Transportation sector: Domestic and International Shipping Emissions

Brett Mayes[1,3], Mark Powell[1,3], Dan Knights[2,3], Max Schofield[1,3] and Ted Mackereth[1,3]
1) OceanMind, 2) University of Minnesota, 3) Climate TRACE

1 Introduction

Greenhouse gas (GHG) emissions from global shipping are estimated to be just below 3% of total anthropogenic GHG emissions (IMO 2021). Despite the significance of this level of emissions, international shipping, along with international aviation, is exempt from the legally binding Paris Climate Agreement (UNFCCC 2016). This exemption means that other authorities must be responsible for decarbonizing shipping. The International Maritime Organization (IMO) is the United Nations (UN) body with authority to regulate international shipping. Positive progress has been made by the IMO to encourage decarbonisation, with adoption of their revised decarbonation strategy in July 2023 (IMO 2023). However, there is contention as to whether this strategy will be sufficient to achieve the level of decarbonisation required to meet the Paris agreement temperature goals, with projections that emissions from shipping could account for 17% of global emissions by 2050 (Halim et al. 2018).

Some shipping companies are demonstrating leadership and taking steps towards decarbonization, despite the lack of any binding requirement to reduce emissions. Voluntary actions include disclosure of GHG and non-GHG emissions, decarbonization pledges, and development of tools and approaches to support decarbonization. But not all shipping companies are proactive, and binding requirements may be necessary to achieve shipping emissions reduction consistent with the Paris Accord. The current lack of transparency regarding shipping emissions hinders efforts by those outside the industry to promote emissions reduction for shipping.

Transparency regarding all emissions from individual vessels and company fleets could provide opportunities for investors and customers of shipping companies, and others, to support leading companies that are voluntarily reducing emissions. The Ship it Zero campaign is an example of efforts to support low emissions shipping (Ship it Zero 2021). Such mechanisms have been effective in providing tangible benefits for industry leaders in other industries that implement socially beneficial policies that may increase costs in the short term (Virdin et al. 2022).

There is no public disclosure of emissions from individual vessels for most of the world's commercial shipping fleet. Studies of emissions from global shipping include comprehensive estimates of total shipping emissions commissioned by the IMO, the UN body that regulates international shipping (IMO 2015; 2021), and studies by others (Olmer et al. 2017). In addition, the European Union has created a monitoring, reporting and verification program and publishes ship emissions data annually for ships that call on EU ports (EMSA n.d.). Results from these studies include global emissions totals for all ships and by shipping subsector (e.g., container ships, oil tankers, vehicle carriers, etc.). These studies also included emissions by operational phase (cruising speed, berth, anchor, manoeuvring) and analysis of factors that can influence emissions such as vessel speed. None of these studies provide emissions estimates for individual vessels.

The best current publicly accessible source of emissions from individual vessels is the European Union Monitoring, Reporting, and Validation (EU MRV) database with data on emissions from individual vessels that call on EU ports, which claims to include about 25% of the world's commercial shipping fleet each year. It is the only large-scale published record of individual vessel emissions, but gaps remain for the remaining 75% of vessels that do not call on EU ports. The MRV data is part of the EU effort to encourage decarbonization by monitoring emissions and potentially establishing a cost for carbon emissions from shipping. To date, the EU MRV has published 6 years of emissions data and is moving towards imposing a price on carbon for ship traffic that departs or arrives from EU ports and loads or offloads cargo or passengers, providing new insight into emissions from shipping.

Information on shipping emissions has also been produced to inform decarbonization and investment decisions. The results of these emissions evaluations have been presented as carbon efficiency levels aggregated at the level of sectors or transit routes (Clean Cargo Working Group (CCWG) 2017; Poseidon Principles 2021; RightShip n.d.). However, these studies also lack ship-specific emissions estimates and have not driven accountability of shipping for GHG emissions. As an alternative, an approach has been developed by OceanMind that addresses the current gap in monitoring shipping emissions.

Recent advances in machine learning (ML), increased access to ship data via Automatic Identification System (AIS), and cloud processing allow for improved estimation of ship emissions for greater transparency. OceanMind has developed an approach that makes use of readily available data and modelling to close the gap in ship emissions reporting. We used the EU MRV data and published emissions models from studies by the IMO and International Council on Clean Transportation to develop our model (IMO 2015; 2021; Olmer et al. 2017). Our model relied on two components. First, a ML model was developed using the EU MRV data that predicted the emissions capacity of a ship in $kgCO_2$/nautical mile (nmi) travelled based on ship characteristics such as vessel size, engine power, year built, and other factors (see section 2.3 below). Second, ship activity data from AIS was used to estimate distance travelled during time intervals between AIS transmissions and an empirical model was used to estimate emissions based on time elapsed between AIS transmissions and ship speed.

Together, these techniques were used to estimate emissions by ship type and reported on the Climate TRACE website as individual trips, as well as aggregated data by ports and countries (https://climatetrace.org/).

2 Methodology

The following section describes the datasets employed and the methodology used by OceanMind to develop a model for estimating GHG emissions from the shipping industry on a monthly basis from 2015 to 2024. The model was developed in two stages:

1. **ML Model** – The ML part of the model was developed to estimate the emissions capacity ($kgCO_2/nmi$) for individual vessels.
2. **Empirical Model** – Combining the emissions capacity with vessel activity data (distance travelled (nmi) and speed over ground (nmi/h)) to arrive at a value for tonnes of CO_2 emitted for a particular journey.

2.1 Datasets employed

OceanMind used a variety of data sources for various tasks throughout the process of developing the model. Data was needed not only to train the ML model, but as a validation tool to ensure the emissions estimates were realistic and sensible. Each of the data sources will be described below, as well as how the data was used in the development process.

2.1.1 IMO 3rd and 4th GHG Reports

The IMO produced models which combine vessel characteristics with their activity to understand the physical processes that affect the output of emissions for vessels, in what they refer to as a 'bottom up' approach. In their GHG reports the IMO published the aggregated results of this methodology, showing the trends across different vessel types and shipping sectors (IMO 2015; 2021). From these studies, OceanMind used the auxiliary engine power demand by ship type, size, and operational phase to inform the emissions of vessels when they are in port and running on auxiliary engine power. Emissions factors for various, non-CO_2, gases were also taken from the IMO 4th GHG report.

2.1.2 Lloyd's List Intelligence Data

Lloyd's List Intelligence provides a multitude of vessel characteristics such as their size, engine power and age (Data and Analytics | Lloyd's List Intelligence (lloydslistintelligence.com)). This data was the primary source of input data for the ML models used by OceanMind. The full list of characteristics used in the model can be found in Table 1.

2.1.3 S&P Global IHS Markit Maritime Database

The IHS Markit database provides characteristics data equivalent to that found in the Lloyd's List Intelligence data. This data source was used in the same way as the Lloyd's List Intelligence data to collect characteristics information on new vessels which were added to our available tracked vessels in the 2024 release. This data source helped to fill gaps for vessels that are smaller than 500 gross tonnes (GT) as well as including newer vessels that have been commissioned since the last refresh of Lloyd's List Intelligence data.

2.1.4 Ship GHG emissions data from the European Union's Monitoring, Reporting, and Validation (EU MRV) dataset

As mentioned previously, the EU MRV dataset is the largest publicly available set of emissions data for individual vessels ([THETIS-MRV (europa.eu)](https://mrv.emsa.europa.eu)). This dataset provides self-reported vessel emissions, as well as the time spent at sea and amount of cargo transported for individual vessels available. The emissions factors, given in units of kgCO_2/nautical mile (nmi) travelled, from this dataset were used as a target label for development of the ML aspect of this model. The full dataset also served as useful validation data for confirming the plausibility of the estimated emissions.

The detailed information is limited to vessels that made port calls in EU ports and involved the loading or unloading of cargo or passengers. This means that estimating emissions for vessels outside of this subset would have to include some form of extrapolation without further data. Also, this dataset only contains a subset of the trips a vessel undergoes, making the identification of a vessel's activity slightly more difficult to cross reference with other data sources.

The EU MRV dataset has yearly releases of data but also has continuous updates to the data, resulting in multiple versions of the same year of data. Where this data has been used, the version numbers will be notated.

2.1.5 Vessel activity and characteristics data from Automatic Identification System (AIS)

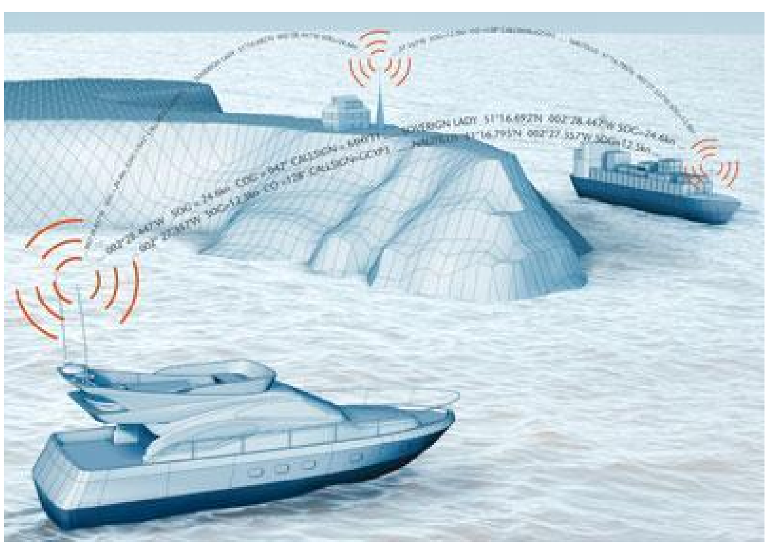

Figure 1 - Showing the process of vessels relaying AIS messages via an in-land tower. ('AIS Transponders' n.d.)

AIS is a system that collects positional data from thousands of ships, originally for the purpose of ship safety, to avoid collisions at sea. Figure 1 shows a representation of vessels relaying the positional data that they get from onboard equipment to each other and to a relay tower in-land. OceanMind has been using this technology to track vessels' activity for many different purposes, in this case to estimate the distance travelled by vessels to inform emissions estimates. AIS systems generally report a vessel's position, speed and heading very frequently, ranging from several times per minute, to several times per day. Using a combination of data from Orbcomm and Spire, a total of over 58,000 unique vessels are tracked using AIS for the purpose of estimating their emissions. Other messages transmitted on the AIS system not only tell the position of the vessel but also include some information about the characteristics of the vessel. This data was used to validate and fill gaps in the characteristics data we had on vessels.

2.1.6 Vessel operational data from ship owners and other sources

Validation data for individual vessel emissions is not widely available but several sources were used in the development process, these included:

- Vessel characteristics, activity and measured GHG emissions from an owner of 2 container vessels and 3 fishing vessels. This data remains confidential and was used only to assist in expanding the knowledge and understanding of the model. This is hereafter referred to as "vessel ground truth data".
- Siglar, an "independent provider of actionable insights leading to substantial reduction in CO2 emissions in the shipping industry" provided operational data for several vessels ('Siglar – Carbon Efficient Chartering' n.d.). Two thirds of this data was used in the development of the model, with the remainder reserved for validation. Siglar also provided consultation on opportunities for model improvements.
- 'Bottom-up' emissions model results from an anonymous source that were used for comparison and validation of the model output.
- Various online sources of vessel characteristics data were used to fill gaps and validate vessel characteristics where appropriate.

2.1.7 Global Fishing Watch (GFW) Emissions Data

GFW conducted a pilot project for the Climate TRACE consortium during 2024 that consisted of them using their emissions algorithms to capture emissions estimates for low information and 'dark' vessels. The outcome of this pilot project is a dataset using a combination of OM and GFW data that covers over 500,000 vessels globally which is discussed in Section 2.5. For information on GFW approach can be found in Climate TRACE's GitHub methodology repository.

2.1.8 Shipping sector inventory data

Another method used for verification of OceanMind's modelled emissions estimates was to compare to other emissions inventories. Where possible, the emissions estimates were adjusted to match the coverage of the relevant inventory, with this process described in section 3.3.2. The following inventories were included for comparison:

- IMO Fourth GHG Study (IMO 2021) compared global emissions from all shipping.
- Emissions Database for Global Atmospheric Research (EDGAR) (European Commission. Joint Research Centre. 2023) compared emissions from all shipping.

2.1.9 Data on port locations

In addition to estimating vessel emissions, port emissions were estimated by aggregating emissions from vessel trips and port stays. The locations of ports were taken from a multitude of sources. These include the World Port Index ('Maritime Safety Information' n.d.), Pew Charitable Trusts ('The Pew Charitable Trusts' n.d.), along with a ports database that OceanMind has been building for many years. OceanMind's Port database is constantly growing and adapting based on new information gained from satellite imagery as well as clustering analysis being carried out on vessel behaviour. An example of port locations and attributions is shown in Figure 2 and 7.

Figure 2 - A) Climate TRACE UI displaying the attribution of 2022 emissions to ports, where each circle represents a port and the size of the circle is dependent on the relative emissions of that port. Red box indicates B) the United Kingdom with known ports highlighted in orange.

2.2 Data preparation

With data coming from various sources, and some of the sources relying on self-reported data, several stages of data preparation were required to make sure the data was reliable and fit for the purpose of training a ML algorithm. Primarily the EU MRV and Lloyd's List Intelligence datasets were used in this stage of development.

2.2.1 Data Cleaning

To assemble a set of vessels that were suitable for training the ML model, the vessels available in the EU MRV's datasets for 2018 (v261), 2019 (v193) and 2020 (v67) reports were collected. This comprises 16,665 unique vessels, with a total of 36,180 unique entries across the three years of reports. Of those 16,665 vessels, 16,516 had data on vessel characteristics available in Lloyd's List Intelligence. The AIS data for these vessels was then obtained, for the purpose of removing vessels from the training set that had poor quality tracking data available, as they would be impossible to validate. Ship entries were removed if they fell into the following categories where annual time at sea of less than 1 week of greater than 1 year and/or annual distance travelled of less than 1000 km.

This filtering removed 1,799 entries. Following this, outlier detection was performed on the vessel characteristics data. Vessels were removed that had values greater than three times the interquartile range above the 75th percentile, or three times the interquartile range below the 25th percentile for the following characteristics: reported fuel efficiency, deadweight, gross tonnage, breadth, maximum engine power, auxiliary engine power, calculated average speed, reported cruising speed, and calculated distance travelled. This resulted in removing 329 additional entries, leaving 33,853 entries from 15,855 unique ships, with 11,198, 11,709, and 10,946 ships reporting fuel efficiency in 2018, 2019, and 2020, respectively.

The final stage of data cleaning was to add three engineered variables to aid in the predictive modelling process. These three variables are as follows (see Table 1):

- **FlagNameBin**: The top 15 most common flags are retained, with any others being binned into an "Other" category.
- **ShipTypeEU**: A binned version of the vessel type from the Lloyd's List Intelligence category i.e., all different types of bulk carrier mapped to the "Bulk carrier" category.
- **Length**: Vessel length taken as the maximum value of the LengthOverallLOA or LengthRegistered in the metadata files.

2.2.2 Imputation of missing information

Once the data had been cleaned and engineered to be fit for purpose, there was another barrier before model development could be carried out. This was that the vessel characteristics metadata was not complete for all vessels. To rectify this, models were trained on vessels with complete records, to be able to predict and impute the values for vessels that had information missing.

The *R* (R Core Team n.d.) package for Multivariate Imputation by Chained Equations a.k.a. *mice* (Buuren and Groothuis-Oudshoorn 2011) was used for this purpose. The nature of the *mice* algorithm allows for iterative improvements to be made on each round of imputation, then the entire process can be repeated and improved further. After some optimisation the number of iterations per repetition and the number of repetitions, were set to 2 and 20 respectively.

Table 1 - List of input variables used for training the ML models, along with the percentage of each variable that was imputed based on OceanMind's total ship catalogue.

Predictor Name	**Description**	**Example**	**% imputed**
FlagName	Flag State (Country which the vessel is registered to)	Japan	
FlagNameBin	Top 15 flags, with remainder grouped as OTH (other)	JPN	
ShipTypeEU	High level ship type group	Oil tanker	
YearOfBuild	Year vessel built	2009	
Deadweight	Vessel Deadweight	63,800 tonnes	2.2 %
GrossTonnage	Vessel GrossTonnage	50,300 tonnes	0 %
Speed	Vessel Service Speed	12.1 knots	14.3 %
Length	Vessel length	96 m	1.6 %
Breadth	Vessel width	16.61 m	10.5 %
Draught	Vessel Draught	7.41 m	2.9 %
PowerKwMax	Power of main engines	1,491 Kw	20.1 %
PowerKwAux	Power of auxiliary engines	830.8 Kw	65.3 %

2.3 Model Development

After preparing the data sufficiently, development of the model could commence. The first stage was selecting a ML algorithm that would be capable of predicting a vessel's emissions factor (kg CO_2 /nmi), followed by validating and improving the estimates using an empirical model. The following sections describe the process carried out to achieve both of these steps.

2.3.1 ML Model

Several ML algorithms were tested with the dataset, to determine which was the best suited to the problem. These algorithms were random forests, extreme gradient boosting, ridge-penalised regression on continuous variables, and linear regression using the PowerKwMax variable. The RStudio packages used to implement these were randomForest, xgboost, caret, and base, respectively.

It was found that a separate model being trained for each vessel type category gave optimal performance during initial testing. Building from this, each algorithm was tuned to optimise its performance. Each package and algorithm have many parameters that can be tuned to adapt the algorithm for specific applications, please refer to the documentation for each named package for an explanation of these tunable parameters.

A process known as 5-fold cross-validation was used in the training process, where the training data is split into 5 subsets and the training is carried out 5 times, reserving one subset of the data for validation in each iteration. This allows validation to be carried out on the fly and improves the generalisation of the algorithm as it is less likely to over train on the training dataset. One third of the total dataset was kept to one side for final testing to validate the results. To quantify the performance of the algorithms, the root-mean-squared-error was used, which is a commonly used performance measure.

The final model used to estimate the CO_2 emissions factor after the entire testing process was a random forest model with 2,000 trees (*ntree*), 8 nodes per tree (*nodesize*) and 15 variables sampled at each branching (*mtry*).

2.3.2 Empirical Adjustments

To transform the emissions factor from the above ML model into CO_2 emissions, one must multiply by the distance travelled by a vessel. When a vessel is transmitting frequently, this can be calculated using the speed over ground reported by AIS, multiplied by the time between transmissions (Equation (1)).

$$Distance\ travelled\ (nmi) = speed\ over\ ground\ (nmi/h) \times time\ (h) \quad (1)$$

However, as time between transmissions gets longer, the likelihood of speed changes increases so the physical distance is calculated using the positional data of the vessel at the beginning and end of the period.

It is known from literature that the emissions of a vessel have some dependence on the speed that the vessel is travelling at (IMO 2021; Olmer et al. 2017). Especially the effect of a vessel travelling at a speed lower than its maximum cruising speed (Degiuli et al. 2021). This effect was also observed when comparing the raw emissions data to the ground truth vessel data provided by vessel owners. To account for this effect, a so-called “speed adjustment factor” (SAF) was developed using the high-resolution ground truth vessel data. It was found during this investigation that the service speed of a vessel did not adequately describe the maximum cruising speed of a vessel so a ‘reference speed’ was derived empirically to produce the best fit between the model’s emissions and the ground truth emissions. The relationship of the ratio between the reference speed of a vessel and the actual speed over ground with the emissions of a vessel is found to follow a square law (Adland, Cariou, and Wolff 2020; Berthelsen and Nielsen 2021), which allows us to arrive at a SAF given by Equation (2).

$$SAF = \left(\frac{speed\ over\ ground}{reference\ speed}\right)^2 \quad (2)$$

This dimensionless factor can then be used to adjust the emissions produced by a vessel when travelling at slow speeds. The SAF is capped at a value of 1 to prevent drastically increasing emissions estimates when travelling above the reference speed.

Another factor that must be accounted for when estimating emissions is the power output of the auxiliary engines of a vessel. As mentioned in 2.1.1, this information was gathered from the 3rd and 4th IMO GHG studies (IMO 2021; 2015) for each vessel type. It was found that the speed had little to no effect on the auxiliary engine usage while at sea, so those emissions were removed from the speed adjustment calculation. The auxiliary engines also run while a vessel is at anchor, and often while it is at berth, which is accounted for in the model by taking the average of the auxiliary engine demand in these two operating phases and multiplying by the amount of time spent in port based on a linear relationship between the auxiliary engine power demand and the emissions from those engines.

Combining these factors together gives a set of emissions estimation formulae (Equations (3) and (4)) that can be applied to a vessel dependent on whether it is in a port or travelling at sea:

$$CO_2\ emissions\ at\ sea\ (tonnes) = \left(\left(\left(distance \times E_c\right) - AUX_s\right) \times SAF\right) + AUX_s \quad (3)$$

$$CO_2\ emissions\ in\ port\ (tonnes) = AUX_p \times time \quad (4)$$

Where E_c represents the emissions capacity, and AUX_s/AUX_p represent the emissions from the auxiliary engines while at sea or in port respectively.

2.3.3 Non-CO_2 emissions estimates

To fully understand the impact of vessels' emissions worldwide, emissions from gases other than CO_2 need to be considered. As the current model does not measure the fuel usage of a vessel but instead based on a CO_2 emissions factor, these estimates could not be baked into the same methodology used for the CO_2 emissions estimates. Instead, using data from the IMO 4th GHG report on emissions factors, conversion factors between CO_2 emissions and other gases were formulated. The IMO 4th GHG report gives fuel-based emissions factors for many pollutants per fuel type in grams of the pollutant per gram of fuel burnt. By taking the ratio of a particular gas and the grams of CO_2 per gram of fuel burnt, an emissions ratio between the gas and CO_2 can be reached. The report includes four fuel types; Heavy Fuel Oils (HFO), Liquefied Natural Gas

(LNG), Marine Diesel Oil (MDO) and Methanol. For each of these fuels, the various factors are shown in Table 2.

Table 2 - Full breakdown of the calculation of emissions factor ratios for all gases per fuel type

Non-CO2 Gases per gCO2								
Fuel Type	**gCH4/gCO2**	**gN2O/gCO2**	**gSOx/gCO2**	**gNOx/gCO2**	**gCO/gCO2**	**gVOCS/gCO2**	**gPM2.5/gCO**	**gPM10/gCO2**
HFO	0.000016	0.000058	0.016323	0.024374	0.000925	0.001028	0.002229	0.002425
LNG	0.004341	0.000036	0.000011	0.004878	0.001441	0.000577	0.000036	0.000040
MDO	0.000016	0.000056	0.000540	0.017689	0.000808	0.000749	0.000259	0.000281
Methanol	0.000000	0.000000	0.016323	0.005797	0.000925	0.001028	0.000027	0.000029

As mentioned, the current OM model doesn't capture the fuel type used, so to effectively use this data, knowledge of the average percentage of vessels using each fuel type in the IMO report as shown in Table 3 was required to formulate a weighted average emissions factor for each gas, shown in Table 4.

Table 3 - Fuel usage per fuel type for international and domestic shipping in the IMO database

	Fuel Usage (kT)		**Fuel Usage (%)**	
Fuel Type	**International**	**Domestic**	**International**	**Domestic**
HFO	188.33	34.58	79.19%	39.60%
LNG	10.90	0.54	4.58%	0.62%
MDO	38.46	52.18	16.17%	59.75%
Methanol	0.13	0.03	0.05%	0.03%
Total	237.82	87.33	100.00%	100.00%

Table 4 - Weighted average emissions factors for all gases reported by OM

Weighted Average Factor							
gCH4/gCO2	**gN2O/gCO2**	**gSOx/gCO2**	**gNOx/gCO2**	**gCO/gCO2**	**gVOCS/gCO2**	**gPM2.5/gCO2**	**gPM10/gCO2**
0.000167	**0.000057**	**0.011315**	**0.021803**	**0.000910**	**0.000933**	**0.001598**	**0.001738**

Comparing these to the ratios between CO_2 and these gases given by the IMO for 2018, each gas total is less than 10% different to the IMO results. At the voyage level the uncertainty on these values will be large as the estimates are based on global averages, but this shows that at the highest levels of aggregation, these factors are consistent with estimates from the IMO.

2.4 Model Deployment

Using vessel data from Lloyd's List Intelligence as well as new vessel information gathered from the IHS Markit database and other sources, any vessels that had sufficient AIS data could be inserted into the model to estimate their emissions. This resulted in emissions estimates on a per trip/port stay basis for over 58,000 unique vessels spanning the period 2015-2024, an improvement of over 10,000 vessels since the 2023 data release. This data will continue to be updated on a monthly basis as more AIS data is processed by our model. As part of the reporting process, these emissions were attributed to ports as described below in section 2.4.1, followed by aggregation to the country level described in section 2.4.2. For a visualisation of the emissions, as well as the option to download the data, please visit the Climate TRACE website. A description of the column definitions can be found in the tables provided in Appendix A.

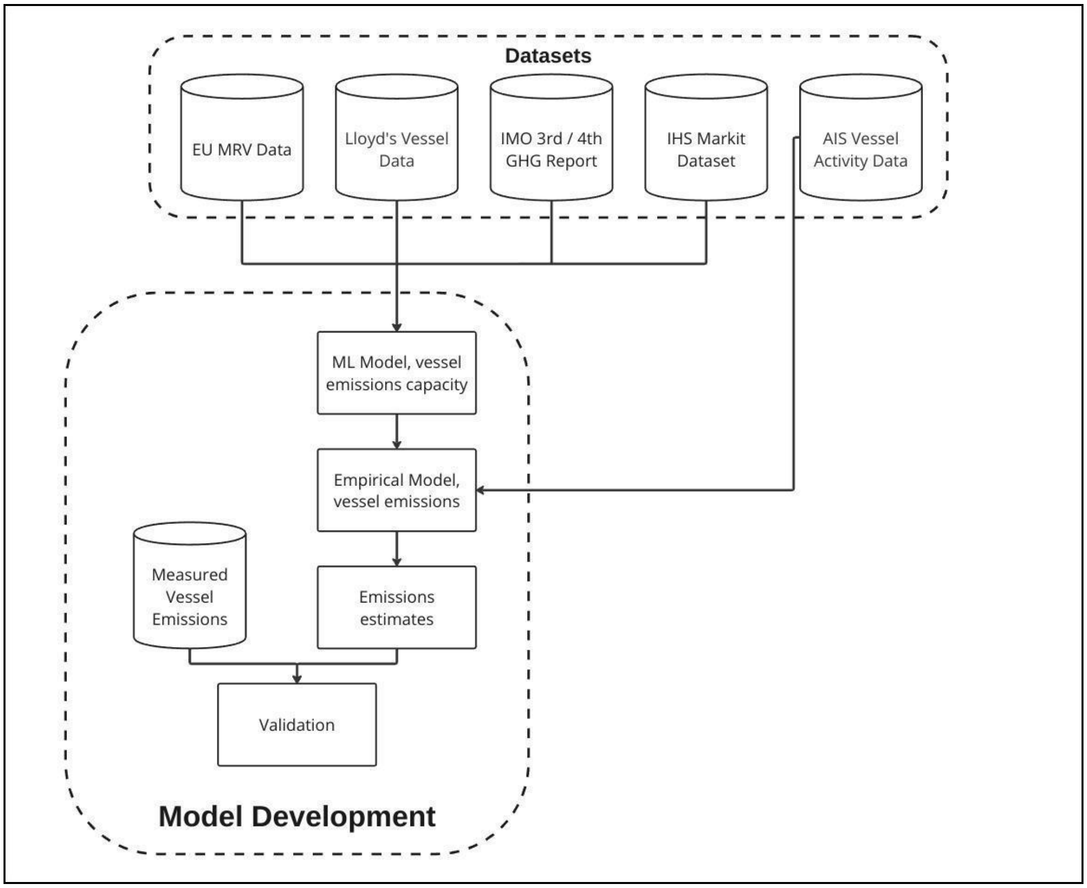

Figure 3 - Flowchart of the model development process, showing the datasets employed in this stage

2.4.1 Port Attribution

To improve the visualisation of geographical emissions estimates, two terms were used to define a monitoring period, relating to whether a vessel has been travelling between ports (trip) or staying at a singular port (port stay):

Trip: A journey that starts with the vessel within one port and ends with a vessel in another port. The emissions from a trip are divided equally between the origin and destination ports.

Port Stay: A period that a vessel spends within a single port region. All emissions from a port stay are attributed to the holding port.

Using the list of ports mentioned in section 2.1.7, a set of heuristics, as follows, must be used to ensure that vessels stopping in unknown ports are not incorrectly classified as still being on a trip. To determine when a vessel is classified as being ‘in port’ a series of rules were made and applied additively to the AIS data.

- Firstly, a vessel must be *moving slowly*, i.e., have a speed over ground of less than 1 nmi/h, to be reasonably considered as being ‘in port’.
    - **First Rule- “*moving slowly*”:** have a speed over ground of < 1 nm/h.
- Secondly, the ships are split into two categories based on the vessel type. Vessel types that are classified as being ‘larger’ are chemical tankers, liquefied natural gas (LNG) carriers, liquefied petroleum gas (LPG) carriers and oil tankers, with anything else classed as ‘smaller’. The reasoning for this is that smaller vessels tend to move around much closer to ports and shores, so the rules must be defined separately based on the size of a vessel. For large (small) vessels, a rule for the maximum distance from a port is defined as being within 12 (5) nm of a known port, or any land mass.
    - Second Rule- *“**be within range of land or a known port**”***:** Be within 12 (5) nm of a known port or land mass, separately applied for large (small) vessels. A vessel is classified as large if it falls into the category of chemical tanker, LNG carrier, LPG carrier or oil tanker.
- Thirdly, there are holding areas worldwide, such as the Panama Canal, that were observed to fit these rules but should not be considered as a port as vessels within these areas are generally still in transit, but simply waiting to pass through choke points. For this reason, holding areas were identified using AIS data and excluded from areas that can be considered as being ‘in port’. If a known port fell within a holding area, it was excluded from these holding area rules and a vessel would be counted as visiting a port if it entered that port area.
    - **Third Rule-** “***not be within a holding area***”**:** Certain areas worldwide are defined as holding areas and should not be considered as ports.

This information is provided at the vessel level, with emissions for trips and port stays outlined in their entirety. Additionally, the port attribution methodology allows the aggregation of these trips and port stays to a port level where a *trip* has half of the emissions assigned to the port of departure and half to the port of arrival, and a *port stay* has all emissions assigned to the port where the stay was carried out. If all three of the heuristics are met but the vessel is not within a known port area, the emissions are instead assigned to the nearest region based on the Global Administrative Areas (GADM) boundaries. Figure 7 in the results section provides an example of the port attribution emissions.

2.4.2 Country aggregation

The final level of aggregation available for vessel emissions is at the country level. To achieve this, all emissions that were assigned to ports were also assigned to the country associated with that port. In the aggregation to country level emissions, gap filling techniques were used to incorporate known gaps in coverage. This process is fully described in section 3.3. It should be noted also that all aggregation is done based on the end date of a trip or port stay.

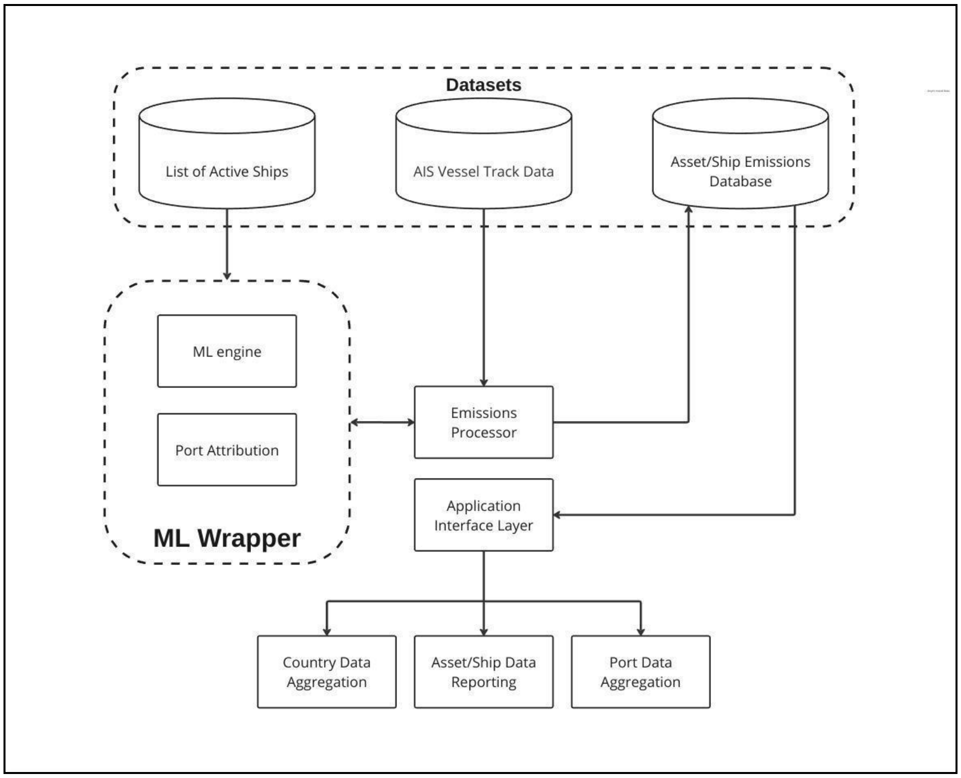

Figure 4 - Flowchart showing the model deployment process, with the datasets employed at the top and the output reports at the bottom

Table 5 - Summary of the Climate TRACE shipping dataset

General Description	**Definition**
Sector definition	***Shipping***
UNFCCC sector equivalent	***1.A.3.d Domestic Navigation, 1.C.3 Other, International Navigation***
Temporal Coverage	*2015 – July 2024, and ongoing (pending monthly releases)*
Temporal Resolution	*Per vessel: trip based temporal resolution. Country/Port: Monthly resolution*
Data format(s)	*CSV*
Coordinate Reference System	*EPSG:4326, decimal degrees*
Number of assets/countries available for download and percent of global emissions (as of 2023)	*58,160 total vessels representing ~67% of this sector's emissions (based on gap filling model). Emissions attributed to 8,113 unique ports.*
Total emissions for 2023	*1,151 million tonnes CO_2e 100yr GWP*
Ownership	*We used ownership data provided by Lloyd's List Intelligence*

General Description	Definition
What emission factors were used?	*Predicted emissions factors (kgCO_2/nmi) using ML techniques. Emissions factors per gram of fuel used from IMO 4th GHG report to estimate other gases.*
What is the difference between a "NULL / none / nan" versus "0" data field?	*"0" values are for true non-existent emissions. If we know that the sector has emissions for that specific gas, but the gas was not modelled, this is represented by "NULL/none/nan"*
total_CO2e_100yrGWP and total_CO2e_20yrGWP conversions	*Climate TRACE uses IPCC AR6 CO_2e GWPs. CO_2e conversion guidelines are here:https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_FullReport_small.pdf*

2.5 Combined dataset with Global Fishing Watch (GFW) data

In 2024 GFW completed a pilot project with CT to attempt to capture emissions from lower information and dark vessels that aren't captured by OM's model. The details of the GFW methodology can be found here or a summary is available from the Climate TRACE GitHub methodology repository in *Transportation*. To integrate their dataset with the OM dataset, first each trip entry was checked to see if its IMO or MMSI number matched any which were already covered in the OM dataset, to minimise the risk of double counting emissions across both datasets. Once the dataset had been deduplicated the start and end points of each trip was then attributed to the closest port in the OM ports database. This combined dataset aims to have increased coverage, covering over 570,000 unique vessels across the 2015-2024 time period. A large number of these vessels are small, low information vessels which have had their type and characteristics estimated using GFW's algorithms.

In 3.1.1 the OM methodology for uncertainty and confidence is discussed. For consistency this methodology was also mostly adopted by GFW for their data, and when combining the confidences of the two datasets at the port level this decreased the confidence on the port level estimates as there is a higher concentration of low information, low confidence vessels on a global scale.

GFW did not initially include estimates for VOCS, $PM_{2.5}$ or PM_{10} so this information was backfilled by OM using the same methodology as outlined in 2.3.3.

As part of the pilot project, GFW also attempted to estimate the emissions of the 'dark' fleet globally, with their methods described in the document linked above. 'Dark' in this context refers to the vessels which are active on the ocean but are not seen to be transmitting on AIS for any given reason. These emissions have been assigned to the iso3 code 'UNK' (unknown) as they cannot be attributed to any specific country.

3 Results and Validation

The output achieved by the model can be found in section 2.4, which explains the various levels of aggregated data that are published by Climate TRACE. Another dataset delivered is the uncertainty and confidence on the reported values of each asset's emissions data, described in section 3.1. Several techniques were then used to validate the results of the model against other data sources, described in section 3.2. Finally, to be able to make direct comparisons with other data sources, the gaps in our dataset must be understood, the investigation into this is detailed in section 3.3.

3.1 Uncertainty and Confidence

Part of the most recent Climate TRACE data release was to provide uncertainty and confidence estimates for the asset level emissions data. The following sections describe the process used by OceanMind to provide this information.

3.1.1 Uncertainty and Confidence Methodology

For a first pass at estimating uncertainty on emissions data, each variable for vessels is assumed to be normally distributed and the provided values for uncertainty are the calculated standard deviation of the distributions. In future, more time is needed to explore the distributions of each variable in depth to improve the uncertainty estimates, which will be discussed in section 4. At the port level, standard error propagation was used to combine the uncertainty values of vessels that contributed to the port's emissions for that particular reporting period.

The confidence levels used are generally defined as a measure of how well the data is understood and how trustworthy the data source might be, i.e.:

- **Very High:** The data is measured directly, from a trustworthy source, or several independent sources agree reasonably well.
- **High:** Interpolated estimates within the bounds of a well trusted dataset i.e., using a ML algorithm with datasets very closely related to the training data.
- **Medium:** Extrapolated from well trusted datasets to assets of a similar nature or measured directly but with no confirmation of trustworthiness or secondary sources to compare with.
- **Low:** Extrapolated largely from the original dataset on an asset that may not be entirely similar or based somewhat on assumptions.
- **Very Low:** Many assumptions made or low possibility of validation.

Using this basis, confidence values were assigned to the data for each asset individually. For vessels, the confidence levels can be seen in Table 6, where values in brackets are for vessels which do not appear in the EU MRV dataset. The difference here is because vessels that appear in the EU MRV dataset are well understood by the underlying ML algorithms as the EU MRV dataset was used for training. Lower confidence was assigned to any vessel that does not appear in the EU MRV dataset for the emissions estimates as it requires the assumption that these vessels follow a similar efficiency to those within the EU MRV dataset, which cannot be confirmed within the bounds of this model.

Table 6 - Describing the confidence and uncertainty for each attribute for vessels in the Climate TRACE dataset. "other_gas" refers to any other gases reported by OM that aren't explicitly listed.

Data attribute	Confidence Definition	Uncertainty Definition
type	Very High	Standard deviation
capacity	High	Standard deviation
capacity_factor	High	Standard deviation
activity	High	Standard deviation
CO2_emissions_factor	High (in EU MRV) Low (not in EU MRV)	Standard deviation
CH4_emissions_factor	Very Low	Standard deviation
N2O_emissions_factor	Very Low	Standard deviation
other_gas_emissions_factor	Very Low	Standard deviation
CO2_emissions	Medium (in EU MRV) Very Low (not in EU MRV)	Standard deviation
CH4_emissions	Very Low	Standard deviation
N2O_emissions	Very Low	Standard deviation
other_gas_emissions	Very Low	Standard deviation
total_CO2e_100yrGWP	Low (in EU MRV) Very Low (not in EU MRV)	Standard deviation
total_CO2e_20yrGWP	Low (in EU MRV) Very Low (not in EU MRV)	Standard deviation

For confidence on ports (Table 7), a similar ethos was used, with ports that lie within countries that are in the European Economic Area (EEA) being assigned slightly higher confidence values. This is because of the higher volume of vessels visiting these ports that are within the EU MRV dataset. The overall confidence levels for ports are lower, as the port level data is an aggregation of a vast number of trips and port stays for thousands of vessels, where the distribution of which vessels visit certain ports is not entirely understood (Figure 7).

Table 7 - Describing the confidence and uncertainty for each attribute for ports in the Climate TRACE dataset

Data attribute	Confidence Definition	Uncertainty Definition
capacity	Medium	N/A
capacity_factor	Medium	N/A
activity	Medium	N/A
CO2_emissions_factor	Medium (in EEA) Very Low (not in EEA)	N/A
CH4_emissions_factor	Very Low	N/A

Data attribute	Confidence Definition	Uncertainty Definition
N2O_emissions_factor	Very Low	N/A
other_gas_emissions_factor	Very Low	N/A
CO2_emissions	Low (in EEA) Very Low (not in EEA)	N/A
CH4_emissions	Very Low	N/A
N2O_emissions	Very Low	N/A
other_gas_emissions	Very Low	N/A
total_CO2e_100yrGWP	Low (in EEA) Very Low (not in EEA)	N/A
total_CO2e_20yrGWP	Low (in EEA) Very Low (not in EEA)	N/A

3.2 Validation

To validate the emissions estimates, a few different techniques were used. Firstly, a comparison to measured values from ship owners (section 3.2.1) was used to test if there was sufficient agreement between the estimated emissions and those measured from highly granular data collected directly from vessels. Secondly, using time series analysis (section 3.2.2) shows that trends in the data match expected results from known events. Finally, to aid in comparisons with other GHG inventories (section 3.3.2), studies to understand the gaps in coverage were carried out (section 3.3).

3.2.1 Comparison to measured values

The measured emissions values described in section 2.1.5 were obtained for the purpose of validating emissions estimates from the model. Siglar provided a sample of 12 vessels for this study, which were compared to OceanMind's first and second model versions. Note that the methodology described above refers to version 2.0 of the model.

Table 8 shows the comparison values for single voyages carried out by each vessel in the sample dataset. The measured values provided by Siglar, the estimates from version 1.0 and version 2.0 of the OceanMind model, and the differences from the measured values for each model are shown. The normalised root mean squared error (nRMSE) is a way to measure the overall difference between two datasets and is displayed based on the total error on all 12 vessels, using the Siglar data as the ground truth. The nRMSE is lower for version 2.0 (13.9%) than version 1.0 (22.1%), meaning that the emissions estimates have improved over the two models.

Table 8 - Comparison of version 1.0 (2021) and version 2.0 (2022) of the OceanMind emissions model with measured emissions from Siglar vessels. nRMSE is the overall error for all 12 vessels

Vessel #	Siglar measured tonnes CO_2	Model 2.0 tonnes CO_2	Difference: Model 2.0 - Siglar	nRMSE model 2.0	Model 1.0 tonnes CO_2	Difference: Model 1.0 - Siglar	nRMSE model 1.0
1	4999	5,333	334	13.9%	5,673	674	22.1%
2	4641	5,116	475		5,543	902	
3	587	598	11		403	-184	
4	1581	1,165	-416		1,253	-328	
5	2018	1,816	-202		2,363	345	
6	1254	923	-331		1,404	150	
7	1483	1,291	-192		1,973	490	
8	2845	2,732	-113		2,737	-108	
9	1112	1,068	-44		1,234	122	
10	261	344	83		274	13	
11	223	279	56		265	42	
12	205	265	60		331	126	

Figure 5 displays a box and whisker plot made using the data from Table 8, with 0 on the y-axis referring to when the estimated emissions match the measured emissions exactly. The boxes represent data that falls into the 2^{nd} and 3^{rd} quartiles, and the whiskers span to the maximum and minimum values. The line and 'x' within each box represent the median and mean values respectively. Model 2.0 has a lower variance, and the mean and median values sit closer to 0, further reinforcing the improvements made to the model.

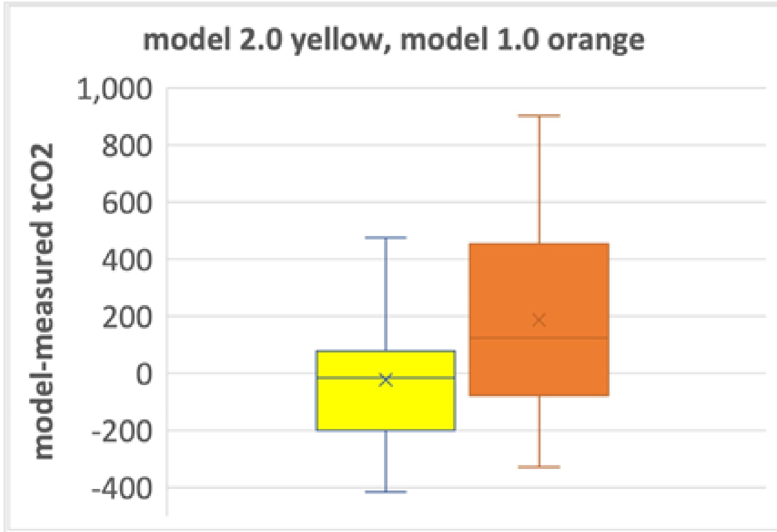

Figure 5 - Box and whisker plot showing the distribution of the differences between measured emissions values for 12 vessels from Siglar and the emissions estimates of the same vessels from OceanMind model v1.0 (orange) and model v2.0 (yellow).

3.2.2 Time Series Analysis

An important aspect of validating the emissions estimates is not only to understand the accuracy of individual voyages' estimates but to show that larger scale trends can be observed in the data. For this purpose, OceanMind carried out a study on cruise shipping during the COVID-19 pandemic. The World Health Organisation (WHO) declared COVID-19 an international pandemic in the week of March 16th, 2020, which coincided with greatly reduced activity from cruise ships (March et al. 2021; Sampson 2020). Figure 6 shows the observed trend of the CO_2 emissions estimates for cruise ships that appear in the Climate TRACE database. The Climate TRACE emissions estimates follow the expected trend, observing the period from March 2020 to June 2021, the monthly total of emissions was around 30% lower on average than previous or later months. This demonstrates the ability of the model to observe trends from significant events that impact the shipping sector.

A more recent example of these trends being visible in the emissions data is shown in Figure 7, which shows a huge downturn in the emissions which were attributed to the port of Baltimore, Maryland following Francis Scott Key Bridge collapse on March 26, 2024 (vertical dashed line in Figure 7). Previously, port emissions were increasing from mid-2020, which then saw a rapid decrease March 2024, due to ships unable to enter the port. The ability to observe these events with such recency can help us to understand the impact on the emissions in the areas surrounding the event.

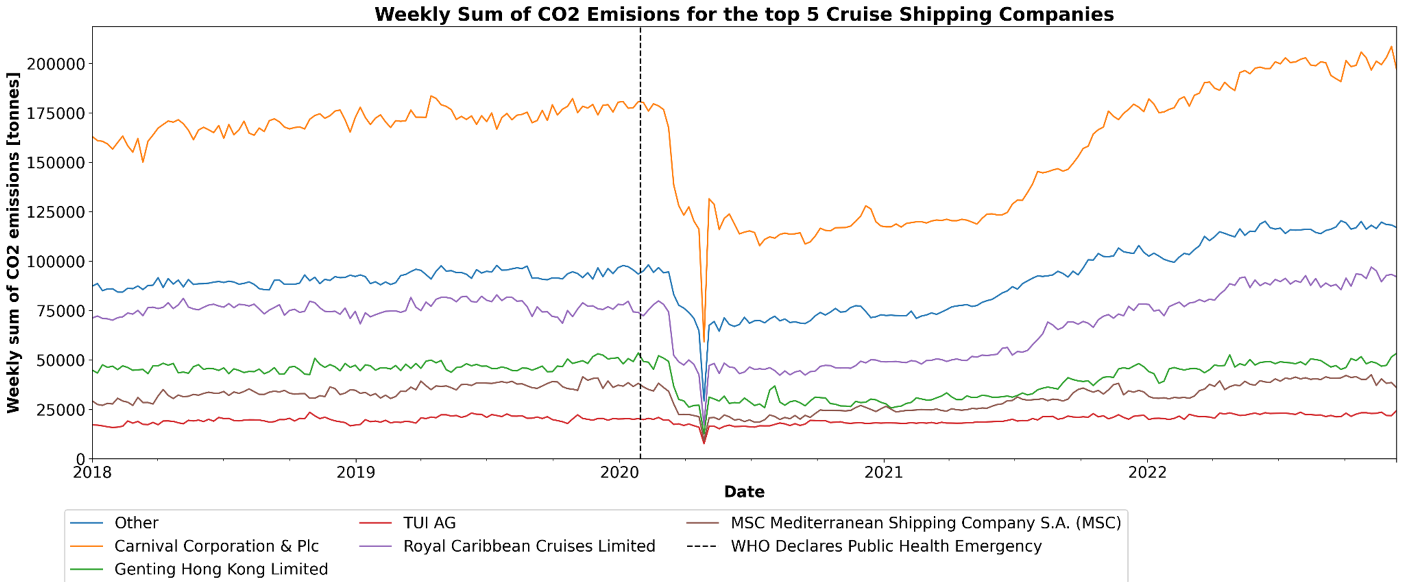

Figure 6 - Weekly sum of estimated CO_2 emissions, in tonnes, from major cruise ship companies during January 2018 to December 2022, based on OceanMind's model. Companies outside the top 5 are grouped as 'Other'. The black dashed line represents the date that the World Health Organization declared COVID-19 a public health emergency.

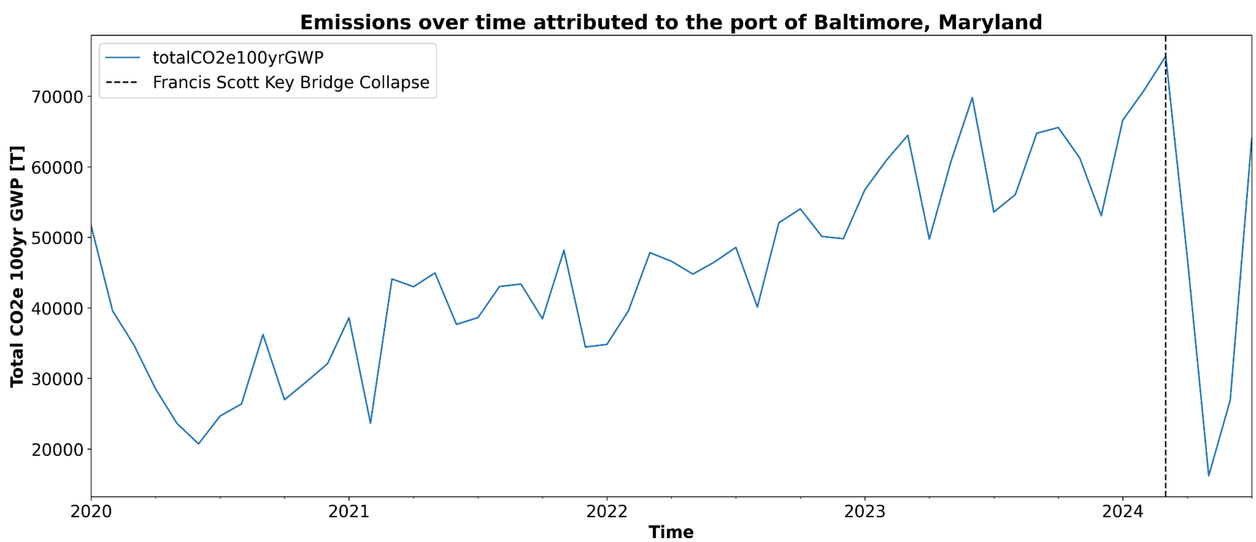

Figure 7 - Monthly emissions totals for the port of Baltimore, Maryland from January 2020 to July 2024 based on the OM emissions model. The black dashed line represents the date of the Francis Scott Key Bridge Collapse (March 26th, 2024).

3.3 Gap Filling

Comparing the overall emissions estimates from the Climate TRACE database to other, well recognised, inventories is an important step in validation. The largest roadblock with carrying this out is that no two inventories agree on the number of vessels worldwide, and often cover different vessels due to differences in data sources. These sections cover the methods used to attempt to 'fill' the known gaps in coverage of the data sources used in the Climate TRACE model, followed by using the gap filled data in comparisons with other inventories. In the Climate

TRACE database, the overall country level aggregated data has been scaled based on the results of this section, where each different scale factor is outlined below.

3.3.1 Sectors included in Climate TRACE database

The first stage in understanding the coverage of the Climate TRACE model was to compare the counts of the types of vessels that are already covered within the model. The vessel counts from the IMO 4th GHG study (IMO 2021) were used for comparison, as this is a trusted source and one of the most in depth GHG reports currently published. Table 9 provides the results of this comparison, with the ratio between the IMO vessel counts to the Climate TRACE vessel counts in the final columns. For any vessel types where Climate TRACE already observes more of that vessel type than the IMO, no scaling was applied when aggregating to the country level. However, for vessel types that the IMO observed more than Climate TRACE, the ratio was used as a scale factor on individual vessel emissions when aggregating to the country level.

Table 9 - Comparing counts of vessel types between the Climate TRACE database and IMO vessel counts, with ratios in the final column. IMO Ratio = IMO Count / Climate TRACE count

Vessel type	Climate TRACE Count	IMO Count	IMO Ratio
Liquid Gas Carrier	2431	1953	0.80
Bulk Carrier	13492	11672	0.87
Chemical Tanker	10777	5506	0.51
Container	6057	5182	0.86
General Cargo	12927	14994	1.16
Oil Tanker	3946	8177	2.07
Passenger	759	4071	5.36
Refrigerated Cargo	775	895	1.15
Ro-Pax	5472	3148	0.58
Ro-Ro	716	2002	2.80
Vehicle Carrier	807	828	1.03

3.3.2 Fishing vessels

The current data sources used in the model only cover merchant vessels, so fishing vessels fall into a sector that is not covered. Again, there is no single source that provides a definitive count of the emissions from fishing vessels. To account for this sector, data was gathered from several data sources:

- **International Council on Clean Transportation (ICCT):** 42 million tonnes CO_2 (2015) (Olmer et al. 2017)
- **IMO:** 40.7 million tonnes CO_2 (2017) (IMO 2021)
- **Global Fishing Watch (GFW):** 29 million tonnes CO_2 (2021), which is based on an emissions file shared to OM from GFW ('Transparency for a Sustainable Ocean | Global Fishing Watch' n.d.)

The mean of each of these sources came to ~37 million tonnes CO_2, which was then compared to the overall emissions from the IMO 4th GHG study and Climate TRACE model. This average value accounts for ~3.5% of all emissions from the IMO study.

3.3.3 Other vessel types

There are also vessel types outside of those mentioned above that contribute a significant amount to the overall global emissions from the shipping sector. Using the emissions estimates from the IMO, and the GFW emissions dataset provided as part of a pilot project for CT the percentage of overall emissions of each sector is described in Table 10. Combining additively the total percentage from the IMO dataset, with the percentage attributed to fishing vessels in section 3.3.2 gives an overall scale factor of 10.9% to account for sectors not covered in the Climate TRACE database. It was decided to continue using the IMO percentages for this scaling factor as the GFW dataset is relatively new and will take more investigation to understand the validity and accuracy of the vessel type mappings.

Table 10 - Percentage of overall emissions in the IMO study and GFW dataset attributed to vessel types not covered so far. Equation 5, below, shows how all vessels and ratios are combined to generate a gap filling model.

	Percentage of overall emissions	
Vessel Type	**IMO**	**GFW**
Tugs	3.8%	1.09%
Yachts	0.5%	0%
Offshore	2.0%	0.6%
Service	1.3%	0.6%
Miscellaneous	0.1%	0.26%
Total	**7.7%**	**2.55%**

3.3.4 Unfinished Trips

The final factor included in the gap filling model is a result of the reporting process described in section 2.4.1 as the reports only include completed trips i.e., trips that have terminated at a port.

This means that when the data is processed, any trips that have not yet reached a port do not get included in the database. To account for this, the total emissions from all unfinished trips was calculated and compared to the overall emissions estimates, which revealed that ~1% of emissions can be attributed to unfinished trips. It is worth noting that some of these "trips" are from vessels that have been repurposed into offshore storage facilities, which will be discussed in section 4.1.

3.3.5 Summary of Gap Filling Model

To summarise, for vessel sectors that are covered by the Climate TRACE model, in the process of aggregating to the country level the emissions are scaled dependent on the ratio of the number of vessels observed in the Climate TRACE model vs. the IMO vessel count. Once the observed emissions have been scaled and aggregated, the overall country level emissions are scaled based on an additive scale factor from other sources of known knowledge gaps. The average of found sources on fishing vessel emissions accounts for 3.5% of emissions, the total percentage of other vessel types in the IMO study account for 7.7% of emissions and unfinished trips account for 1% of emissions, combining for an overall scale factor of 11.9%. Combining all of this gives Equation 5, where E_x represents the emissions from a given type of vessel x corresponding to the categories in Table 9.

$$Adjusted\ CT\ emissions\ =\ (E_{LGC} + E_{BC} + E_{CT} + E_{C} + (E_{GC} \times 1.16) + (E_{OT} \times 2.07) + (E_{P} \times 5.36) + (E_{RC} \times 1.15) + E_{RP} + (E_{RR} \times 2.80) + (E_{VC} \times 1.03)) \times 1.119 \tag{5}$$

3.3.6 Comparison to Other Inventories

Table 11 shows the total estimated global emissions from the Climate TRACE model, including international and domestic emissions, after applying the gap filling model. To compare similar emission estimates from other sources, most recent years available for several inventories were used.

Table 11 - Climate TRACE estimated global emissions from international and domestic shipping after using the gap filling model with the percentage increase compared to the previous year for the years 2015-2023.

Year	Modelled Million tonnes CO_2e 100yr GWP	Gap Filled Million tonnes CO_2e 100yr GWP	Percentage change
2015	425.6	644.6	-
2016	484.6	730.2	13.3%
2017	537.4	805.2	10.3%
2018	584.1	873.4	8.5%
2019	635.5	954.7	9.3%
2020	647.6	933.2	-2.3%
2021	747.7	1082.2	16%
2022	790.0	1178.5	8.9%
2023	764.7	1151.1	-2.3%

The results of these comparisons in Table 12 show that there is fair agreement between the OM dataset to both the EDGAR and GFW datasets with positive percentages showing the proportion of emissions more the inventory estimated over the OM inventory for a given year, and vice versa for negative percentages. Given that the gap filling methodology was known to be rudimentary and, in some cases, based on comparisons of vessel counts to a single source, these comparisons show that the Climate TRACE emissions estimates are often within 10-20% of other inventories' estimates for a given year. There is a slightly larger gap to the IMO dataset, however this difference is improving for every year where the IMO has available data. It is likely the closer agreement is due to improved coverage of AIS meaning the vessels that are covered in the OM dataset are more consistently tracked. In general, the IMO estimates are higher than the other inventories compared, but this could be due to a multitude of reasons, thus it is difficult to know which inventory is the 'correct' answer. These discrepancies show why it is crucial for these numbers to be as transparent as possible, and why Climate TRACE is such a useful source to shed light on where these inventories could be over or under estimating.

Table 12 - Table showing inventory comparisons of OM only gap filled data to IMO, EDGAR and GFW's own datasets for years 2015-2023 where data is available for each given inventory.

	Millions of Tonnes CO_2e 100yr GWP						
Year	**OM**	**IMO**	**% difference**	**EDGAR**	**% difference**	**GFW**	**% difference**
2015	644.6	1,008	36.1%	830.4	22.4%	668.2	3.54%
2016	730.2	1,045	30.1%	849.1	14.0%	763.8	4.41%
2017	805.2	1,083	25.7%	887.3	9.2%	834.0	3.45%
2018	873.4	1,076	18.8%	893.0	2.2%	857.8	-1.82%
2019	954.7			882.4	-8.2%	870.9	-9.63%
2020	933.2			802.8	-16.2%	865.8	-7.78%
2021	1082.2			860.7	-25.7%	933.1	-15.98%
2022	1178.5			914.7	-28.8%	993.5	-18.62%
2023	1151.1					1,038.9	-10.80%

4 Discussion and Conclusion

Following the previous release of Climate TRACE data, the OceanMind model for estimating shipping emissions has implemented port attribution, as well as preliminary studies into understanding the uncertainty, confidence, and coverage of the model. Section 4.1 aims to discuss the known limitations of the model, and improvements that are being considered for future development. Following this, section 4.2 summarises and concludes the overall findings of the OceanMind methodological process.

4.1 Model Limitations and Improvements

An important part of developing any estimation model is to be able to reflect on the work carried out and realise where it could have been done differently. There are many factors that could contribute to uncertainties in emissions estimation, and this section aims to highlight and summarise those that are known to OceanMind at the time of writing. Where possible and applicable, improvements that could mitigate or eliminate these limitations are also discussed.

- The model used to estimate the vessels' emissions factors is trained only on data from EU vessels that appear in the EU MRV dataset. This means that estimating the emissions factor for vessels outside of this dataset requires the assumption that EU vessels are representative of the global fleet. In studies by the IMO (IMO 2021), it was found to be the case that those vessels were reasonably representative of the global fleet, but more work would be needed to fully understand this. Another possibility for improvement is to move to an alternative data source that does not rely on vessel location.

- A limitation of the emissions factors in the EU MRV dataset is that they are calculated based on the total CO_2 emissions of a vessel, divided by the total distance travelled by a vessel. This means that vessels that spend a lot of time in port, with their auxiliary engines switched on, may have disproportionately high emissions factors due to a lower activity value. This propagates into the estimation model as it uses these emissions factors as a target label, and it is non-trivial to extract to what degree this has an effect. Reconsidering the ML target label would be the main way to overcome this limitation.
- The current AIS processing pipeline uses a vessel's Maritime Mobile Service Identity (MMSI) number to match with the characteristics database, which is inadequate as vessels are prone to changing MMSI number for reasons such as reflagging. This means that vessels drop out of the model's coverage when they change their MMSI number for any reason. In future, the model will be changed to use the IMO number of a vessel instead as these should never change for an individual vessel.
- A difficulty encountered when assigning emissions to ports, was that there are many areas where vessels will move slowly near land that are not close to any ports known to OceanMind. Efforts to increase the number of known ports has added hundreds, if not thousands of new ports to the OceanMind database but this list is not yet complete. Any emissions that cannot be assigned to a known port are instead assigned to the nearest region as defined by the Global Data Lab regions shapefile ('Area Database - Global Data Lab' n.d.), thus reducing the global resolution of locational emissions. Continued efforts to improve the OceanMind port database will mitigate this limitation.
- As mentioned in section 3.3.4, there are a certain number of unfinished trips that are from vessels that never go to a port or have never been observed to visit a port in the 8.5 years of data available. There are many possibilities for explaining this behaviour, one of which is that some large tanker vessels have been repurposed into offshore storage facilities. Some of these static facilities are now defined as ports in the OM port database, but there is still work to be done to fully rectify this matter.
- This release included a preliminary look into the uncertainty associated with the emissions estimates (section 3.1). It is known that this was a naïve approach, and more work is to be done to understand the actual underlying distributions and uncertainty associated with each layer of the model.
- The speed adjustment factor described in section 2.3.2 was based on a small set of ground truth data. Thus, although good agreement with this data was observed, extrapolation was inevitable when applying this factor to the whole dataset. In future, more investigations into the effect of speed on emissions will be carried out.
- Military vessels are still not covered by the model due to lack of transparency in the sector.

- If there are large data gaps, algorithms used by OM to parse AIS tracks will interpolate between the data points. This however leads to some inaccurate data when erroneous AIS messages are separated from actual transmissions by enough time. Several different things can cause flawed AIS messages, for example, vessels spoofing their MMSI or location, and data corruption. These erroneous AIS messages are easy to spot when a vessel is actively transmitting as they produce impossible speeds and can be removed. However, with enough temporal separation this separation is not so trivial. This is an ongoing issue which is being worked on and will be fixed in a future version of the model.
- Occasionally, two AIS devices will transmit on the same MMSI. Again, these simultaneous tracks can often be separated, although it is sometimes difficult to distinguish which is the 'correct' track. A symptom of this is that occasionally a trip will have incredibly high activity recorded, however this does not affect the emissions estimated for these journeys.
- For the new calculations of capacity factor and emissions factor as outlined by CT (described in the tables in Appendix A) vessels that were staying in ports with an activity value of 0 caused issues in the calculations due to dividing by 0. Because of this, activity in ports is set to 1 nautical mile if it is less than 1 nautical mile. This does not affect the emissions.

4.2 Conclusions

OceanMind has developed a model using a combination of ML and empirical techniques to estimate emissions for over 58,000 vessels over 8.5 years of activity, with monthly updates to continue through 2024 and into 2025. This dataset is the first of its kind to offer free, publicly available emissions data that is granular down to the individual vessel trips. Another unique aspect of our model is that it provides flexible attribution to individual vessels, vessel owners, flag states, areas of operation, or other aspects of shipping. This allows data users to customise emissions estimates to suit their purpose. Climate TRACE aims to be able to produce an inventory that can give a global picture of GHG emissions from the entire shipping sector. Work is being carried out to improve estimations of the uncertainty and coverage gaps so the emissions data can be as reliable as possible. Additionally, Climate TRACE aims to be able to provide up to date information that is consistently updated which allows, for example, policy makers to be able to start reacting to these trends and can address them promptly, rather than retrospectively. A combined dataset of OM and GFW estimations has also been gathered, covering over 570,000 vessels globally. This is a step towards 100% coverage which will hopefully be achievable with model progressions from both parties.

To improve our model, next steps include improved accuracy in estimates for individual ships and developing new ways to provide the full details of model results to data users. This includes more frequent updates of emissions for over 58,000 vessels. Model accuracy will be improved through acquisition of more training and validation data to improve the reliability and accuracy of the shipping model. Model coverage will be improved by acquiring more data on the number of vessels not covered by the datasets employed by the model.

The Climate TRACE dataset is available freely so that users can leverage the dataset to directly make impacts within the shipping sector. Table 11 (in section 3.3.6) shows the emissions increases by year for the Climate TRACE dataset, with an almost 79% increase in emissions from 2016-2022. This large increase could also be related to the increase in AIS coverage over the years, but even in the last few years (2019-2023) an increase of 21% has been seen. Organisations like the IMO are turning their focus towards attempting to implement regulations on shipping emissions to bring them into line with the Paris Accord, but currently there are no mandatory regulations in place. The hope is that our dataset can be used to help inform and drive conversation on this topic, highlighting the rapid increase in the emissions from global shipping to encourage policy change.

Bibliography

1. Adland, Roar, Pierre Cariou, and Francois-Charles 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* 140 (August): 101972. https://doi.org/10.1016/j.tre.2020.101972.
2. 'AIS Transponders'. n.d. Accessed 1 November 2023. https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx.
3. 'Area Database - Global Data Lab'. n.d. Accessed 31 October 2023. https://globaldatalab.org/areadata/.
4. Berthelsen, Frederik H., and Ulrik D. Nielsen. 2021. 'Prediction of Ships' Speed-Power Relationship at Speed Intervals below the Design Speed'. *Transportation Research Part D: Transport and Environment* 99 (October): 102996. https://doi.org/10.1016/j.trd.2021.102996.
5. Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. 'Mice: Multivariate Imputation fuelby Chained Equations in R'. *Journal of Statistical Software* 45 (December): 1–67. https://doi.org/10.18637/jss.v045.i03.
6. 'Carbon Monitor'. n.d. Accessed 19 October 2023. https://carbonmonitor.org/.
7. Clean Cargo Working Group (CCWG). 2017. '2016 Global Maritime Trade Lane Emissions Factors'. https://www.bsr.org/reports/BSR_CCWG_2016_Global_Maritime_Trade_Lane_Emissions_Factors.pdf.
8. Degiuli, Nastia, Ivana Martić, Andrea Farkas, and Ivan Gospić. 2021. 'The Impact of Slow Steaming on Reducing CO2 Emissions in the Mediterranean Sea'. *Energy Reports* 7 (November): 8131–41. https://doi.org/10.1016/j.egyr.2021.02.046.
9. EMSA. n.d. 'EU MRV Dataset'. Accessed 6 July 2023. https://mrv.emsa.europa.eu/#public/eumrv.

10. European Commission. Joint Research Centre. 2023. *GHG Emissions of All World Countries: 2023.* LU: Publications Office. https://data.europa.eu/doi/10.2760/953322.
11. GHG Protocol. n.d. 'Corporate Value Chain (Scope 3) Standard | GHG Protocol'. Accessed 25 October 2023. https://ghgprotocol.org/corporate-value-chain-scope-3-standard.
12. Greer, Krista, Dirk Zeller, Jessika Woroniak, Angie Coulter, Maeve Winchester, M. L. Deng Palomares, and Daniel Pauly. 2019. 'Global Trends in Carbon Dioxide (CO2) Emissions from Fuel Combustion in Marine Fisheries from 1950 to 2016'. *Marine Policy* 107 (September): 103382. https://doi.org/10.1016/j.marpol.2018.12.001.
13. Halim, Ronald A., Lucie Kirstein, Olaf Merk, and Luis M. Martinez. 2018. 'Decarbonization Pathways for International Maritime Transport: A Model-Based Policy Impact Assessment'. *Sustainability* 10 (7): 2243. https://doi.org/10.3390/su10072243.
14. IMO. 2015. 'Third IMO GHG Study 2014'. 2015. https://www.imo.org/en/ourwork/environment/pages/greenhouse-gas-studies-2014.aspx.
15. ———. 2021. 'Fourth Greenhouse Gas Study'. 2021. https://www.imo.org/en/ourwork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx.
16. ———. 2023. 'Revised GHG Reduction Strategy for Global Shipping Adopted'. 7 July 2023. https://www.imo.org/en/MediaCentre/PressBriefings/pages/Revised-GHG-reduction-strategy-for-global-shipping-adopted-.aspx.
17. March, David, Kristian Metcalfe, Joaquin Tintoré, and Brendan J. Godley. 2021. 'Tracking the Global Reduction of Marine Traffic during the COVID-19 Pandemic'. *Nature Communications* 12 (1): 2415. https://doi.org/10.1038/s41467-021-22423-6.
18. 'Maritime Safety Information'. n.d. Accessed 24 October 2023. https://msi.nga.mil/Publications/WPI.
19. Olmer, Naya, Bryan Comer, Biswajoy Roy, Xiaoli Mao, and Dan Rutherford. 2017. 'Greenhouse Gas Emissions from Global Shipping, 2013–2015', October.
20. Parker, Robert, Julia Blanchard, Caleb Gardner, Bridget Green, Klaas Hartmann, Peter Tyedmers, and Reg Watson. 2018. 'Fuel Use and Greenhouse Gas Emissions of World Fisheries'. *Nature Climate Change* 8 (April): 333–37. https://doi.org/10.1038/s41558-018-0117-x.
21. Poseidon Principles. 2021. 'Annual Disclosure Report'. 2021. https://www.poseidonprinciples.org/finance/wp-content/uploads/2021/12/Poseidon-Principles-Annual-Disclosure-Report-2021.pdf.
22. R Core Team. n.d. 'R: The R Project for Statistical Computing'. Accessed 6 July 2023. https://www.r-project.org/.
23. RightShip. n.d. 'Technical Information'. Accessed 6 July 2023. https://rightship.com/technical-information?section=rightship-s-ghg-rating.
24. Sampson, Hannah. 2020. 'All Major Cruise Lines Halt Sailings Temporarily in Response to Coronavirus'. *Washington Post*, 14 March 2020.

https://www.washingtonpost.com/travel/2020/03/12/princess-viking-cruise-lines-halt-all-sailings-temporarily-response-coronavirus/.
25. Ship it Zero. 2021. 'Home - Ship It Zero'. 2021. https://shipitzero.org/.
26. 'Siglar – Carbon Efficient Chartering'. n.d. Accessed 19 October 2023. https://www.siglarcarbon.com/.
27. 'The Pew Charitable Trusts'. n.d. Accessed 24 October 2023. https://www.pewtrusts.org/en?id=298.
28. 'Transparency for a Sustainable Ocean | Global Fishing Watch'. n.d. Accessed 31 October 2023. https://globalfishingwatch.org/.
29. UNFCCC. 2016. 'Shipping Aviation and Paris'. 18 May 2016. https://unfccc.int/news/shipping-aviation-and-paris.
30. Virdin, John, Tibor Vegh, Blake Ratcliff, Elizabeth Havice, Jack Daly, and Jack Stuart. 2022. 'Combatting Illegal Fishing through Transparency Initiatives : Lessons Learned from Timber Supply Chains'. *Marine Policy* 138 (April): 104984. https://doi.org/10.1016/j.marpol.2022.104984.

Data citation format: Mayes, B., Powell, M., Knights, D., Schofield, M., and Mackereth, T. (2024). *Transportation sector- Domestic and International Shipping Emissions.* OceanMind, UK and the University of Minnesota, USA, Climate TRACE Emissions Inventory. https://Climate TRACE.org [Accessed date].

Geographic boundaries and names (iso3_country data attribute): The depiction and use of boundaries, geographic names and related data shown on maps and included in lists, tables, documents, and databases on Climate TRACE are generated from the Global Administrative Areas (GADM) project (Version 4.1 released on 16 July 2022) along with their corresponding ISO3 codes, and with the following adaptations:

- HKG (China, Hong Kong Special Administrative Region) and MAC (China, Macao Special Administrative Region) are reported at GADM level 0 (country/national);
- Kosovo has been assigned the ISO3 code 'XKX';
- XCA (Caspian Sea) has been removed from GADM level 0 and the area assigned to countries based on the extent of their territorial waters;
- XAD (Akrotiri and Dhekelia), XCL (Clipperton Island), XPI (Paracel Islands) and XSP (Spratly Islands) are not included in the Climate TRACE dataset;
- ZNC name changed to 'Turkish Republic of Northern Cyprus' at GADM level 0;
- The borders between India, Pakistan and China have been assigned to these countries based on GADM codes Z01 to Z09.

The above usage is not warranted to be error free and does not imply the expression of any opinion whatsoever on the part of Climate TRACE Coalition and its partners concerning the legal status of any country, area, or territory or of its authorities, or concerning the delimitation of its borders.

Disclaimer: The emissions provided for this sector are our current best estimates of emissions, and we are committed to continually increasing the accuracy of the models on all levels. Please review our terms of use and the sector-specific methodology documentation before using the data. If you identify an error or would like to participate in our data validation process, please contact us.

1. Appendix A: Metadata descriptions

Table A.1 - Asset level metadata description for asset-climate_trace_domestic (international)-shipping-ship_MMDDYY

Data attribute	Definition
sector	Shipping
asset_sub-sector_name	Domestic/International Shipping
asset definition	Vessels
start_date	The date and time of the start of a trip / port stay (YYYY-MM-DD HH:mm:ss)
end_date	The date and time of the end of a trip / port stay (YYYY-MM-DD HH:mm:ss)
model_number	Version of model used to produce data
asset_identifier	Unique identification number for a vessel
asset_name	Vessel name
iso3_country	ISO3 country code with which the vessel is flagged
location	Lat/Lon of the departure and arrival ports given as a linestring
type	Vessel type
capacity_description	Gross Tonnage of the vessel
capacity_units	Gross Tonnes
capacity_factor_description	Ratio of distance travelled to gross tonnage (Activity / Capacity)

Data attribute	Definition
capacity_factor_units	Nautical Miles per Gross Tonnes
activity_description	The distance travelled by a vessel
activity_units	Nautical Miles
CO2_emissions_factor	The emissions per nautical mile calculated from CO2 emissions divided by activity
CH4_emissions_factor	The emissions per nautical mile calculated from CH4 emissions divided by activity
N2O_emissions_factor	The emissions per nautical mile calculated from N2O emissions divided by activity
other_gas_emissions_factor	The emissions per nautical mile calculated from the emissions of the relevant gas divided by activity (including SO_x, NO_x, VOCS, $PM_{2.5}$, PM_{10}, and CO)
CO2_emissions	CO_2 emissions estimated using the model
CH4_emissions	CH_4 emissions estimated using the model
N20_emissions	N_2O emissions estimated using the model
other_gas_emissions	Other gas emissions estimated using the model (including SO_x, NO_x, VOCS, $PM_{2.5}$, PM_{10}, and CO)
total_CO2e_100yrGWP	The effective CO_2 emissions over the next 100 year period
total_CO2e_20yrGWP	The effective CO_2 emissions over the next 20 year period
other1_description	Duration of the trip / port stay
other1_units	Hours : Minutes : Seconds
other2_description	Departure Port Name
other2_units	N/A
other3_description	Arrival Port Name
other3_units	N/A
other4_description	Departure Port ISO3 Country Code
other4_units	N/A
other5_description	Arrival Port ISO3 Country Code
other5_units	N/A

Data attribute	Definition
other6_description	Departure Port Unique ID
other6_units	N/A
other7_description	Arrival Port Unique ID
other7_units	N/A
other8_description	'isTrip' flag, True if trip, False if Port Stay
other8_units	Boolean
other9_description	Departure Port EEZ
other9_units	N/A
other10_description	Arrival Port EEZ
other10_units	N/A
other11_description	Deadweight of vessel (where available)
other11_units	Tonnes
other12_description	CO_2 emissions factor from the machine learning model
other12_units	kg CO_2 / Nautical Mile

Table A.2 - Port level metadata description for asset-climate_trace_domestic (international)-shipping-port_MMDDYY.

Data attribute	Definition
sector	Shipping
asset_sub-sector_name	Domestic / International Shipping
asset definition	Ports
start_date	The date and time of the start of the reporting period (YYYY-MM-DD HH:mm:ss)
end_date	The date and time of the end of the reporting period (YYYY-MM-DD HH:mm:ss)
asset_identifier	Unique Port ID generated by OceanMind
asset_name	Port name where available
iso3_country	ISO 3 country code of the country the port is associated with

Data attribute	Definition
location	Lat / Lon of the centre of the port location
type	N/A
capacity_description	Number of vessels that visited the port during the recording period
capacity_units	Integer
capacity_factor_description	Nautical miles per trip / port stay during the reporting period
capacity_factor_units	Nautical miles / Number of trips or stays
activity_description	Sum of the distance travelled by the vessels that visited the port during the recording period
activity_units	Nautical miles
CO2_emissions_factor	Total CO_2 emissions from all trips port stays divided by the total activity of all vessels that visited during the reporting period
CH4_emissions_factor	Total CH_4 emissions from all trips port stays divided by the total activity of all vessels that visited during the reporting period
N2O_emissions_factor	Total N_2O emissions from all trips port stays divided by the total activity of all vessels that visited during the reporting period
other_gas_emissions_factor	Total emissions of other gases from all trips port stays divided by the total activity of all vessels that visited during the reporting period (including SO_x, NO_x, VOCS, $PM_{2.5}$, PM_{10}, and CO)
CO2_emissions	Total CO_2 emissions attributed to the port. 50% of vessel trips that originate/arrive at this port, as well as 100% of port stays that occur at this port.
CH4_emissions	Total CH_4 emissions attributed to the port. 50% of vessel trips that originate/arrive at this port, as well as 100% of port stays that occur at this port.
N20_emissions	Total N_2O emissions attributed to the port. 50% of vessel trips that originate/arrive at this port, as well as 100% of port stays that occur at this port.
other_gas_emissions	Total emissions of other gases attributed to the port. 50% of vessel trips that originate/arrive at this port, as well as 100% of port stays that occur at this port.as (including SO_x, NO_x, VOCS, $PM_{2.5}$, PM_{10}, and CO)
total_CO2e_100yrGWP	The effective CO_2 emissions over the next 100 year period
total_CO2e_20yrGWP	The effective CO_2 emissions over the next 20 year period
other1_description	Exclusive Economic Zone associated with the port
other1_units	N/A

Table A.3 - Country level metadata description for country-climate_trace_domestic (international)-shipping_MMDDYY.

Data attribute	Definition
sector	Shipping
start_date	The date and time of the start of the reporting period (YYYY-MM-DD HH:mm:ss)
end_date	The date and time of the end of the reporting period (YYYY-MM-DD HH:mm:ss)
iso3_country	ISO3 code for the country reported on
CO2_emissions_tonnes	Total CO_2 emissions attributed to this country (after scaling)
CH4_emissions_tonnes	Total CH_4 emissions attributed to this country (after scaling)
N20_emissions_tonnes	Total N_2O emissions attributed to this country (after scaling)
other_gas_emissions_tonnes	Total emissions from other gases attributed to this country (after scaling) (including SO_x, NO_x, VOCS, $PM_{2.5}$, PM_{10}, and CO)
total_CO2e_100yrGWP	The effective CO_2 emissions over the next 100 year period
total_CO2e_20yrGWP	The effective CO_2 emissions over the next 20 year period