Improving voyage efficiency in the shipping 4.0 decarbonization era

Takis J. Varelas
Danaos Research Center
Piraeus Greece
drc@danaos.gr

Dimitrios Kaklis
Dep. of Informatics and Telematics
Harokopio University of Athens,
Danaos Research Center,
NCSR Demokritos
Athens, Greece
drc@danaos.com

Iraklis Varlamis
Department of Informatics and Telematics
Harokopio University of Athens
Athens, Greece
varlamis@hua.gr

Artemis Flori
Danaos Research Center
Piraeus Greece
drc@danaos.gr

***Abstract*—The objective of this work is to pave the way toward a carbon-neutral and efficient operational blueprint for the waterborne sector, through the lens of the Industry 4.0 era. In this direction, we demonstrate a cutting-edge integrated ecosystem (ARTeMIS) for operational efficiency and environmental compliance and focus on the respective building block comprising the envisaged platform. ARTeMIS incorporates an IoT suite responsible for data acquisition, as well as a multi-purpose processing pipeline for CI/CD (continuous integration/deployment) of simulation models concerning operational optimization. In the context of this work, the proposed framework was adapted accordingly to capture, analyze and continuously predict the Fuel Oil Consumption (FOC), Effective HorsePower (EHP), Speed, and pollutant concentrations, of the vessel. Utilizing this streamlined procedure we are able to assess the scrubbing efficiency, as well as the environmental footprint of the vessel (CII - Carbon Intensity Indicator) in order to further optimize the vessel operation. Furthermore, the paper argues on how generic models like STEAM (Ship Traffic Emissions Assessment Method) should be transformed radically with the utilization of shipping 4.0 driven frameworks like the one proposed in the context of this work. Finally, the unleashed highly accurate prediction potential of the proposed system concerning operational efficiency, anomaly detection, and environmental compliance is demonstrated. The onboard proof of concept and the assimilated results so far, are also depicted to illustrate the feasibility and potential of the proposed approach.**

***Keywords*-IoT, regression. ANN, MADM, digitalization, decarbonization**

I. Introduction

The *4th Industrial revolution*, welcomed recently by the maritime industry, is often termed in pertinent literature as Shipping 4.0. The potential of digitalization in the Shipping 4.0 era is based on and realized through innovative technologies and techniques that aim to optimize operational efficiency, environmental protection, and decision-making based on a set of state-of-the-art prediction models. Big data analytics and deep learning techniques could be used to exploit the vast amount of data captured in real-time from sensors, in conjunction with digital or analog measurement instruments and consequently upgrade the onboard network of devices to an integrated IoT framework. In order to evaluate the theoretical potential of these technologies in a practical use case, we implement a comparative analysis between the real world (living lab) and a digitalized twin to continuously measure, simulate and assess how we can improve the operational efficiency and environmental compliance of vessels.

Air pollution from vessels constitutes a serious problem according to official reports (1). Maritime transport emits around 940 million tonnes of CO_2 annually and is responsible for about 2.5% of Global Greenhouse Gas (GHG) emissions. These emissions are projected to increase significantly if mitigation measures are not put in place swiftly. IMO, the EU, Port authorities, and institutes constantly issue several directives and procedures (SEEMP, MRV, EEXI, etc.) that aim to provide a strict regulatory frame to control the global or local emissions' impact and develop mitigation plans on the road to decarbonization and environmental protection. The efficient operation of vessels in terms of fuel consumption per mile is in close relation to air emissions. More precisely, energy consumption instead of fuel oil consumption per mile should be used as a term since the industry is currently evaluating other energy sources (renewables, electricity, nuclear) and alternate fuels (LNG; H2; NH3; C2H5OH). Energy consumption reduction could be also achieved with the adoption of applied retrofits such as bulbous modification, propeller design, and/or best maintenance and operation practices (trim, weather routing, slow steam operation).

The main pillars of this paper concern mainly, emission control - FOC approximation as well as the demonstration of a SOTA digital platform for vessel monitoring and operational optimization. More specifically the main contributions of this work consist of:

- A novel method that couples standard marine engineering theory with data-driven models to provide a robust FOC approximation method.

- A streamlined procedure for continuous monitoring of the vessel. Causal analysis and event recognition methods utilizing SOTA algorithms are demonstrated in the context of a streamlined algorithmic procedure for data acquisition, processing, and curation.
- A consolidated approach that incorporates the aforementioned methods to realize a holistic platform for operational efficiency and environmental compliance.

Section II that follows analyzes pertinent literature on optimal route planning emissions control and consequently FOC estimation. Section III demonstrates the streamlined procedure from data acquisition and curation to model deployment. Section III illustrates the developed ARTeMIS systems and elaborates on the specific methodology adopted for FOC approximation. Finally, section IV assimilates the contributions of this work and outlines possible pathways to further expand and enrich its findings.

