2,000 Digital Ships: A Scalable Method for Emissions Benchmarking and Independent Validation

Erik Verboom, AlbatrosDigital, Rotterdam/The Netherlands, erik@albatros.digital
Nico van der Kolk, AlbatrosDigital, Rotterdam/The Netherlands, nico@albatros.digital
Joris de Vroom, Hermess, Emmeloord/The Netherlands, joris.devroom@hermess.nl
Jos de Laat, Royal Netherlands Meteorological Institute, The Netherlands, jos.de.laat@knmi.nl

Abstract

This paper presents a scalable method for constructing large numbers of digital ship models from public specifications and AIS-derived operating profiles. Models are compared against EU MRV data to assess the match for predicting fuel use and CO_2 output. We also investigate correlations with satellite-based emission estimates, offering pathways for independent verification. Combining naval architecture, big data, and remote sensing, the approach supports benchmarking, compliance, and technology evaluation. Results prove a good match between simulation results and MRV reporting at only 2.9% median discrepancy with a 1σ spread of 31.9% on an individual ship-level. Ship-type and age-specific correlations are explored as further methods for reducing the spread.

1. Introduction

The objective of this study is to evaluate the consistency of physics-based ship performance modelling when scaled from single-vessel applications to fleet level. While individual ship performance analysis is well established, its industrialisation across thousands of vessels introduces challenges in data quality, automation robustness, and methodological comparability. This paper investigates how such a pipeline performs when applied systematically to multi-year operational histories across a heterogeneous fleet.

We present a modular modelling workflow that integrates AIS-derived vessel tracks, publicly available ship particulars, and hindcast metocean data to estimate fuel consumption over multi-year periods. The pipeline is designed for reproducibility and minimal ship-specific tuning, enabling large-scale application without manual intervention per vessel.

To evaluate modelling behaviour at scale, results are benchmarked against EU THETIS-MRV data. MRV reporting is treated as a standardized regulatory reference dataset rather than absolute ground truth. The purpose of the comparison is therefore not to certify individual ship emissions, but to assess systematic agreement patterns, bias structure, and dispersion characteristics across a statistically meaningful fleet sample.

The analysis focuses on fleet-level accuracy characteristics over a six-year period (2018–2023), covering more than 1600 ship-years of simulation. By examining aggregate bias, spread, and structural correlations with ship type and age, the study provides insight into where large-scale physics-based modelling performs consistently and where systematic deviations remain.

The contribution of this work lies not in replacing regulatory reporting, but in demonstrating that fleet-scale digital ship modelling can produce statistically coherent results when benchmarked against standardized emissions data. The findings inform both the robustness of large-scale modelling approaches and the limitations that must be considered when interpreting such comparisons.

2. Data sources

To enable the complete modelling pipeline, several external data sources will be used. Where possible, open-source datasets were used, but when necessary, commercially available datasets were used as a more complete fallback option.

- EU THETIS-MRV: ship-year CO_2/fuel, distance, time at sea/berth, transport work; used strictly for external validation.
- AIS: Historic location and navigation status data per ship; used to infer operational states and reconstruct voyages.
- Technical registers: Main particulars of vessels. Used for vessel modelling and scaling from similar known models.
- Metocean: Global hindcast data on wind, waves and currents. Used to model voyages and estimate the fuel consumption at location – timestamp combinations.
- Port locations: An open-source list of 4000 ports worldwide with their locations.

2.1. EU THETIS-MRV

THE¬TIS-MRV is the EU system/platform run by EMSA (European Maritime Safety Agency) for Monitoring, Reporting, and Verification (MRV) of fuel consumption and CO_2 (and other GHGs) emitted by large ships calling at EU/EEA ports. It supports companies (shipowners/operators), verifiers, and administering authorities in fulfilling their legal obligations under the EU MRV Regulation (Regulation (EU) 2015/757) and its amendments. It involves Monitoring Plans, Emission Reports, verification, and public reporting. THETIS-MRV reporting is mandatory since 2018 for each vessel that has had at least one port call in a given year. Originally, this was for all vessels larger than 5000 GT, as per 2025 some vessels between 400 and 5000 GT are also required to report their emissions.

THETIS-MRV data is reported to EMSA on a yearly basis. Reporting of fuel consumption and emissions data within this framework is done after third-party validation. Nevertheless, one long term goal of this research is to enable independent validation of reported emissions data, for example by comparing the reporting to satellite observations. Relevant to this study are reporting details such as how a voyage is defined (port-to-port), which voyages are reported (at least one port inside the EU), which ports count as EU ports, how to handle voyages that start in one year and finish in another, etc. Section 3.2 provides further details on the methodology used enable replication of MRV reported data.

2.2. AIS Tracks

AIS (Automatic Identification System) is a vessel tracking technology that was introduced to improve maritime safety and collision avoidance. The system is mandatory for all passenger ships and all cargo ships greater than 500 GT on all voyages. In practice, this means that all vessels of interest to this study have AIS transceivers installed.

AIS messages are transmitted by the ship using VHF radio and can be picked up (and forwarded) by other ship- and shore station (Terrestrial AIS) as well as satellites (Satellite AIS). Combined, a global location history of a vessel can be reconstructed, with a new location every few minutes on average. Commercial datasets with vessel histories going back to the early 2000s for the entire merchant fleet are available and have been used for this study. Their completeness and correctness have been evaluated based on individual sample ships in the dataset of this study.

