Voyage performance evaluation based on a digital twin model

M Liu¹, Q Zhou¹, X Wang², C Yu³, and M Kang³
¹Seastel Marine System (Canada) Inc., 1250 René-Lévesque Blvd W Suite 2200, Montreal, QC H3B 4W8, Canada
²Newcastle University in Singapore, SIT Building @ Ngee Ann Polytechnic, 537 Clementi Road #06-01, 599493, Singapore
³Seastel Marine System (Shanghai) Co. Ltd., Room 7-102, 555 Dongchuan Road, Shanghai, 200241, China

Abstract. Using digital twins in voyage performance evaluation is becoming critical for ocean vessels to reduce GHG emissions. A novel GBM approach is proposed in this paper to establish a digital twin model for voyage performance prediction. The weather hindcast data are introduced to enrich noon reports (NR) and automatic identification system (AIS) datasets, which are split into training and validation sets to develop GBM. The NR and AIS datasets collected from a 57000DWT bulk carrier are used to demonstrate the fidelity and capability of the proposed GBM. The voyage performance prediction from the GBM shows better accuracy than those from pure WBM or pure BBMs. An arrival time forecast and a weather routing showcase are also presented to demonstrate the application effects of GBM. The proposed GBM provides a satisfying prediction of ship speed and fuel consumption without mandatory sensor-collected data, thus applicable for a variety of vessels. In those cases where more sensors are available onboard, the proposed approach can incorporate sensor data to improve the model accuracy further.

1. Introduction
In recent years, voyage performance evaluation and operational optimization based on digital twin models are becoming increasingly critical to reduce Greenhouse Gas (GHG) emissions in shipping industry. There are three mainstream modeling techniques to build the digital twins, namely white, black, and grey-box models[1].

A white-box model (WBM) expresses the system mechanism involving physical principles and numerical/experimental results. Many attempts have been made to improve the model accuracy by modifying empirical formulas or simulation methods. Moreno-Gutiérrez compared four existing methods for calculating energy consumption and emissions[2]. The advantages of those methods were discussed and combined into a new approach, which demonstrated better accuracy in predicting the energy consumption and emissions for a particular scenario. Nielson developed several WBMs to simulate the required power and fuel consumption of an academic hull in the time domain[3]. Comparisons were conducted to investigate the effects of operational and environmental factors on voyage performance.

A black-box model (BBM) describes input-output relation based on measured data. Bialystocki used a statistical approach to analyze a dataset from noon reports (NR)[4]. The impact of the ship's draft, displacement, hull roughness, and weather condition on ship's fuel consumption was investigated. A simple prediction model with acceptable accuracy was proposed. Safaei established a prediction model of fuel consumption based on multiple linear regression[5]. The NR and AIS data of four VLCCs were used in the model establishment. Gkerekos compared the performance of different machine learning models on predicting ship fuel consumption[6]. The various methods such as support vector machines,
random forest regressors, extra trees regressors, artificial neural networks, and ensemble methods were analysed in the study with the data from NR and Automated Data Logging & Monitoring (ADLM) system. The comparison results revealed that the artificial neural networks (ANN) model is preferred and high-quality training data from ADLM could significantly improve model performance. Pedersen and Larsen established a nonlinear neural network model to predict ship propulsion performance [7]. The sensor dataset collected from a tanker with high data frequency contained ship speed, relative wind speed, and the temperature of air and seawater. Liang compared the performance between the traditional machine learning method and deep learning method on predicting ship propulsion power [8]. AIS data, ship propulsion power measurements, and weather data were employed in the study. Farag developed a ship performance model based on ANN and multi-regression techniques to predict the main engine power and fuel consumption in varying operating conditions [9]. The sensor datasets collected from a VLCC with high data frequency were deployed with AIS data for model establishment. Coraddu created a ship digital twin to estimate the speed loss caused by fouling [10]. A data-driven approach based on deep extreme learning machines was proposed with sensor datasets from two tankers.

A grey-box model (GBM) combines the physical principles with data-driven modeling results. Leifsson established GBMs with both serial and parallel approaches to estimate the speed and fuel consumption of a container vessel [11]. The results showed that GBM could significantly improve the prediction of fuel consumption compared with pure WBM and pure BBM. The real sensor data is necessary for evaluating voyage performance and optimizing ship operations.

Since high-quality sensor data are not complete in most cases, the commonly recorded noon reports (NR) and automatic identification system (AIS) data are valuable for digital twin modelling. However, NR data is not regarded as high-quality data due to the possible human errors and low frequency of collection. Some attempts have been made to overcome these problems. Safaei employed the statistical methods to correct and enrich the NR data of four VLCC [12]. The raw NR data were refined, and the newly generated data were statistically in harmony with the original data. Based on the refined NR and AIS data, the fuel consumption prediction model was successfully established [5].

In this paper, a novel GBM approach is proposed to establish a digital twin model for voyage performance evaluation. The weather hindcast data are introduced to enrich the NR and AIS data, which are split into training and validation sets to develop GBM. An arrival time forecast and a weather routing showcase of a 57000DWT bulk carrier are presented to demonstrate the fidelity and capability of the proposed GBM. The paper proceeds as follows. Section 2 describes the structure of the GBM model and the enrichment of the NR and AIS datasets. Section 3 establishes the GBM of a 57000DWT bulk carrier for prediction of ship speed and fuel consumption. Section 4 presents both validation and application results of the GBM. Section 5 gives the conclusions and discussions.

2. Modelling

2.1 GBM Structure
The GBM combines both WBM and BBM to utilize the prior knowledge, physical principles, and data-driven results to form the digital twin model. Both WBM and BBM are treated as sub-models in a GBM. There are mainly two types of GBM structure, namely serial and parallel [11]. In a GBM with serial structure, two or more sub-models are connected in serial order, and the output of each sub-model can be used as the input of the following sub-models. The output of a serial GBM is the output of its last sub-model. In a GBM with a parallel structure, two or more sub-models are distributed parallel with the independent input data. The output of a parallel GBM is the combination of all sub-model outputs.

A novel GBM with the mixed structure, as shown in figure 1, is proposed to build a digital twin model for voyage performance prediction. The input variables of the GBM are weather information and propeller revolution rate, while the output variables are corresponding ship speed and the accumulative fuel consumption over a certain period. The sub-model WBM_Voyage serves as the GBM backbone to calculate the ship speed and fuel consumption with calm water effect, current effect, and wind effect. The sub-model BBM_Speed predicts the wave-caused speed loss $Speed_{BBM}$, while sub-model
BBM_Oil predicts the wave-caused fuel consumption increase Oil_BBMs. The outputs from the WBM and BBMs are added together to form the GBM outputs Ship Speed and Fuel Consumption, just like a normal parallel GBM. Meanwhile, the outputs from the WBM are also directed into the BBMs as part of inputs, just like a normal serial GBM.

The sub-model WBM_Voyage also serves as a measure to enrich the NR and AIS datasets via weather hindcast data, which is critical to recover the weather change process and ship sailing process during voyage. The enriched datasets are split into training and validation sets to develop the sub-model BBMs.

![Figure 1. Illustration of GBM Structure.](image)

### 2.2 WBM_Voyage Sub-model

The sub-model WBM_Voyage serves as the GBM backbone to calculate the ship speed and fuel consumption with calm water effect, current effect, and wind effect. Among the total ship resistance components, only the wave added resistance is excluded from resistance calculation in WBM[13].

The calm water resistance $R_{calm}$ could be calculated following Holtrop and Mennen[14]. The current effect is included by replacing the absolute ship speed with the relative ship speed.

The wind resistance $F_{wind}$ could be calculated by equation (2.1). $A_f$ is the frontal area of the superstructure exposed in the wind, and $V_{wind}$ indicates the wind speed relative to the ship. The wind load coefficient $C_x$ could be obtained from Blendermann's method[15].

$$F_{wind} = \frac{1}{2} C_x \cdot A_f \cdot \rho \cdot V_{wind}^2$$  \hspace{1cm} (2.1)

The propeller thrust $T$ and propeller torque $Q$ under open water condition are expressed as equations (2.2) and (2.3). The interaction between the propeller and ship hull is represented by the wake coefficient $\omega_p$ and thrust deduction coefficient $t_p[13]$

$$T = k_r \rho n^2 D^4$$ \hspace{1cm} (2.2)

$$Q = k_q \rho n^2 D^5$$ \hspace{1cm} (2.3)

For a specific propeller revolution rate $n$ retrieved from NR dataset, the corresponding steady speed could be easily determined from the balance between the propeller thrust $T$ and ship resistance[15]. This steady speed $Speed_WBM$ is one output of the sub-model WBM_Voyage.
At the steady speed, the propeller torque $Q$ is counterbalanced by the driving torque from the main engine. The engine brake power $P_B$ is determined by the propeller delivered power and the shaft efficiency $\eta_s$. Once the Specific Fuel Oil Consumption (SFOC) of the engine is known, the engine fuel consumption rate (FCR) could be calculated by equation (2.4). