overview of the EEXI and CII metrics, including their definitions, methodologies for calculating them, and their role in measuring carbon efficiency.

The literature review will also examine the current research on the relationship between EEXI, CII, and carbon emissions in maritime shipping, with a particular focus on the use of big data analysis to measure and improve carbon efficiency. It will explore the potential for big data analysis to provide more accurate and comprehensive data on carbon emissions, and to identify opportunities for operational and technological improvements.

Overall, this literature review will provide a comprehensive overview of the current state of research on carbon emissions in maritime shipping, with a focus on the EEXI and CII metrics and the potential for big data analysis to guide and inform strategies for improving carbon efficiency in the industry.

Review by Issa, Ilinca, and Martini [8] shows that Maritime shipping is a crucial aspect of global trade and the global economy, with over 85% of the volume of global trade in goods transported by sea. However, maritime transport also has significant environmental impacts, including carbon emissions. Approximately 3.3% of the world's carbon dioxide (CO₂) emissions are attributable to maritime transport, with emissions from marine diesel oil (MDO), marine fuel oil (MFO), and heavy fuel oil (HFO) all contributing to the problem. Reducing carbon emissions in the maritime shipping industry is a significant challenge, but there are a range of strategies that can be used to achieve this goal. Alternative fuels, energy efficiency improvements, and operational measures all have the potential to reduce emissions, but they also have significant economic and resource constraints.

Paper by Grzelakowski, Herdzik, and Skiba [6] mentions that Despite its significant contribution to global economic growth, maritime transport also generates negative externalities, primarily in the form of greenhouse gas (GHG) emissions. They discuss how digitalization and the use of artificial intelligence (AI) are being explored as potential ways to reduce emissions in maritime shipping. AI algorithms can optimize shipping routes, reduce fuel consumption, and minimize emissions. Additionally, digitalization can enable better data collection and analysis, which can facilitate more accurate emissions reporting and monitoring.

According to Kao, Chung, and Chen [11], the use of automatic identification system (AIS) to estimate ship emissions, which is an advantage due to the system's ability to provide real-time navigational information. Studies have been conducted utilizing AIS data to estimate ship emissions in different regions, such as Hong Kong and the Pearl River Delta, Las Palmas Port, Qingdao Port, Tianjin port, Naples port, and unidentified vessels with missing ship parameters. The studies have focused on macro-scale spatial and temporal resolution, high-resolution ship emission inventory, high temporal-spatial ship emission inventory, higher spatial-temporal resolution, and real-time ship emission monitoring. It proposes simulation model based on AIS data, specification and what-if scenarios as shown in Figure 2.2

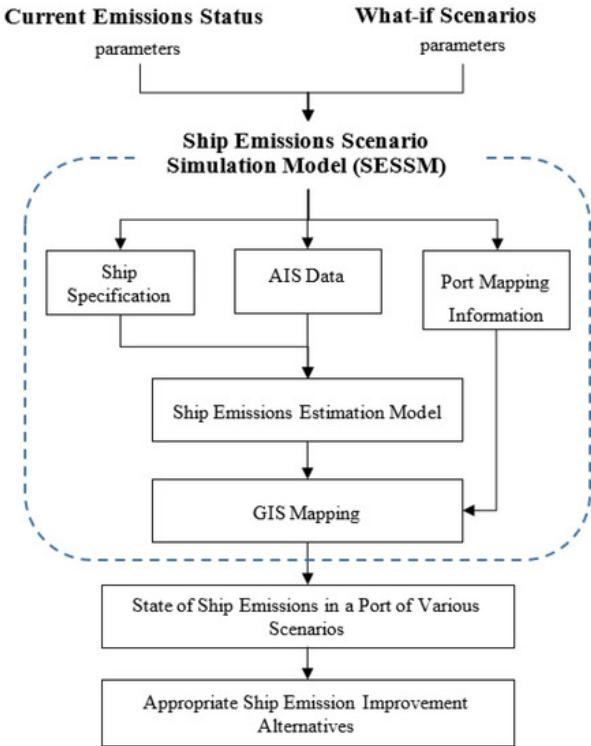


Figure 2.2: Simulation framework

Research by Sou et al. [23], discusses the need for carbon intensity indicators (CIIs) as performance monitoring tools in the shipping industry, particularly for tracking energy efficiency trends and progress towards climate targets. The review highlights the lack of consensus on suitable CIIs, as proposed by various countries to the International Maritime Organization (IMO), and the need for a more comprehensive understanding of global progress towards carbon intensity targets from both demand and supply side indicators. The study aims to address this issue by analyzing CIIs for shipping and the factors that influence the carbon intensity of shipping at the global level. Index decomposition analysis (IDA) is used to quantify the contribution of various factors, including energy efficiency, to changes in carbon intensity from 2012 to 2018.

According to report by Stevenson [25], The International Maritime Organization will introduce Energy Efficiency eXisting ship Index (EEXI) and Carbon Intensity Indicator (CII) regulations in 2023 as part of the wider decarbonisation goals for shipping. More than three-quarters of the existing fleet will not initially meet EEXI baselines and will need to take action to achieve compliance, with overridable engine power limitations (oEPL) expected to be a popular option. However, the effect on vessel operations over a year will be quite small due to the relatively small number of hours where steaming speeds would exceed oEPL limits. The compliance with EEXI can result in modest improvements in AER, CII, and annual CO₂ emissions.

The International Maritime Organization (IMO) has made maritime decarbonization a priority, setting targets to reduce greenhouse gas emissions from ships. To achieve these targets, the IMO has adopted mandatory measures, including the carbon intensity indicator (CII), which measures carbon emissions per unit transport work for each ship. But Wang, Psaraftis,

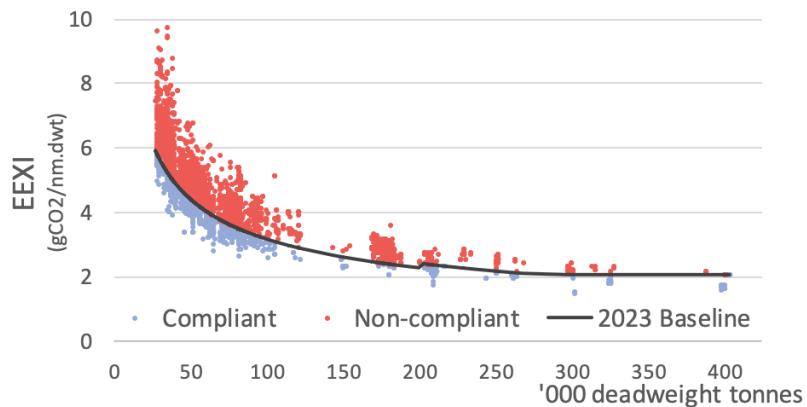


Figure 2.3: EEXI BULKER ESTIMATES VS. 2023 BASELINE

and Qi [31] argues there are potential paradoxes with the CII, as it may increase carbon emissions in some situations. There are at least four potential versions of the CII, including supply-based, demand-based, distance-based, and sailing time-based, but the IMO has not yet agreed on which to use. More elaborate models and indicators should be developed to analyze the potential impacts of the CII and achieve utmost carbon emissions reduction.

In “Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions” [14], author Munim et al. explains how Big data and artificial intelligence (AI) have become essential components of data-driven decision-making in most industries. However, the maritime industry still relies on intuition more than on data, mainly because of the vast size of its network and planning problems. The maritime industry generates large amounts of data that, if appropriately utilised in decision-making, can improve maritime safety, reduce environmental impacts, and minimise cost. In this review, we focus on studies that deal with big data and AI applications within the maritime context to map the conceptual structure of the field and identify future research avenues. AIS data to investigate the impact of speed reduction on fuel consumption and carbon emissions in the shipping industry. The study found that a 10% reduction in ship speed could result in a 17% reduction in fuel consumption and a corresponding reduction in carbon emissions. The authors suggested that reducing ship speed is an effective way to reduce fuel consumption and carbon emissions in the maritime industry.

The study by Acomi and Acomi [1], delves into the realm of maritime environmental conservation, emphasizing the Energy Efficiency Operational Index (EEOI) as a pivotal tool. Addressing concerns of marine pollution encompassing water and air aspects, the paper underscores international efforts for emission reduction in shipping. The EEOI, designed to measure carbon emissions per unit of transport work, aids ship-owners and operators in enhancing energy efficiency during operational voyages. The research demonstrates how commercial software and a custom-developed program estimate EEOI values pre-voyage and onboard, revealing the influence of unpredictable factors on energy efficiency. This analysis not only showcases the EEOI's significance in curbing emissions but also highlights its vital role in the broader context of maritime sustainability and environmental protection.