An AIS message contains several fields regarding vessel information: it's current voyage and navigational status. It is worth noting that commercial providers of AIS datasets will often divide the complete set of transmitted data into several packages, to be sold separately. Another drawback of AIS is that the data is self-reported, and for some fields manually entered, leading to inaccurate values and significant differences between individual ships.

For this study, the fields of relevance are: Latitude, Longitude, Timestamp, Speed, Draught, and Navigational status code.

Within the context of this study, AIS data has been used to construct a history for each vessel within the timeframe of interest. Speed and navigational status have been used to estimate when separate voyages start and end. Furthermore, the draught field has been used to correctly model each voyage and provide accurate fuel consumption estimates.

2.3. Vessel Data

One of the goals of this study is to get an idea of vessel modelling accuracy based on a minimal set of vessel-specific information. The vessel specific datapoints that have been used for this study are: Length over all, Beam, and Draft.

The rest of the required modelling is obtained and scaled within the AlbatrosDigital modelling engine from comparable ships of the same type.

2.4. Metocean data

The metocean data considered for this study consists of wind data, wave data and current data. For each, different models have been used to provide the most accurate hindcast weather.

2.4.1. Wind / Waves

The global hindcast was generated by Oceanum with the third-generation spectral wave model WAVEWATCH III, *Tolman (1991),* with forcing the ERA5 dataset to produce a global wave hindcast of unprecedented quality. The hindcast was run using the recently released version 6.07 of WAVEWATCH III. ST4 source terms were used *Ardhuin et al. (2010)*, which have been established over the past decade as the most suitable for global application, e.g. *Stopa et al. (2015)*. The model uses ERA5 hindcast wind-data obtained from ECMWF.

In addition to the global model several regional hindcasts are generated several regional grids have been generated to ensure high quality data that covers the globe (https://datasets.oceanum.io/). The regional hindcasts are generated with SWAN ST6 with a resolution of 0.05° x 0.05° and are calibrated and validated against satellite measurements. The models are forced with ERA5 wind fields.

The regional SWAN models are nested in our 0.5° global WaveWatch III model. This way the swell from the oceans is properly taken into account. The specifications for regional SWAN models are:

- WW3 Spectral boundaries
    - 0.5°
    - ERA5 forcing
    - Extremes correction
- Spectral output typically at 0.05° x 0.05° resolution
- Validated against satellites

2.4.2. Currents

Total surface current data are obtained by linearly superimposing tidal current components on residual ocean currents. For tidal data, the global Oceanum tidal dataset with a spatial resolution of 400 m is used. Here, the tidal constituents have been downscaled from TPXO9 global tidal solution using the OTIS model of barotropic ocean tides, *Egbert and Erofeeva (2002)*. Current residual velocities, sea surface elevation, temperature and salinity will be extracted from GLORYS, a global 1/12 ° reanalysis product released by the EU-funded Copernicus Project. Atmospheric forcing data, including 10 m wind speed, mean sea-level pressure, precipitation, temperature, humidity and solar radiation will be sourced from the ERA5 reanalysis product.

Non-tidal (residual) currents and water level will be obtained from the Mercator 9 km Global reanalysis. The reanalysis includes daily means of temperature, salinity, currents, sea level, mixed layer depth and ice parameters from the top to the bottom over the global ocean. It is based largely on the current real-time global forecasting CMEMS system. The model component is the NEMO platform driven by ECMWF winds. Observations are assimilated by means of a reduced-order Kalman filter. Along track altimeter data (Sea Level Anomaly), Satellite Sea Surface Temperature, Sea Ice Concentration and in-situ Temperature and Salinity vertical Profiles are jointly assimilated. Moreover, a 3D-VAR scheme provides a correction for the slowly evolving large-scale biases in temperature and salinity.

2.5. Port locations

An extensive and open-source list of about 4000 worldwide port locations has been used for voyage identification. The complete list can be found on https://github.com/tayljordan/ports.

3. Methodology

The modelling for this study takes place in multiple steps. An individual model of each ship will be made. This model will contain a fuel table detailing the expected – time-averaged – fuel consumption in every possible operational state (speed, draft, metocean parameters).

With these individual vessel models, voyages are simulated over an analysed timeframe of 6 years (2018 to 2023). The historical vessel tracks are divided into separate voyages. For each voyage, the fuel consumption at each track point is calculated and added up to aggregate over desired timeframes.

At last, aggregated numbers on distances sailed, days at sea and fuel consumption can be computed. Great care is taken to follow the methodology as prescribed by the EU regulations to ensure that results can be reliably compared to reported in the EU THETIS-MRV datasets.

3.1. AlbatrosDigital vessel modelling

AlbatrosDigitals software is a ship simulation tool based on a performance prediction program. Around that, a cloud-based modelling suite of naval architecture and marine engineering tools has been developed aggregating vessel information, metocean data and vessel history. At its core, the software uses a constrained-optimization solver that computes a ship's response to maintain a steady course in 4-dof under given conditions such as draft, ship speed, wind conditions and wave conditions. An adjustable level of simulation physics is available to match the modelling purpose, or available ship information. The AlbatrosDigital software is highly customizable, allowing users to easily include their own specific models and requirements.

