Organisation for Economic Co-operation and Development

SDD/DOC(2023)4

Unclassified

English - Or. English

7 September 2023

STATISTICS AND DATA DIRECTORATE

CO2 emissions from global shipping – a new experimental database

SDD Working Paper No. 118

JT03524824

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

CO_2 emissions from global shipping – a new experimental database

Daniel CLARKE

Philip CHAN

Matthew DEQUELJOE

Yuri KIM

Sarah BARAHONA

OECD STATISTICS WORKING PAPER SERIES

The OECD Statistics Working Paper Series – managed by the OECD Statistics and Data Directorate – is designed to make available in a timely fashion and to a wider readership selected studies prepared by OECD staff or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the Organisation. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

Working Papers describe preliminary results or research in progress by the authors and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed and may be sent to the Statistics and Data Directorate, OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The release of this working paper has been authorised by Paul Schreyer, OECD Chief Statistician and Director of the OECD Statistics and Data Directorate.

http://www.oecd.org/sdd/publicationsdocuments/workingpapers

Abstract

The shipping industry is essential for international trade, but it is also an important source of CO_2 emissions. To make progress towards climate targets, countries need to monitor CO_2 emissions from vessels owned by their ship operator companies. However, most shipping activity takes place outside national borders, making it more difficult to monitor than activity taking place within countries. The OECD's experimental database on *OECD.stat* provides a new source of data for CO_2 emissions from global shipping, which is available monthly in near real time. This data will help national statistics producers to compile their Air Emission Accounts (AEAs) for the System of Environmental Economic Accounting (SEEA). This *Working Paper* presents some initial results from the new data source and describes how they were produced. The method is based on granular and timely ship-level data provided by the United Nations Global Platform, and it uses a bottom-up estimation approach to produce results broken down by country and type of ship.

Résumé

L'industrie du transport maritime est essentielle pour le commerce international, mais elle est aussi une source importante d'émissions de CO_2. Pour progresser dans la réalisation des objectifs climatiques, les pays doivent assurer un suivi des émissions de CO_2 des navires appartenant à leurs compagnies maritimes. Cependant, la plupart des activités de transport maritime se déroulent en dehors des frontières nationales, ce qui les rend plus difficiles à surveiller. La base de données expérimentale de l'OCDE disponible chaque mois en quasi-temps réel sur OECD.stat, constitue une nouvelle source de données sur les émissions de CO_2 du transport maritime mondial. Ces données aideront les producteurs de statistiques nationales à compiler leurs comptes d'émissions atmosphériques (AEA) pour le système de comptabilité économique et environnementale (SEEA). Ce document de travail présente quelques résultats initiaux de cette nouvelle source de données et décrit la manière dont ils ont été produits. La méthode se fonde sur des données granulaires et récentes relatives aux navires, fournies par la Plateforme Mondiale des Nations Unies. Elle utilise une approche d'estimation ascendante afin de produire des résultats ventilés par pays et par type de navire.

Keywords: Environmental-economic accounting, transport, greenhouse gas emissions, climate, net zero.
JEL Classification: L91, Q56

Acknowledgements

We would like to thank the United Nations Global Platform for making this research possible by providing access to Automated Identification System (AIS) ship-tracking dataset developed by International Maritime Organization (IMO) and related datasets. We would also like to acknowledge the technical support, advice and encouragement provided by the team at the United Nations Statistics Division responsible for managing AIS data on the platform, including: Markie Muryawan, Melissa Paca, and Sean Lovell.

The research involved technical discussions with many national statistics producers. We thank everyone involved, in particular: Emmanuelle Pages, Adrien Mercier, Benjamin Cuniasse, Nicolas Carnot and Guillaume Houriez (France); Konstantinos Papandreou and Dionisis Katsikopoulos (Greece); Redbad Mosterd and Sjoerd Schenau (Netherlands); Trude Melby Bothner, Malin Bjelvert, and Evan Hart (Norway); Gemma Thomas, Ian Townsend, Craig Mclaren, Annie Maw, Richard Heys and Kirsten Newton (United Kingdom).

We would also like to extend our sincere thanks to all those who helped use with the research and provided helpful comments on the drafts: Arturo de la Fuente and Stephan Moll (Eurostat); Thomas Dewilde (Wärtsilä Corporation); and Chloe Acas, Emilie Berger, Laurent Daniel, Nathalie Girouard, Bram Groenewoud, Ivan Hascic, Asa Johansson, James Jolliffe, Claire Jolly, Myriam Linster, Florian Mante, Olaf Merk, Annabelle Mourougane, Graham Pilgrim, Rodrigo Pizarro, Roberta Quadrelli, Paul Schreyer, Ashley Ward, Colin Webb and Norihiko Yamano (OECD).

Table of contents

CO_2 emissions from global shipping – a new experimental database 2

OECD STATISTICS WORKING PAPER SERIES 3

Acknowledgements 5

Summary 9

1 Introduction 10
  Motivation for the study 10
  Relevance 11

2 Key results 12

3 Study design and data sources 16
  Scope of the study 16
  Allocation of ships to countries 17
  Special cases 18
  Data sources and preparation 19

4 Modelling methods 22
  Choosing the modelling approach 22
  Overview of the model and workflow 23
  Feature identification from data sources (input matrix) 24
  Running the model and checking its performance 25
  Aggregation by ship type and over time 28

5 Assessing and presenting results 30
  Comparison with other global studies 30
  Comparison with official AEAs 31
  Protecting confidentiality 33

6 Conclusion and future work 35
  Conclusion 35
  Future work 35

Bibliography 38
Annex A. 42
Annex B. 43

FIGURES

Figure 2.1. Shares of CO_2 emissions from global shipping (%) by ship type, 2019-2022 12
Figure 2.2. CO_2 emissions from global shipping by country, OECD, 2022 13
Figure 2.3. Monthly CO_2 emissions index for container and cruise ships 14
Figure 2.4. CO_2 emissions from air and maritime transport, 2019 to 2022 15
Figure 3.1. Shares of ship types in global estimates by DWT 16
Figure 3.2. Ship-level information and allocation to countries in the national accounts 17
Figure 4.1. Model workflow 23
Figure 4.2. MAPE by ship type 27
Figure 4.3. Estimated efficiencies compared to IMO averages 28
Figure 4.4. Number of national operators per OECD country, 2019-22 29
Figure 5.1. Annual estimates for the global fleet compared 30
Figure 5.2. Comparison of OECD estimates with official AEAs for Germany and Spain 31
Figure 5.3. Comparison of OECD estimates with official AEAs for Greece and Netherlands 32
Figure 5.4. Comparison of OECD estimates with official AEAs for Italy 32
Figure 5.5. Emissions concentration: HHI in OECD countries 34

TABLES

Table 4.1. Vessel characteristics variables used in the random forest model 25
Table 4.2. Test set and k-fold cross-validation metrics 26
Table 4.3. Top 10 features by permutation importance 27

Table A.1. Vessel Population Summary Statistics: OECD estimates 2019-2022 compared with IMO figures for 2018 42
Table B.1. AIS vessel coverage 43
Table B.2. Imputation of ship characteristics 43
Table B.3. Use of a predicted emissions efficiency ratio, by ship type 44
Table B.4. Use of a predicted emissions efficiency ratio, by country (OECD countries) 45

Summary

Global shipping is essential for international trade, but it is also an important source of CO_2 emissions. To make progress towards climate targets, countries need to monitor CO_2 emissions from vessels owned by their ship operator companies. However, most shipping activity takes place outside national borders, making it difficult to monitor. The OECD's new experimental database on *OECD.stat* provides a source of monthly data for CO_2 emissions from global shipping, updated each quarter with results broken down by country and type of ship. The data will help national statistics producers to compile their Air Emission Accounts (AEAs) for the System of Environmental Economic Accounting (SEEA).

The OECD estimates that there were 858 million tonnes of CO_2 emissions globally from the shipping industry in 2022, compared with 739 million tonnes of CO_2 emissions from air transport (flights). The OECD countries with the largest CO_2 emissions from shipping in 2022, based on the country of residence of the ship operator, were Greece, Japan and the United States. The new data also shows that the reduction in CO_2 emissions from shipping during the COVID-19 pandemic was less than for flights.

This Working Paper describes how the OECD database and estimates of CO_2 emissions from shipping were produced using a ship-tracking dataset known as the Automatic Identification System (AIS), which was developed by International Maritime Organization (IMO) of the United Nations (UN) and is available on the UN Global Platform. The AIS tracks individual ships as they move around the world. The OECD estimated CO_2 emissions for each voyage and allocated them to the country of the ship operator in line with the "residence principles" of the international System of National Accounts and the SEEA.

The study adopted a bottom-up approach that uses each ship's AIS transmissions and information on its characteristics from the Information Handling Services (IHS) ship register. The study also used information from previous IMO studies and the European Union's monitoring, reporting and verification data system (EU-MRV). The combined IHS and AIS data is linked to the EU-MRV using the IMO identification number to create a training dataset which is used to predict an emissions efficiency ratio for each ship using a random forest regression model. The random forest model was chosen because it strikes a balance between computational efficiency and granular modelling, and it has greater accuracy than the alternative regression models that were tested. The emissions efficiency ratio is then multiplied by the distance travelled by the ship to obtain its CO_2 emissions.

Comparisons of results of this study with those of the International Energy Agency (IEA) and IMO suggest that the estimation model is robust at the global scale. However, when compared with individual country AEAs, there are differences. Although there are several explanations for these, the paper recommends that compilers of official AEAs compare their data with ours, as it may help to fill gaps in the information available to them. The OECD plans to continue working with national statistics producers on improving alignment, as well as using this new data source to contribute to the development of AEA "bridging items".

1 Introduction

Motivation for the study

The shipping industry is essential for international trade, but it is also an important source of global CO_2 emissions. The International Maritime Organization (IMO)'s fourth study of greenhouse gas (GHG) emissions (IMO, 2020[1]) estimated that in 2018 maritime transport accounted for 2.9% of total GHG emissions originating in human activity.

Transport is one of the areas of economic activity for which information is needed to monitor countries' progress towards climate targets under the 2015 Paris Agreement, which is a treaty under the United Nations Framework Convention on Climate Change (UNFCCC), and the agreements reached at subsequent Conferences of the Parties (COPs). However, reporting to the UNFCCC is on a "territory" basis, which means that it covers activities taking place within the geographic borders of each country.[1] Transport activities that take place outside national borders – in the air or on the sea – are more difficult for countries to monitor than those taking place within their borders.

In 2022, the OECD published a near-real-time database of CO2 emissions from air transport. It has now developed a data source for emissions from global shipping, also available monthly in near real time. The new "experimental" database for CO2 emissions frommaritime transport is derived from location tracking devices used for all types of ships making international voyages. The data from these devices feeds into a ship-tracking dataset, known as the Automatic Identification System (AIS), which was developed by International Maritime Organization (IMO) of the United Nations (UN) and is available on the UN Global Platform.[2] Allocation to countries in the database is based on the residence of the companies that operate the ships, known as ship operators[3] (see Allocation of ships to countries).

A number of advantages were identified compared with related statistics[4] and the information for the water transport sector currently available in the Air Emissions Accounts (AEAs) compiled by countries:

- Coverage: the new database has global coverage, with vessels covering operators from over 190 countries and territories. AEAs that include figures for the water transport sector are currently only available for 38 countries.[5]

---

[1] For shipping, this usually means where the ships are re-fuelled.

[2] https://unstats.un.org/wiki/display/AIS/AIS+data+at+the+UN+Global+Platform

[3] The OECD's country-level estimates are on a "residence" basis (not the "territory" basis used for UNFCCC reporting). The residence principle is used both in national accounting and in the System of Environmental-Economic Accounting (SEEA). Emissions are allocated to the country of residence of the ship operator.

[4] Top-down (fuel statistics-based) emissions for water transport are available on an annual basis for selected countries from the IEA (CO_2 Emissions from Fuel Combustion) and from UNFCCC inventories reporting (https://unfccc.int/topics/mitigation/resources/registry-and-data/ghg-data-from-unfccc).

[5] Official AEAs are available at: https://stats.oecd.org/Index.aspx?DataSetCode=AEA.

- Frequency and timeliness: other sources of CO_2 emissions are usually available on an annual basis only and have a publication time lag of between one and two years. The new OECD database of CO_2 emissions from maritime transport contains monthly data, updated each quarter.
- Accuracy: the methodology used (see Modelling methods) is a bottom-up (vessel-level) method based on detailed information from variables contained within or linked to AIS data via the shipping vessels' unique identifier (e.g. IMO number). This methodology facilitates opportunities for integration with other information for granular analyses and adjustments for special cases, making it possible to produce more accurate measures including when aggregated for type of vessels and countries.

Relevance

A key objective for this piece of work was to help countries improve their estimates for CO_2 emissions from international shipping, in particular their official AEA estimates which are part of the System of Environmental-Economic Accounts (SEEA). Using a bottom-up approach to CO_2 emissions estimation makes it possible to compare the results with national estimates of AEAs (see Aligning with the AEAs). This contributes to the work of national compilers of the AEAs and helps countries to monitor their progress towards meeting climate targets.

The OECD database for CO2 emissions frommaritime transport can also be used by international organisations wishing to assess progress towards meeting agreed targets for shipping. For example, the IMO's 2023 Strategy on Reduction of GHG Emissions from Ships envisages a reduction in carbon intensity of international shipping by at least 40% by 2030 compared with 2008 and achievement of climate neutral ("Net Zero") GHG emissions from international shipping "by or around" 2050.[6]

The European Union (EU) aims to reduce GHG by at least 55% by 2030, and to achieve Net Zero by 2050. With its capacity to track emissions from shipping at regional/country level, the OECD's new database can be used to monitor progress towards these targets. Analyses by vessel type could also be useful in relation to the EU Emissions Trading System, which includes CO_2 emissions from ships with over 5000 gross tonnage utilising European ports.

The data potentially has many other applications for climate change mitigation policies. In the context of integrating GHG emissions data with statistics on economic activities, further work could be done linking this information to analyses of trade, employment, and impacts of new policies, and incentives or technologies designed to increase transport efficiency.

In theory, it would also be possible to produce comparative indicators such as CO_2 emissions per unit of value added for the water transport industry. However, due to the experimental nature of the new database on CO_2 emissions from maritime transport (explained in the following sections), we have not included such indicators in this paper.