In conclusion, the literature reviewed emphasizes the importance of addressing the sig-

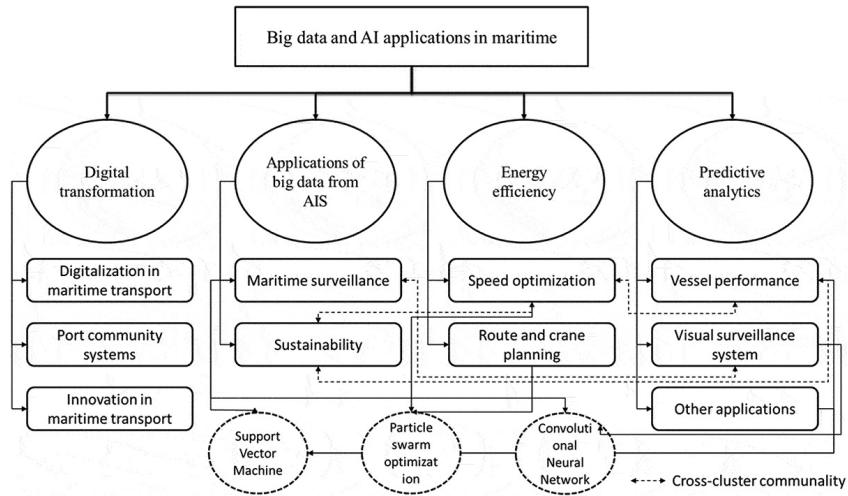


Figure 2.4: Application of Big data and AI in maritime industry

nificant environmental impact of carbon emissions in the maritime shipping industry. While there are various strategies to reduce emissions, such as alternative fuels, energy efficiency improvements, and operational measures, they have significant economic and resource constraints. Digitalization and the use of artificial intelligence (AI) are being explored as potential ways to reduce emissions by optimizing shipping routes, reducing fuel consumption, and minimizing emissions. Furthermore, the use of automatic identification system (AIS) data can facilitate real-time emissions monitoring, while carbon intensity indicators (CIIIs) can be used as performance monitoring tools. The International Maritime Organization (IMO) has made maritime decarbonization a priority by setting targets to reduce greenhouse gas emissions from ships and introducing regulations like EEXI and CII. However, potential paradoxes with the CII and lack of consensus on suitable CIIIs highlight the need for more elaborate models and indicators to achieve utmost carbon emissions reduction. Big data and AI applications have the potential to improve maritime safety, reduce environmental impacts, and minimize costs. Overall, more research is needed to address the challenges of monitoring and reducing carbon emissions in the maritime shipping industry while meeting global trade demands.

Chapter 3

Data Acquisition and Preprocessing

Data collection and understanding play a pivotal role in extracting meaningful insights and deriving accurate conclusions. The success of any analytical endeavor heavily relies on the quality, comprehensiveness, and relevance of the data used. In the case of carbon emissions in maritime shipping, gathering and understanding the data is crucial for capturing the intricacies of this complex domain. By exploring various data sources from industry databases, a comprehensive dataset can be acquired, encompassing diverse dimensions of carbon emissions in the maritime sector. Furthermore, understanding the structure, variables, and limitations of the collected data is essential for ensuring the validity and reliability of subsequent analyses. This includes examining the completeness of data, identifying any biases or data gaps, and verifying the accuracy of measurements. Ultimately, a thorough and informed understanding of the data sets the foundation for conducting rigorous data analysis and generating actionable insights for addressing carbon emission challenges in maritime shipping.

Collaboration with Astrup Fearnleys Code has been instrumental in enhancing my research on carbon emissions in maritime shipping. As a leading firm in the maritime shipping industry, their expertise and industry connections have provided me with invaluable support and access to crucial datasets. Specifically, through their collaboration, I was able to gain access to two significant datasets: Automatic Identification System (AIS) data and IHS dataset. Collaborating with Astrup Fearnleys Code and leveraging their industry expertise has not only provided me with valuable datasets but also allowed me to gain insights into the complex dynamics of the maritime shipping industry.

3.1 IHS Markit dataset

The IHS Markit dataset provides valuable vessel specification data for big data analysis in maritime shipping. This dataset offers comprehensive information on vessel characteristics, including dimensions, tonnage, engine capacity, and ownership. By leveraging this dataset, researchers can gain insights into the diverse specifications of vessels operating in the maritime industry.

Analyzing vessel specifications from the IHS Markit dataset allows for a deeper understanding of the maritime shipping landscape. Researchers can explore correlations between vessel characteristics and various factors such as fuel efficiency, cargo capacity, or operational performance. These insights can aid in optimizing vessel selection, fleet management, and decision-making processes related to vessel operations.

By utilizing the vessel specification data from the IHS Markit dataset, this research contributes to enhancing operational efficiency and optimizing vessel-related decisions in the maritime shipping industry. This dataset equips this research to identify trends and patterns in the maritime shipping industry related to vessel characteristics. This comprehensive dataset serves as a robust foundation for my research, enabling me to draw meaningful conclusions and make data-driven recommendations for the future of the industry.

RTFS SERVICIOS ESPECIALIZADOS S.A. DE C.V.		VESSEL SPECIFICATION SHEET	
ASME DEMO VESSEL		JOB NUMBER: 00012000	
CUSTOMER: ASME DEMO		PAGE: 1 OF 1	
PROJECT: ASME DEMO		PREPARED BY:	
PLANT LOCATION:		ITEM NO.: RTFS-1	
ITEM NO.: RTFS-1		MODEL / SERIAL: RTFS-1	
SERVICE: OIL STORAGE TANK		UNITS: US CUSTOMARY POSITION: HORIZONTAL	
PROCESS DATA			
PIPE NAME: CB	VAPOR FLOW RATE: -----	H30	MASS DENSITY: -----
STD LIQUID FLOW: -----	LIQUID FLOW RATE: -----	H30	MASS DENSITY: -----
TEMPERATURE: OPERATION: 100 °F	MAXIMUM: 120 °F	DESIGN: 200 °F	
PRESSURE: OPERATION: 90 PSIG	MAXIMUM: 80 PSIG	DESIGN: 150 PSIG	
LEVELS: LEVEL NORMAL = IN	LAW: -----	LAL = -----	LAL = -----
	LAW: -----	LAL = -----	LAL = -----
DESIGN DATA			
CODE: ASME SECTION VIII DIV. 3 2007 EDITION, 2008 ADDENDA		PIPELINEAR DIMENSIONS	
OTHERS: -----		LENGTH FT: 50'	DIA: 20 in OUTSIDE DIAMETER
KIND OF "U": DO NOT CLOSE REPAIR		VOLUME FT: 80000	THICKNESS: SHELL: 3/8 in STD HEADS: 3/8 in NOM. STD.
HEAD TYPE: WELD CAPS 30MPA SCH. STD.		NOT REQUIRED	ESTIMATED WEIGHTS
POSTWELD HEAT TREATMENT: -----		SPOT WT3	EMPTY: 912 TEST: 1112
RADIOPGRAPHY: SPOT WT3		OPERATION:	SURFACE PREPARATION & PAINTING
JOINT EFFICIENCY: SHELL: 0.95 HEAD: 0.95		100% PIG	PLATE: FINISH
HYDROTEST: 100% PIG		PRIME: -----	PIPE: FINISH COLOR: -----
PHYSICAL TEST (IN 100%): NONE		INTERMEDIATE: -----	INTERMEDIATE: -----
CORROSION ALLOWANCE: 0"		PRELIMINARY NOZZLE LOCATION	
WIND LOAD DESIGN: NONE			
NONE			
ELEVATED DESIGN CODE: NONE FACTOR: NONE			
PRODUCTION OUTSIDE NOZZLES: STANDARD TOLERANCE			
STIFFENING RINGS			
MANHOLE COVERS			
INSULATION			
THERMOGRAPHIC			
NOZZLE SCHEDULE			
ITEM SIZE	RATING	PROJECTION OUT	PURPOSE
1" NPS	300X150 NPT	MIN	CUTLET
1/2" NPS	200X150 NPT	MIN	CUTLET
1/2" NPS	200X150 NPT	MIN	MAN
1/2" NPS	300X150 NPT	MIN	SAFETY VALVE
1/2" NPS	300X150 NPT	MIN	DRAIN
MANUFACTURE			
SHELL	SA-106-B	COUPLINGS: NONE	
HEADS	SA-234-WPS	COUPLINGS: SA-106	
WEARING PLATE	NONE	LIFTING LUG: SA-26	
SADDLES	SA-26	BAFFLE: -----	
SHIRT	---	BOLTS: SA-103-B7	
FLANGE	SA-105	NUTS: -----	
NOZZLE	SA-105	NAMEPLATE: SA-240-304	
REVIEWS	1	2	3
INITIAL	AL002309		
REVIEWED	DRS		

Figure 3.1: Vessel Specification Sheet [9]

3.2 The Automatic Identification System (AIS) dataset

AIS (Automatic Identification System) was developed in the 1990s to enhance navigation safety and prevent ship collisions. It allows ships equipped with AIS to communicate with each other and coastal authorities through VHF transmissions. The International Maritime Organization (IMO) mandates that all international voyage ships above 300 gross tonnage and all passenger ships must have an AIS transmitter. Governments and authorities in different nations also enforce AIS applications to improve safety and security.

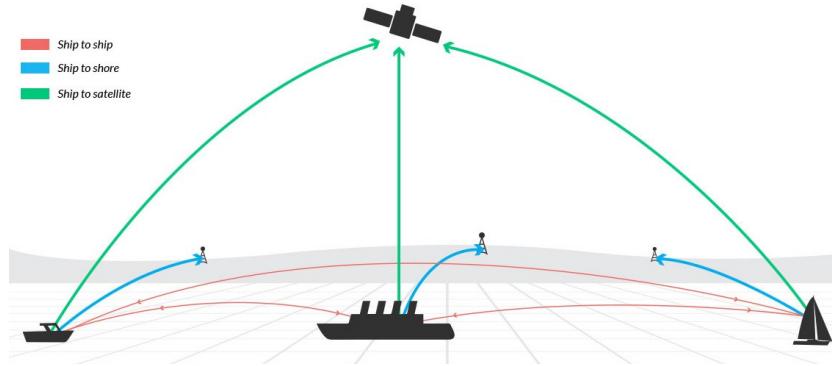


Figure 3.2: Working of AIS

There are two types of AIS transceivers: Class A and Class B. Class A transceivers broadcast more datafields and have higher reporting frequencies. The information broadcasted by a Class A transceiver can be categorized into static information, dynamic information, and voyage-related information. Dynamic information is automatically transmitted every 2-10 seconds when the ship is underway and every 3 minutes when anchored. Static and voyage-related information are broadcasted every 6 minutes regardless of navigational status. Class B transponders transmit a reduced set of data and have sparser reporting intervals compared to Class A transponders.