For the presented study, the challenge was to work with a minimal set of ship-specific information but keep the modelling accuracy as high as possible. This has been done by working with a total of 48 reference ships of various ship types and sizes. Detailed ship characteristics such as calm water resistance curves, propulsive efficiencies and engine/fuel types for each individual model have been taken from the closest reference ship and scaled according to the main dimensions where appropriate. A sufficient similarity score between reference ship and individual model has always been ensured.

3.2. EU THETIS-MRV definitions and details

To be able to compare simulation results to the reported MRV data, it is essential to follow the methodology prescribed by the regulation. This encompasses the definition of exactly which fuel consumption and emissions are reported, the definition of a “voyage”, which voyages do count for the reporting, and many more details.

3.2.1. Voyage definition

A “voyage” according to EU MRV regulation is always berth to berth. Each voyage will start and end with a port call. Only port calls where cargo is loaded/unloaded and/or where passengers embark/ disembark are counted. Events like anchoring, refuelling, pilot boarding, changes of crew etc do not break a voyage, even if this happens in a port. Anchoring time, while it does not break a voyage, is not counted as “hours at sea”, a metric that is also reported on a in the MRV reports, https://www.dnv.com/maritime/insights/topics/mrv/FAQs-EU-MRV/.

A voyage is always reported entirely in the year in which the start date of the voyage falls. For example, a voyage that starts in December 2018 but ends in January 2019, should be reported in 2018.

3.2.2. Voyage reporting

Voyages must be reported if the start and/or end port lie within EU territory. EU territory is defined as ‘mainland’ EU and several overseas territories: Açores, Madeira, Canarias, Guadeloupe, French Guyana, Martinique, Mayotte, Saint Martin and Reunion. Next to that, also Norway (except Svalbard) and Iceland ports are also included in the EU MRV reporting.

If a voyage starts in one year and finishes in the next, it shall be reported in the first year, the year of departure, https://www.lr.org/en/services/statutory-compliance/eu-ets-and-eu-mrv/eu-mrv-regulation/.

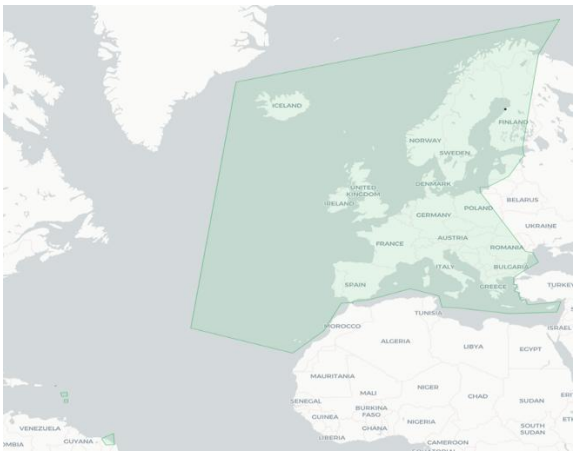


Fig.1: EU geofence for identifying EU-connected voyages. EU mainland and overseas territories in the Caribbeans

3.2.3. Fields of interest and definitions

The public MRV reports have multiple fields that are of interest in this study, and a few assumptions need to be made to be able to make sensible comparisons to simulations:

- Total fuel consumption [t]: This is the total fuel that is consumed within the Thetis MRV jurisdiction (EU connected voyages + EU ports)
- Annual average Fuel consumption per distance [kg/nm]
- Total CO_2 emissions [t]: Total emissions: in EU ports + EU connected voyages
- CO_2 emissions which occurred within ports under a MS jurisdiction at berth [t]: in-port emissions

It is important to note that both the fuel consumption and CO_2 fields refer to the total amounts for the entire ship, including main engine and any generators that might be operating. To make the simulation vs reporting comparison that this study tries to achieve, the total distance sailed annually is calculated based on the first two fields.

$$Distance\ sailed\ [nm] = \frac{FC\ total\ [m\ tonnes]}{FC\ per\ nm\ [\frac{m\ tonnes}{nm}]}$$

In a second step, to estimate the fuel consumption attributed to voyages only, we look at the CO_2 emissions in-port vs total:

$$FC\ at\ Sea\ [m\ tonnes] = FC\ total * (1 - \frac{In\ Port\ CO2\ [m\ tonnes]}{Total\ CO2\ [m\ tonnes]})$$

This second step makes the assumption that the emissions per kg of burned fuel are identical in port and at sea, which is a fairly dramatic simplification glossing over things like shore power, dual-fuel systems and much more. However, for the goals of this study, this crude approximation should serve well.

3.3. AIS tracks

One of the main challenges for this study is to correctly interpret a vessels history from its AIS data. Specifically, the AIS track needs to be separated into voyages, voyages need to be categorised into 'intra-EU', 'entering EU' and 'leaving EU'. Each voyage can be simulated in the AlbatrosDigital Virtual Voyage in order query the correct historic metocean data and calculate aggregated variables such as fuel consumption and voyage duration.

3.3.1. Down-sampling

To keep the total amount of points manageable and to keep the required compute within the budget, all AIS tracks have been down sampled to one point per hour. Larger gaps that occurred within tracks were also filled up with one point per hour, assuming a straight-line trajectory of the vessel.