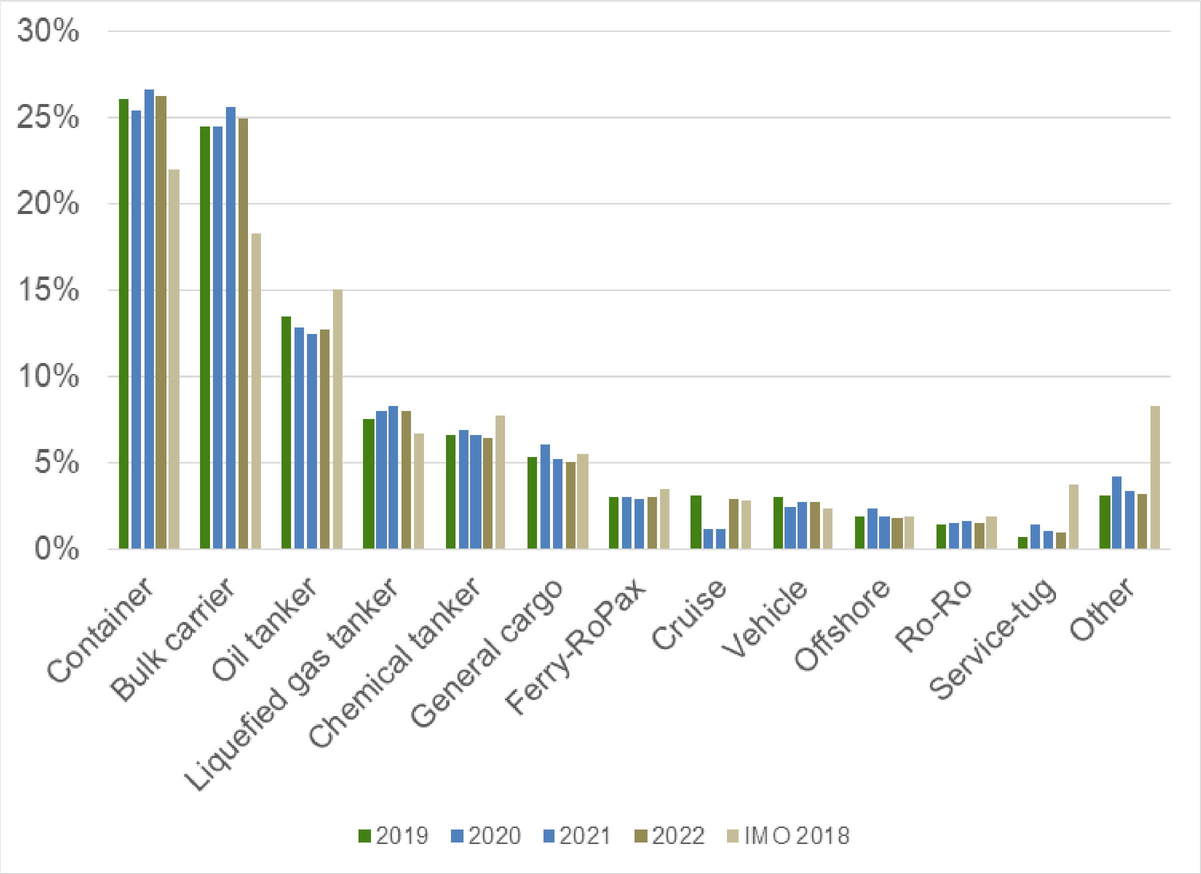
---

[6] https://www.imo.org/en/MediaCentre/PressBriefings/pages/Revised-GHG-reduction-strategy-for-global-shipping-adopted-.aspx.

2 Key results

The Automatic Identification System (AIS) data used to produce the new estimates is collected from transmitters onboard ships making international voyages approximately every six seconds and is available from January 2019. This data is combined with other information by the OECD and results are produced using an estimation model (see Modelling methods). They show that around half of total emissions are from container ships and bulk carriers, the first two ship type categories in Figure 2.1. Another one-fifth are from the transport of fossil fuels (oil and liquefied natural gas tankers). The smaller types of maritime vessels (cruise ships, for example) have relatively little impact on the total emissions.

Figure 2.1. Shares of CO_2 emissions from global shipping (%) by ship type, 2019-2022



Note:

1. OECD estimates are available from 2019. Estimates for 2018 from the latest IMO greenhouse gas study (IMO, 2020[1]) are shown for comparison.
2. Ship type categories follow (IMO, 2020[1]) based on IMO Safety Regulations. Ro-Ro = roll-on/roll-off ship. Ro-Pax = roll-on/roll-off passenger ship.

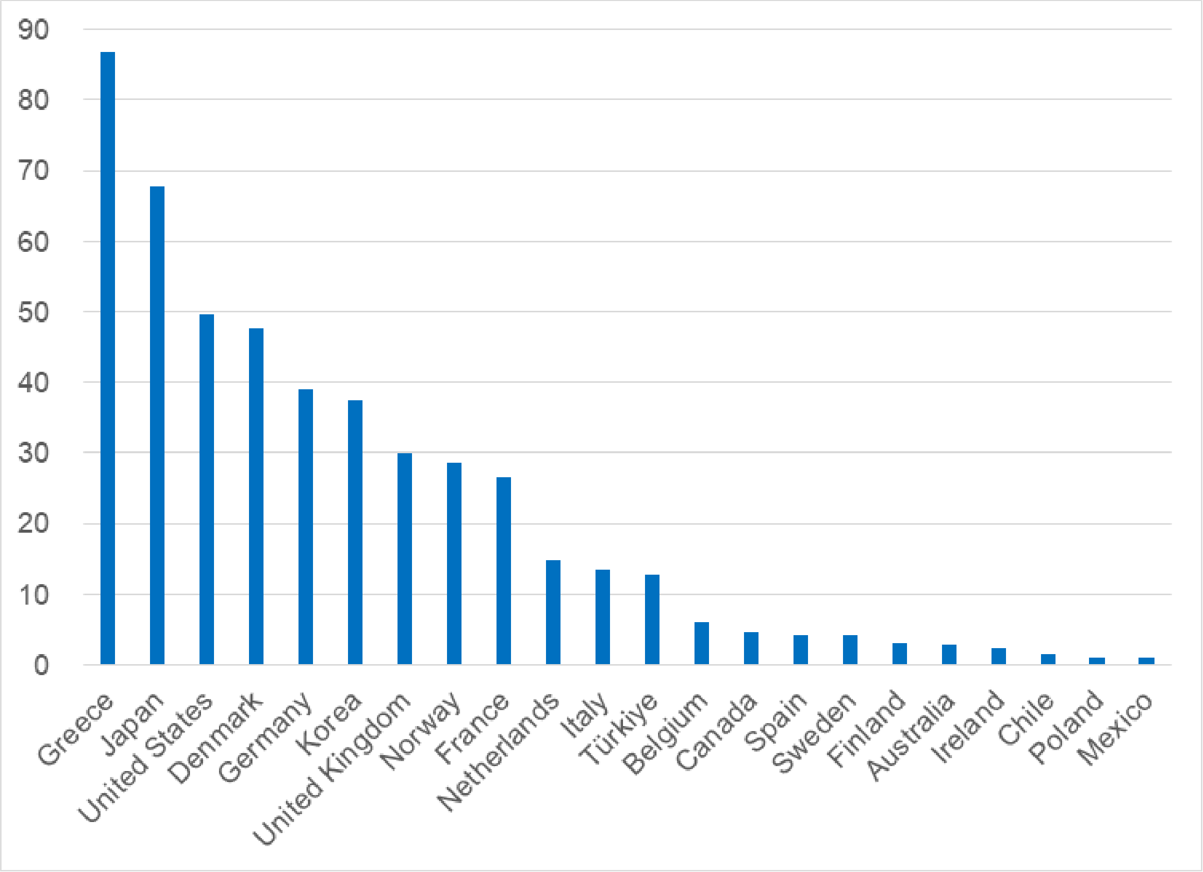
Source: OECD database and (IMO, 2020[1]).

In 2022, according to our estimates, there were 858 million tonnes of CO_2 emissions globally from the shipping industry, compared with 739 million tonnes of CO_2 emissions from air transport (domestic and international flights); and 63% of emissions from global shipping came from vessels operated by companies based in OECD countries. The level of CO_2 emissions from global shipping in 2022 may have been affected by disruption to shipping as a result of Russia's invasion of Ukraine and related trade sanctions against Russia.

Figure 2.2 shows the countries with the largest CO_2 emissions from global shipping in 2022, based on the country of residence of the ship operator. The top three countries were Greece, with 87 million tonnes of emissions (10% of the world total), Japan with 68 million tonnes and the United States with 50 million tonnes. Emissions are highly concentrated among countries, with the top ten countries accounting for 50% of the world total. It should be noted that the country breakdowns are "experimental" partly because although the OECD has already discussed the estimates with several countries, further discussions may be needed as compilers of official Air Emissions Accounts (AEAs) continue to review the results (see Aligning with the AEAs).

Figure 2.2. CO_2 emissions from global shipping by country, OECD, 2022

Million tonnes of CO_2



Notes:

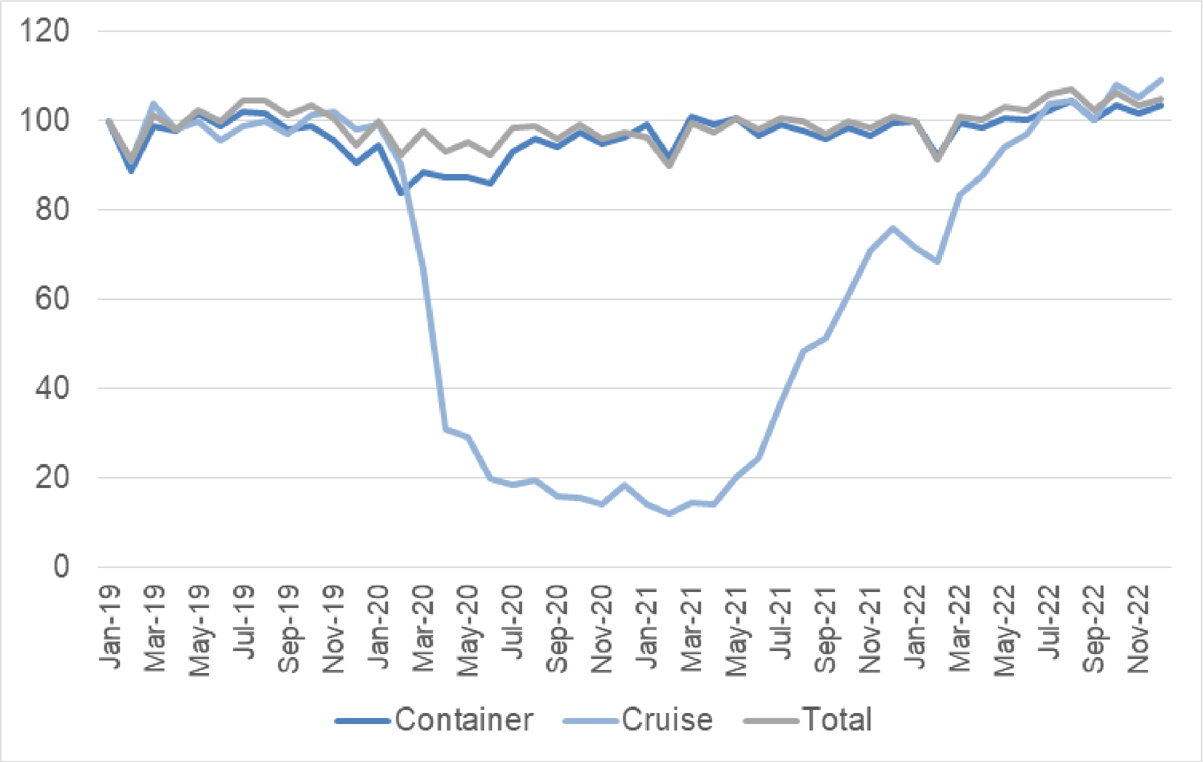
1. Figures for OECD countries with less than 1 million tonnes of CO_2 emissions are not shown.
2. Excludes overseas territories.

Source: OECD database.

The new database can also be used to explore the impact on CO_2 emissions from shipping of events such as the COVID-19 pandemic. The OECD's estimation model starts from the year before the pandemic. The impact of the pandemic on CO_2 emissions from global shipping can be seen in March-April 2020 and the following months (Figure 2.3). Cruise ships were particularly affected, and this was reflected in a sharp fall in their CO_2 emissions in March and April 2020. Emissions from cruise ships only reached pre-pandemic levels again in the mid-year high season of 2022. On the other hand, container ships were less affected at the start of the pandemic, and by August 2020 their emissions had returned to the levels recorded in January. The trend for total CO_2 emissions from global shipping is driven by large vessels such as container ships.

Figure 2.3. Monthly CO_2 emissions index for container and cruise ships

Jan 2019=100



Source: OECD database.

Figure 2.4 compares the annual figures for CO_2 emissions from air and maritime transport over the period of the COVID-19 pandemic. Global CO_2 emissions from flights were affected by domestic and international travel restrictions,[7] they fell by 46% between 2019 and 2020 and remained 21% below 2019 levels in 2022. On the other hand, CO_2 emissions from global shipping fell by only 4% in 2020, and by 2022[8] they were 2% higher than in 2019. A similar pattern can be seen for air and maritime transport emissions from OECD countries.

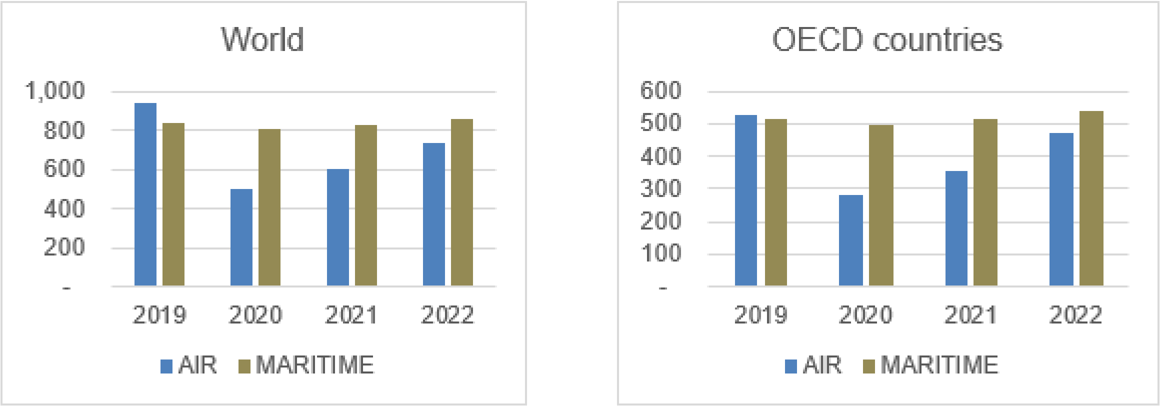
---

[7] See https://ourworldindata.org/covid-international-domestic-travel from (Mathieu et al., 2020[52]).