Data field	Type	Description
AIS identity and location	Static	Maritime Mobile Service Identity (MMSI) and the location of the system's antenna on board
Ship identity	Static	Ship name, IMO number, type, and call sign of the ship
Ship size	Static	Length and width of the ship
Ship position	Dynamic	Latitude and longitude (up to 0.0001 min accuracy)
Speed	Dynamic	Ranging from 0 knot to 102 knots (0.1 knot resolution)
Rate of turn	Dynamic	Right or left (ranging from 0 to 720° per minute)
Timestamp	Dynamic	Timestamp of the message in UTC

Table 3.1: AIS message data fields [16]

Table 3.1 shows the data fields transmitted by AIS messages. Combining AIS data with other databases can provide additional information. For example, port-to-port average speed

can be calculated based on voyage distance and time stamps reported at the two ports. Cargo weight can be estimated using draught and ship sizes. Technical ship specifications, such as DWT (deadweight tonnage), capacity, design speed, and design draught, can be obtained from fleet databases using the IMO number. Port-to-port bunker consumption can be estimated based on speed, distance, and technical ship specifications like DWT and capacity [32].

3.3 AIS Trade Flow System

AIS Trade Flow systems utilize data transmitted by vessels worldwide to furnish both real-time and historical insights into maritime trade operations [7]. The primary aim is to establish a framework that delineates trade activities between ports, leveraging AIS signals. Designating port areas within AIS Trade Flow systems is a multifaceted undertaking encompassing geographical and operational factors. From a geographical perspective, a port area is typically demarcated by a set of coordinates outlining the physical confines of the port and its adjacent waters. This encompassing region might encompass berths, anchorages, and at times, even approach channels. Identifying whether a ship resides within a port area often necessitates comparing its current AIS-reported coordinates with the established port boundaries. Should the ship's location fall within these boundaries, it's deemed situated within the port area. Nonetheless, the analysis transcends mere geographical coordinates. For instance, a ship might exist within a port's geographical bounds; however, if it's merely passing through without halting or participating in port operations, it might not qualify as being 'in port' from an operational standpoint.

Addressing these intricacies, AIS Trade Flow systems frequently employ advanced algorithms to precisely ascertain port boundaries and classify vessel conduct. This assessment takes into account parameters like the vessel's velocity, heading, and historical behavioral trends. Through the integration of these diverse data facets, these systems can render an exceptionally precise portrayal of port undertakings and vessel trajectories. AF Code has constructed such a system by adhering to overarching principles, resulting in the creation of a dataset incorporating AIS signals alongside information about port visits, cargo loading, and unloading.

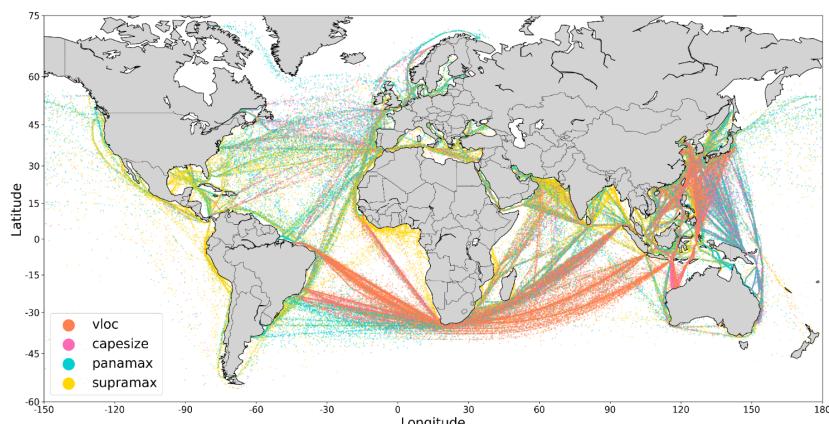


Figure 3.3: Trade flow from 150,000 AIS sampled signals.

To achieve this, the AIS data is processed to identify port-to-port voyages and then stored them in a database based on segments as shown in the Figure 3.4

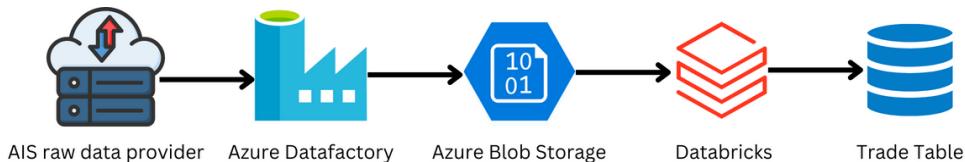


Figure 3.4: AIS Raw Data Processing

3.3.1 Vessel Voyages

Trade flow systems are designed to identify port-to-port voyages. Each of these voyages is either Laden voyage or Ballast voyage from one port to another. At port vessel can load or unload cargo or just refuel and continue to another port. It is important to determine port stops and port area to identify the type of voyage. This section will describe how these voyages are identified and classified.

Laden Voyages

The laden voyage concept centers on cargo movement, distinct from the vessel's own traversal. This characterization holds particular significance in evaluating logistical chains and cargo dynamics within maritime transportation. A laden voyage pertains to the route spanning from the cargo's loading point at the departure port to its unloading point at the arrival port.

Ballast Voyages

The ballast voyage concept centers on the vessel's movement without cargo, contrasting with its movement when laden. A ballast voyage is characterized as the route from the port of cargo unloading, serving as the departure port, to the port of cargo loading, functioning as the arrival port.

3.3.2 Port Stop and Port Area

The initial key phase involves obtaining the coordinates of the targeted port and subsequently establishing a polygon encompassing its vicinity. The overarching approach revolves around quantifying unique visits within this polygon, using a geo-fence strategy. Given the utilization of S-AIS data, there exist temporal intervals during which satellites may not receive ship messages due to blind zones. This introduces the potential for ships to enter and exit the designated port area without satellite detection. To mitigate this, a larger polygon around the port can decrease the likelihood of overlooked visits. However, such an expansive polygon could

result in the inclusion of vessels merely passing through, particularly if the port is positioned near a major shipping route. Hence, the ideal polygon size is contingent on the port's location, it should be sizable enough to capture all visits yet sufficiently compact to exclude vessels in transit. This is illustrated by Figure 3.5, showcasing LNG terminals in Trinidad and Algeria enclosed by identical polygons. Visual inspection reveals message flow towards the port in Trinidad (5.4a) and a substantial portion of vessel flow bypassing the port in Algeria (5.4b). Additionally, supplementary criteria like capping maximum speed and requiring navigational statuses of *anchored* or *moored* can be incorporated to ensure vessels are genuinely engaged in loading activities [7].

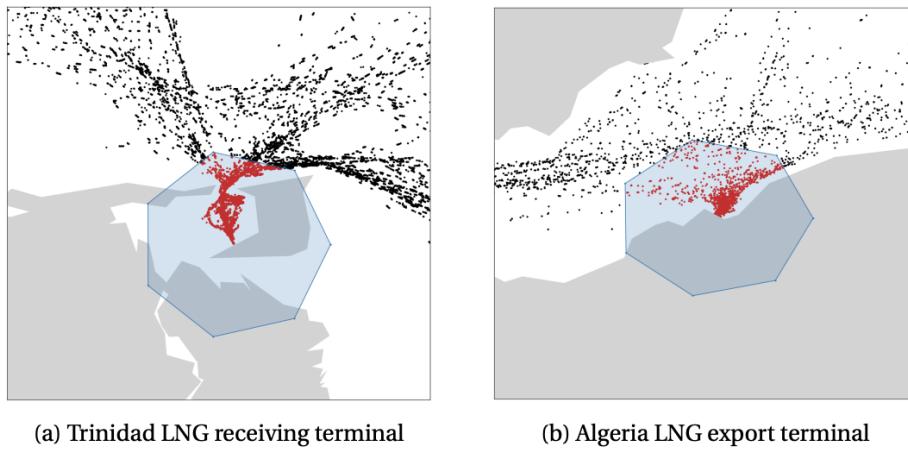


Figure 3.5: Port Area

By using the trade flow system described in this section, we can figure out the trades made by ships in 2022. If we assume that these ships will continue with the same trading patterns in 2023, we can use the trade data to make an estimate of emissions for 2023.

Chapter 4

Tools and Data Analysis Techniques

In this section, we will use tradeflow data for the year 2022, along with specifications from the IHS dataset, to perform data analysis.

4.1 Tools

To perform data analysis, we will utilize the following tools:

- Python
- Databricks
- Apache Spark
- MSSQL
- Pandas
- Astrup Fearnleys Code Emission Calculator API

4.1.1 Databricks

Databricks is a leading platform for big data analysis, specifically designed to handle large-scale datasets efficiently and effectively. Leveraging Apache Spark as its core engine, Databricks provides a distributed computing framework that enables processing massive volumes of data in parallel across multiple nodes. This distributed architecture allows for significant performance improvements, reducing the time it takes to process and analyze big data compared to traditional approaches. With its ability to handle both batch and real-time data processing, Databricks empowers data engineers and analysts to extract valuable insights from vast datasets, enabling them to make data-driven decisions at scale.

One of the key advantages of using Databricks for big data analysis is its ease of use and collaborative features. The platform offers interactive notebooks (Figure 4.1) that allow data professionals to write, execute, and share code seamlessly. This enables collaborative data exploration and simplifies the iterative process of data analysis and model development. Moreover, Databricks provides a rich set of built-in libraries and integrations with popular big data tools and machine learning frameworks, streamlining the development and deployment of complex data pipelines and advanced analytical models. By abstracting the complexities of distributed data processing, Databricks empowers data teams to focus on the analysis and interpretation of results, accelerating the time-to-insight for big data projects and ultimately driving business growth and innovation. One of the key benefits of using Databricks for big data analysis is its ability to handle large volumes of data with ease. Whether you're working with terabytes, petabytes, or even exabytes of data, Databricks can scale to meet your needs without sacrificing performance or reliability. This makes it an ideal solution for organizations looking to perform complex big data analysis tasks, such as machine learning, data warehousing, and real-time streaming analytics [5].

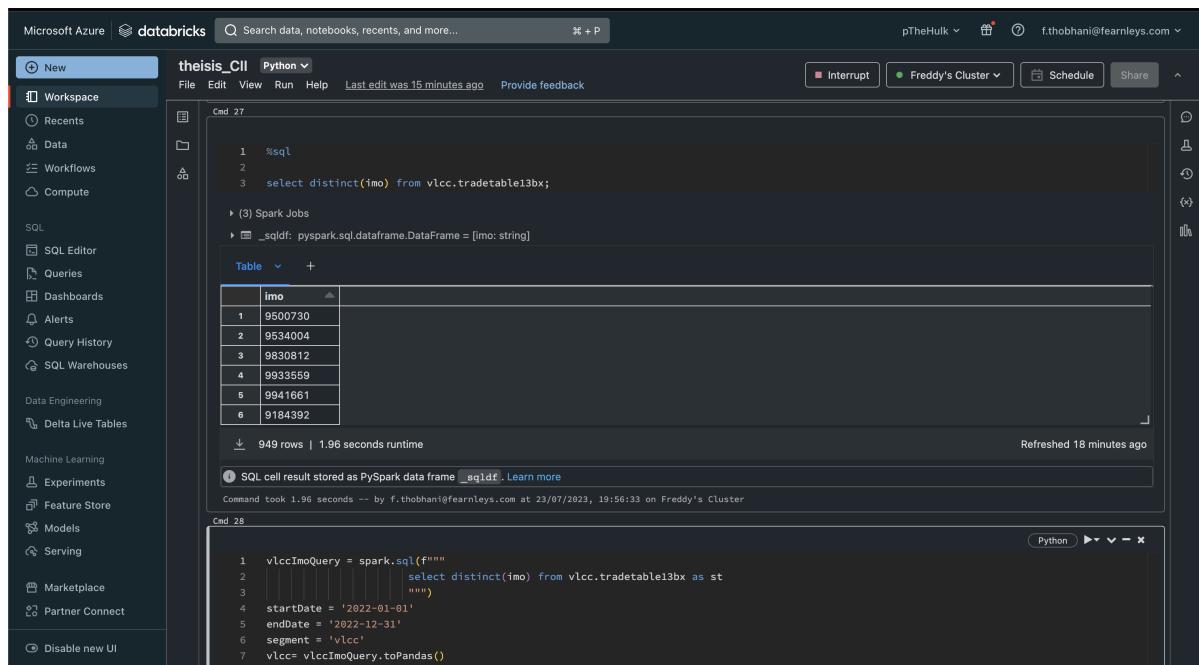


Figure 4.1: Databricks Notebook

Overall, Databricks is a highly capable platform for big data analysis that is well-suited to a wide range of use cases, from simple data processing tasks to complex machine learning and statistical analysis.

The entire analysis will be conducted on Databricks with the following configuration:

- Databricks Runtime Version: 11.3 LTS (includes Apache Spark 3.3.0, Scala 2.12)
- 8GB Memory, 4 Cores with 1 driver and 1 worker node.
- Python 3
- Elastic Disk

The screenshot shows the Databricks Configuration interface for a cluster named "Freddy's Cluster". The left side of the interface contains several configuration sections:

- Policy:** Set to "Unrestricted". Options include "Multi node" (selected) and "Single node".
- Access mode:** Set to "Single user access". Options include "Single user" (selected) and "Multiple users".
- Performance:** Shows the Databricks Runtime Version as "11.3 LTS (includes Apache Spark 3.3.0, Scala 2.12)". Other options include "Use Photon Acceleration" (unchecked).
- Worker type:** Set to "Standard_F4" (8 GB Memory, 4 Cores). It shows 1 worker and 1 current instance.
- Driver type:** Set to "Standard_F4" (8 GB Memory, 4 Cores).
- Autoscaling:** Options include "Enable autoscaling" (unchecked) and "Terminate after 60 minutes of inactivity" (checked).

The right side of the interface features a summary panel with the following details:

Summary	
1 Worker	8 GB Memory 4 Cores
1 Driver	8 GB Memory, 4 Cores
Runtime	11.3.x-scala2.12
Standard_F4	
1 DBU/h	

Figure 4.2: Databricks Configuration

4.1.2 Apache Spark

Apache Spark is an open-source distributed computing framework designed for processing and analyzing large-scale datasets in a highly efficient and parallel manner. It provides a unified platform that supports various data processing tasks, including batch processing, real-time streaming, machine learning, and graph processing. Spark's core abstraction is a resilient distributed dataset (RDD), which allows data to be distributed across multiple nodes in a cluster, enabling data processing operations to be executed in parallel [24].

Databricks leverages Apache Spark as its underlying engine to offer a powerful and scalable data analytics platform. By integrating Spark into its infrastructure, Databricks provides users with a seamless and interactive environment for collaborative data engineering and data science tasks. Databricks' interactive notebooks allow data professionals to write and execute Spark-based code, making it easier to perform data manipulations, transformations, and analysis in real-time. The platform also offers support for various programming languages such as Python, Scala, R, and SQL, providing users with flexibility and familiarity in their preferred language.

Furthermore, Databricks enhances Apache Spark by providing additional features and optimizations to improve performance and ease of use. The platform offers auto-scaling capabilities, enabling resources to be automatically allocated and released based on the workload demand, ensuring optimal performance and cost-efficiency. Databricks also provides built-in libraries and tools that simplify complex tasks such as machine learning and data visualization, allowing data scientists to focus on building and deploying advanced analytical models with ease.

In summary, Apache Spark forms the backbone of Databricks, enabling the platform to handle large-scale datasets efficiently and deliver a collaborative and user-friendly environment for data analysis, making it a popular choice for organizations seeking to harness the

potential of big data analytics.

The screenshot shows the Apache Spark Master UI at `spark://10.139.64.4:7077`. It displays the following information:

- Spark Version:** 3.3.0
- URL:** `spark://10.139.64.4:7077`
- Alive Workers:** 1
- Cores in use:** 4 Total, 4 Used
- Memory in use:** 3.9 GiB Total, 3.1 GiB Used
- Resources in use:**
- Applications:** 1 Running, 0 Completed
- Drivers:** 0 Running, 0 Completed
- Status:** ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory	Resources
worker-20230723171149-10.139.64.5-36533	10.139.64.5:36533	ALIVE	4 (4 Used)	3.9 GiB (3.1 GiB Used)	

Running Applications (1)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20230723171146-0000	(kill) Databricks Shell	4	3157		2023/07/23 17:11:46	root	RUNNING	1.4 h

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration

Figure 4.3: Apache Spark Configuration

4.1.3 Astrup Fearnleys Code Emission Calculator API

Astrup Fearnleys Code Emission Calculator API is a service that provides a simple interface for calculating the emissions of a given vessel. The API is built on top of the Astrup Fearnleys Code Emission Calculator. It is an internal tool developed to provide a simple interface for calculating the emissions of a given vessel for researchers.

It requires the following parameters to be passed as input:

Parameter	Type	Description
imo	integer	IMO of a ship that will require emission calculations.
avg_speed_kn	number	The vessel's intended average sailing speed in knots on the route. Optional if duration_h is given.
distance_nm	number	Expected length of the voyage in Nautical Miles.
cargo_unit	string	Cargo unit used to submit cargo_amt. Currently, the only possible value is tons. Used to calculate EEOI and transport_work.
cargo_amt	number	Amount of cargo the ship is carrying (in the given cargo_unit). Used to calculate EEOI and transport_work.
me_fuel_type	string	Fuel type to be considered for the main engine consumption for the emissions estimate. Overwriting the model assumption.
ae_fuel_type	string	Fuel type to be considered for the auxiliary engine consumption for the emissions estimate. Overwriting the model assumption.
boiler_fuel_type	string	Fuel type to be considered for the boiler consumption for the emissions estimate. Overwriting the model assumption.
duration_h	number	Expected duration of the voyage in hours. Optional if avg_speed_kn is given.
load_cond	string	Loading condition of the vessel to be considered in the power estimation.
me_co2_factor	number	CO2 factor for the main engine fuel. Main engine fuel type must be specified.
ae_co2_factor	number	CO2 factor for the auxiliary engine fuel. Auxiliary engine fuel type must be specified.
boiler_co2_factor	number	CO2 factor for the boiler engine fuel. Boiler engine fuel type must be specified.

Table 4.1: Request Parameters for Emission Calculations

In return, it provides the JSON output in following format:

```

1  {
2      "imo": 1234567,
3      "avg_speed_kn": 10.045,
4      "load_cond": "ballast",
5      "cargo_unit": "tons",
6      "cargo_amt": 544385.543,
7      "me_fuel_type": "MGO",
8      "ae_fuel_type": "MGO",
9      "boiler_fuel_type": "MGO",
10     "duration_h": 354.5,
11     "me_fuel_cons_metric_metric_tons": 0.1,
```

```

12     "ae_fuel_cons_metric_tons": 0.1,
13     "boiler_fuel_cons_metric_tons": 0.1,
14     "total_co2_emission_metric_tons": 0.1,
15     "total_so2_emission_metric_tons": 0.1,
16     "total_particulate_matter_emission_metric_tons": 0.1,
17     "total_nox_emission_metric_tons": 0.1,
18     "total_nmvoe_emission_metric_tons": 0.1,
19     "total_ch4_emission_metric_tons": 0.1,
20     "total_n2o_emission_metric_tons": 0.1,
21     "total_co_emission_metric_tons": 0.1,
22     "total_black_carbon_emission_metric_tons": 0.1,
23     "total_organic_carbon_emission_metric_tons": 0.1,
24     "transport_work": 0.1,
25     "eeoi": 0.1,
26     "me_co2_factor": 2.1,
27     "ParamValidation": {
28         "status": "OK"
29     },
30     "LimitMessage": {
31         "message_description": "You are within your call limit"
32     }
33 }
```

Listing 4.1: JSON output format from emission api

4.2 Methodology

In this section, using tradeflow data for the year 2022, along with specifications from the IHS dataset, we will perform data analysis.

Tradetable data is available in the trade flow dataset. For each segments there is trade flow and therefore we will have to perform the analysis for each segment separately.

The steps involved in the analysis are as follows:

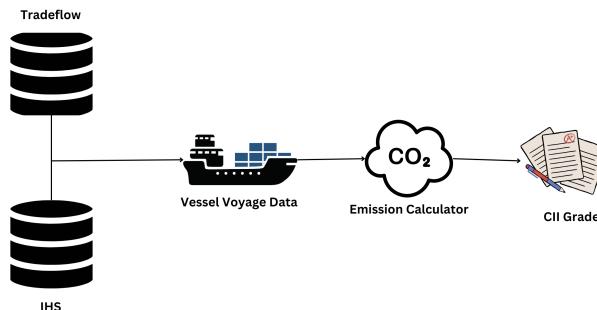


Figure 4.4: Methodology Overview

1. Getting distinct IMO numbers from trade flow

Distinct IMO (International Maritime Organization) numbers were obtained from the trade flow dataset by querying the textttimo column. This process helped in identifying all the unique ships present in the dataset, forming the foundation for further analysis.

2. Getting necessary information for each IMO number from IHS and trade flow

For each distinct IMO number identified in the previous step, the IHS dataset was queried or joined with trade flow to gather relevant information about each ship. This included details such as specifications, fuel type, engine types, and more. The process consisted of the following steps:

- **Estimating Cargo Based on Draught:** Cargo was estimated using the draught noted on ports and the maximum draught from the IHS dataset. This allowed for an understanding of the cargo load for different vessels.
- **Determining Laden or Ballast Trade Flows:** The trade flows were categorized into laden or ballast based on the ballast threshold percentage from segment parameters. This categorization assisted in differentiating the vessel's operational conditions.
- **Converting to Pandas DataFrame:** The obtained shipTrade data was converted into a Pandas DataFrame, facilitating further processing and analysis.
- **Dividing into DataFrames for Laden and Ballast Trade Flows:** The trade data was divided into separate DataFrames for laden and ballast trade flows. This segmentation enabled focused analysis on different operational conditions.
- **Extracting Information on Distance, Time, and Average Speed:** Information such as distance, time, and average speed was extracted for both ballast and laden trade flows. These parameters provided insights into the efficiency and operation of the vessels.
- **Calculating Cargo:** The total cargo was calculated based on laden trade data, contributing to the overall understanding of the vessels' capacity and utilization.

3. Cleaning data

Data cleaning is a critical step in the methodology, ensuring that the dataset is ready for analysis. This involves handling missing values, removing duplicates, standardizing units, and correcting any inconsistencies. Specific aspects of the data cleaning process included:

- **Handling Missing Speed Values:** In some cases, the average speed determined from AIS (Automatic Identification System) data might be missing or unusually high for specific trade flows. To address this, the distance and time values were used to calculate the average speed, ensuring that the analysis was based on consistent and realistic speed data.

- **Excluding Trades with Missing Distance:** Any trades with missing average distance values were excluded from the analysis. This helped maintain the integrity and accuracy of the dataset.

Sometimes for certain trades average distance was missing, this trades were excluded.

4. Calculating emissions for laden and ballast tradeflows.

The calculation of emissions for laden and ballast trade flows was a central component of the methodology. Utilizing the information extracted in the previous steps, the emissions were calculated using the AFC (Automated Flight Control) emission calculator API. Here's how the process was executed:

- **Emission Calculation Function:** A specific function, `getEmission`, was employed to calculate emissions. This function took essential parameters like IMO number, cargo amount, distance, load condition, and estimated speed as input.
- **API Request:** The function made a request to the AFC emission calculator API, passing the required parameters. The response from the API included detailed emission data, providing insights into various emission aspects such as CO₂, SO₂, NO₂, and others.
-
- **Error Handling:** The process included robust error handling to manage any issues with the API request, ensuring that the analysis was not disrupted by unexpected errors or inconsistencies.

5. Calculating CII, CII required, and CII reference.

Carbon Intensity Indicator (CII) is calculated as the ratio of emissions (CO₂) to transport work (distance traveled × cargo amount). CII required is based on regulatory standards, while CII reference can be derived from historical data or benchmarks. Calculate these indicators for each tradeflow using the emission data and transport work values.

CII reference is calculated using the following formula:

$$CII_{ref} = a \cdot \text{Capacity}^{-C} \quad (4.1)$$

where a and C are constants according to Figure 4.5, and Capacity is the deadweight of the ship.

CII required can be calculated using the following formula:

Ship type		capacity	a	c
Bulk Carrier	DWT \geq 279,000	279,000	4745	0.622
	DWT < 279,000	DWT	4745	0.622
Gas Carrier	DWT \geq 65,000	DWT	14405E+7	2.071
	DWT < 65,000	DWT	8104	0.639
Tanker		DWT	5247	0.610
Container ship		DWT	1984	0.489
General cargo ship	DWT \geq 20,000	DWT	31948	0.792
	DWT < 20,000	DWT	588	0.3885
Refrigerated cargo carrier		DWT	4600	0.557
Combination carrier		DWT	5119	0.622
LNG Carrier	DWT \geq 100,000	DWT	9.827	0
	100,000 > DWT \geq 65,000	DWT	14479E+10	2.673
	DWT < 65,000	65,000	14479E+10	2.673
Ro-ro cargo ship (VC)	GT \geq 57,700	57,700	3627	0.590
	57,700 > GT \geq 30,000	GT	3627	0.590
	GT < 30,000	GT	330	0.329
Ro-ro cargo ship		GT	1967	0.485
Ro-ro passenger ship	Ro-ro passenger ship	GT	2023	0.460
	High-speed craft	GT	4196	0.460
Cruise passenger ship		GT	930	0.383

Figure 4.5: a and C constants

$$\text{Requited CII} = \left(\frac{100 - Z}{100} \right) \times \text{CIIRef} \quad (4.2)$$

In Equation 5.2, the reduction factor Z represents the initial value of 5% in the year 2023, with an annual increment of 2%. Additionally, for the years 2027 to 2030, the Z factors are subject to enhancement and refinement, guided by the evaluation of the short-term measure's effectiveness.

Year	Reduction Factor (Z)
2023	5%
2024	7%
2025	9%
2026	11%
2027	**
2028	**
2029	**
2030	**

Table 4.2: Reduction Factors Over the Years

6. Calculating CII grade.

Based on the calculated CII and CII required, determine the CII grade for each tradeflow, indicating its compliance with emissions regulations. This can be done using predefined thresholds or standards to categorize tradeflows into different CII grades.

Taking into account that vessels will experience similar voyages in 2023 as they did in 2022, the CII grade for the end of 2023 can be estimated with a reduction factor Z set at 5.

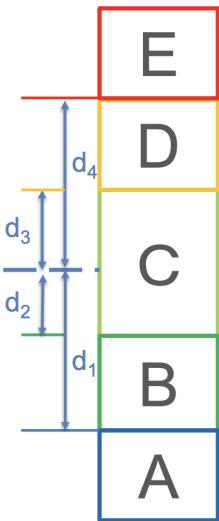
The "dd" vector is established by calculating the ratio between Attained CII and Required CII. This ratio is subsequently compared against the thresholds for CII grades to ascertain the appropriate grade. The larger the deviation of the ratio, the lower the grade assigned, with A representing the best and E signifying the poorest performance.

Ship type	d1	d2	d3	d4
Bulk Carrier	0.86	0.94	1.06	1.18
Gas Carrier	0.81	0.91	1.12	1.44
	<65,000DWT	0.85	0.95	1.06
Tanker	0.82	0.93	1.08	1.28
Container ship	0.83	0.94	1.07	1.19
General cargo ship	0.83	0.94	1.06	1.19
Refrigerated cargo carrier	0.78	0.91	1.07	1.20
Combination carrier	0.87	0.96	1.06	1.14
LNG Carrier	0.89	0.98	1.06	1.13
	<100000DWT	0.78	0.92	1.10
Ro-ro cargo ship (VC)	0.86	0.94	1.06	1.16
Ro-ro cargo ship	0.76	0.89	1.08	1.27
Ro-ro passenger ship	0.76	0.92	1.14	1.30
Cruise passenger ship	0.87	0.95	1.06	1.16

Figure 4.6: dd vectors for determining the rating boundaries of ship types

Based on figure 4.6, the CII grade can be determined as represented in figure 4.7.

Following above 6 steps, we can calculate emissions, CII, CII required, CII reference and CII grade for each vessel. Results can be stored in a SQL table for further analysis. For each segment, we can perform above steps as showing in Listing 4.2

Figure 4.7: CII grading based on dd vector

```

1 capsizesImoQuery = spark.sql("""
2     select distinct(imo) from capesize.tradeFlow as st
3     """
4 )
5 startDate = '2022-01-01'
6 endDate = '2022-12-31'
7 segment = 'capesize'
8 capeSize = capsizeImoQuery.toPandas()
9 for imo in capeSize['imo']:
10     imoCIIData = getCIIGradeByImo(imo, startDate, endDate, segment, 5)
11     df = pd.DataFrame([imoCIIData])
12     spark.createDataFrame(df).write.mode("append").saveAsTable("emissions
    ↛ .capesize_cii_2022_v3")
    time.sleep(5)

```

Listing 4.2: Analysis for capsize segment

In listing 4.2, we are getting distinct IMO numbers for capsize segment. Then for each IMO number, we are calling `getCIIGradeByImo` function to get emissions, CII, CII required, CII reference and CII grade. Using `spark.createDataFrame(df).write()` we are storing results in SQL table.

Based on the data obtained so far, we will analysis the results in the next chapter.

Chapter 5

Results and Discussion

Based on the preceding analysis, following parameters were derived for each vessel, assuming they will engage in similar trade flows as observed in 2022, but in the year 2023. The vessels have been categorized into segments, facilitating a more intricate examination of the outcomes. The segments encompassed within this report include Capesize, Panamax, Supramax, Aframax, Suezmax, VLGC, and VLCC.

The analysis encompassed a total of 9,406 vessels. The parameters acquired for each vessel comprise an array of emissions and indices, contributing to a comprehensive assessment. These parameters encompass CO₂ emissions, SO₂ emissions, NOx emissions, PM emissions, NMVOC, CH₄, N₂O, CO, Black Carbon, Organic Carbon, EEOI, CII, and CII grade.

5.1 Grade Distribution

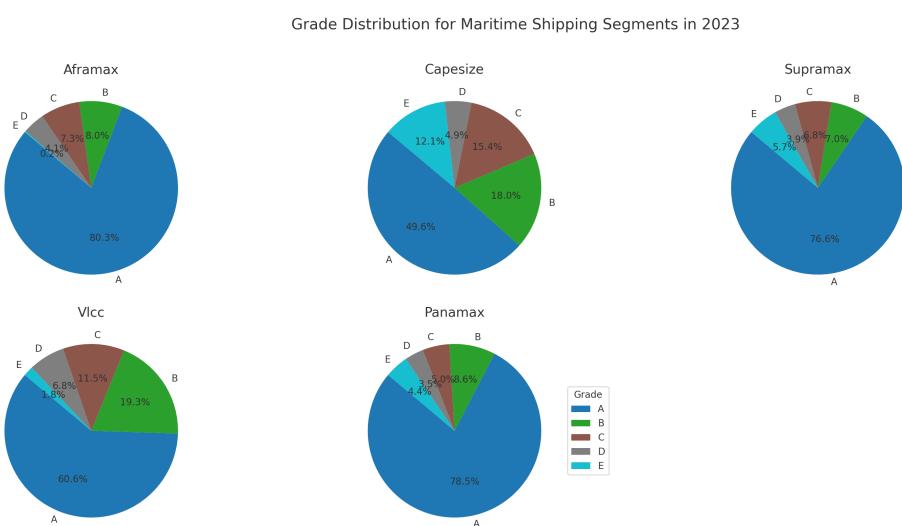


Figure 5.1: Grade Distribution

The pie chart analysis of vessel grades across different maritime shipping segments in

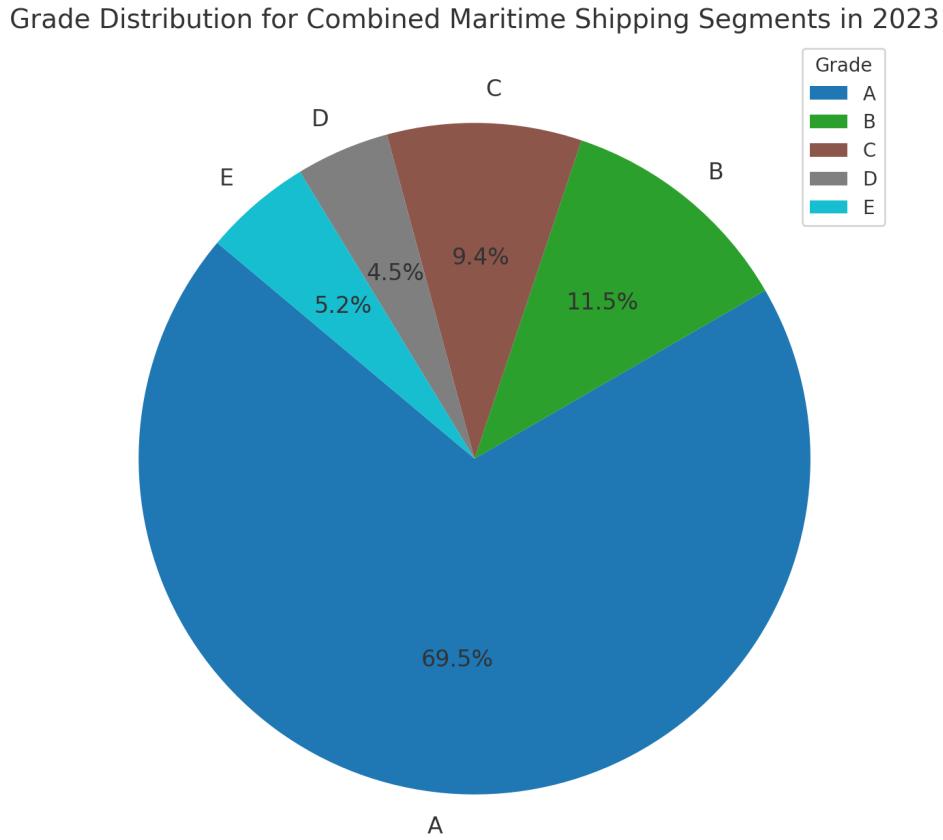


Figure 5.2: Combined Grade

Figure 5.1, along with the combined view in Figure 5.2, offers a nuanced understanding of the quality and compliance distribution within the industry as of 2023.

Most of segments shows similar distribution of grades, with Grade A being in between 60% to 80% while Grade D and E being around 10%. This can be reflected in combined chart too. Though capesize is an exception with only about 50% of vessels in Grade A and almost 16% with grade D and E.

The combined pie chart encapsulates the overall landscape, showing 69.5% of vessels in Grade A, and 10% in Grade D and E.

The segment-specific insights reveal subtle variations that may be influenced by factors such as vessel size, purpose, technological capabilities, and market demands. Understanding these variations is essential for targeted interventions and strategies to further enhance quality and compliance.

In conclusion, the pie chart analysis paints a vivid picture of the maritime shipping industry's commitment to high standards as reflected in the grades. It also emphasizes the need for continuous innovation, particularly in the face of stricter grading systems in the future. The segment-wise insights provide valuable guidance for industry stakeholders, regulators, and policymakers, facilitating data-driven decision-making and fostering a culture of continuous improvement and adaptation.

5.2 Projected Grade Trend

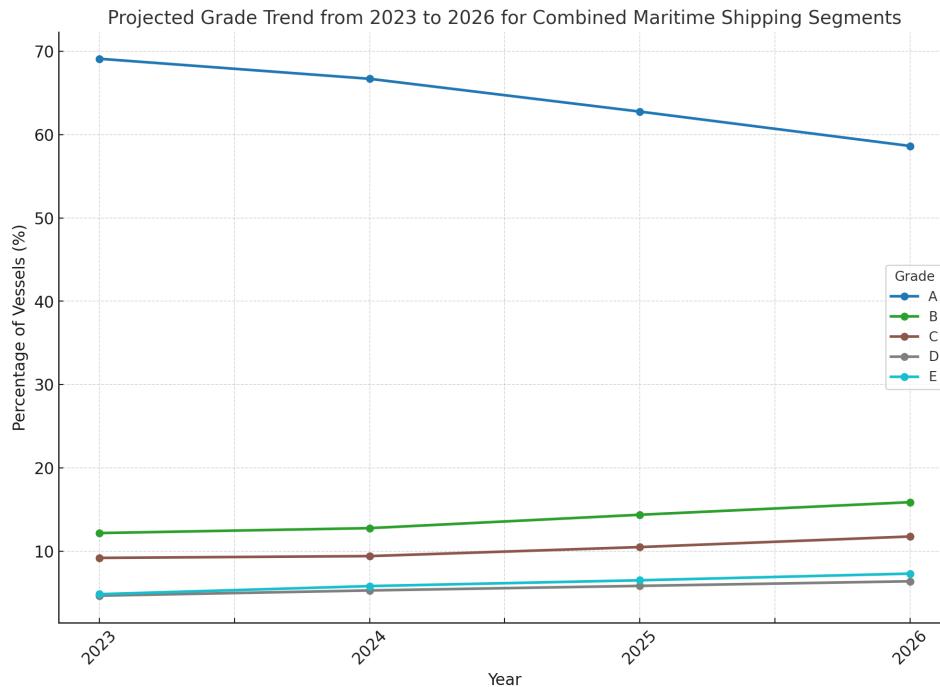


Figure 5.3: Projected Grade Trend

The analysis of the projected grade trend from 2023 to 2026 in plot 5.3 unveils a compelling narrative of the industry's response to an increasingly stringent grading system. The most pronounced observation is the decline in the percentage of vessels achieving Grade A, falling from 69% in 2023 to less than 60% in 2026. This reduction is not indicative of a decline in quality or compliance but rather a reflection of the escalating standards and criteria set by the grading authorities.

This trend underscores the industry's challenge to keep pace with evolving regulations and the imperative for continuous innovation and improvement. As the grading system becomes more demanding, the criteria for achieving Grade A are elevated, pushing some vessels into the B and C categories. The growth in Grades B and C, from about 12.18% to 16% and 9% to 12% respectively, illustrates this shift.

Even the slight increases in Grades D and E, though representing a smaller fraction of the fleet, emphasize the pressing need for maritime shipping to adapt, innovate, and invest in new technologies and practices. The trend towards more rigorous grading is likely driven by global efforts to enhance environmental sustainability, safety standards, and overall efficiency within the maritime sector.

The observed grade trend serves as both a challenge and an opportunity for maritime shipping. It calls for a proactive and strategic approach, where shipping companies must embrace research and development, adopt advanced technologies, and foster a culture of excellence. The industry must recognize that the path to high grades is dynamic and requires ongoing commitment to align with the progressive standards.

In conclusion, the projected grade trend from 2023 to 2026 is a testament to the maritime shipping industry's evolving landscape. It emphasizes the critical role of innovation in meeting the future demands of a stricter grading system. This trend offers valuable insights for policymakers, regulators, and industry leaders, guiding them towards collaborative efforts that ensure the maritime shipping industry not only maintains its current standards but continually strives to reach new heights of quality and responsibility.

