Ship-Level Fuel Consumption from Bulk Shipping

Oliver Xu

March 1, 2024

1 Engineering Estimation of Fuel Consumption

A vessel's hourly fuel consumption (FC) is specific to each of the three types of on-board machinery: the main engine (ME), auxiliary engine (AE), and auxiliary boiler (AB). In 2012, these machinery contribute 71%, 25% and 3.7% of total fuel respectively. Overall, they can be all be described by the same engineering estimation method, which can be divided into two parts. Depending on the given engine machinery, a vessel's hourly fuel consumption is the product of its hourly engine power demand (W_i) and specific fuel consumption (SFC_i).

$$FC_i = W_i \cdot SFC_i.$$

1.1 Estimation of the hourly engine power demand

1.1.1 Main Engine Power (Admiralty formula)

The main engine power required by a vessel, also known as the propulsive power, is calculated using the **Admiralty formula**. This formula takes into account the ship's speed and its resistance. The resistance of a ship is the force that the ship needs to overcome to move through the water. It's composed of hydrodynamic resistance (resistance due to water) and aerodynamic resistance (resistance due to air).

- **Hydrodynamic and Aerodynamic Resistance:** Hydrodynamic resistance is formed by the frictional resistance (due to the interaction of the ship's hull with water) and the residual or wave-making resistance (due to the formation of waves when the ship moves in the water). Aerodynamic resistance is caused by the ship's exposed surfaces moving through the air. Both these forces are modified by weather conditions, which can change the speed, direction, and frequency of winds and waves.
- **Hull Surface Conditions:** The condition of the ship's hull also affects the hydrodynamic resistance. Changes on the hull surface, such as coating deterioration, fouling growth, and plating deformation due to wear and tear, can significantly increase the frictional resistance.

- **Additional Variables:** Other variables that need to be considered for the ship's propulsive needs are the ship loading condition, the weather modifier to the ship's propulsive efficiency, and the fouling modifier. A correction factor is applied to certain ship types and sizes to adjust the speed-power relationship.

- **Quantify propulsive power demanded:** To sum up, the following equation takes account of all effects mentioned above:

$$W_i = \frac{\delta_w \cdot W_{\text{ref}} \cdot \left(\frac{t_i}{t_{\text{ref}}}\right)^m \cdot \left(\frac{v_i}{v_{\text{ref}}}\right)^n}{\eta_w \cdot \eta_f}$$

  - W_i: The power demand for each hourly observation of the given engine system.
  - δ_w: The speed-power correction factor, assumed to be 1 (except large container and cruise).
  - W_{ref}: The reference power as given in the WFR dataset. (extracted from main engine details column)
  - t_i and v_i: The instantaneous draughts and speeds respectively, given by the AIS dataset.
  - t_{ref} and v_{ref}: The reference draught and speed from the WFR dataset. (maximum speed and draught)
  - m: The draught ratio exponent, assumed to be 0.66.
  - n: The speed ratio exponent, assumed to be 3.
  - η_w: The weather modifier to the ship's propulsive efficiency, assumed to be 0.909 for mainly small ships and 0.867 for all other ship types and sizes.
  - η_f: The fouling modifier, assumed to be 0.917 for all ship types and sizes.

  The equation is further divided into three parts for calculation purposes:

$$W_i' = \frac{\delta_w}{\eta_w \cdot \eta_f \cdot {t_{\text{ref}}}^m \cdot {v_{\text{ref}}}^n} \cdot W_{\text{ref}} \cdot (t_i^m \cdot v_i^m)$$

  - $\frac{\delta_w}{\eta_w \cdot \eta_f \cdot {t_{\text{ref}}}^m \cdot {v_{\text{ref}}}^n}$: this part is derived from the WFR dataset.
  - W_{ref}: this part is derived directly from the WFR dataset (text extraction from main engine details).
  - $(t_i^m \cdot v_i^m)$: this part is derived from the AIS dataset.

1.1.2 Auxiliary Machinery Power (vessel size & operational phase)

Power demand by the auxiliary engine and boiler systems per ship type, size, and operational mode are scarce data. Unlike the main engine, the power demand of auxiliary engines and boilers is typically estimated using a different methodology that is more suited to the specific characteristics and operational modes of them. Auxiliary engines are used to generate electrical power for various systems on board a ship, such as lighting, heating, cooling, and other electrical systems. They are also used to power the ship's propulsion system when the main engines are not in use, such as when the ship is at berth or anchored.

- **Operational Phase Assignment:**

  The power demand of the auxiliary engines can vary depending on the operational mode of the ship. For example, when a ship is at sea and its main engines are in use, the power demand of the auxiliary engines might be lower. Conversely, when the ship is at berth or anchored and the main engines are not in use, the power demand of the auxiliary engines might be higher.

  According to the IMO report, they had access to on-board data for fuel consumption and power output, including for auxiliary engines, at different operational modes. This data was used to classify the auxiliary engine power output per operational mode. They also assume that boilers are not used during open-ocean operations (i.e., at sea operation mode) since the ships are assumed to have a Waste Heat Boiler (WHB) installed on-board that reuses the waste heat coming from the main engine and fully covers the heating demand. In our study, we use Table 16 from the IMO report to assign operational phase for each vessel's hourly observation and Table 17 from the IMO report to assign hourly power demand for auxiliary machinery.

- **Auxiliary Engine and Boiler Power Tables:**

  - *Table 16: Operational phase assignment decision matrix*

    In our study, we simplified Table 16 by using only a ship's proximity to land and speed over ground to determine its phase. If the SOG is less or equal to 3 knots, then the ship is anchored regardless of its distance from the coast. If the SOG ranges between 3 and 5 knots (incl. 5), then ship is manoeuvring if its coast distance is less than 5 nautical miles and the ship is at sea otherwise. If the SOG is above 5 knots, then the ship is at sea regardless of its distance from the coast.

  - *Table 17: Auxiliary engine and boiler power output, by ship type, size and operational mode*

    - When main engine power is between 0 and 150 kW then auxiliary engine and boiler are set to zero;

- When main engine power is between 150 and 500 kW then the auxiliary engine is set to 5% of the main engine installed power while the boiler power output is based on Table 17;
- When the main engine power is larger than 500 kW then the auxiliary engine and boiler values shown in Table 17 are used.

1.2 Computation of Specific Fuel Consumption (SFC)

1.2.1 Baseline Specific Fuel Consumption

The concept of baseline SFC is used to derive SFC. The SFC base is the main engine, auxiliary engine, and auxiliary boiler's lowest SFC seen in their loading curve – in other words, the most fuel-efficient point. The SFC base varies based on engine age, fuel type, engine type, and system.

The main engine SFC is assumed to vary as a function of its load in a parabolic manner: at low loads, the SFC tends to be at its highest level, to then decreases until it reaches a minimum (e.g. 75% MCR), and finally, after this point, the SFC begins to rise again. In other words, it is modeled as the following:

$$SFC_{ME,i} = SFC_{base} \cdot (0.455 \cdot Load_i^2 - 0.710 \cdot Load_i + 1.280)$$

For auxiliary engines and boilers, it is assumed that they are not dependent on their load and, hence, are not corrected by CFL. Therefore, their SFC is simply their SFC base.

$$SFC_{AE|BO,i} = SFC_{base}$$

1.2.2 Derivation of Baseline SFC

In our study, we use Table 19 from the IMO report to assign Baseline SFC for each vessel, which depends on its engine type, fuel type, and year of built. The table has been updated with the latest sources available. Pavlenko et al. (2020), while researching the climate implications of using LNG as a marine fuel, included an extensive literature review on the fuel consumption of LNG-fueled engines and the most recent SSD and MSD engines. It is assumed that the dual-fuel LNG engines always operate on LNG as their primary fuel while the mass of pilot fuel injected remains constant across engine loads. Although we lack specific data pertaining to the engine types of our vessels, we categorize their engines using Fuel Types, Number of Strokes, RPM and Engine Model Name:

- **Oil Engines:** These engines are powered by diesel cycles. They are further classified based on their engine speed:
  - **Slow-Speed Diesel (SSD):** Any two-stroke diesel engine with an engine speed lower than or equal to 300 RPM.

- **Medium-Speed Diesel (MSD):** Any diesel engine with an engine speed ranging from 300 to 900 RPM.
- **High-Speed Diesel (HSD):** Any diesel engine with an engine speed above 900 RPM.

- **LNG Engines:** The study considers various internal combustion engine types that can be fueled by LNG. These engines are further subdivided based on their operating cycle, speed, and fuel injection method
  - **LNG-Otto SS:** Any two-stroke LNG engine below or equal 300 RPM, with model names including "WinGD" or "Wartsila" as these engines have been sold as WinGD engines built by Wärtsilä up to date. However, recently termed "ME-GA" by MAN Energy Solutions for the year 2020 and later.
  - **LNG-Otto MS:** Any four-stroke LNG engine above 300 RPM, without including "LBSI" in the model name.
  - **LNG-Diesel:** Any two-stroke LNG engine below or equal 300 RPM, with model names including "ME" or "MAN Energy Solutions" or "MAN B.&W." as these engines have so far only been built by MAN Energy Solutions.
  - **LBSI:** Any four-stroke LNG engine above 300 RPM, with model names including "LBSI". These engines are mainly built by Rolls-Royce/Bergen.

2 Interpolation of Missing Ship Movements

2.1 Improving Emissions Estimations

The integrity of our estimations rely on the accuracy of ship location over time. Utilizing linear interpolation should result in more accurate estimates because it allows for a more accurate application of location-dependent emission factors.

2.1.1 Haversine Formula

The formula is:

$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cdot\cos(\phi_2)\cdot\sin^2\left(\frac{\Delta\lambda}{2}\right) \tag{1}$$

$$c = 2\cdot\arctan 2\left(\sqrt{a}, \sqrt{1-a}\right) \tag{2}$$

$$d = R\cdot c \tag{3}$$

Wherein:

- ϕ denotes latitude in radians,
- λ denotes longitude in radians,

- R stands for the Earth's radius,
- $\Delta\phi$ is the latitude differential,
- $\Delta\lambda$ is the longitude differential,
- a represents the half-chord length squared between two points,
- c quantifies the angular distance in radians,
- d signifies the computed distance over the Earth's surface.

2.1.2 Hourly Interpolation Procedure

Considering the Earth's spherical shape, we employ the Haversine distance formula to interpolate between each pair of known GPS coordinates. For each hour that lacks data between these points, the interpolation method estimates new coordinates that are evenly spaced along the great-circle route, which connects the observed locations. This ensures that the interpolated points are distributed at consistent intervals, maintaining the natural arc of the Earth's curvature in the data set.

2.1.3 Potential Underestimation Bias

Although this method is robust for calculating direct distances, it can occasionally yield atypical outcomes, such as suggesting that ship trajectories cross over landmasses. In practice, vessels must navigate around terrestrial obstacles, which typically leads to longer routes, increased fuel consumption, and consequently higher emissions. When such direct paths over land are generated, they do not account for the additional distance that must be traversed around coastlines and obstacles, nor for the complex maneuvers ships must perform to ensure safe passage. As a result, the direct distance is often shorter than the actual navigable distance, which in turn affects the precision of our estimations.

3 Annual Aggregation of Ship-Level Stats

We match reported annual ship-level fuel consumption from a European Union Emissions Monitoring, Reporting and Verification (MRV) Maritime Regulation System with AIS tracking data and technical characteristics from World Fleet Register (WFR) of dry bulk ships. As a baseline, we follow industry standard procedures as discussed above to calculate hourly fuel consumption estimates, then aggregate them to obtain engineering estimates of annual ship-level fuel consumption.

3.1 Yearly Aggregated Engine Power Demand

This part presents the logarithmic transformation of the yearly aggregated W_{year}, where W_i is the hourly statistics:

$$W_i = \frac{\delta_w}{\eta_w \cdot \eta_f \cdot {t_{\text{ref}}}^m \cdot {v_{\text{ref}}}^n} \cdot W_{\text{ref}} \cdot (t_i^m \cdot v_i^n) \tag{4}$$

We begin by defining constants C' and C as follows:

$$C' = \frac{\delta_w}{\eta_w \cdot \eta_f \cdot {t_{\text{ref}}}^m \cdot {v_{\text{ref}}}^n} \tag{5}$$

$$C = W_{\text{ref}} \cdot C' \tag{6}$$

which allow us to rewrite the original equation as:

$$W_i = C \cdot (t_i^m \cdot v_i^n) \tag{7}$$

The yearly aggregated W_{year} is given by:

$$W_{\text{year}} = \sum_{i=1}^{N} W_{\text{ref}} \cdot C' \cdot (t_i^m \cdot v_i^n) \tag{8}$$

where N is the number of hourly observations over the year. Equivalently, this can be written as:

$$W_{\text{year}} = W_{\text{ref}} \cdot C' \cdot \sum_{i=1}^{N} (t_i^m \cdot v_i^n) \tag{9}$$

We then take the natural logarithm of both sides to get:

$$\ln(W_{\text{year}}) = \ln\left(W_{\text{ref}} \cdot C' \cdot \sum_{i=1}^{N} (t_i^m \cdot v_i^n)\right) \tag{10}$$

Applying logarithm properties, we can rewrite the above as:

$$\ln(W_{\text{year}}) = \ln(W_{\text{ref}}) + \ln(C') + \ln\left(\sum_{i=1}^{N} (t_i^m \cdot v_i^n)\right) \tag{11}$$

Thus, we introduced a set of newly engineered annual variables into our future machine learning framework to improve model performance. These variables are constructed based on the decomposition above and are aimed at capturing more relationshipss that current variables may miss.

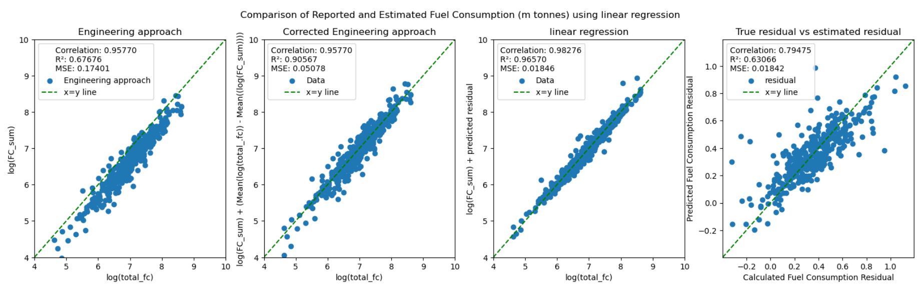
We have generated five time-variant variables for the model: $\sum_{i=1}^{N} t_i^m \cdot v_i^n$, $\sum_{i=1}^{N} \frac{t_i^m}{t_{\text{ref}}^m}$, $\sum_{i=1}^{N} \frac{v_i^n}{v_{\text{ref}}^n}$, $\sum_{i=1}^{N} \frac{t_i}{t_{\text{ref}}}$, and $\sum_{i=1}^{N} \frac{v_i}{v_{\text{ref}}}$, and two time-invariant variables: C' and W_{ref}. Additionally, three bias-offsetting annual variables were introduced: the vessel's port-dwell percentage, the longest Haversine distance between GPS points, and the ratio of missing hourly movement during sailings.