II. Related work

A. The past: Vessel dependent solutions

During the last decade, due to the poor availability of real measurements and vessel operational profile details, the estimation of vessel emissions was performed using generic statistical models and semi-empirical formulas, without taking into consideration vessel-specific particulars, status, and operational data. Reasonably, the estimates were deviating from the actual values and thus such models witnessed limited usability in the shipping operational optimization context. A successful emissions estimation model is the generic STEAM (2; 3). Based on the ships' vessel particulars (taken from the IHS Fairplay database[1]), information from the engine manufacturers, and AIS data for the vessel, this model is able to assess the power consumption, the load of the engine and therefore the fuel-oil consumption (FOC) of the vessel. Consequently, the model estimates the emissions of NOx, SOx, CO, CO2, PMs as a function of time and location. The initial model has been revised several times to improve its accuracy. In the third revision, the load during voyages could be determined based on the ratio of the vessel speed and the calculated resistance that the ship was required to overcome at a specified speed (4).

The techniques employed in the literature for estimating the FOC and therefore the emissions of the vessel are based on vessel characteristics and/or environmental conditions, and can be grouped into the following categories:

- White box models, where analytical equations and approximation methods (e.g. Computational Fluid Dynamic equation - CFD), take advantage of a variety of vessel-specific variables and hydrodynamic principles to model the added resistance of the hull of a specific vessel (5). The admiralty constant, as well as the resistance constant ($RL/\Delta V^2$), where R is the total resistance and L is the length overall of the vessel, have been utilized in the past (6) to replicate the hydrodynamic behavior of new ship designs by using only the design point values of the corresponding parameters (speed (V), displacement (D), length (L), breadth (B) and draught (T)), rather than the whole spectrum of the operational domain. Thus, it can be easily inferred that the purpose of the Admiralty constant was to provide an initial baseline for comparing the hydrodynamic performance of different ships in their respective design conditions, rather than monitoring their operational state. This makes it evident that the admiralty constant was neither intended nor demonstrated to be a suitable operational hydrodynamic performance indicator for a ship.
- Data-oriented approaches that combine vessel-trajectory data gathered from onboard sensors, coastal or satellite receivers (AIS data), or Noon Reports, with Machine and Deep-Learning algorithms. These techniques are ranging from simple Regression analysis, using standalone models like Support Vector Regression (SVR), Lasso Regression (LR), Polynomial Regression, etc., to ensemble non-parametric schemes like Random Forest regression (RF), Decision Trees or AdaBoost, where the approximation power of each model is appropriately combined in order to infer the underlying function, and Deep Learning approaches (7). Some methods also deal with the problem of deteriorating performance as new batches of data corresponding to arbitrary distributions are introduced to the estimator (8).
- Grey-box model (GBM) approaches (9; 10) that combine machine learning (ML) methods, also known as black-box models (BBM), with analytical models, known as white-box models (WBM), in order to increase the prediction accuracy.

ANNs (Artificial Neural Networks) have been at the center of attention lately in many research areas. As far as the FOC of the vessel is concerned not many studies utilize the computational power of ANN's to approximate FOC, mainly due to the problem of missing historical data. The studies found in pertinent literature dealing with FOC estimation from a deep-learning perspective are briefly presented below. Some studies experimented with baseline sequential ANN's by applying a dropout in the weights in order to achieve better generalization error (11) or by tuning a number of hyperparameters (learning rate, number of neurons, number of layers, activation function) utilizing brute force methods like randomized grid search ((12), (13)). (14), employed a Recurrent NN in order to estimate FOC, but without further research as far as the architecture of the network is concerned.

B. The present: IoT-driven solutions

During the past few years, several IoT-related systems have been developed, installed, and utilized successfully on hundreds of vessels. Indicatively Greek innovative technology companies are providing awarded AI/IoT empowered solutions: Danaos with waves, Prisma with Laros, DeepSea with Casandra, and METIS with Metis-Space provide dedicated

[1]https://www.acml-egypt.com/Fairplay.html

frameworks concerning total emissions monitoring and reporting that helps shipping companies to ensure regulatory compliance and reduce their environmental footprint. The main benefits for a shipping company that has deployed a system that monitors CO_2 emissions are rule compliance, continuous monitoring through a versatile Ship Energy Efficiency Management Plan (SEEMP), adaptive solutions for anomaly detection as well as tailor-made dedicated voyage planning solutions for operational efficiency and environmental compliance. The aim of this work is to encourage and motivate the waterborne sector (shipowners, external vendors, charterers, suppliers, end-users) to go one step further and incorporate in their existing workflow the broad range of functionalities offered, by the proposed envisaged digital ecosystem in order to vastly automate the decision-making procedure towards a carbon neutral and efficient operational blueprint.