5.3 EEOI vs Grade

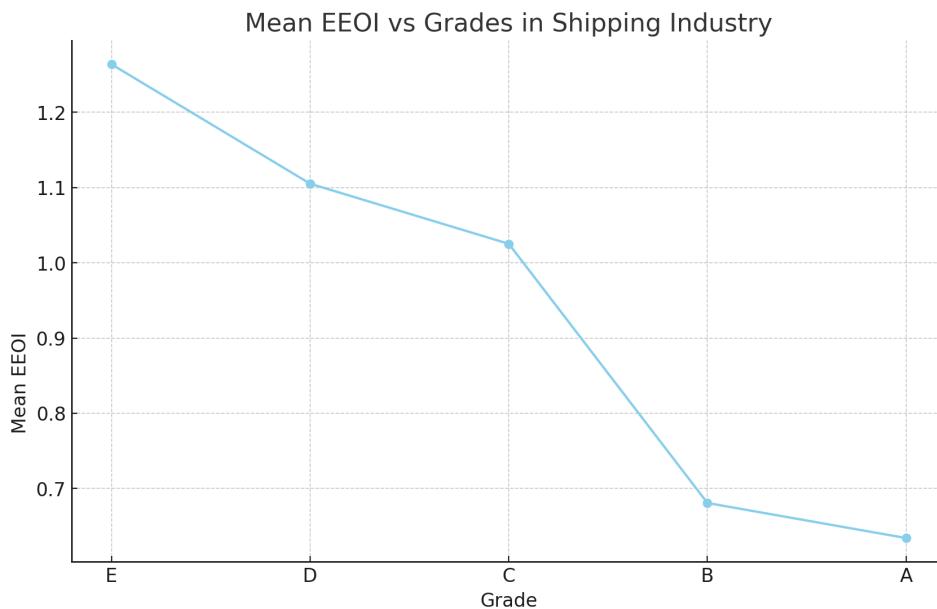


Figure 5.4: Mean EEOI vs Grade

The above Figure 5.4 represents relationship between EEOI and grade.

Lower Efficiency Operational Indicator (EEOI) means vessels uses less fuel to achieve same amount of work. The relationship between the Energy EEOI and vessel grades in the maritime shipping industry offers a logical and meaningful correlation. Lower EEOI values, signifying better energy efficiency, correspond to higher vessel grades (A being the highest). The grading system incentivizes energy-efficient practices and provides a transparent metric for evaluating vessel quality and environmental responsibility.

5.4 Average speed vs Grade

To understand if speed has any impact on grade, scatter plot of average speed vs grade is plotted. From the plot 5.5 it is noticed that vessels with lower speed have better grade. It is very clear that almost all the vessels with speed 12.5 knots or more have grade C or below.

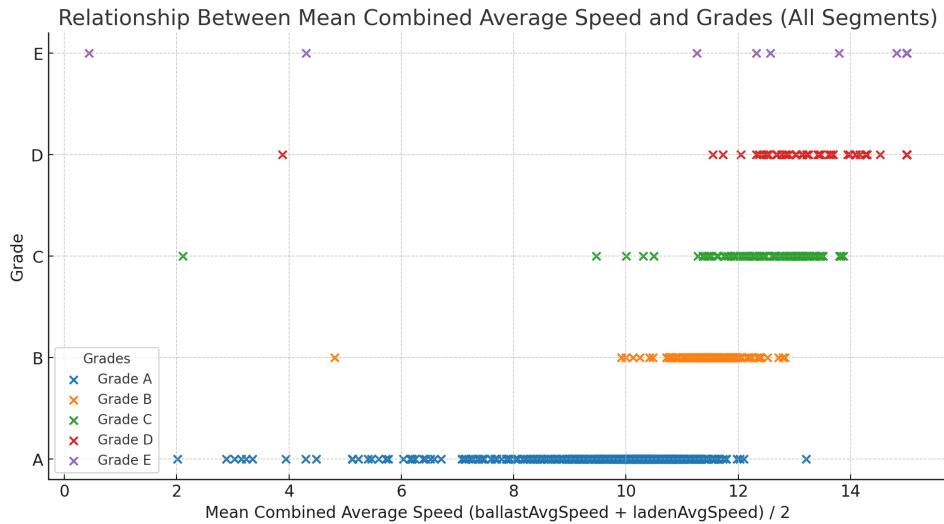


Figure 5.5: Average Speed vs Grade

5.5 Grade by Build Country

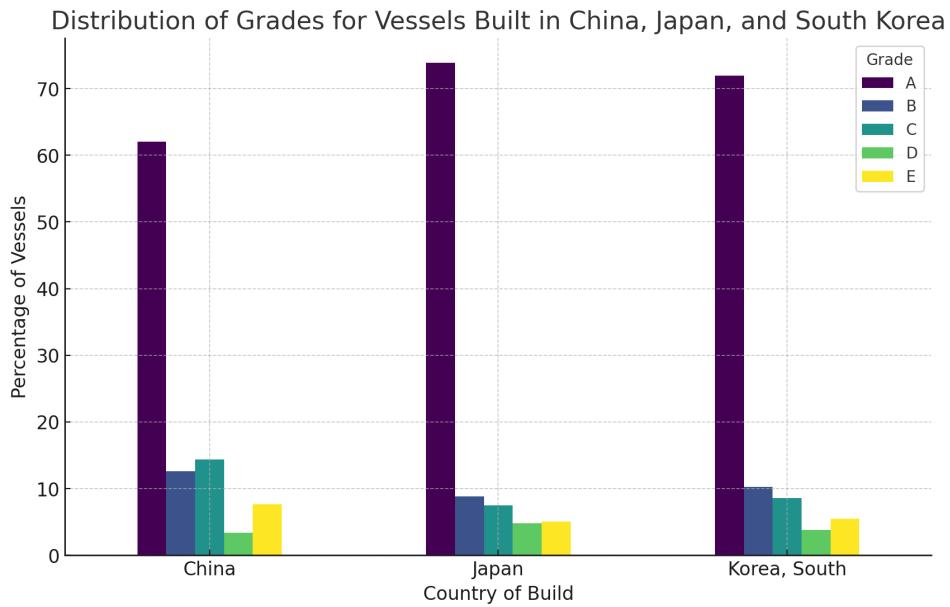


Figure 5.6: Grade By Build Country

Country where vessels are built can effect grade significantly as many parameters like materials, technology, regulations etc. can vary from country to country. To verify this hypothesis, grade distribution by build country is plotted in Figure 5.6. To make the plot more readable, only top 3 countries China, Japan and South Korea were considered.

From plot 5.6 it is clear that vessels built in Japan have better grade than vessels built in China or South Korea.

5.6 Addressing Low Rating

When a ship's Carbon Intensity Indicator (CII) receives a low grade of D for three consecutive years or E for one year, specific measures must be undertaken to address and improve the situation. As per the regulations, a corrective action plan must be submitted and approved.

Some of the strategies include the use of alternative fuels, which may require significant investment but have high effectiveness; the application of low friction paint and Energy Saving Devices (ESD) to improve hardware; voyage optimization and fleet optimization, which may require preliminary examination but have a low to middle cost and depend on the ship's specific situation; and slow steaming, which requires an analysis of monitoring results and estimation of cost-effectiveness. These methods indicate a multifaceted approach that considers cost, effectiveness, and specific ship characteristics to overcome the challenges associated with a low CII rating. The strategies emphasize both technological improvements and operational adjustments, reflecting an integrated approach to enhance environmental performance.

5.7 Emissions

Based on AF code emission API and vessel parameters, various emissions are calculated for each vessel grouped by segment.

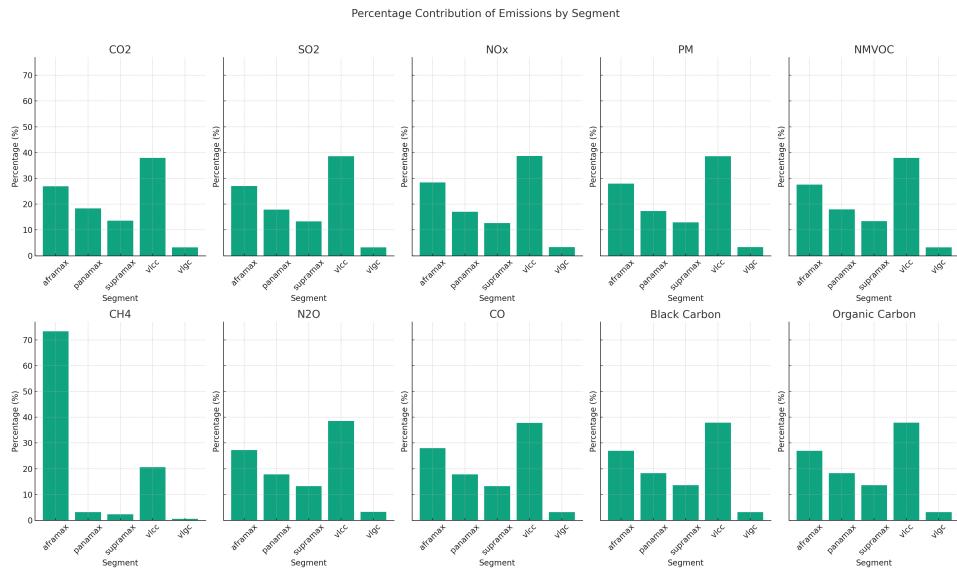


Figure 5.7: Percentage of Emissions by Segment

Figure 5.7 shows percentage of emissions by segment. The analysis reveals a diverse emission profile across segments such as VLCC, VLG, Aframax, Panamax, Capesize, and Supramax. Notably, the VLCC segment emerges as a significant contributor to emissions such as CO₂, SO₂, and NO_x, highlighting potential areas for emission reduction strategies.

Chapter 6

Conclusion

This study delved into a comprehensive investigation of carbon emissions in maritime shipping, a crucial yet intricate facet of the industry. Through the course of this research, several key questions were addressed:

1. **Understanding Complexity with Simplicity:** The analysis demonstrated that even the multifaceted nature of emissions could be understood through a well-structured and comprehensible system. By breaking down complex elements, the thesis facilitated a clearer grasp of emissions dynamics.
2. **Categorization of Vessels:** The categorization of vessels into distinct segments such as Capesize, Panamax, Supramax, and others enabled a more nuanced examination of carbon emissions. This segmentation provided unique insights into each category's characteristics and emissions behavior.
3. **Factors Affecting Carbon Emissions:** A comprehensive analysis was conducted to identify the multitude of factors affecting a vessel's carbon emissions. This understanding forms a foundation for more targeted and effective emissions control strategies.
4. **Realistic Emissions Reduction:** The thesis explored various cost-effective and practical measures to reduce carbon emissions. These findings not only contribute to environmental sustainability but also offer realistic paths for industry implementation.

6.1 Future Work

Looking ahead, there are several promising avenues for future research. The consideration of port emissions alongside voyage emissions presents an opportunity for a more holistic view of maritime carbon footprints. Additionally, analyzing more vessel segments will further refine our understanding and provide more detailed insights. These future efforts align with the ongoing pursuit of environmental responsibility and efficiency in the maritime shipping industry.

In conclusion, this thesis has made a significant contribution to the field by unraveling the complexities of maritime shipping's carbon emissions. It has provided valuable insights, practical solutions, and set the stage for future research that continues to push the boundaries of our understanding and capabilities.

Bibliography

- [1] Nicoleta Acomi and Ovidiu Cristian Acomi. “Improving the voyage energy efficiency by using EEOI”. In: *Procedia-Social and Behavioral Sciences* 138 (2014), pp. 531–536.
- [2] Lai Fatt Chuah et al. “Implementation of the energy efficiency existing ship index and carbon intensity indicator on domestic ship for marine environmental protection”. In: *Environmental research* 222 (2023), p. 115348.
- [3] Lai Fatt Chuah et al. “Implementation of the energy efficiency existing ship index and carbon intensity indicator on domestic ship for marine environmental protection”. In: *Environmental Research* 222 (2023), p. 115348. ISSN: 0013-9351. DOI: <https://doi.org/10.1016/j.envres.2023.115348>. URL: <https://www.sciencedirect.com/science/article/pii/S0013935123001408>.
- [4] Ernest Czermański et al. “Implementation of the Energy Efficiency Existing Ship Index: An important but costly step towards ocean protection”. In: *Marine Policy* 145 (2022), p. 105259.
- [5] *Data Lakehouse Architecture and AI Company - Databricks*. URL: <https://www.databricks.com/>.
- [6] Andrzej S. Grzelakowski, Jerzy Herdzik, and Sławomir Skiba. “Maritime Shipping Decarbonization: Roadmap to Meet Zero-Emission Target in Shipping as a Link in the Global Supply Chains”. In: *Energies* 15.17 (2022). ISSN: 1996-1073. DOI: [10.3390/en15176150](https://doi.org/10.3390/en15176150). URL: <https://www.mdpi.com/1996-1073/15/17/6150>.
- [7] Torjus Halden. “Estimation of Trade Flows With the Use of AIS Data: The Case of LNG Shipping”. MA thesis. NTNU, 2019.
- [8] Mohamad Issa, Adrian Ilinca, and Fahed Martini. “Ship Energy Efficiency and Maritime Sector Initiatives to Reduce Carbon Emissions”. In: *Energies* 15.21 (2022). ISSN: 1996-1073. DOI: [10.3390/en15217910](https://doi.org/10.3390/en15217910). URL: <https://www.mdpi.com/1996-1073/15/21/7910>.
- [9] Daniellee Gaspard Itur. *F100 vessel specification sheet*. URL: <https://www.scribd.com/document/391050652/f100-Vessel-Specification-Sheet#>.
- [10] Veronica Jaramillo Jimenez, Hyungju Kim, and Ziaul Haque Munim. “A review of ship energy efficiency research and directions towards emission reduction in the maritime industry”. In: *Journal of Cleaner Production* 366 (2022), p. 132888. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2022.132888>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652622024817>.

- [11] Sheng-Long Kao, Wu-Hsun Chung, and Chao-Wei Chen. “AIS-Based Scenario Simulation for the Control and Improvement of Ship Emissions in Ports”. In: *Journal of Marine Science and Engineering* 10.2 (2022). ISSN: 2077-1312. DOI: [10.3390/jmse10020129](https://doi.org/10.3390/jmse10020129). URL: <https://www.mdpi.com/2077-1312/10/2/129>.
- [12] Anthony King. *Emissions-free sailing is full steam ahead for ocean-going shipping*. <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/emissions-free-sailing-full-steam-ahead-ocean-going-shipping>. Sept. 2022.
- [13] Tatiana Lukoianova and Victoria Rubin. “Veracity Roadmap: Is Big Data Objective, Truthful and Credible?” In: *ASIST 2013: 76th Annual Meeting of the Association for Information Science and Technology: Beyond the Cloud: Rethinking Information Boundaries* 24 (Nov. 2013). DOI: [10.7152/acro.v24i1.14671](https://doi.org/10.7152/acro.v24i1.14671).
- [14] Ziaul Haque Munim et al. “Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions”. In: *Maritime Policy & Management* 47.5 (2020), pp. 577–597. DOI: [10.1080/03088839.2020.1788731](https://doi.org/10.1080/03088839.2020.1788731). eprint: <https://doi.org/10.1080/03088839.2020.1788731>. URL: <https://doi.org/10.1080/03088839.2020.1788731>.
- [15] Harry E Pence. “What is big data and why is it important?” In: *Journal of Educational Technology Systems* 43.2 (2014), pp. 159–171.
- [16] Heather M Perez et al. “Automatic identification systems (AIS) data use in marine vessel emission estimation”. In: *18th Annual International Emission Inventory Conference*. Vol. 14. 2009, e17.
- [17] Katarzyna Prill et al. “A new method of determining energy efficiency operational indicator for specialized ships”. In: *Energies* 13.5 (2020), p. 1082.
- [18] V Rajaraman. “Big data analytics”. In: *Resonance* 21 (2016), pp. 695–716.
- [19] Huilin Ren, Yu Ding, and Congbiao Sui. “Influence of EEDI (Energy Efficiency Design Index) on ship–engine–propeller matching”. In: *Journal of Marine Science and Engineering* 7.12 (2019), p. 425.
- [20] Miguel Rodriguez, Mario Pansera, and Pablo Cabanelas Lorenzo. “Do indicators have politics? A review of the use of energy and carbon intensity indicators in public debates”. In: *Journal of Cleaner Production* 243 (2020), p. 118602.
- [21] Dan Rutherford, Xiaoli Mao, and Bryan Comer. “Potential CO₂ reductions under the energy efficiency existing ship index”. In: *International council on clean transportation. Working paper* 27 (2020).
- [22] Marco Siddi et al. “The European Green Deal: Assessing its current state and future implementation”. In: *UPI REPORT* 114 (2020).
- [23] Weng Sut Sou et al. “Reducing the carbon intensity of international shipping – The impact of energy efficiency measures”. In: *Energy Policy* 170 (2022), p. 113239. ISSN: 0301-4215. DOI: <https://doi.org/10.1016/j.enpol.2022.113239>. URL: <https://www.sciencedirect.com/science/article/pii/S030142152200458X>.
- [24] *SparkTM - Unified Engine for large-scale data analytics*. URL: <https://spark.apache.org/>.
- [25] Alastair Stevenson. *Bulker EEXI compliance boosts CII ratings*. SSY online, June 2021. URL: <https://www.ssyonline.com/media/2165/eexi-effects-bulkers-jun22rev-1.pdf>.

- [26] Aydin TOKUŞLU. “Analyzing the Energy Efficiency Design Index (EEDI) performance of a container ship”. In: *International Journal of Environment and Geoinformatics* 7.2 (2020), pp. 114–119.
- [27] Tien Anh Tran. “A research on the energy efficiency operational indicator EEOI calculation tool on M/V NSU JUSTICE of VINIC transportation company, Vietnam”. In: *Journal of Ocean Engineering and Science* 2.1 (2017), pp. 55–60.
- [28] Yuh-Ming Tsai and Cherng-Yuan Lin. “Effects of the Carbon Intensity Index Rating System on the Development of the Northeast Passage”. In: *Journal of Marine Science and Engineering* 11.7 (2023), p. 1341.
- [29] *UN body adopts climate change strategy for shipping*. imo.org, 2018. URL: <https://www.imo.org/en/MediaCentre/PressBriefings/Pages/06GHGinitialstrategy.aspx>.
- [30] *UN body adopts climate change strategy for shipping*. imo.org. URL: <https://www.imo.org/en/MediaCentre/PressBriefings/Pages/06GHGinitialstrategy.aspx>.
- [31] Shuaian Wang, Harilaos N. Psaraftis, and Jingwen Qi. “Paradox of international maritime organization’s carbon intensity indicator”. In: *Communications in Transportation Research* 1 (2021), p. 100005. ISSN: 2772-4247. DOI: <https://doi.org/10.1016/j.commtr.2021.100005>. URL: <https://www.sciencedirect.com/science/article/pii/S2772424721000056>.
- [32] Dong Yang et al. “How big data enriches maritime research—a critical review of Automatic Identification System (AIS) data applications”. In: *Transport Reviews* 39.6 (2019), pp. 755–773.
- [33] Ibna Zaman et al. “Challenges and Opportunities of Big Data Analytics for Upcoming Regulations and Future Transformation of the Shipping Industry”. In: *Procedia Engineering* 194 (2017). 10th International Conference on Marine Technology, MARTEC 2016, pp. 537–544. ISSN: 1877-7058. DOI: <https://doi.org/10.1016/j.proeng.2017.08.182>. URL: <https://www.sciencedirect.com/science/article/pii/S1877705817333386>.
- [34] Shuang Zhang et al. “An alternative benchmarking tool for operational energy efficiency of ships and its policy implications”. In: *Journal of Cleaner Production* 240 (2019), p. 118223. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2019.118223>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652619330938>.