3.3.2. Identifying voyages

The first step is to divide the AIS track into separate voyages. Several methods have been investigated to determine whether a new voyage has been started. Looking at the reported speed, destination, changes in draft and navigation status, or any combination of the above have been tried out. It turns out that the accuracy of reported (AIS) data can be very questionable, Fig.2.

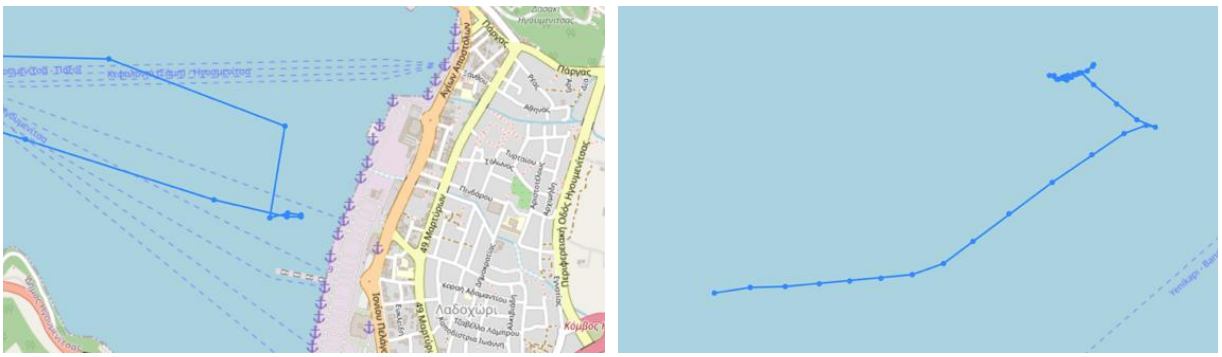


Fig.2: Discrepancies in navigation status: A ferry berthing for about 1.5 h, but never reporting status 5 (left), 1 h of sailing while the status says "anchoring". (right)

In the end, to keep the algorithm simple, fast, and applicable to datasets from various sources, we focussed on a combination of navigation status, reported speed and harbour proximity. To validate the voyage identification, resulting yearly sailed miles have been compared to the MRV reports. The mismatch between yearly reported distance and yearly sailed distance can be seen below. This has been used as one of the main criteria to decide which vessels to include in the results.

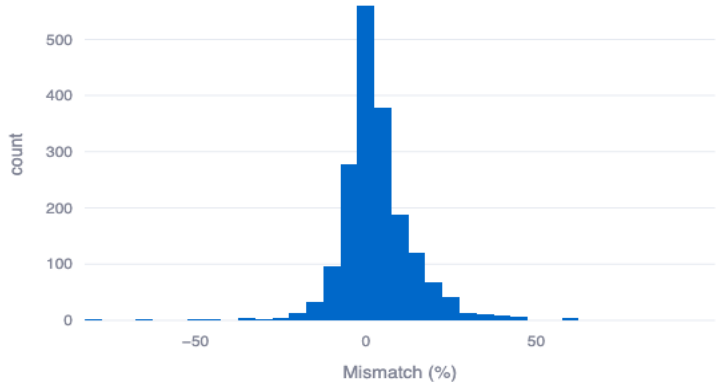


Fig.3: Matching sailed miles to reported miles per ship over a 6-year period (2018 to 2023). Y-Axis counts ships, X-Axis the percentage difference between reported and simulated voyages. Half of the vessels (959 / 1848) are below 5% mismatch.

4. Vessel Selection

While the title of the study says 2000 digital ships, the reality is a bit more complicated than that. Starting from 21.063 ships that are present in the MRV reporting data, through multiple steps and due to various boundary conditions, the set of ships has been reduced to 270 ships that have been modelled for the complete 6-year period. The objective of this selection process was to keep the ship set representative of the MRV fleet while complying with hard technical constraints and minimising the influence of data or pre-processing errors. This way the study can provide a complete and fair indication of the match that can be obtained between simulation data and reported data.

4.1. Ship type

To be able to create a model with only publicly available data of a ship, the Digital Shipyard needs to take the ship type into account. Reality is messy and the ship types reported in the Thetis MRV dataset will contain exotic values. Also, it turns out that – at least according to the reporting – ships seem to change type rather frequently. One year a vessel might be a general cargo ship, and next year it suddenly becomes a Bulk carrier. After cleaning up the ship type categorization, a complete table of ship types and the number of times they occur in the dataset can be found in Table I.

Table I: Ships in MRV dataset, by ship type

Ship type	**Number**	**Normalized**	**To Model**
Bulk carrier	7.365	35 %	Yes
Oil tanker	3.388	16 %	Yes
Container ship	2.967	14 %	Yes
Chemical tanker	2.110	10 %	Yes
General cargo ship	1.691	8 %	Yes
Gas carrier	626	3 %	Yes
Vehicle carrier	615	3 %	Yes
LNG carrier	492	2 %	Yes
Ro-pax ship	468	2 %	Yes
Passenger ship	290	1 %	No
Ro-ro ship	288	1 %	Yes
Refrigerated cargo carrier	200	1 %	No
Container/ro-ro cargo ship	70	0 %	Yes
Combination carrier	15	0 %	No
Passenger ship (Cruise Passenger ship)	10	0 %	No
Other ship types	468	2	No
Totals	**21.063**		**20.080**

About 983 ships, or about 4.6% of the ships from the dataset are excluded due to a ship type that can't easily be created with AlbatrosDigitals fully automated modelling approach. For the modelled fleet that will be modelled, we will try to approach the percentages per ship type found in Table II.

4.2. Reporting periods

A second factor for vessel selection is the total number of sailed miles that a vessel has reported and the number reported data points. It is relevant to have a sufficiently long timespan to compare simulation to reported data for each ship. This is to ensure that reporting errors and data anomalies will be averaged out instead of skewing the study results. For this reason, ships that have many miles reported are preferred.

The choice has been made to only consider ships that reported to MRV-THETIS every single year between 2018 and 2023. These will be ships that have a large reported milage on one hand, and a total of six datapoints for comparison (one for each year). This criterium leaves 4.219 ships for this study.

4.3. Reference models

To be able to create a detailed and accurate model of each individual ship based off only main particulars, AlbatrosDigital works with a system of reference ships. To ensure model accuracy, a similarity score between the ship to model and the reference ship is calculated based on the main particulars. If this similarity score is below a certain threshold, no model will be created. This restriction further limits the number of ships suitable for this study. Especially small general cargo ships show a wide variance in design choices, making them less suitable for this approach. In combination with the desire to keep ship sizes and ship-types representative of the entire fleet, this limits the ships suitable for this study to 1.848.

4.4. Sailed distance

A check is done comparing the reported miles sailed in each year to the calculated miles sailed each year for each ship. Due to the specific definitions and reporting rules of THETIS-MRV and inaccuracies in the AIS data and the voyage identification algorithm, a good match on sailed distance was not always possible.

Since a mismatch in sailed distance will inevitably lead to differences in fuel consumption, a threshold has been set at a maximum of 5% discrepancy over the 6-year period. If this threshold is exceeded, the ship has been excluded from this study. This leaves 959 vessels for this study.

4.5. Simulation time

Finally, the compute budget for the study is limited, making it impossible to simulate all 6 years for all 959 vessels. The selection of ships and/or simulation timeframe needs to be reduced further. In the end the choice has been made to limit the simulations to 270 vessels but keep the entire 6-year period, covering over all 1620 ship-years of tracks.

5. Results

As mentioned before, the simulation runs for this study cover 1620 ship-years of sailing. For each ship-year there are two fuel consumption estimates available: the yearly MRV reported value and the value that results from the simulation. This study does not aim to take any of these sources as the "ground truth". Nevertheless, the good match – looking at mean and median – suggests that both fuel consumption estimates yield plausible results. As expected for such a large sample size and real-world data, there is a decent spread in data-match. The results presented in this paper will focus on analysing the mean, the spread and various interesting correlations that can be found in the data.

5.1. Data match

The first thing for analysing the results of this study is to have a look at the over-all match between reported and simulated fuel consumption in Figure 4. The simulation results match the reported data well: The simulation mean predicts a 5.18% higher fuel consumption than reported, with the median discrepancy being 2.89%. Except for some explainable outliers, all datapoints lie within the bandwidth of ±50%. It is worth noting that the spike at 100% is caused by several outliers. This bin collects anything > 100% error. Outliers that are collected in this bin and that skew the mean can be explained mostly by clear misreporting over one specific year (see section 6.3.2 for an example).

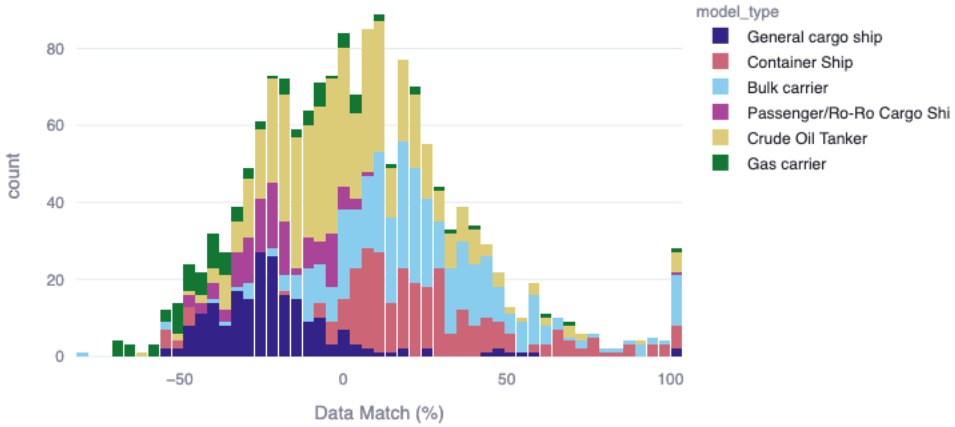


Fig.4: Match between simulated and reported fuel consumption. Median deviation is 2.89% while the 1σ spread is 31.94%.

5.2. Correlations

To understand the source of this spread further and potentially allow further analysis and improvement of modelling, correlations between various variables and the fuel consumption have been analysed. The most clear and strong correlations are discussed below.

5.2.1. Sailed distance

As expected, the mismatch in sailed distance between simulation and reported data translates pretty much on-to-one to a mismatch in predicted fuel consumption mismatch, accounting for about 10% of the observed spread in the data-match.



Fig.5: Correlation between sailed distance error and fuel consumption mismatch

5.2.2. Ship type

As can be seen in Fig.4, both the mismatch in fuel consumption and the spread between individual ships strongly depends on the ship type. The simulation seems to underpredict for Bulk carriers and Container vessels, while it overpredicts for General cargo ships, Ro-Ro ships and Gas carriers. Digging into the cause of this effect deeper is a clear recommendation for further research. Currently, the data does not provide satisfactory explanation for these trends. Some hypothesis might be that there are variations in hotel-load that are not modelled correctly. Another hypothesis is that the chosen hull shapes that represent each ship type differ quality.

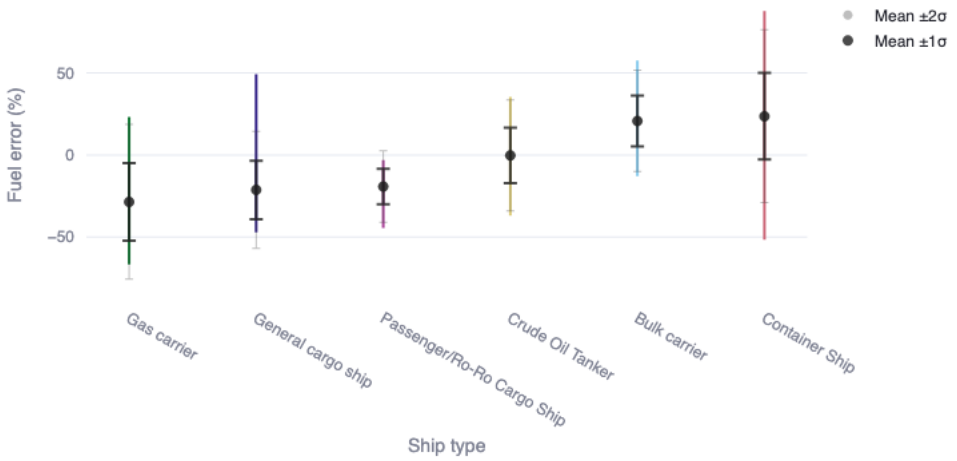


Fig.6: Spread and mean error per ship type

5.2.3. Year of build

Another Interesting correlation is the year of build compared to the fuel consumption mismatch. The modelling did not consider any penalty for older ships. This shows that – on average – newer ships are indeed more efficient than older ships. This effect seems to be quite significant over longer time periods.

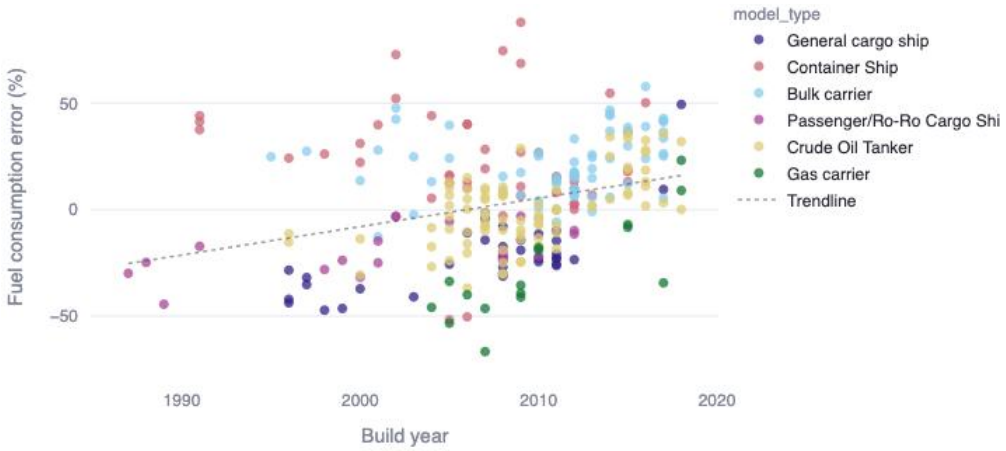


Fig.7: Correlation between year of build and data match. The Pearson correlation is about 0.3, indicating weak to moderately linear relationship.

5.3. Outliers

Apart from the discussed structural conclusions drawn from this study, it is worth to have a quick look at extreme outliers. These cases provide some interesting insights into the data quality, and where we might find which failure modes exist for such a large-scale study. While the provided plots are

extreme examples, various more subtle cases and outliers are found in the AIS dataset and reporting. Such extreme outliers could be filtered out by the data-pipeline before any computationally heavy simulation steps are taken. Meaning that these kinds of outliers do not skew the over-all results of the fuel consumption comparison between simulations and MRV reporting.

5.3.1. Misreporting

Below is the track of a single vessel for one year. According to the MRV reporting data. This vessel only sailed 400 nm of voyages that started or ended inside the EU. This kind of under-reporting was found in a hand full of cases. Due to the vessel selection criteria only looking at the total reported mileage, not on a per-year level, this specific case, as well as a handful of similar cases are still part of the reported results and explain the >100% discrepancy bin in Fig.4.

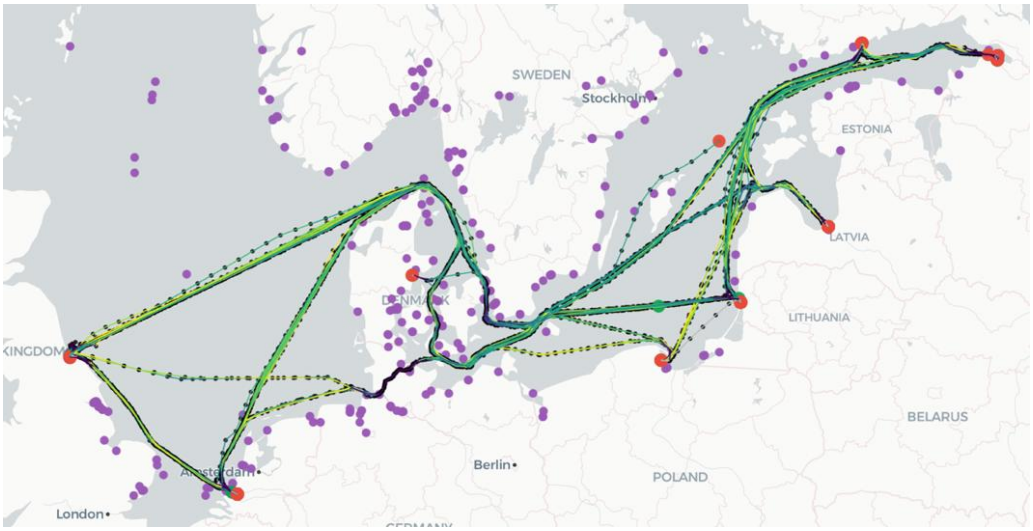


Fig.8: An underreported track: This is the 2018 track of a ship that reported only 400 nm sailed in that year.

5.3.2. Faulty AIS signals

It should come as no surprise that the AIS sailed distance of this vessel is far larger than the reported mileage. In fact, over all six years, the difference amounts to about 300%. Such extreme cases were filtered out of the reported results due to the large discrepancy in reported vs AIS track miles.

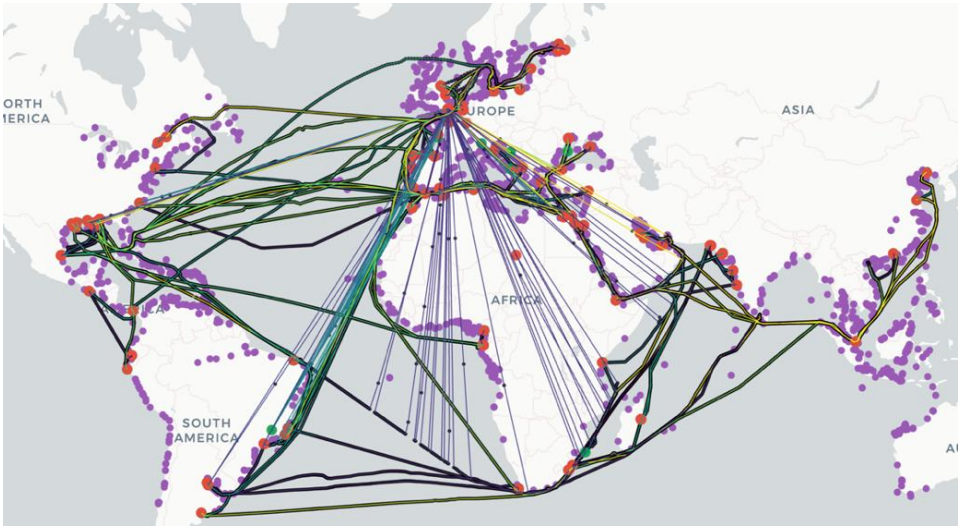


Fig.9: Faulty location signals lead to largely overestimated sailed distances based on AIS tracks

6. Satellite measurements

Satellite observations of nitrogen dioxide (NO2) have a proven history of detection of shipping lanes, and more recent advances in earth observation – in particular the ESA Sentinel-5p satellite and its TROPOMI instrument - have brought detection of emission plumes from individual (large seafaring) ships within reach, Fig.10, see also *Kurchuba et al. (2024)*. As shipping NO_2 emissions from Sentinel-5p are a reliable proxy for shipping CO_2 emissions this opens the envelope for monitoring ship performance and regulation compliance. Monitoring currently consists of incidental daily snap shots when a ship emission plume is sampled under the right viewing conditions. That allows for building a large database of ships and emissions, which can be resampled according to certain ship specifications for which – due to regulations - differences in emissions are expected, *Riess et al. (2025)*. Lack of such differences strongly indicates that regulations have not been effective.

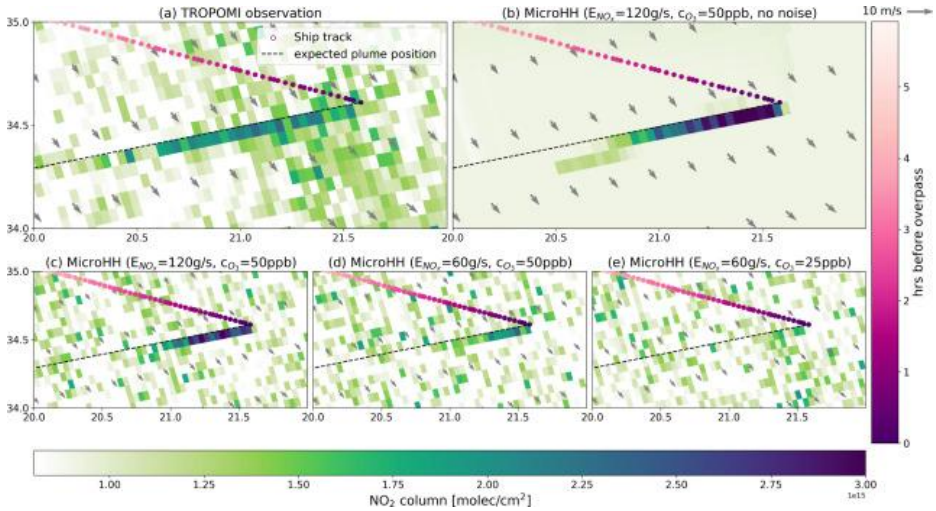


Fig.10: NO_2 plume of a ship sailing in the Mediterranean on 5 June 2019. (a) Observed TROPOMI data, (b) MicroHH simulation resampled at TROPOMI grid using an emission strength of 120 g s^{-1}, (c) like (b) but with realistic instrument noise, (d) like (c) but with half the emission flux and (e) like (d) but with half the background ozone. The pink dots indicate the AIS position of the ships in the three hours before overpass at 11:42 UTC as indicated by the right color bar, the black dashed line shows the wind-shifted track, which is the ship location displaced the 10 m wind and time difference between the TROPOMI measurement and ship location. The choice for halving the emission is motivated by the emission limit of Tier III ship, the choice for halving the background ozone is motivated by the difference in ozone levels between the eastern Mediterranean and the North Sea. The NO_2 emissions given here and reported elsewhere in this manuscript are given as mass of NO_2, meaning that an emission of 46 g s^{-1} equals 1 mol s^{-1}. Source: *Riess et al. (2025).*

7. Conclusions and recommendations

In conclusion, this study has evaluated more than 1600 ship-years of ship histories, a total of about 82 thousand voyages and 73 million nm of sailed distance. The first conclusion that can be drawn is that the quality of various data sources – AIS tracks, metocean hindcast, vessel particulars, etc. – is sufficient for such large scale and multi-year simulations. The data pipeline from raw data, automated cleaning and filtering to the final, processable vessel histories yields a dataset that is representative of a heterogeneous fleet across many ship types, operating conditions and years.

Secondly, studying physics-based ship modelling on a fleet level leads to very usable results. The average and median of computed fuel consumption lie very close to what is being reported, suggesting good high-level assumptions, good average match between assumed and actual vessel characteristics and good modelling methodology.

Finally, it can also be concluded that the fleet-wide spread is still significant. While averages and means match well, the 2-sigma level boundary does approach 50% mismatch between simulation and reporting. To reduce this spread, this study provides multiple correlations that can serve as a starting point to further improve fleet-wide modelling. Looking closer at ship-type specific factors such as hotel-load, hull shape and engine characteristics should enable a better match for individual ship types. Introducing a built-year factor to account for newer ships becoming more efficient should further reduce the fleet-wide spread. These are clear recommendations for follow-up research to significantly reduce the spread in fleet-wide studies.

Another question is how accurate the two discussed methods for estimating vessel efficiency are. The fact that the discussed methods match well hints at a decent accuracy, however both – simulations and reporting – come with their own set of challenges, and each can be susceptible to different kinds of errors and biases. In this respect, a very promising recommendation for further work is to look into measuring ship emissions using Satellite observations as discussed in section 6. This could provide a third point of calibration, enabling well-founded conclusions on over-all biases between different efficiency estimation methodologies.

References

ARDHUIN, F.; HERBERS, T.H.C.; JESSEN, P.F.; O'REILLY, W.C. (2003), *Swell transformation across the continental shelf. Part II: validation of a spectral energy balance equation*, J. Phys. Oceanogr. 33, pp.1940-1953

EGBERD, G.D.; EROFEEVA, S.Y. (2002), *Efficient Inverse Modeling of Barotropic Ocean Tides*, J. Atmospheric and Oceanic Technology 19, pp. 183-204

KURCHABA, S.; et al. (2024), Sensitivity analysis for the detection of NO_2 plumes from ships using TROPOMI, Remote Sensing of Environment

RIESS, T.C.V.W.; et al. (2025), Emissions of individual ships from TROPOMI NO_2 plumes, Remote Sensing of Environment

STOPA, J.E.; ARDHUIN, F.; BABANIN, A.; ZIEGER, S. (2015), *Comparison and validation of physical wave parameterizations in spectral wave models*, Ocean Model 103, pp.2-17

TOLMAN, H.L. (1991), *A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents*, J. Phys. Oceanogr. 21, pp.782-797

VAN DER KOLK, N.; TARDIF, T.; STELLA, F.; JACOBI, G. (2025), *Aero-hydrodynamic inter-actions for wind powered ships: retrofit of a high-speed mega RoRo vessel*, HIPER Symp., Tullamore