III. The Proposed Platform for Vessel Monitoring and Energy Optimization

There are several operational optimization frameworks and platforms for energy efficiency on vessels that consequently reduce negative environmental footprint. These methods correspond mainly to power management by modeling the propulsion system (main engine-propeller) of the vessel, alternative fuels (LNG, LPG, methanol, biofuel, and hydrogen), photo voltaic (PV) installments, waste heat recovery systems retrofit or new design solutions. The proposed system, which is depicted in Figure 1, combines **A**pplied **R**esearch **Te**chniques for **M**onitoring, **I**dentifying, and **S**uggesting (ARTeMIS) to offer a versatile platform for energy efficiency to vessel owners. ARTeMIS is envisioned and designed as a real-time monitoring application of ship operations by employing a digital replica of the en-route vessel. It aims to optimize a variety of multivariate objective functions corresponding to a variety of different dependent variables. These variables range, from emissions reduction KPIs (FOC approximation) to operational efficiency (Time Charter Equivalent (TCE)) and regulatory compliance (CII, EEOI) indicators. These are just some of the parameters that ARTeMIS incorporates into its workflow to provide tailor-made mitigation solutions toward a carbon-neutral operational blueprint. The core functionality of the envisaged platform entails a multidisciplinary Decision Support System (DSS) consisting of various operational optimization approaches and mitigation strategies, like, Routing optimization, Causal Analysis - Event recognition - Predictive Maintenance (biofouling, corrosion - degradation of the hull, charter party compliance, elicit goods identification), safety (identification of parametric roll, prevention of cargo loss) and emissions monitoring and projection. Furthermore, this versatile ecosystem aims to provide a solution to a major shortcoming of the maritime industry namely, **slow steaming** that jeopardizes vessels utilization. This problem is being addressed by implementing a *Multiple Attribute Nonlinear Decision-Making (MADM)* support mechanism that aims to optimize, jointly the trade-off between *time saving* and emissions reduction, during a voyage, as it is depicted in Table I.

Table I: MADM nonlinear DSS model factsheet for mitigating slow steeming

Symbol	Description: Formula	units
d	Voyage Distance	Miles
v	Speed	Miles/day
T	Voyage duration:d/v	days
R	Voyage running cost per day (proportional) excluding bunkering	\$/day
b	Bunkering cost: Fuel oil consumption cost: tones per day · fuel ton-cost $foc_d \cdot \overline{f_C}$	\$/day
F	fixed voyage expenses: canal dues etc.	\$
E	voyage cost expenses: $F + (R + b) \cdot T$	\$
I	Income	\$
P	Profit: $I - F - (R + \overline{foc_d} \cdot \overline{f_C}) \cdot T$	\$
cii	$\overline{cii} = \frac{foc \cdot \overline{Ef}}{\nabla \cdot s} = \frac{c \cdot \overline{foc_d}}{v}$	gr_{CO_2}/tonmiles
TCE	$tce = \frac{I - F - (R + \overline{foc_d} \cdot \overline{f_C}) \cdot T}{T}$ $tce = \frac{(I-F) \cdot v - (R + \overline{foc_d} \cdot \overline{f_C}) \cdot s}{s}$	\$/day
objective	$max(tce)$	
subject to	$\overline{CII} < goal, t <= JIT, v > v(EPL)$	

The last two rows of table summarize the ultimate goal of the proposed Decision Support System, which is to maximize TCE during a voyage while on the same time adhere to regulatory and Charter Party agreements and indicators. These indicators refer mainly to CII (Carbon Intensity Indicator), JIT (Just In Time Arrival), and speed (v) adopted during a voyage, that can be adjusted accordingly to either comply or exceed EPL (Engine Power Limitation) depending on the specific KPIs established by the shipowner that outline the ”*performance*” of a particular voyage.

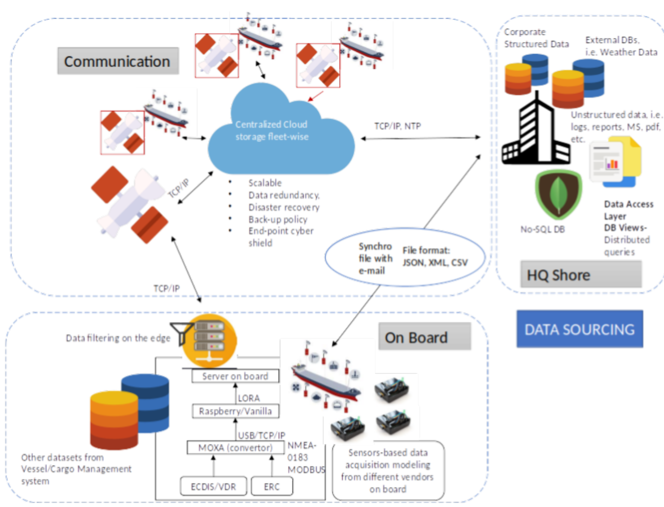

Figure 1: Internal topology of ARTeMIS ecosystem

A. Pre-processing pipeline

1) Data acquisition: Data acquisition is being handled by an onboard IoT infrastructure consisting of both analog and digital measuring devices that are connected via wired and wireless collector nodes on an intranet network. This network is designed to be *Available*, *Reliable*, and *Maintainable (ARMed)*, to ensure continuous operation. Standard interface protocols for hardware and software are supported. For instance, the main Engine flow meter is installed in the control engine room, where the temperature and mass/minute fuel flow signals are wirelessly transmitted to a collector device with two digital ports and two 2.0 mA ports. Similarly, the Torque meter, which is also located in the engine control room, sends EHP and RPM signals to a wireless collector with a serial port that can be configured to RS232, RS422, or RS485 mode using the NMEA 0183 protocol. This broad architecture of the IoT framework enables the network to horizontally expand by adding new devices, such as the continuous monitoring system of scrubber emissions using an interface unit, which is typically achieved through field bus communication, usually realized via the *Mod-bus protocol*.

2) Data cleaning: Raw data, collected from the sensors of the vessel, are in time-series form in the minute granularity and tend to be "noisy" (i.e. high variance, high standard deviation from the mean), and in some cases even erroneous. In order to remove noise, we employed a fit & filter technique (10) that effectively "cleaned" the data but at the same time kept the bulk of information needed for training robust predictive models. Data filtering was implemented in two stages. First, assuming that our dataset follows a normal-like distribution, we filter out data points that lie outside of the 99% confidence interval and keep values for each feature that lie within the 3-times standard deviation band from the mean value. Then a transformation of the dataset into 15 min rolling window averages is applied to further smoothen out any spikes and outliers that occur in the feature set from the onboard sensors.

For further dataset cleaning, an algorithm based on DBSCAN (10) is applied. DBSCAN is an unsupervised machine learning technique used to identify clusters of varying shapes in a data set (15). DBSCAN can identify clusters in large spatial datasets by looking at the local density of the data points. The core functionality of DBSCAN clustering lies in its robustness to outliers. It also does not require the number of clusters to be decided a priori, in contrast with centroid-based approaches such as K-Means. However, DBSCAN requires two parameters: epsilon and minPoints. Epsilon is the radius of the circle around each data point that defines the desired density, along with minPoints which is the minimum number of data points required within the epsilon radius for that data point to be classified as a cluster. The FOC and STW as they are collected from sensors (raw data), the quasi-steady and DBSCAN cleaned version are visualized in Figures 2 and 3.

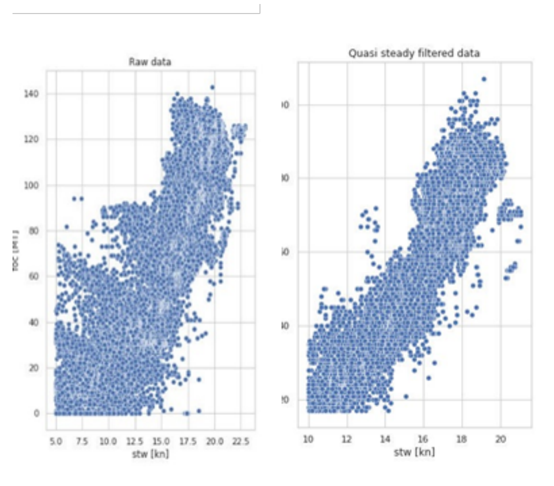

Figure 2: Raw & quasi steady filtered data (STW - FOC)

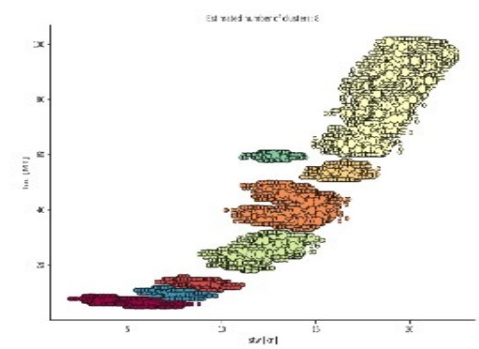

Figure 3: "Cleansed" version of STW vs FOC with DBSCAN

B. FOC predictive module

The maritime industry involves various stakeholders, who have invested a considerable amount of time in fuel oil consumption approximation. Each party involved strives to measure, monitor and predict, the energy of the en-route vessel while minimizing negative environmental impact. Although measurement and monitoring solutions exist (as previously mentioned: Danaos-Waves[2], Prisma-Laros[3], DeepSea-Casandra[4] and METIS-Metis-Space[5]), the prediction of fuel consumption remains a pivotal point for optimal voyage planning, maintenance, scheduling, and decision-making. Theoretical calculations mainly rely on the admiralty coefficient, which concerns the preliminary estimation of the power required in a new design to achieve the desired speed. This coefficient is calculated using the formula first demonstrated in (16) and is described as follows:

$$k = \frac{\Delta^{2/3} V^3}{EHP}, \quad (1)$$

The admiralty constant as well as the resistance constant, proposed by Telfer (17), were used to model the variation in hydrodynamic performance of different ship designs. Therefore, it is obvious that they used only the design point values of the included parameters but not the whole range of the operational domain. In other words, the velocity (v),

[2]https://www.danaosshipping.gr/news/innovation/waves_performance_dashboards/

[3]https://www.laros.gr/

[4]https://www.deepsea.ai/cassandra/

[5]https://www.metis.tech/metis-space/

displacement (Δ), and so on, were corresponding actually to the design point values derived from open sea trials. Thus, it was never intended to use these constants to monitor the operational performance of an individual ship but rather just to compare the hydrodynamic performance of different ships in their respective design conditions. It is also noteworthy that the originally proposed admiralty constant was a function of "effective" horsepower (EHP) while the modern admiralty coefficient uses shaft power instead (ITTC (2017)). Thus, the two, clearly, differ by the factor of propulsive efficiency of the ship, which is known to be varying for different operational conditions. Thus, it can be clearly concluded that the originally proposed admiralty constant or any of its variations were actually neither intended nor proven to be an appropriate operational hydrodynamic performance indicator for a ship. Obviously, the admiralty coefficient was rather developed to compare the hydrodynamic performance of different ship designs. In any case, the idea of summarizing the calm-water speed-power curve into a singular or a few constant values can still be realized using a simple statistical analysis of the operational data recorded onboard a ship.

Speed exponent (n), outlines the relationship between speed and power, which is widely accepted as $EHP \propto v \cdot n$, with $n = 3$ according to the admiralty coefficient. From a physics point of view, the value of $n = 3$ is quite appropriate for the low-speed range when the total resistance coefficient remains constant (and therefore, independent of ship speed) due to negligible wave resistance. Kristiansen (18) used a computer model based on updated Guldhammer and Harvald's method ((19)) to estimate the value of n for container ships of different sizes and service speeds. He concluded that the cubic relationship is only valid for container ships in the low-speed range.

C. White Box modeling: Proposed $sfoc$ formulation and prediction based on generalized admiralty coefficient

We can clearly conclude that the originally proposed admiralty constant or any of its variations were actually neither intended nor proven to be an appropriate operational hydrodynamic performance indicator for a ship. *ARTeMIS* aims to exploit and enhance previous approaches that calculate the mass flow rate of CO_2, NO_x, and SO_x emissions utilizing vessel particulars and ship's main engine attributes (SFOC curves, MCR, Nominal Power produced, etc.). It uses features acquired in real-time from onboard sensor installments (e.g. FOC, draft, speed, EHP, and Deadweight) instead of theoretical calculations and vessel-dependent variables and gets a good approximation of vessel emissions.

The *ARTeMIS* ecosystem initiates its streamlined procedure by identifying variables that have a high correlation factor with the independent variable, which in our case is FOC. To achieve this, the well-established Pearson method is employed to assess the correlation between all possible pairs of the dependent variables. Correlation coefficients are used to measure the extent to which two measurement variables "vary together," similar to covariance. However, unlike covariance, the correlation coefficient is scaled to ensure its value is not dependent on the units of measurement.

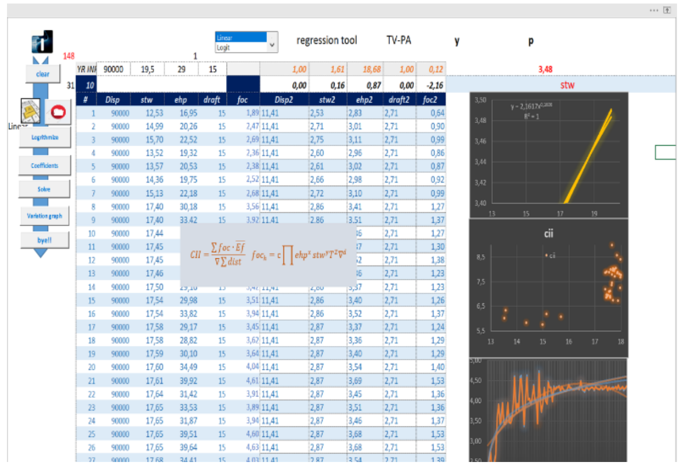

Figure 4: ARTeMIS application layer for FOC approximation

From a physics point of view, the generalized admiralty coefficient defines a log-linear relationship between speed through water (STW) and shaft power (Ps). Thus, the proposed theoretical model is employed, utilizing a multiplicative exponential formula comprising the top four features (Power, Speed, Draft, and Displacement) and is described as follows:

$$y(x_i, \ldots, x_n) = c_0 \prod_{i=1}^{n} x_i^{c_i} \tag{2}$$

where y is the FOC of the vessel, x_1, ..., x_n are the aforementioned variables, and the coefficient c_i refers to the constant and variables' exponents that are expressed as characteristics of the vessel's shape and/or hull roughness and should be approximated. In order to conclude with a linear combination of the variables comprising equation e1 we apply the natural logarithm (i.e. ln) on both sides of the equation. Then equation 2 is transformed to 3.

$$ln(y(x_i, \ldots, x_n)) = ln(c_0) + \sum_{i=1}^{n} c_i ln(x_i) \tag{3}$$

Coefficients (c_i) are generated by training a baseline regression model described by equation 3 utilizing historical data (FOC, EHP, speed) that are acquired from onboard IoT systems. The aforementioned formula was thoroughly evaluated on a test set of $\simeq 10^4$ observations, exploiting the vast amount of data collected. The implemented white box predictive model for known parameters such as Power, Speed, Draft, and Displacement, utilizes the regression *linest* function (20) which calculates the R^2 score (coefficient of determination) for a predicted "best fit "poly-line by using the "least squares" method. The model interface screenshot is depicted in Figure 4.

D. Grey box modeling: Proposed FOC prediction based on ANN and regression polynomials

Gray box models (GBM) a term first introduced in (21) combine theoretical models (WBM) with data-driven approaches (BBM). In their simplest version, GBM attempt to integrate prior knowledge extracted from a theoretical model as a new feature to the training process of a Black Box Model (BBM).

As it is evident from pertinent marine engineering literature, and also known and accepted amongst marine practitioners, *EHP* (power absorbed by the ship propulsion system) is strongly connected with STW (speed). Towards this direction, we aim to train a simple ANN, in terms of computational cost and complexity, that utilizes as input, one variable, the speed of the vessel, and has one output, the EHP.

$$y_{out} = \mathcal{F}(x, \mathcal{W}) \tag{4}$$

where x here is the speed of the vessel and W is the learned weight vector.

The specifics of the architecture adopted to build the model are not in the scope of this work. Nevertheless, it is crucial to designate, that the number of hidden layers and interconnecting nodes of the network have been defined on the basis of brute force principles, by utilizing an *exhaustive grid search* algorithm. A variety of other hyper-parameters like the activation function of the model have been determined by leveraging certain laws of physics that govern the relationship between the two variables, with computational fluid dynamics and marine engineering theory. The problem at hand (EHP approximation from STW) is nonlinear and the optimal solution requires the appropriate utilization of state-of-the-art optimization methods. In order to exploit the approximation capabilities of the model we utilized the generalized reduced gradient method to appropriately update the weights and converge eventually to a global optimum. The gradient of the underlying objective function is updated by incorporating into the model, the squared deviation, between the actual and predicted value from the model (least squares minimization).

By utilizing as input to the GBM, the output of the WBM (equation 3) we employ an enriched feature set, that can be utilized as input for the GBM in the form:

$$\mathcal{D}_n = \left\{ \left(\begin{bmatrix} x_1 \\ f_{WBM(x_1)} \end{bmatrix}, y_1 \right), \ldots, \left(\begin{bmatrix} x_n \\ f_{WBM(x_n)} \end{bmatrix}, y_n \right) \right\}$$

Incorporating this method to ARTeMIS workflow, we are able to enhance the predictive capabilities of the BBM and approximate more "accurately" the *EHP* and therefore the FOC of the vessel for any given speed value. Figure 5 presents the actual data as well as the corresponding calculated output from the NN.

E. Real world experimental evaluation: the IoT bridge

The evaluation of the proposed methodology and platform has been performed in real conditions on a real-world experiment that is described in the following. In the context of emissions reporting and evaluation, a research team

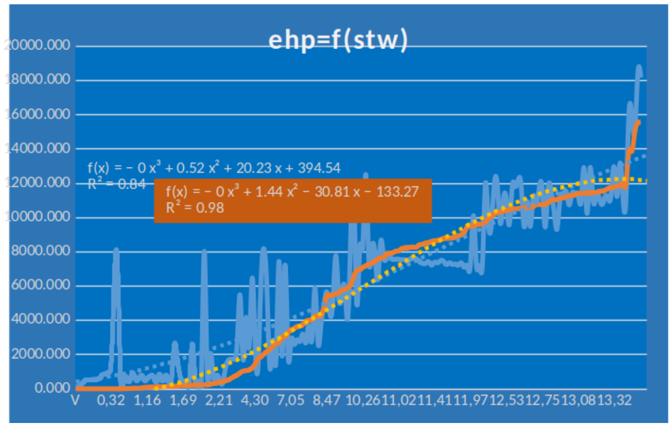

Figure 5: EXP=f(STW) Orange: ANN predictions, Blue: real measurements

(S.Team) embarked onboard a Danaos vessel for a 10-day trip with a Horiba PG-350 portable multi-gas analyzer. Exhaust concentrations of CO, CO_2, NO_x, and SO_x emissions were measured following the ISO 8178-2 protocol. The PG-350 utilizes NDIR (Non-dispersive infrared) detectors to measure CO, CO_2, and SO_2 concentrations.

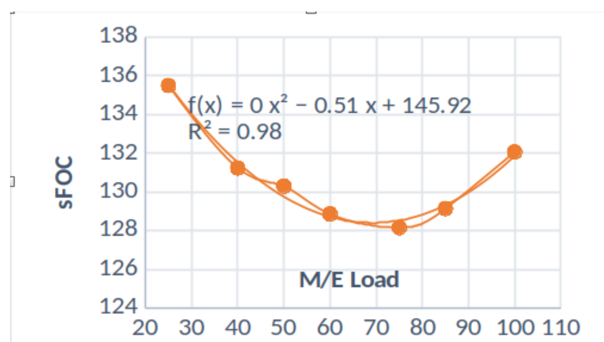

Figure 6: SFOC curve based on SHOP data

The goal of the S.Team was to evaluate the STEAM methodology by approximating the fuel consumption of the vessel utilizing the SFOC (Specific Fuel Oil Consumption) curve that outlines SFOC - Main Engine load (M/E load) relationship on different engine load rates, provided by the engine manufacturer (Figure 6). In order to interpolate (or extrapolate) for the whole operational domain (various engine load values) we utilized a second-order best-fit function $\mathcal{F}$: $\mathcal{F}(load) = SFOC$, fitted on the discredited SOC - M/E load values (Figure 6). Utilizing $\mathcal{F}$, the S.Team calculated the FOC which at the same time is measured by the flowmeter sensor installed on board. The flowmeter measurements were used to train the *GBM* presented in section III-D In Figure 7 below we demonstrate the approximation capabilities of STEAM vs the GBM proposed.

The results presented in Figure 7 clearly show that the GBM, a physics-informed neural network, which was trained using a customized feature set that takes into account the

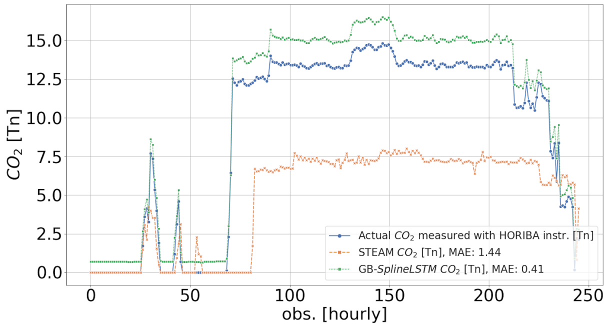

Figure 7: Actual vs STEAM vs GBM CO_2 emissions

operational state of the en-route vessel, is capable of predicting CO_2 emissions with a higher degree of precision compared to the STEAM methodology (a significant deviation is observed of approximately $\simeq$ 20% on average between the STEAM methodology and the GBM).

The experimental results presented above, outline the importance for maritime companies to adopt a fully digitized and automated operational blueprint, towards an informed decision making procedure that aims to exploit, data-driven solutions, by leveraging the approximation capabilities of deep learning algorithms with real-time measurements, acquired from on-board sensor installments. This vessel- agnostic streamlined procedure will enable shipowners and externals stakeholders to achieve efficiency in fleet management with tangible benefits in terms of emission reduction, environmental compliance and protection of crew safety onboard.

IV. Conclusions

Digitalization is a highly promising tool for the decarbonization of the shipping industry, without adversely affecting operational efficiency. To achieve this, it is recommended to apply multiple attribute decision-making (MADM) techniques that analyze big data from IoT while cleaning the noise (high veracity) attributed to the data inherently, with state-of-the-art processing algorithms. While general models are available, it is essential to customize and adjust them for individual cases. The STEAM methodology should be adapted to the new digital era, leveraging innovative technologies to explore more efficient alternative solutions instead of relying on low STEAMing.

Acknowledgment

This Publication was supported partially by the program of Industrial Scholarships of Stavros Niarchos Foundation, and partially by the SmartSea project, an European Union's Horizon 2020 research and Innovation programme under the Marie Skłodowska-Curie Grant Agreement No 612198. Access to industrial data has been provided by Danaos Shipping Co.

References

[1] I. OECD *et al.*, "Energy and air pollution: world energy outlook special report 2016," 2016.

[2] N. Moussiopoulos, R. Sokhi, T. George, F. Evangelia, C. Eleftherios, and P. Ioannis, "Environmental impacts of shipping: from global to local scales," *12th International Conference on Air Quality, Science and Application*, vol. 143, 2020.

[3] J.-P. Jalkanen, J. Lasse, J. Kukkonen, A. Brink, J. Kalli, and S. Tapani, "Extension of an assessment model of ship traffic exhaust emissions for particulate matter and carbon monoxide," *12th International Conference on Air Quality, Science and Application*, vol. 12, pp. 2651–265, 2012.

[4] Hollenbach, "Estimating resistance and propulsions for single screw and twin screw ships," *12th International Conference on Air Quality, Science and Application*, pp. 72–7, 1998.

[5] T. Varelas, S. Archontaki, J. Dimotikalis, O. Turan, I. Lazakis, and O. Varelas, "Optimizing ship routing to maximize fleet revenue at danaos," *Interfaces*, vol. 43, no. 1, pp. 37–47, 2013.

[6] E. Telfer, "The design presentation of ship model resistance data," *International Shipbuilding Progress*, vol. 10, no. 106, pp. 215–232, 1963.

[7] D. Kaklis, G. Giannakopoulos, I. Varlamis, C. D. Spyropoulos, and T. J. Varelas, "A data mining approach for predicting main-engine rotational speed from vessel-data measurements," in *Proceedings of the 23rd International Database Applications & Engineering Symposium*, 2019, pp. 1–10.

[8] I. Kaklis, Dimitrios Varlamis, G. Giannakopoulos, C. Spyropoulos, and T. J. Varelas, "Online training for fuel oil consumption estimation: A data driven approach," 2022.

[9] A. Coraddu, L. Oneto, F. Baldi, and D. Anguita, "Vessels fuel consumption forecast and trim optimisation: a data analytics perspective," *Ocean Engineering*, vol. 130, pp. 351–370, 2017.

[10] D. Kaklis, I. Varlamis, G. Giannakopoulos, Spyropoulos, Constantine, and T. Varelas, "From steam to machine: Emissions control in the shipping 4.0 era," *The 8th International Symposium on Ship Operations, Management & Economics (SOME)*, 2022.

[11] C. Gkerekos, I. Lazakis, and G. Theotokatos, "Machine learning models for predicting ship main engine fuel oil consumption: A comparative study," *Ocean Engineering*, vol. 188, p. 106282, 2019.

[12] P. Christos and A. Ziakopoulos, "Predicting vlcc fuel consumption with machine learning using operationally available sensor data," *Ocean Engineering*, vol. 111, 2020.

[13] R. Savitha, A. Al Mamun *et al.*, "Regional ocean wave height prediction using sequential learning neural networks," *Ocean Engineering*, vol. 129, pp. 605–612, 2017.

[14] Y. Z. Yongjie, Zhu and T. Li, "Predicting ship fuel consumption based on lstm neural network," *7th International Conference on Information, Cybernetics, and Computational Social Systems*, vol. 111, pp. 310–313,

2020.

[15] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, vol. 96, no. 34, 1996, pp. 226–231.

[16] P. Gupta, B. Taskar, S. Steen, and A. Rasheed, "Statistical modeling of ship's hydrodynamic performance indicator," *Applied Ocean Research*, vol. 111, 2021.

[17] E. Telfer, "The design presentation of ship model resistance data," *International Shipbuilding Progress*, vol. 10, no. 106, pp. 215–232, 1963.

[18] H. O. Holmegaard Kristensen, "Model for environmental assessment of container ship transport," in *SNAME Maritime Convention*. OnePetro, 2010.

[19] GuldHammer and Harvald, "Ship resistance."

[20] F. A. Morrison, "Obtaining uncertainty measures on slope and intercept of a least squares fit with excel's linest," *Houghton, MI: Department of Chemical Engineering, Michigan Technological University. Retrieved August*, vol. 6, p. 2015, 2014.

[21] Bohlin, "Interactive system identification: Prospects and pitfalls," 1992.