[8] The level of CO_2 emissions from global shipping in 2022 may have been affected by disruption to shipping as a result of Russia's invasion of Ukraine and related trade sanctions against Russia.

Figure 2.4. CO_2 emissions from air and maritime transport, 2019 to 2022

Million tonnes of CO_2



Source: OECD database.

3 Study design and data sources

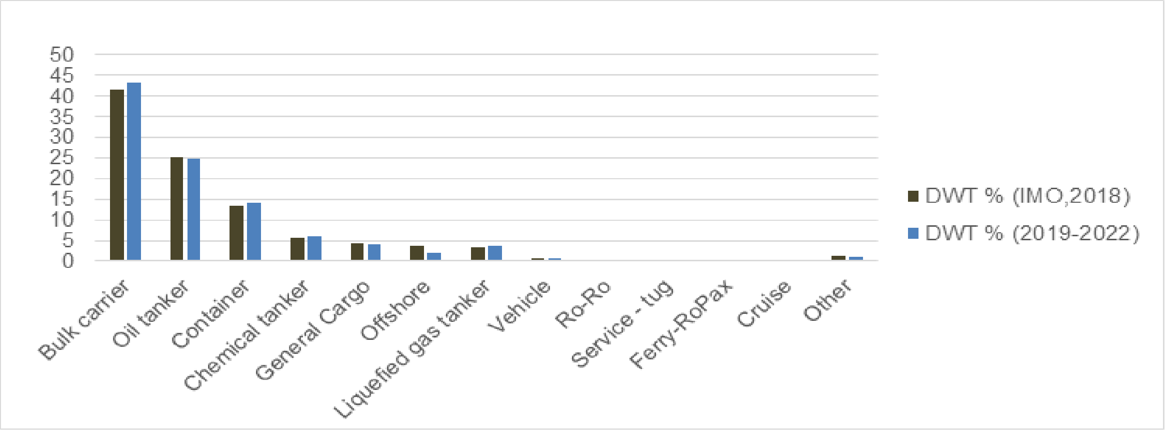
Scope of the study

The scope of this study was limited to emissions of carbon dioxide (CO_2) only. Other studies have also covered other greenhouse gases and air pollutants, but according to the latest IMO greenhouse gas (GHG) study (IMO, 2020[1]), CO_2 emissions account for 91% of global warming potential of the shipping industry.

The scope of activity for the estimation was the global shipping fleet defined by the vessels included in the Automated Identification System (AIS) ship-tracking system. The AIS is compulsory for international commercial ships with gross tonnage (GT) of 300 or more tonnes, and all passenger ships regardless of size.[9] It excludes some fishing vessels, and it may include some large ships on inland waterways as well as seagoing vessels.

The population of vessels derived from the AIS dataset (2019 onwards) contains fewer oil and chemical tankers than that of the IMO 2020 study (which uses 2018 data). However, in terms of deadweight tonnage (DWT)[10] by ship type, the share for bulk carriers, tankers and container ships is similar (Figure 3.1). Annex A provides further details.

Figure 3.1. Shares of ship types in global estimates by DWT



Notes:

1. DWT % is the proportion of DWT of all ships that is in this ship type category. Estimates for 2019-2022 are an average of the four years of estimates produced by the OECD.
2. Ship type categories follow IMO (2020) based on IMO Safety Regulations. Ro-Ro = roll-on/roll-off ship. Ro-Pax = roll-on/roll-off passenger ship.

Source: Authors' calculations and IMO (2020).

---

[9] IMO (1974/1980). International convention for the Safety of Life at Sea (SOLAS), Chapter V: Safety of Navigation, Regulation 19.

[10] DWT is a standard measure for measuring overall load or carrying capacity of vessels in terms of mass or weight.

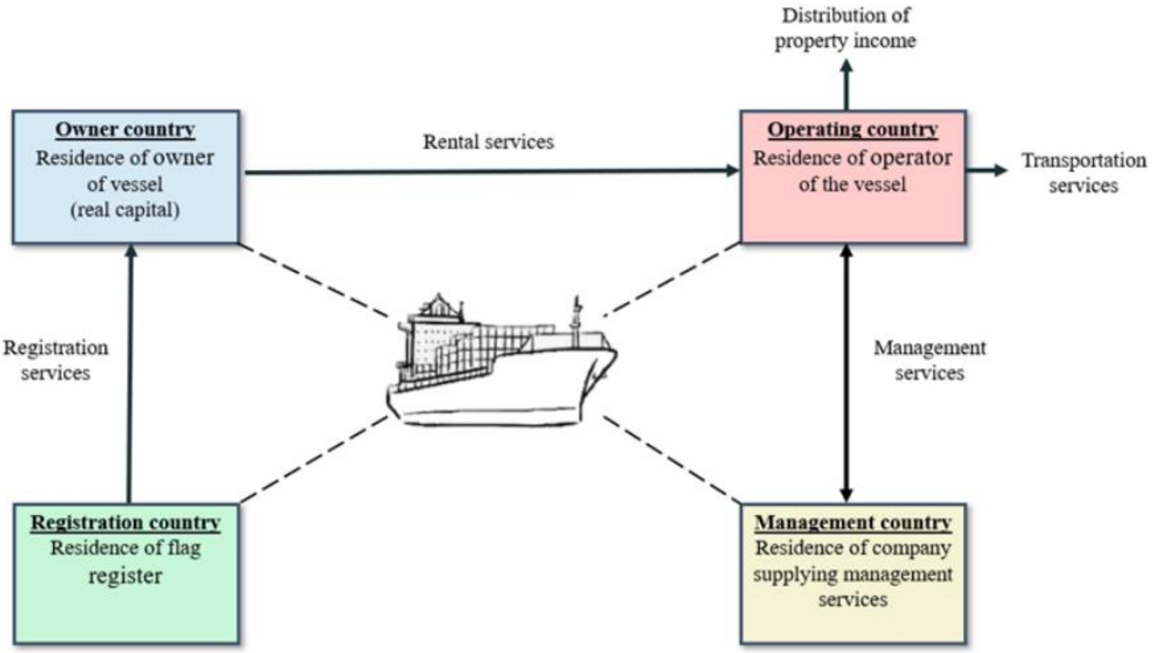
Allocation of ships to countries

This study adopts the residence principles of the international System of National Accounts (SNA) and the System of Environmental-Economic Accounting (SEEA), which follows the SNA approach. The national accounts report on economic activity by resident enterprises, which, along with households, are the basic unit for measurement of economic activity. Each vessel in the global fleet is associated with a ship operator[11] in the AIS database, with information on the company's country of residence.

Residence is determined in the SNA not by a legal or citizenship criterion but according to the centre of economic interest for the business. Statistics Norway developed Figure 3.2 to show how economic activity such as production of transport services and the income associated with each ship is associated with the ship operator and the "Operating country" (top right-hand corner of Figure 3.2). The same approach is used to allocate the CO_2 emissions from each ship to the country of residence of its ship operator.

Figure 3.2 also shows that ships can be linked to the country where an owner lives ("Owner country"), the country of a company that supplies management services ("Management country") or to the country of the ship's flag ("Registration country"). However, these aspects are not used to determine a ship operator's residence and do not affect the allocation of CO_2 emissions from ships to countries.

Figure 3.2. Ship-level information and allocation to countries in the national accounts



Source: Statistics Norway case study, chapter 9.13 in (Eurostat, 2020[2]).

---

[11] The data source used in this study (ShipData - IHS Markit, via the UN Global Platform) defines the ship operator as "The company responsible for the commercial decisions concerning the employment of a ship and therefore who decides how and where that asset is employed. The direct beneficiary of the profits from the operations of the ship, this company may also be responsible for purchasing decisions on bunkers and port services. A medium to long-term time or bareboat charterer is considered to be the operator of the ship. Vessels within commercial pools are considered to be operated by the pool".

The allocation for economic residence purposes can be modified where additional information is available. For example, for cases of ship operators with operations and office locations across the globe, there may be special treatments and existing agreements between national statistical systems for allocation of the companies (see Special cases). The aim of any adjustments made in allocation to countries, is to align with the compilation of the AEAs and other economic statistics.

The national accounts and the SEEA classify data on productive activities according to primary activity of firms using the International Standard Industrial Classification of All Economic Activities (ISIC). Maritime vessels (or fleet of vessels) are usually classified under the water transport industry, which is Section H Division 50 (H50) in the ISIC. Although this was generally the case for ships in our database, it should be noted that the dataset also includes cases of ship operators that may also be classified in another economic activity. For example, some ship operators are classified by national statistics offices (NSOs) as transport logistics companies (H52 in the ISIC). The dataset also includes a small number of companies involved in maritime transport for which the primarily business is considered by the NSO to be natural resources extraction (e.g. offshore oil or natural gas extraction).

Special cases

Multi-national enterprises

An institutional unit can be a resident of only one economic territory at a time, so the national aggregation of emissions is mutually exclusive across countries. However, the shipping industry often involves vertical or horizontal integration of business activities and given its trans-boundary nature, it is naturally compatible with multi-national enterprise (MNE) structures. Many ship operators have offices in multiple countries and some are part of companies that also engage in other economic activities.

When compiling the national accounts, NSOs from several countries where an MNE operates may reach agreements about which parts of the MNE should be considered resident in each country. Where such agreements exist, the estimates of emissions from ships operated by the relevant MNE may be adjusted to align the OECD estimates with the treatment in the national accounts. The approach mirrors the methodology used for the OECD database of CO_2 emissions for air transport.[12] Guidance can be found in the UNECE *Guide to Measurement of Global Production* (UNECE, 2015[3]) and in Eurostat's *Handbook on the compilation of statistics on sea and air transport in national accounts and balance of payments — 2020 edition* (Eurostat, 2020[2]).

Overseas territories

Overseas territories are another important consideration for allocation of international shipping to countries. Overseas territories may be included in the economic statistics of the "parent" country or they may be have separate accounts. This is only relevant for a small number of countries, but the impact on the results for CO_2 emissions from maritime transport by country was found to be significant in some cases.[13] As

---

[12] See for example the case of SAS Airlines, https://www.oecd.org/sdd/co2-emissions-from-air-transport-ecc9f16b-en.htm.

[13] Denmark (Greenland and Faeroe Islands); France (Guadeloupe, French Guiana, Martinique, Reunion Island, and Mayotte, Saint-Pierre and Miquelon, Wallis and Futuna, French Polynesia, Saint-Barthélemy, and Saint-Martin); Netherlands (Curacao, Sint Maartin, Aruba, Bonaire, Sint Eustatius and Saba); Portugal (Azores and Madeira):Spain (Canary Islands, Balearic Islands);United Kingdom (Overseas territories: Gibralter, Bermuda, Cayman Islands, Falkland Islsndd, and Monsserrat, and UK Virgin Islands, Crown dependencies: Guernsey, Jersey, Isle of Man); United States (Puerto Rico, American Samoa, US Virgin Islands, Guam, and the Commonwealth the Northern Mariana Islands).

practices vary, the approach to treatment of overseas territories was handled on a case-by-case basis. The OECD database of CO_2 emissions from maritime transport aims to replicate the practices followed in by national statistical offices. However, for the sake of consistency, in the OECD database emissions for overseas territories are shown separately.

Data sources and preparation

The data sources used in the estimation of CO_2 involve real-time movement of ships and their physical characteristics, as well as information on their technical specifications, operator nationality, and service status. This section starts by providing an overview of the sources. Then we outline how the sources are linked to construct the input data for our estimation model.

AIS ship movement data

The first data source is the AIS, a real-time geo-tracking and identification system for maritime vessels. The system contains static and dynamic vessel information, electronically exchanged between AIS-receiving stations (onboard, ashore or satellite). It uses Global Positioning Systems (GPS) in conjunction with shipboard sensors and digital Very High Frequency (VHF) radio communication equipment to automatically exchange navigation information electronically.[14] AIS transceivers are required for most large shipping vessels in the global fleet by the International Convention for the Safety of Life at Sea (SOLAS).[15]

The AIS tracking system was originally developed to complement marine radar technology for collision avoidance. It contains information on the coordinate, speed, draught, and navigational status of the vessel and its engines which are transmitted every few seconds.

In recent years there has been a growing interest in applying the data to other domains. Li et al. (2016[4]) used AIS data to estimate ship emissions in the Pearl River Delta region in China, while Leong et al. (2015[5]) did so for the Singapore Straits, one of the busiest shipping areas in the world. A study on the various uses[16] of AIS signals in maritime research was conducted by Svanberg et al. (2019[6]). With the blossoming of this new research literature, and a growing user demand, the AIS dataset in near-real time is made available to researchers on the UN Global Platform as part of its goal to increase accessibility of big data sources for policy-relevant research.[17]

This OECD study uses several types of information from AIS transmissions to estimate CO_2 emissions. We calculate the distance travelled by each vessel between each coordinate location transmitted by successive AIS signals using the geodesic distance, following Karney's method (Karney, 2013[7]). This measurement, which takes into account the Earth's surface curvature, produces daily total distances travelled for all vessels with a valid or potentially valid IMO number in the database. The IMO number

---

[14] UN Statistics Division, “Overview of AIS Dataset”, https://unstats.un.org/wiki/display/AIS/Overview+of+AIS+dataset.

[15] Chapter V (Safety of Navigation) Regulation 19, of the SOLAS Convention requires fitting an AIS transceiver for nearly all ships above 300 gross tonnage (GT), including for ships that are not engaged in international voyages. Some exceptions are made depending, for example, on ship type or year of construction. Coverage of inland waterways varies by country and the SOLAS regulation does not broadly apply for fishing vessels, pleasure craft, support vessels and inland waterway vessels.

[16] In addition, Coello et al (2015[15]) used AIS data to estimate the emissions of the United Kingdom fishing fleet. Arslanalp et al. (2019[47]) leveraged AIS data to nowcast trade activity in Malta, while Furukawa et al. (2022[48]) have produced nowcasts of exports in Japan. Fiorini et al. (2016[49]) undertook route analysis from raw AIS data for maritime spatial planning.

[17] https://unstats.un.org/bigdata/un-global-platform.cshtml.

consists of seven digits and can be verified by checking that the remainder of a weighted combination of the digits divided by ten equals the last digit of the IMO identification.[18] The database contains a small number of erroneous IMO numbers; only valid IMO numbers are used for the estimation (see Future work). Valid matches for IMO identification were achieved for 93% of the vessels in the database (average 2019-2022, see Table B.1 in Annex B). Thus 7% of the global fleet with AIS transmitters is excluded from the "experimental" results (those currently available) due to insufficient data.

We also extract the vessel's draught and speed-over-ground (SOG) directly from AIS as they are transmitted in tandem with the coordinate data. Draught is a measure of a vessel's depth (i.e. distance between the waterline and the ship's keel) and thus is an indicator of the ship's load, while SOG is a measure of the speed relative to distance of surface area for the globe. Both variables are expected to influence fuel usage of the vessel and are employed in bottom-up approaches to calculating emissions such as Goldsworthy et al. (2015[8]) and Smith et al. (2013[9]). We calculate the daily average of the draught and SOG recorded for each ship, later aggregated to an average of the period in question (e.g. month or year).

It should be noted that the AIS data contains errors, which need to be corrected before it can be used. There are a number of factors that can impact the quality of AIS messages emitted by ships. The signals may be partially lost due to meteorological elements or magnetic interference. There may also be decryption errors or technical errors with the AIS device that lead to invalid values for certain variables. For instance, invalid latitude and longitude coordinates can yield locations that are far away from the ship's actual location. AIS message receivers have specific timeslots for data reception which may mean that in crowded areas not all ship data will fit, leading to data losses. Land-based receivers have a range of 40 nautical miles which can also limit vessel coverage. Ships can also turn off their AIS transponders, leaving gaps in the data. Also, some variables are inputted manually, such as draught or destination, which can give rise to human errors or intentional misreporting.[19]

Implausible observations are filtered out in several ways. First, latitude and longitude coordinates that fall outside their defined ranges of -90 to +90 and -180 to +180 degrees respectively are discarded. Then the voyage zone of a ship is identified in order to discard invalid positions that can arise due to device signalling errors. This is done by identifying the most common H3 index resolution level 1 value across the ship's messages sent for a given day under the assumption that this contains the set of valid messages.[20] Any index value that is not found in the two surrounding rings of neighbouring hexagons to this value is then discarded. In this way, erroneous or implausible location coordinates in AIS data are identified and filtered without losing the rest of the (correct) data for the same vessels and distances can still be calculated over time. Lastly, when calculating the distance travelled between two points, the implied speed in knots is calculated by considering the length of time between the two messages. If the implied speed exceeds 60 knots then the observation is discarded as implausible.

IHS ship register database

The second data source used in this study is the Information Handling Services (IHS) ship register,[21] available from the UN Global Platform, which contains information on ships' characteristics. This information is updated in the UN Platform on a monthly basis. It includes information on the ship's operator, including the name and country of domicile (or main headquarters), which is used as the first criterion for identifying the vessel operator's economic residence (see Allocation of ships to countries). The

---

[18] https://www.imo.org/en/OurWork/IIIS/Pages/IMO-Identification-Number-Schemes.aspx.

[19] https://unstats.un.org/wiki/display/AIS/Overview+of+AIS+dataset.

[20] H3 Index is a hexagonal geospatial referencing system applicable at multiple levels of resolution (0-15).

[21] The register is compiled by IHS Markit, which merged with S&P Global group in 2022.

register also provides information on the technical specifications of vessels, including the physical dimensions of the ship, the engines onboard for propulsion and electricity generation, fuel types, design speed, ship type and other features. These feed into the pool of candidate features used in the OECD's emissions model (see Modelling methods).

IMO greenhouse gas studies

The IMO studies are important sources of expert research on GHG emissions from the global fleet and was used as a reference for several aspects of the work, including the classification of vessel types and as an input of information into the model on efficiency of vessels with respect to fuel consumption and CO_2, given its type and size.

EU-MRV

The OECD's CO_2 estimation model also uses the of the EU's monitoring, reporting and verification data system (EU-MRV).[22] The EU MRV system was developed for the EU strategy (European Union, 2015[10]) on reducing GHG emissions from the shipping sector and data currently covers annual CO_2 emissions from 2018 onwards for all vessels over 5,000 GT loading or unloading cargo or passengers at ports in the European Economic Area (EEA). The IMO's Fourth GHG Study (IMO, 2020[1]) found that the operations and vessel coverage of the EU-MRV dataset was highly representative of global shipping activity.

We use the EU-MRV both for its reported emissions efficiency measure, which is available for around one-fifth of ships in the AIS dataset, and to estimate an efficiency measure for the four-fifths of ships that are not in the EU-MRV dataset. Specifically, we take the variable *average CO_2 emissions per nautical mile* from this dataset as the target variable to train the regression model (see Modelling methods).

Linking across datasets

Starting with the IHS ship register, characteristics of recorded vessels are linked to the ship type and weight class as defined in the IMO's Fourth GHG Study (IMO, 2020[1]) using the first two digits of StatsCode5 from the IHS register. This identification permits several advantages. First, it facilitates comparisons of our estimates with the IMO's results, which we take as benchmark. Second, using the predefined IMO ship types, the average of each type and weight class is used to impute missing or implausible values in the IHS variables. Table B.2 in Annex B shows the proportion of observations where imputation is required for the ship characteristics used in the model. While most variables have a relatively low proportion of missing values, variables such as "total bunker capacity" and "displacement" require imputation for more than one-third of all observations. Real-time variables derived from the AIS dataset are then added to include the distances travelled, draught and SOG.

To obtain the training dataset, the combined IHS and AIS data is linked to the EU-MRV using the IMO identification number. We retain only the vessels that are in the EU-MRV as well as the AIS database and IHS ship register. In this way, we capture all vessels which are in the ship register and transmit AIS signals in the year in question. This is around 18% of vessels, which are used to predict the emissions efficiency ratio in our regression model (see Overview of the model and workflow).

---

[22] https://mrv.emsa.europa.eu/#public/eumrv.

4 Modelling methods

Commercial ships emit pollutants in two main ways. First, exhausts are emitted from the main engine during propulsion. The amount of emission depends on the engine size, fuel used, operating speed and load. These four factors form the basis of the emission modelling strategy in much of the literature favouring the activity-based approach. In addition, auxiliary engines also emit pollutants in generating electricity used for lighting, heating, pumps, refrigeration, ventilation, and other services. Second, while at berth, the auxiliary engine often stays on to maintain these activities. Depending on the port and local authority, the type of fuel used when a vessel is close to port might be regulated to limit the amount of harmful pollutants exposed to inhabitants. For instance, the International Convention for the Prevention of Pollution from Ships (MARPOL) created emission control areas off the Pacific coasts of the United States, Canada, Baltic Sea and North Sea area.[23] Regulations in other countries were reviewed by (OECD, 2017[11]) and more recently by the maritime insurer Gard.[24]

Choosing the modelling approach

The activity-based approach to estimating maritime emissions is the most common in the literature. This bottom-up approach posits that the emission E of a vessel is given by a multiplicative relationship involving the installed power of engines P, the load factor LF, operating time for of each engine T, and the emission factor EF of all the pollutants emitted by the vessel, or

$$E_{i,j,k,l} = P_j * LF_{j,l} * T_{j,k,l} * \ EF_{i,j,k} \tag{1}$$

for all pollutant i, engine type j, fuel type k, and operational mode l.

One of the most comprehensive methods based on this approach was developed by Olmer et al. (2017[12]) for the International Council on Clean Transport (ICCT), where the authors augmented the model to include external variables such as weather and travel conditions. Olmer et al. (2017[12]) was also a reference for the IMO's Fourth Greenhouse Gas (GHG) Study (IMO, 2020[1]). Other studies that used an activity-based approach include Goldsworthy et al. (2015[8]), Olesen et al. (2010[13]), Jalkanen et al. (2017[14]), Coello et al. (2015[15]), Ng et al. (2013[16]), Chen et al. (2016[4]), Johansson et al. (2017[14]), Leong et al. (2015[5]). One drawback, however, is the complexity of such detailed modelling, such that emissions estimates were in most cases limited at the port or regional level. It was also difficult to calculate time series.

Another common approach is a top-down approach using bunker fuel sales or consumption to estimate ship emissions. This approach has been studied by Olivier et al. (1999[17]), Endresen et al. (2003[18]), and Mao et al. (2022[19]), among others. Essentially, it multiplies the volume of fuel consumption by an emission factor of the corresponding fuel, for all pollutants, to obtain a bound on the level of emissions that is possible. Fuel consumption could be obtained, for instance, by a bunker fuel delivery note and periodic stocktakes of fuel tanks and monitoring fuel tank levels on board (European Union, 2015[20]). While this

---

[23] https://www.imo.org/en/OurWork/Environment/Pages/Emission-Control-Areas-(ECAs)-designated-under-regulation-13-of-MARPOL-Annex-VI-(NOx-emission-control).aspx.

[24] https://www.gard.no/web/updates/content/29212584/regional-sulphur-emission-limits-at-a-glance.

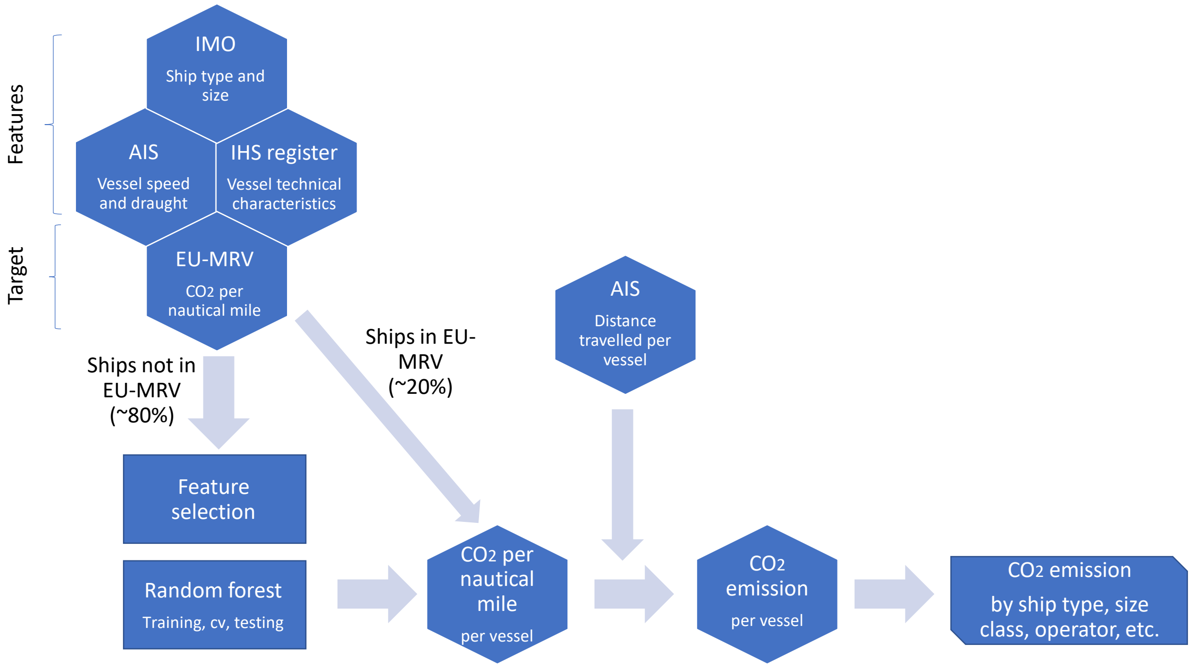
approach enjoys relative ease of implementation, bunker sales and actual fuel consumption data are not publicly available. Moreover, there is debate in the literature on the usefulness of such statistics (Eyring et al., 2010[21]).

This OECD study chose a bottom-up approach that differs from the traditional activity-based bottom-up approach outlined above (Equation 1) in that it exploits vessel-level information, as the traditional method does, but simplifies the modelling of ship specification and movement. This avoids many of the issues associated with full activity-based models which have relatively heavy requirements in terms of data inputs, data processing and shipping expertise. Moreover, it allows aggregation and breakdowns according to the national accounting principle of economic residence and permits estimates to be produced in a more frequent and timely fashion than those of the IMO GHG studies. The choice of a random forest model (see Overview of the model and workflow) was part of the strategy to strike the right balance between computational efficiency and granular modelling.

Overview of the model and workflow

In the OECD approach, CO_2 emissions are estimated for each vessel in the global Automated Identification System (AIS) dataset on the UN Global Platform in a two-step approach. First, we estimate, for each ship for which such information is not available from the EU-MRV dataset, an auxiliary variable known as the emissions efficiency ratio which is defined as the *average CO_2 emissions per nautical mile* (kg CO_2 / n mile). This is done with random forest regression. A value is estimated for around 82% of all ships in the study (see Table B.3 in Annex B). Second, this ratio is multiplied by the distance travelled by the ship to obtain its CO_2 emissions. Emissions by country and ship are then aggregated from daily averages. Figure 4.1 provides an overview of this strategy, starting from the data preparation step.

Figure 4.1. Model workflow



Source: The authors.

A random forest model was selected due to its ability to capture complex and non-linear relationships which are expected of ships' emissions. A ship's speed and fuel consumption, and thus emissions, is conventionally related through the "cubic law" which states that fuel consumption is proportional to the speed of the ship through water raised to the third power (Psaraftis and Lagouvardou, 2023[22]). Nonlinearity of other variables is further discussed and tested empirically by Adland et al. (2020[23]). An additional advantage of random forest regression is that it is an ensemble technique which averages over many models to provide higher levels of accuracy than one model alone. Details of the method may be found in Hastie, Tibshirani and Friedman (2009[24]).

In the initial stages of the work, lasso and ridge regressions were also considered. However, results indicated that they behaved largely as basic regressions as the optimal constraints were near zero. In our exploratory analysis, lasso and ridge recorded accuracy scores that were significantly lower than that observed in the random forest model, and so were discarded in favour of the more complex model.

The emissions efficiency ratio variable *average CO_2 emissions per nautical mile* from the EU-MRV dataset was chosen as the target variable for the random forest model due to having high coverage in the EU-MRV dataset, and the ability to adapt it according to distances travelled, a variable that we can obtain using the AIS. In addition, ratio indicators provide more stability than level indicators, for example total CO_2 emissions. In this case, efficiency as defined by emissions per distance travelled also benefits from ease of interpretation.

The assumptions underpinning this approach are that the technical specifications of a vessel stay constant in the short term, that differences in emissions efficiency during a voyage average out, and that changes in a ship's emissions come principally from the amount of distance covered in a given time period. In this way, we exploit ship characteristics to determine emission patterns, while using real-time movement information of the AIS dataset to calculate timely and frequent estimates of ships' emissions. The random forest model departs from equation (1) above by learning emission patterns such that a ship's full real-time development, including its engines' status and operational mode as provided by AIS, do not need to be explicitly modelled. This approach builds on a similar strategy used by team Blue Carbon from the Wärtsilä Corporation[25] in the 2020 UN Hackathon for AIS data.

Feature identification from data sources (input matrix)

Of the 167 features contained in the full AIS-IHS dataset, only a subset was chosen for use in the random forest model. We used the Pearson, Spearman, and Kendall correlation tests to assess features which could provide the highest predictive power in relation to the target variable (the ship's emissions efficiency ratio) while remaining plausible. The results are reported in Table 4.1. Most variables were found to be moderately or strongly correlated with the target variable. The highest correlation was for Total Kilowatts of Main Engines at 0.70 (Spearman); the lowest correlation for Total Horsepower of Auxiliary Generator at 0.32 (Kendall).

---

[25] https://www.wartsila.com/media/news/23-09-2020-wartsila-ranked-first-in-un-challenge-to-fight-climate-change-with-big-data-2786975.

Table 4.1. Vessel characteristics variables used in the random forest model

Variable	Description (IHS definitions unless calculated)
Breadth Moulded	The maximum breadth (metres) of a ship's hull measured in the middle of the ship to the moulded line of the frame.
Cubic Meter	Calculated as Breadth Moulded * Depth * Length Overall LOA
Deadweight	Deadweight (dwt) - The weight in tonnes (1000 kg) of cargo, stores, fuel, passengers and crew carried by the ship when loaded to her maximum summer draught.
Depth	Moulded depth in metres. Height from the lowest point on the keel to the uppermost continuous deck
Displacement	The total weight of a vessel in metric tonnes, of both ship and its contents, when loaded to its maximum draught.
Draught (average)	Annual average of Draught calculated with AIS data
Fuel Type 2 Capacity	Fuel Capacity (cubic metres) of second lightest type of fuel (mainly residual oil)
Gross Tonnage	Gross Tonnage (GT) - Gross Tonnage is a function of the moulded volume of all enclosed spaces of the ship as per the 1969 International convention on tonnage measurement of ships.
Length Overall LOA	The distance (meters), parallel to the centre line between the extreme points at the bow and stern of a vessel outside of the main hull.
Light Displacement Tonnage	Derived from deadweight and displacement and is used to calculate Steel value at scrapping.
Net Tonnage	Net Tonnage (Nt) - Net Tonnage is derived the gross tonnage (gt) of a vessel as per the 1969 International convention on tonnage measurement of ships, with adjustments for depth, draught and passengers.
Power bhp ihp ship max	Prime mover maximum continuous rating power.
Power kw service	Prime mover service speed power, This is normally determined as the continuous service rating (CSR) about 85% of MCR.
Speed over ground (average)	Annual average of SOG calculated with AIS data
Total Bunker Capacity	Indicates the total bunker capacity (fuel volume) onboard in cubic metres
Total Horsepower of Auxiliary Generators	Total Power of Auxiliary Generators recorded in brake horsepower.
Total Power Of All Engines	Total power in Kilowatts of both prime movers and all auxiliary engines

Source: IHS Markit metadata on the UN Global Platform.

Running the model and checking its performance

After processing, matching with ship characteristics data, and cleaning/editing the input AIS data as described in Data sources and preparation, for each year of data a random-forest regression model is run on the vessel-level dataset that combines data from EU-MRV, IHS, and AIS to predict the target efficiency variable for all vessels for which this information is not available from the EU-MRV. The programming language used for this work is Python. The model is run using 1000 decision trees chosen based on the results of a randomized search cross-validation for parameter selection. Three quarters (75%) of the data is used as the training set, and the remainder (25%) is used as the test set. This method selects the optimal choice of hyperparameters given a defined set of values by evaluating a series of estimation models using randomly selected combinations. The current model result for 2022 relied on EU-MRV data for 2021 due to lag time in data availability.

Cross-validation

The performance of the model is assessed in two ways: firstly, using the test dataset; and secondly, running a k-fold cross-validation with k = 5. In the second procedure, the data is split into 5 separate subsets (folds) and then the model is run 5 times using each group as the test set and the remaining data as the training set. This provides an indication of how the models generalise to unseen vessel data. Our research used the Mean Absolute Error (MAE), which provides the average of the differences between the predicted and reported values in the same unit as the model's target variable (kg of CO_2 per nautical mile). As carbon emission efficiency values can vary widely among different ship types, the Mean Absolute Percentage Error (MAPE) is also examined. This is a measure of the model's prediction accuracy.

Table 4.2 displays the test and cross-validation results found for each year. The R-squared values for the test dataset show that the model explains between 76% (2020) to 87% (2022) of the variation in carbon emissions efficiency values. With the test dataset, the MAPE shows that the predictions generated from the model are on average around 12% away from the reported values and thus have a prediction accuracy of 88%. The cross-validated estimates have a MAPE of around 14%, giving a prediction accuracy of 86%.

It is important to note that the overall emissions results in this study are more accurate than this analysis suggests because 18% of vessels use EU-MRV reported values for emissions efficiency rather than predicted values (the remaining 82% generated by the model). For country-level results accuracy can vary depending on the coverage of vessels in the EU-MRV data. **Table B.4** in **Annex B** shows the percentage of vessels in each OECD country for which the study uses a predicted emissions efficiency value. EU countries use higher proportions of reported values and lower proportions of predicted values compared with non-EU countries, so accuracy will be higher.

Table 4.2 also shows the average MAE scores over the 5 folds from the cross-validation procedure compared with MAE scores from the test set predictions. Except for the 2019 model, the MAE scores found with the test set are close to one standard deviation away from the average MAE score found from the cross-validation (the average MAE minus the standard deviation). This suggests that the model can adequately generalize to unseen ship data. The relatively high variability in the cross-validation MAE scores may be partly explained by the fact that not all folds may contain a representative set of observations across all ship types and for certain ship types such as refrigerated bulk ships and offshore vessels some folds will have zero instances due to the low number of observations.

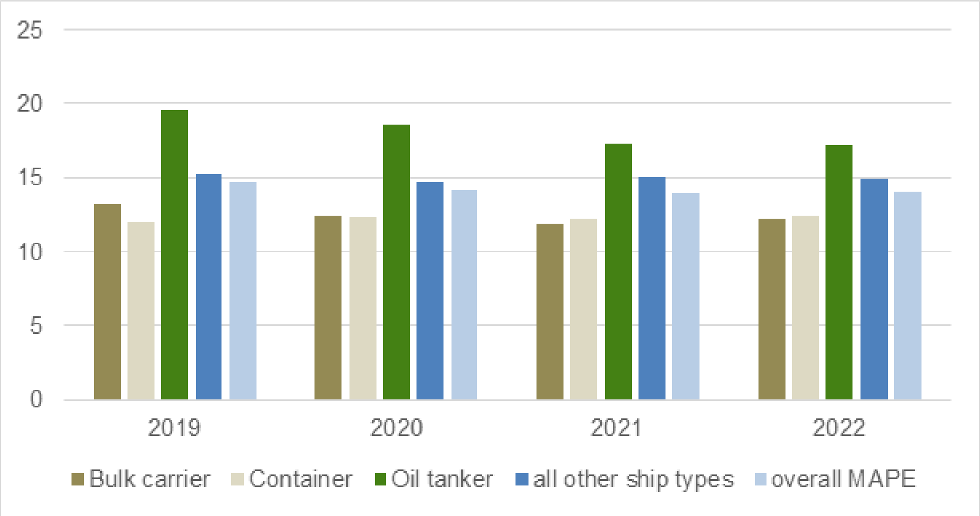
Table 4.2. Test set and k-fold cross-validation metrics

	test dataset			cross-validation		
year	**R-squared**	**MAPE (%)**	**MAE (kg CO2 / n mile)**	**MAPE (%)**	**Average MAE (kg CO2 / n mile)**	**std dev average MAE**
2019	0.86	12.38	43.91	14.73	57.33	8.10
2020	0.76	11.96	46.17	14.18	55.93	9.59
2021	0.83	11.87	43.50	13.92	51.58	7.75
2022	0.87	12.43	43.66	13.99	51.78	6.80

Source: Authors' calculations.

With the cross-validation estimates the MAPE by ship type is also examined. Figure 4.2 displays the MAPE found for the three ship types that contribute the most to total emissions, bulk carriers, containers, and oil tankers, along with the MAPE for all other ship types and the overall MAPE. For bulk carriers and containers, the MAPE is lower than the overall MAPE across all four years. This may be in part due to the fact that these ship types are well-represented in the dataset making up around 44% of all observations across the four years. For oil tankers the MAPE is higher than the overall MAPE over all four years with the highest value of 19.5% seen in 2019. This indicates that the model is less accurate at predicted emissions efficiency for this ship type and future iterations of the model should look into improving this.

Figure 4.2. MAPE by ship type



Source: Authors' calculations.

Permutation importance

The permutation importance for each year is examined to provide insight on which features are contributing the most to the models' predictions. This method measures the decrease in model performance as measured by the R-squared when the values for the given feature are randomly shuffled. Table 4.3 shows the top ten features ranked by permutation importance for each year. Features related to the power of the vessels such as "powerkwmax", "Powerbhpihpshpmax" and "TotalHorsepowerofAuxiliaryGenerators" rank at or near the top across all four years. The variation in the permutation importance rankings across the years may be due to changes in the interactions between features, as neither the data distribution of the technical characteristics nor the variation in relationships with the target variable are expected to change much year to year.

Table 4.3. Top 10 features by permutation importance

2019	2020	2021	2022
TotalHorsepowerofAuxiliaryGenerators (0.31)	Powerkwmax (0.24)	Powerkwmax (0.12)	Powerbhpihpshpmax (0.14)
TotalPowerOfAllEngines (0.21)	Powerbhpihpshpmax (0.2)	Powerbhpihpshpmax (0.1)	Powerkwmax (0.13)
Powerbhpihpshpmax (0.18)	sog_avg (0.08)	GrossTonnage (0.04)	GrossTonnage (0.05)
Powerkwmax (0.15)	TotalHorsepowerofAuxiliaryGenerators (0.07)	TotalPowerOfAllEngines (0.03)	TotalHorsepowerofAuxiliaryGenerators (0.03)
draught_avg (0.09)	Ship type (IMO)_Cruise (0.04)	LightDisplacementTonnage (0.02)	LengthOverallLOA (0.03)
GrossTonnage (0.07)	GrossTonnage (0.04)	NetTonnage (0.02)	draught_avg (0.03)
sog_avg (0.04)	BreadthMoulded (0.02)	TotalHorsepowerofAuxiliaryGenerators (0.01)	TotalPowerOfAllEngines (0.03)
LengthOverallLOA (0.03)	Depth (0.02)	BreadthMoulded (0.01)	NetTonnage (0.02)
Deadweight (0.03)	Deadweight (0.02)	draught_avg (0.01)	sog_avg (0.02)
TotalBunkerCapacity (0.02)	TotalPowerOfAllEngines (0.02)	LengthOverallLOA (0.01)	LightDisplacementTonnage (0.02)

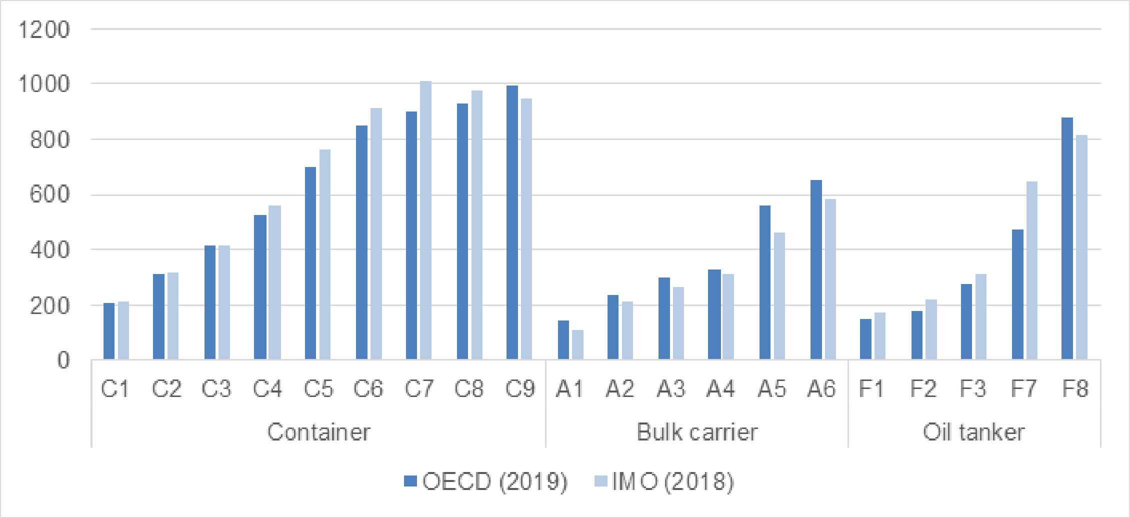
Source: Authors' calculations.

Aggregation by ship type and over time

The next step is to summarise initial vessel-level results by vessel type and size class and by vessel operators. The vessel type and size classes were adopted from the IMO's Fourth GHG Study (IMO, 2020[1]). Ship type and size classes serve multiple purposes in the methodology. First, the vessel type and size classes are used to link vessel level data with information on average energy use, as reported in (IMO, 2020[1]), as additional inputs into the model. Second, the ship type categories are used for analyses of the results. This provides further insight into the nature of activities underpinning the emissions. The average efficiencies across ship type and size classes, calculated based on the results from our model, compare well with the IMO results (for example in Figure 4.3).

Figure 4.3. Estimated efficiencies compared to IMO averages

Tonnes CO_2 per nautical mile



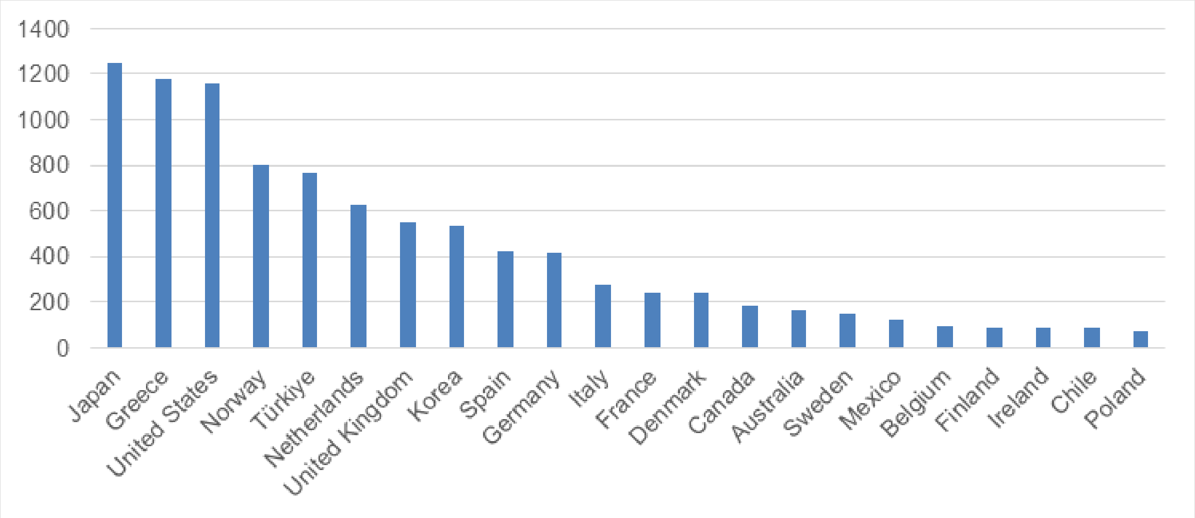
Notes: Estimates for 2018 from the latest IMO GHG study (IMO, 2020[1]) are shown for comparison.
Source: Authors' calculations and (IMO, 2020[1]).

Finally, the vessel-level emissions are then allocated to each ship operator and then from ship operators to their countries of residence. Japan, the United States and Greece are the country of residence for over one thousand operators of vessels in the global fleet (Figure 4.4). Nine of the countries shown in Figure 4.4 have fewer than 200 ship operators. Another 13 countries that are not shown[26] have fewer than 100 ship operators, while the Czech Republic and Hungary have no ship operators in the database.

---

[26] Aligning with Figure 2.2 which does not show countries with less than 1 million tonnes of CO2 emissions in 2022.

Figure 4.4. Number of national operators per OECD country, 2019-22



Notes:

1. Shows the same countries as in Figure 2.2.
2. Number of companies operating in at least one year in the 2019-22 period.

Source: Authors' calculations based on IHS data.

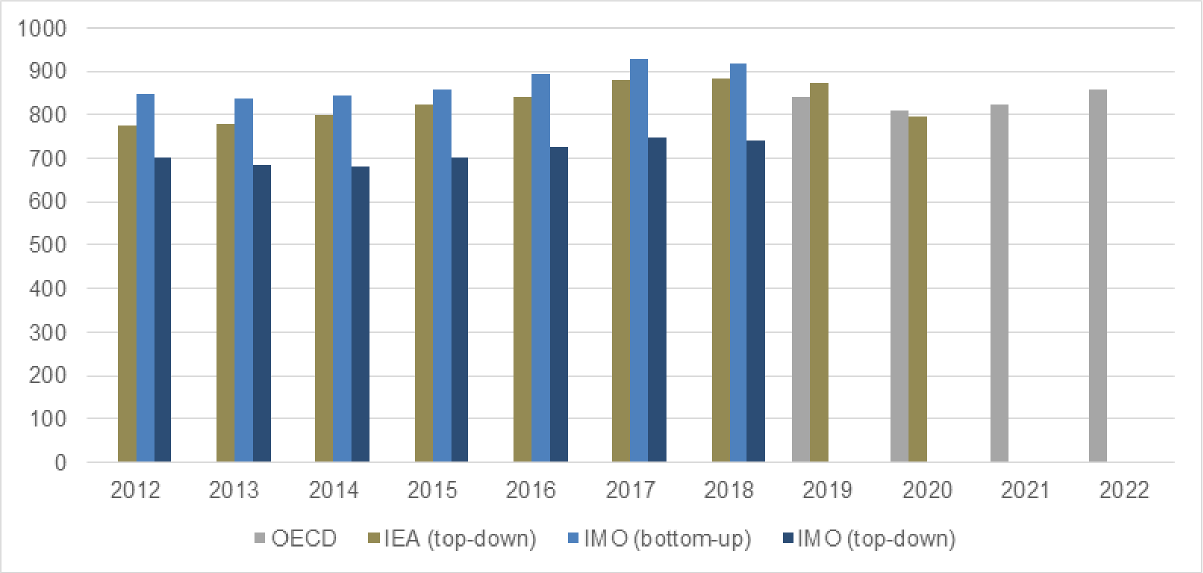
5 Assessing and presenting results

Comparison with other global studies

Comparisons were conducted to assess and improve the quality of the estimates. The IMO's Fourth Greenhouse Gas Study (IMO, 2020[1]) is a useful reference point as it calculated CO_2 emissions for the global shipping fleet using both top-down and traditional activity-based bottom-up methods.[27] Another important source for global emissions data is the International Energy Agency (IEA), which uses a top-down method. The results of global comparisons with IEA and IMO suggest that the estimation model is robust at the global scale (Figure 5.1).

Figure 5.1. Annual estimates for the global fleet compared

Million tonnes of CO_2



Notes: IEA estimates (top-down) are based on all fuel consumed by shipping: both domestic and international. The IMO top-down estimate is voyage-based and covers all international shipping excluding fishing. The IMO bottom-up estimate is vessel-based and uses AIS data.
Sources: OECD database, IEA Detailed CO2 Estimates, (IMO, 2020[1]).

---

[27] For an explanation of top-down and bottom-up methods see Choosing the modelling approach.

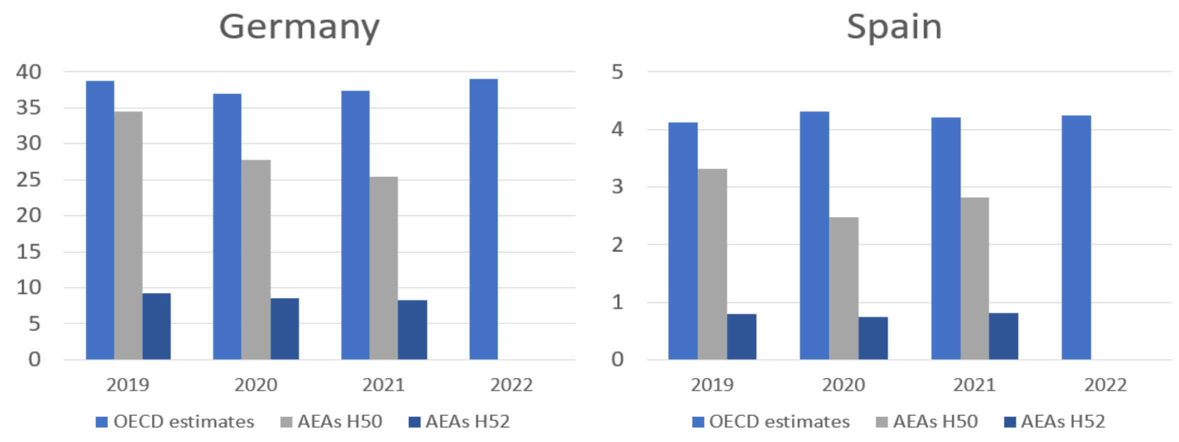
Comparison with official AEAs

National estimates of CO_2 emissions from shipping aligned with the System of Environmental-Economic Accounts (SEEA) Air Emissions Accounts (AEAs) are one of the key outputs from this study. This is because a key objective was to help countries to improve their AEA estimates (see Relevance).

The AEAs follow the residence principles in the national accounts and SEEA, and maritime vessels (or fleet of vessels) are usually classified under the water transport industry, which is Section H Division 50 (H50) in the International Standard Industrial Classification of All Economic Activities (ISIC). However, some ship operators in our database were found to be classified by countries as H52: transport logistics companies (see Allocation of ships to countries). The figures below show some examples comparing our results with those from the official AEAs. Figure 5.2 shows the examples of two countries – Germany and Spain – were our estimates of CO_2 emissions and the official AEAs are of a similar order of magnitude, particularly when taking into account the sum of H50 and H52. In this case, the main utility of the OECD estimates for these countries is that they are available with higher frequency (monthly vs. annual) and are more timely – currently extending to the end of 2022.Our estimates may also be used by such countries to check trends. For example, in Figure 5.2 the official AEAs from both countries suggest a significant reduction in CO_2 emissions from shipping during the COVID-19 pandemic. Our estimates indicate only a small reduction for Germany and none for Spain.

Figure 5.2. Comparison of OECD estimates with official AEAs for Germany and Spain

Million tonnes of CO_2



Notes: AEAs H50 and H52 refer to officially reported CO_2 emissions for water transport services and for transport logistics respectively
Source: OECD database, Eurostat AEAs database.

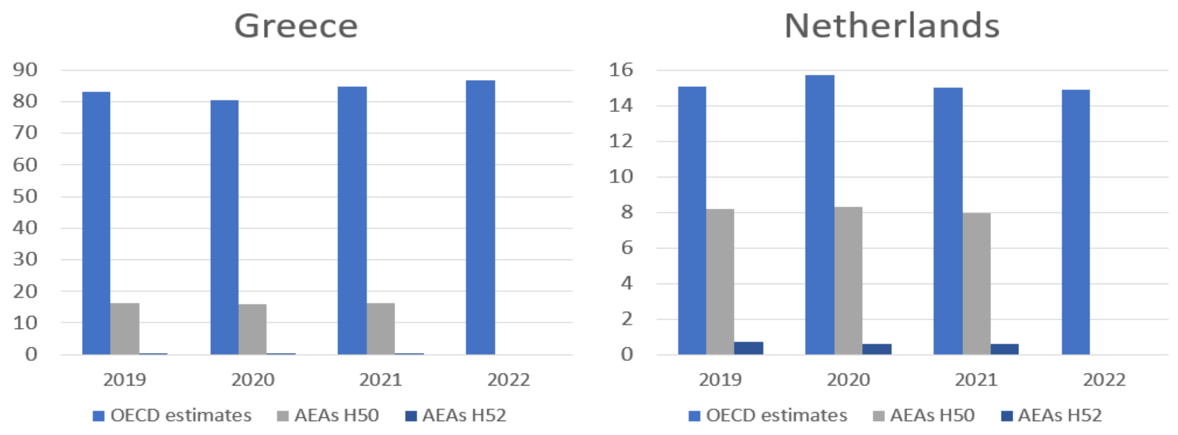
For the countries shown in Figure 5.3 (Greece and the Netherlands), our estimates and the official AEAs are not of a similar order of magnitude: the countries' estimates of CO_2 emissions are much lower than the OECD estimates. This is the case for several other OECD countries. In Figure 5.4 (Italy), the country's estimates are higher than the OECD estimates. This is less common among OECD countries.

In the case of Greece, the gap between our estimates and the official AEAs is particularly high. The OECD estimates of CO_2 emissions from shipping are five times those of the official AEAs for H50 and H52 combined, and higher than Greece's official estimates of CO_2 emissions for all economic activities. Obtaining information for shipping activities outside the national territory is challenging, and this may explain, at least in part, the low estimates for CO2 emissions from shipping in the official AEAs relative to our estimates.

In such cases, it may be helpful for compilers of official AEAs to compare their data with ours. Over the past year, the OECD compared its results to all cases where official AEAs with estimates for H50 are publicly available (currently 35 countries). We then reached out to several member countries, including Greece, to explore the reasons for the differences with a view to improving the estimates on both sides. This programme of work is continuing, and countries are encouraged to participate.

Figure 5.3. Comparison of OECD estimates with official AEAs for Greece and Netherlands

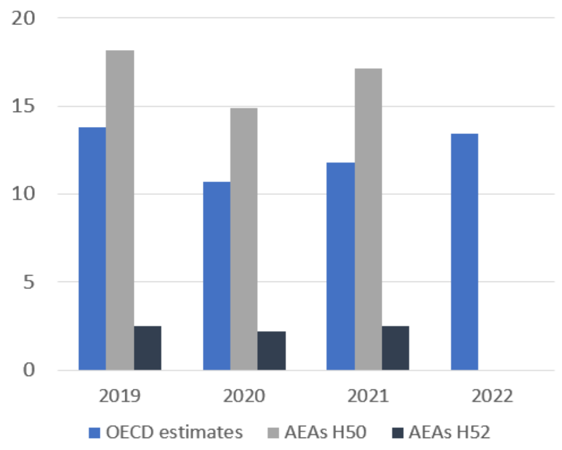
Million tonnes of CO2



Notes: AEAs H50 and H52 refer to officially reported CO_2 emissions for water transport services and for transport logistics respectively.
Source: OECD database, Eurostat AEAs database.

Figure 5.4. Comparison of OECD estimates with official AEAs for Italy

Million tonnes of CO2



Notes: AEAs H50 and H52 refer to officially reported CO_2 emissions for water transport services and for transport logistics respectively.
Source: OECD database, Eurostat AEAs database.

Reasons for the differences

Our work on identifying reasons for the differences between OECD estimates for countries and official AEAs has found a number of possible explanations:

1. Countries may report CO_2 emissions from shipping under other ISIC sectors such as H52 or natural resources extraction, not just H50. Where possible, we have also examined H52 (as in Figure 5.2, Figure 5.3 and Figure 5.4) but this is less feasible for other ISIC divisions if we do not have information about which divisions shipping operators might be classified to in particular cases.
2. One or more of the large resident shipping operators (according to the information in the AIS database) may not be considered resident by the country in question and may appear in another county's national accounts. With multi-national enterprises (MNEs) it may be difficult to determine whether the whole MNE should be allocated to one country, based on the location of the company's headquarters, or whether some part(s) should be considered resident in other countries. The OECD can play a role in identifying such cases and (if appropriate) agreeing on a change with the countries in question; or we may adjust our own estimates to reflect existing agreements between countries.
3. It may be difficult for national statistical offices to obtain information on emissions taking place abroad. If an activity takes place outside of the national territory it is not included in United Nations Framework Convention on Climate Change (UNFCCC) emissions inventories, which are compiled on a territory basis. Here, the OECD estimates may be particularly helpful in providing the extra information required to bridge from the territory to the residence basis (see Development of bridging items).
4. If ship operators that are domiciled in one country register their ships in other countries, this might lead to underestimates of emissions in the official AEAs. The flag of registration would not be used by countries to indicate the residence of the ship operator when compiling the official AEAs. Nevertheless, differences between country of domicile and country of registration may make it harder in practice to obtain the information needed for compiling the estimates.

Some of these discrepancies are the subject of ongoing discussions between those involved in this study and specific countries. These discussions have already resolved some of the issues. For example, discussions with Statistics Norway threw light on key pieces of information that helped both the OECD and Norway to make improvements. Further collaboration with countries is expected to lead towards increased convergence between the OECD's estimates and national reporting in the AEAs. This is one of the reasons why the new database for CO2 emissions from maritime transport has been released as "experimental".

Protecting confidentiality

Due to the relatively small number of shipping operators in most countries (see Aggregation by ship type and over time), we have put in place "disclosure control" measures to protect confidentiality. This means that results for some countries cannot be shown separately, although they are included in the OECD and World totals on the database. We have also decided not to publish breakdowns by country and ship type: ship type breakdowns are published for the OECD and World totals only.

Concentration (or diversity of sources) of the emissions of each country is one element that needs to be taken into account in "disclosure control" processes. Concentration (or diversity) can be analysed using the Herfindahl–Hirschman index (HHI):

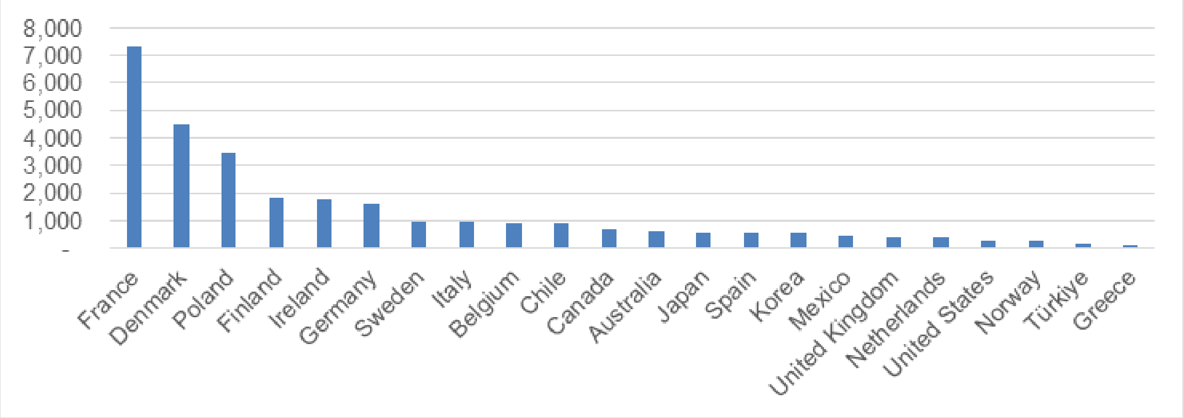
$$HHI_i = S_1^2 + S_2^2 + \cdots S_n^2 \qquad (2)$$

for all ship operators *n* in country *i*, where S = ship operator's percentage share of total emissions (expressed as a whole number, not a decimal)

The HHI is a measure of diversity in a market often used in competition analysis. A high HHI score means that a small number of ship operators account for most of the country's CO_2 emissions from shipping. A score of 10,000 would be equivalent to a perfect concentration, or only one company accounting for all of the emissions.

Figure 5.5 shows that, of the countries with more than one million tonnes of CO_2 emissions from shipping, France has the highest HHI (highest concentration) and Greece has the lowest HHI (greatest diversity). Greece also the highest volume of CO_2 emissions in the OECD (Figure 2.2) and the second highest number of operators, second only to Japan (Figure 4.4).

Figure 5.5. Emissions concentration: HHI in OECD countries



Notes: Shows the same countries as in Figure 2.2 and Figure 4.4.
Sources: Authors' calculations.

6 Conclusion and future work

Conclusion

This paper has presented the work done by the OECD to produce a new database for CO2 emissions from maritime transport with global coverage that is more timely than previous sources, currently providing monthly, quarterly and annual estimates up to the end of 2022. The new estimates are also highly relevant and produced using a method that has been developed to balance efficient use of resources and high levels of accuracy. The robustness of the results has been demonstrated through careful testing of the performance of the model as well as by comparing the results with existing sources.

The estimates on our database are currently labelled "experimental" because of three areas where further work is planned:

1. Improving the model as we learn from experience.
2. Dealing with missing information in the Automated Identification System (AIS) dataset.
3. Improving the alignment of our estimates and national estimates of CO_2 emissions from shipping, in collaboration with national statistical offices.

In addition, there is one area where more work would be helpful to make the OECD estimates more useful. This relates to developing bridging items to link the official Air Emissions Accounts (AEAs) and the greenhouse gas emissions inventories reported by countries to the United Nations Framework Convention on Climate Change (UNFCCC).

These four areas of work are briefly outlined below.

Future work

Improving the model

A number of improvements or extensions to the results and related calculations were identified during the course of the project. For instance, it may be possible to improve the performance of the prediction model for certain ship types. Oil tankers, for example, performed less well than other ship types in the estimation results cross-validation and further investigation into the global fleet of oil tankers and their activity, according to the data, could reveal potential improvements.

Some variables available to the study were not used in our model but could be considered as additional useful elements in future. One example is the navigational status of a vessel, relating the different phases of a voyage. Detailed analysis by the IMO (2020[1]) separates the phases of voyage and considers all engine and fuel factors involved in each phase, while this study only focuses on the distances travelled, draught and speed-over-ground. Moreover, the use of averages for speed-over-ground information in a compromise which masks periods of intense acceleration or deceleration.

Also, while they are at berth ships often turn off their main engine while keeping auxiliary engines on to maintain essential services. Regulations at certain ports also require certain fuels to be used to comply with emissions regulations. Our model does not account for potential emissions-reduction activities and measures near shore. As the research team continue gather knowledge and expertise on the datasets and the behaviour for the vessels, it is expected that improvements to the scope of factors incorporated into the model could eventually lead to improvements in accuracy.

The modelling strategy could also be improved by analysis of changes in the efficiency coefficient. Currently, the efficiency coefficient, *CO_2 per nautical mile*, is assumed to be constant over the year. In reality, it might decline or improve over time as vessels age, undergo maintenance or are equipped with emissions-abatement technology.

Dealing with missing information

There is some evidence that the model may be underestimating emissions because of missing information in the AIS dataset. Specifically, in the dataset there are cases where it is not possible to complete the matching of vessel information and AIS signal, either because an AIS signal could not be found in the data or because the identifier from the AIS data does not match with the available registry of ship characteristics information. Also, some vessels may intentionally disable AIS transmissions, for instance military vessels and vessels undertaking illegal activities.

The current modelling workflow includes some basic steps to maximise the coverage of the global fleet included in the estimates, but further experience with the datasets and additional investigation into unmatched vessel identifiers (for example by incorporating other data sources) could help to expand the overall coverage and thus total volume of calculated emissions. One possible method to match these missing ships is to use the Maritime Mobile Service Identity (MMSI) number.[28] Verifying the ship register with each country's maritime authority could also help to identify missing or incorrectly categorised ships in the AIS and IHS datasets. If this is not possible, statistical techniques could be used to "impute" a response so that the missing information does not affect the aggregate estimates.

Aligning with the AEAs

As discussed in the previous section: Comparison with official AEAs, there are some differences between the OECD country-level estimates and CO_2 emissions from the shipping industry reported in official AEAs. Work will continue in these areas, in particular to resolve:

- Differences in estimates due to different shipping operators and associated vessels being included in estimates for a particular country. For example, official sources may follow a different approach from the AIS database for multi-national enterprises. In some cases, coordination between countries may be desirable, and the OECD can play a facilitating role.
- In what part of the International Standard Industrial Classification of All Economic Activities (ISIC) countries classify shipping operators. While H50 (water transport) is the primary classification for shipping, other ISIC divisions may also be used in some countries.
- Improving information available to national statistical offices (NSOs) on the activities and associated CO_2 emissions of their companies outside of the national territory (needed for compiling the AEA estimates on a residence basis).

Conversations with official statistics compilers are therefore an important part of the OECD's future work to maximise, as much as possible, alignment of our results with the official AEAs. This is envisaged as taking place both through improving mutual understanding about why there are differences in the estimates

---

[28] However, as the MMSI number can change over the life span of a given vessel, this would require access to archives for using the correct vintage of the ship register database to make accurate matches.

and though assisting NSOs in gaining more information about CO_2 emissions by their companies outside the national territory.

Development of bridging items

Greenhouse gas (GHG) emissions inventories, including from domestic and international transport (as memo items) are produced by many countries for reporting to the UNFCCC, although they are only required annually for UNFCCC Annex I countries.[29]

The Intergovernmental Panel on Climate Change (IPCC) provides guidance for countries on reporting GHG emissions inventories to the UNFCCC[30]. The relevant reporting item in the IPCC guidance is item *1.A.3.d Water-borne navigation and international waterborne navigation (marine bunkers)*. The scope of the estimates from the study presented in this paper aligns well with this item.[31]

However, the results in this study are on a residence basis, reflecting the approach of the national accounts and the System of Environmental-Economic Accounts (SEEA), see Allocation of ships to countries, whereas the IPCC guidance requires use of the territory basis for reporting inventories data, reflecting activities taking place within the borders of each country. For shipping, these territory-based estimates are compiled by countries using the location of origin for each domestic and international voyage. In practice, this is based on information about the purchase and delivery of fuel.

Countries frequently use their territory-based emissions inventories as part of the process for compiling their AEAs. The GHG emissions inventories reported by countries are linked to the AEAs using “bridging items” in the AEAs. These are a useful component of the AEAs because they allow analysis of the emissions data on both residence and territory bases using the same table.

As noted above, it would be helpful for the “bridging” process if countries could obtain better information on the activities of resident ship operators outside their territories. The OECD is currently exploring whether the database developed for this study could be used to improve the bridging items in the AEAs by linking our results to information on ships’ fuelling behaviour.

In the AIS database there is no data on ships’ port visits or fuelling behaviour; but work under way at the OECD may help to address this gap by building models that indirectly calculate (or “infer”) when and where vessels call into port. This information could then be used to estimate bridging items. However, any bridging analysis would also need to consider whether the refuelling behaviour of ships is accurately reflected by the port visits. Some assumptions on refuelling might be needed.

---

[29] All OECD countries except Chile, Colombia, Costa Rica, Israel, Korea and Mexico are UNFCCC Annex I countries. Annex I countries report inventories data annually in line with 2006 IPCC Guidelines for National Greenhouse Gas Inventories (IPCC, 2006). Non-Annex I counties are not required to report annually but may still do so.

[30] 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

[31] The 2006 IPCC Guidelines and the AEAs classify activity of fishing vessels separately from transport services. Emissions from fishing vessels are excluded because fishing activities are compiled under ISIC Section A: Agriculture, Forestry and Fishing rather than Section H50: Water transport. Military vessels are also classified separately.

Bibliography

Adland, R., P. Cariou and F. Wolff (2020), “Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet”, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 140, p. 101972, https://doi.org/10.1016/j.tre.2020.101972. [23]

Arslanalp, S., M. Marini and P. Tumbarello (2019), “Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time”, *IMF Working Papers*, Vol. 2019/275, https://doi.org/10.5089/9781513521121.001.A001. [47]

Arslanalp, S., M. Marini and P. Tumbarello (2019), *Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time IMF Working Paper Statistics Department Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time*. [33]

Banerjee, O. et al. (2021), *Gross domestic product alone provides misleading policy guidance for post-conflict land use trajectories in Colombia*, Elsevier B.V. [36]

Chen, D. et al. (2016), *Estimating ship emissions based on AIS data for port of Tianjin, China*, Elsevier Ltd. [37]

Clarke, D. et al. (2022), “CO2 Emissions from air transport: A near-real-time global database for policy analysis”, *OECD Statistics Working Papers*, https://doi.org/10.1787/ecc9f16b-en. [51]

Coello, J. et al. (2015), “An AIS-based approach to calculate atmospheric emissions from the UK fishing fleet”, *Atmospheric Environment*, Vol. 114, pp. 1-7, https://doi.org/10.1016/J.ATMOSENV.2015.05.011. [15]

Eide, M. et al. (2009), *Cost-effectiveness assessment of CO2 reducing measures in shipping*. [40]

Eide, M. et al. (2011), *Future cost scenarios for reduction of ship CO2 emissions*. [43]

Endresen, Ø. (2003), “Emission from international sea transportation and environmental impact”, *Journal of Geophysical Research*, Vol. 108/D17, https://doi.org/10.1029/2002jd002898. [18]

European Union (2015), *EU strategy to reduce CO2 emissions from shipping*. [10]

European Union (2015), “Regulation (EU) 2015/757 of the European Parliament and of the Council of 29 April 2015”, *Official Journal of the European Union*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R0757 (accessed on 10 July 2023). [20]

Eurostat (2020), *Handbook on the Compilation of Statistics on Sea and Air Transport in National Accounts and Balance of Payments - 2020 Edition*, European Union, https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-20-007. [2]

Eurostat (2020), *Interim technical report concerning the conditions for using the data, the methodology and the procedures*. [46]

Eurostat, U. (ed.) (2010), *Illustrated Glossary of Transport Statistics - 4th Edition*, European Commission, https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-10-028. [50]

Eyring, V. et al. (2010), "Transport impacts on atmosphere and climate: Shipping", *Atmospheric Environment*, Vol. 44/37, pp. 4735-4771, https://doi.org/10.1016/J.ATMOSENV.2009.04.059. [21]

Fiorini, M., A. Capata and D. Bloisi (2016), "AIS Data Visualization for Maritime Spatial Planning (MSP)", *International Journal of e-Navigation and Maritime Economy*, Vol. 5, pp. 45-60, https://doi.org/10.1016/J.ENAVI.2016.12.004. [49]

Food and Agriculture Organization of the United Nations et al. (n.d.), *System of environmental-economic accounting 2012 : applications and extensions.*. [32]

Franzese, P. et al. (2017), *Natural capital accounting in marine protected areas: The case of the Islands of Ventotene and S. Stefano (Central Italy)*, Elsevier B.V. [34]

Furukawa, K. and R. Hisano (2022), "A Nowcasting Model of Exports Using Maritime Big Data", Bank of Japan, https://www.boj.or.jp/en/research/wps_rev/wps_2022/wp22e19.htm (accessed on 10 July 2023). [48]

Goldsworthy, L. and B. Goldsworthy (2015), "Modelling of ship engine exhaust emissions in ports and extensive coastal waters based on terrestrial AIS data – An Australian case study", *Environmental Modelling & Software*, Vol. 63, pp. 45-60, https://doi.org/10.1016/J.ENVSOFT.2014.09.009. [8]

Hastie, T., R. Tibshirani and J. Friedman (2009), *The Elements of Statistical Learning*, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-84858-7. [24]