4 Machine Learning Algorithm Application

4.1 Motivation

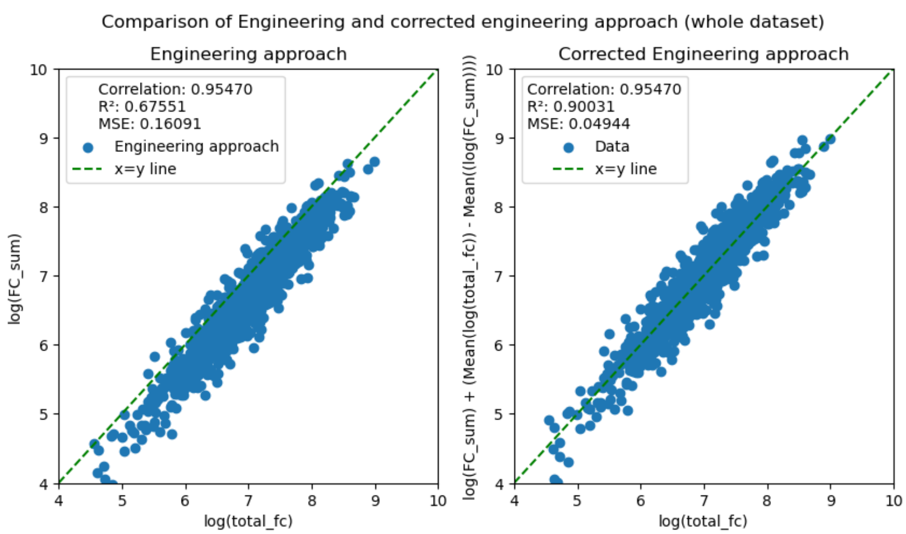
We aim to train various machine learning algorithms on the residuals: the discrepancy between log-transformed reported and calculated fuel consumption, to improve out-of-sample prediction accuracy and correct systemic bias.

4.2 One-Layer Benchmark



12/12/2023 Meeting: Need to re-draft

4.2.1 Comparative Analysis of Systemic Bias in Estimation Approaches



12/12/2023 Meeting: Need to re-draft

4.3 Double-Layer Framework

We employ a two-layered approach that begins with an initial selection of relevant variables, followed by an extensive analysis of various non-parametric tuned models, a crucial step for identifying models with meaningful performance. Central to our methodology is the use of all available resources, divided into two layers: linear regression and tree-based models.

The first layer involves linear regression to establish basic linear relationships, focusing on the residuals, particularly the logarithmic difference between reported and estimated fuel consumption, thus isolating linear dynamics in our dataset. Progressing to the second layer, tree-based models are employed to explore potential non-linear relationships, recalculating residuals by subtracting the predicted residuals from the linear regression level from the actual residuals. Our approach stems from the limitations of linear regression in capturing only linear relationships, prompting us to investigate the possibility of underlying nonlinear dynamics.

The process starts with an 80-20 data split into training and test sets, with the training phase incorporating a unique application of cross-validation on the 80% training data to generate the response variable for the second layer. This involves dividing the training data into 10 folds, iteratively training the linear model on 9 folds and generating predicted residuals on the remaining fold, a strategy that serves to both generate predicted residuals for the entire training set and compute a second set of residuals for the tree-based model. Once we have the second set of residuals, we proceed with the actual training of both models using the full 80% of the training data. The linear regression model is first trained to refine its ability to capture linear relationships, followed by the tree-based model, which focuses on modeling non-linear patterns.

The final stage integrates these insights to predict maritime shipping fuel consumption, combining engineering-based estimated fuel consumption with both sets of predicted residuals. This robust and innovative method provides a comprehensive prediction by capturing both linear and non-linear dynamics within the dataset, offering a more accurate and holistic understanding of fuel consumption patterns in maritime shipping.