Hoffmann, P., M. Eide and Ø. EndresenØyvind (2012), *Effect of proposed CO2 emission reduction scenarios on capital expenditure*. [38]

IMO (2020), *Fourth IMO GHG Study*, https://www.imo.org/en/ourwork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx. [1]

IMO (2015), *Third IMO GHG Study*, https://www.imo.org/en/OurWork/Environment/Pages/IMO-GHG-studies.aspx. [53]

Johansson, L., J. Jalkanen and J. Kukkonen (2017), "Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution", *Atmospheric Environment*, Vol. 167, pp. 403-415, https://doi.org/10.1016/J.ATMOSENV.2017.08.042. [14]

Jolliffe, J., C. Jolly and B. Stevens (2021), "Blueprint for improved measurement of the international ocean economy: An exploration of satellite accounting for ocean economic activity"*, OECD Science, Technology and Industry Working Papers*, No. 2021/04, OECD Publishing, Paris, https://doi.org/10.1787/aff5375b-en. [26]

Karney, C. (2013), "Algorithms for geodesics", *J Geod*, Vol. 87, pp. 43-55, https://doi.org/10.1007/s00190-012-0578-z. [7]

Leong, S. et al. (2015), “Estimation of CO2 emission from marine traffic in Singapore Straits using Automatic Identification Systems data”, *Environmental Science and Information Application Technology - Proceedings of the 2014 5th International Conference on Environmental Science and Information Application Technology, ESIAT 2014*, pp. 177-183, https://doi.org/10.1201/B18559-30/ESTIMATION-CO2-EMISSION-MARINE-TRAFFIC-SINGAPORE-STRAITS-USING-AUTOMATIC-IDENTIFICATION-SYSTEMS-DATA-SEE-CHUAN-LEONG-CAROL-ANNE-HARGREAVES-PRATEEK-SINGHAL-JUN-YUAN. [5]

Li, C. et al. (2016), “An AIS-based high-resolution ship emission inventory and its uncertainty in Pearl River Delta region, China”, *Science of The Total Environment*, Vol. 573, pp. 1-10, https://doi.org/10.1016/J.SCITOTENV.2016.07.219. [4]

Maes, M. et al. (2020), “Accounting for natural capital has cross-cutting relevance for UK public sector decision-making”, *Ecosystem Services*, Vol. 44, p. 101127, https://doi.org/10.1016/j.ecoser.2020.101127. [25]

Mao, X. et al. (2022), “Exporting emissions: marine fuel sales at the port of Singapore”, The International Council on Clean Transportation, https://theicct.org/wp-content/uploads/2022/07/Singapore-exporting_FINAL.pdf (accessed on 10 July 2023). [19]

Mathieu, E. et al. (2020), *Coronavirus Pandemic (COVID-19)*, https://ourworldindata.org/coronavirus. [52]

McGrath, L. and S. Hynes (2020), *Approaches to accounting for our natural capital: Applications across Ireland*, Royal Irish Academy. [41]

McGrath, L. and S. Hynes (2020), *Approaches to accounting for our natural capital: Applications across Ireland*, Royal Irish Academy. [42]

Merk, O. (2014), *Shipping Emissions in Ports*, International Transport Forum, http://www.internationaltransportforum.org/jtrc/DiscussionPapers/jtrcpapers.html. [44]

Ng, S. et al. (2013), “Policy change driven by an AIS-assisted marine emission inventory in Hong Kong and the Pearl River Delta”, *Atmospheric Environment*, Vol. 76, pp. 102-112, https://doi.org/10.1016/J.ATMOSENV.2012.07.070. [16]

OECD (2017), *Analysis of selected measures promoting the construction and operation of greener ships*, OECD. [11]

Olesen, H. et al. (2010), *Ship emissions and air pollution in Denmark*, Danish Ministry of the Environment, https://www2.mst.dk/udgiv/publikationer/2009/978-87-92548-77-1/pdf/978-87-92548-78-8.pdf. [13]

Olivier, J. and J. Peters (1999), *International marine and aviation bunker fuel: trends, ranking of countries and comparison with national CO2 emissions*, National Institute for Public Health and the Environment, Bilthoven, https://www.pbl.nl/sites/default/files/downloads/773301002.pdf (accessed on 10 July 2023). [17]

Olmer, N. et al. (2017), *Greenhouse Gas Emission From Global Shipping, 2013-2015*, https://theicct.org/sites/default/files/publications/Global-shipping-GHG-emissions-2013-2015_ICCT-Report_17102017_vF.pdf. [12]

Psaraftis, H. and S. Lagouvardou (2023), "Ship speed vs power or fuel consumption: Are laws of physics still valid? Regression analysis pitfalls and misguided policy implications", *Cleaner Logistics and Supply Chain*, Vol. 7, p. 100111, https://doi.org/10.1016/j.clscn.2023.100111. [22]

Rypdal, K. et al. (n.d.), *Category Title NFR 1.A*. [31]

Smith Statistics Canada Ottawa, R. (n.d.), *SESSION 8.2: Challenge in Data Needs-Scientific Research 54*, http://www.verypdf.com. [29]

Smith, R. (2007), *Development of the SEEA 2003 and its implementation*. [39]

Smith, R. (n.d.), *Users and Uses of Environmental Accounts A Review of Select Developed Countries*, http://www.wavespartnership.org. [30]

Smith, T. et al. (2013), *Assessment of Shipping's Efficiency Using Satellite AIS data*, UCL Energy Institute, https://theicct.org/sites/default/files/publications/UCL_ship_efficiency_forICCT_2013.pdf (accessed on 9 June 2023). [9]

Svanberg, M. et al. (2019), "AIS in maritime research", *Marine Policy*, Vol. 106, p. 103520, https://doi.org/10.1016/J.MARPOL.2019.103520. [6]

UNECE (2015), *Guide to Measuring Global Production*, United Nations Economic Commission for Europe (UNECE), https://unece.org/statistics/publications/guide-measuring-global-production-2015. [3]

UNFCCC (2013), *Decision 24/CP.19 Revision of the UNFCCC reporting guidelines on annual inventories for Parties included in Annex I to the Convention*, https://unfccc.int/resource/docs/2013/cop19/eng/10a03.pdf#page=2. [27]

Wang, C., J. Corbett and J. Firestone (2007), *Modeling energy use and emissions from North American shipping: Application of the ship traffic, energy, and environment model*. [35]

Wang, X. et al. (2021), *Trade-linked shipping CO2 emissions*, Nature Climate Change, https://doi.org/10.1038/s41558-021-01176-6. [45]

Zhang, Y., J. Chan and A. Lau (n.d.), *Estimation of Marine Emissions Inventory from all AIS-installed Vessels*. [28]

Annex A.

Table A.1. Vessel Population Summary Statistics: OECD estimates 2019-2022 compared with IMO figures for 2018

IMO Type	Count % (IMO, 2018)	Count % (2019-2022)	DWT % (IMO, 2018)	DWT % (2019-2022)	IMO, 2018 - count of ships	OECD estimates – count of ships			
						2019	2020	2021	2022
Bulk Carrier	9.8	17.8	41.5	43.3	11,672	11,214	11,451	11,779	12,040
Oil Tanker	6.8	9.5	25.1	24.9	8,177	6,164	6,181	6,245	6,191
Container	4.3	8.1	13.4	14.1	5,182	5,172	5,198	5,283	5,430
Chemical Tanker	4.6	7.7	5.6	5.9	5,506	4,855	5,023	5,138	5,210
General Cargo	12.5	12.9	4.2	3.9	14,994	8,349	8,348	8,419	8,572
Offshore	6.3	8.4	3.8	1.9	7,555	5,405	5,398	5,468	5,610
Liquefied Gas Tanker	1.6	3.1	3.3	3.7	1,953	1,893	1,964	2,059	2,120
Vehicle	0.7	1.3	0.7	0.7	828	848	828	809	802
Service - other	5.2	0.2	0.6	0.1	6,180	148	147	160	164
Ro-Ro	1.7	1.6	0.3	0.4	2,002	1,011	1,021	1,043	1,065
Service - tug	16.9	9.0	0.3	0.0	20,251	5,610	5,819	6,015	6,055
Miscellaneous - fishing	20.0	6.6	0.2	0.1	23,911	3,898	4,312	4,437	4,529
refrigerated Bulk	0.7	0.9	0.2	0.2	897	621	596	594	566
Ferry-RoPax	2.6	3.1	0.2	0.2	3,148	1,956	1,992	2,070	2,090
Miscellaneous - other	0.5	5.0	0.2	0.4	645	3,156	3,207	3,311	3,395
Cruise	0.5	0.7	0.1	0.1	612	494	477	483	502
Other liquids tankers	0.1	0.1	0.0	0.0	179	45	44	44	43
Ferry-pax only	2.9	1.6	0.0	0.0	3,459	1,042	1,024	1,024	1,100
Yacht	2.1	2.5	0.0	0.0	2,477	1,552	1,573	1,687	1,782

Note: Count % is the proportion of total ships that is in this ship type category. DWT % is the proportion of DWT of all ships that is in this ship type category. 2019-22 is calculated as: sum of all 4 years for the ship type divided by sum of all 4 years for all ship types.
Source: Authors' calculations and (IMO, 2020[1]).

Annex B.

Table B.1. AIS vessel coverage

% of total vessels available in AIS data

	AIS data extraction	Final results
2019	94.24	93.37
2020	93.52	92.59
2021	94.50	93.70
2022	93.20	92.33
average	93.86	93.00

Note: The column “AIS data extraction” includes AIS vessels with a valid or potentially valid IMO number. The column “Final results” contains the set of vessels after data matching and outlier removals.
Source: Authors’ calculations.

Table B.2. Imputation of ship characteristics

% of observations requiring an imputed value

Variable	2019	2020	2021	2022
BreadthMoulded	1.65	1.63	1.58	1.57
Deadweight	8.25	8.56	8.73	8.99
Depth	0.89	0.94	0.97	1.01
Displacement	37.01	37.60	38.28	39.47
Draught	3.44	3.71	3.77	4.03
FuelType2Capacity	53.42	54.22	54.97	55.87
GrossTonnage	0.03	0.03	0.03	0.03
LengthOverallLOA	1.72	1.73	1.66	1.62
LightDisplacementTonnage	37.26	38.09	38.91	40.30
NetTonnage	4.12	4.39	4.56	4.87
Powerbhpihpshpmax	1.13	1.28	1.35	1.52
Powerkwmax	1.13	1.28	1.35	1.52
Speed	9.05	9.65	9.96	10.74
Speedservice	12.58	13.24	13.63	14.33
TotalBunkerCapacity	49.15	49.34	49.42	49.86
TotalHorsepowerofAuxiliaryGenerators	18.85	19.40	19.95	21.24
TotalPowerOfAllEngines	0.95	1.06	1.14	1.29

Source: Authors’ calculations.

Table B.3. Use of a predicted emissions efficiency ratio, by ship type

% of results using prediction in each ship type category

	2019	2020	2021	2022
Bulk carrier	69.38	73.49	70.37	71.01
Chemical tanker	63.56	66.22	64.79	65.36
Container	64.81	65.78	65.61	66.56
Cruise	62.75	79.87	79.92	80.68
Ferry-pax only	100.00	100.00	100.00	100.00
Ferry-RoPax	80.42	82.28	82.66	82.82
General cargo	84.17	85.25	84.49	84.78
Liquefied gas tanker	69.15	70.01	70.91	71.84
Miscellaneous - fishing	100.00	100.00	100.00	100.00
Miscellaneous - other	99.84	99.88	99.88	99.88
Offshore	99.70	99.81	99.78	99.80
Oil tanker	75.02	77.32	77.89	77.74
Other liquids tankers	75.56	75.00	70.45	69.77
Refrigerated bulk	78.42	77.35	77.10	75.97
Ro-Ro	72.70	73.36	75.36	75.87
Service - other	100.00	100.00	100.00	100.00
Service - tug	100.00	100.00	100.00	100.00
Vehicle	39.50	41.91	39.80	39.28
Yacht	100.00	100.00	100.00	100.00
All ships combined	**81.10**	**82.82**	**82.19**	**82.52**

Note: This table shows the proportion of each ship type category where random forest regression is used to predict the emissions efficiency ratio, defined as *average* CO_2 *emissions per nautical mile* (kg CO_2 / n mile).
Source: Authors' calculations.

Table B.4. Use of a predicted emissions efficiency ratio, by country (OECD countries)

% of results using prediction in each country

	2019	2020	2021	2022
Australia	98.73	98.76	98.79	98.82
Belgium	56.90	61.13	62.63	63.68
Canada	84.92	84.68	84.30	84.68
Chile	95.68	97.10	95.76	95.80
Colombia	98.21	96.67	96.72	96.72
Costa Rica	71.43	73.33	73.33	66.67
Denmark	55.09	56.69	56.79	57.22
Estonia	82.35	82.35	81.63	82.12
Finland	65.69	67.36	69.11	70.08
France	65.29	64.40	64.55	67.02
Germany	58.10	59.97	59.00	59.69
Greece	58.73	64.05	60.77	61.77
Iceland	87.61	85.96	88.33	87.04
Ireland	64.16	62.39	62.55	65.61
Israel	67.14	72.46	78.26	78.26
Italy	66.92	68.60	68.16	69.00
Japan	88.09	88.05	86.78	87.49
Korea	86.64	88.62	87.96	88.07
Latvia	89.23	92.48	89.39	87.16
Lithuania	94.94	94.81	95.00	94.67
Luxembourg	98.02	94.95	97.09	97.35
Mexico	99.71	100.00	99.72	99.70
Netherlands	72.70	74.46	73.66	74.33
New Zealand	97.14	97.22	98.63	98.70
Norway	77.06	80.94	77.67	78.60
Poland	67.02	67.02	65.61	66.15
Portugal	85.11	85.11	85.86	86.41
Slovakia	100.00	100.00	100.00	100.00
Slovenia	100.00	90.00	90.00	88.89
Spain	88.57	88.52	89.08	89.27
Sweden	66.67	69.19	66.32	68.81
Switzerland	41.26	43.99	45.24	45.87
Türkiye	74.19	76.49	72.87	73.47
United Kingdom	73.51	76.87	76.60	77.10
United States	89.35	90.16	89.82	89.94
All EU countries	**64.05**	**66.54**	**65.31**	**66.20**
All OECD countries	**73.87**	**76.00**	**74.85**	**75.48**
All countries (world)	**81.10**	**82.82**	**82.19**	**82.52**

Note: The proportion of vessels for which emissions efficiency uses EU-MRV reported values is 100 *minus* the figure shown in the table.
Source: Authors' calculations.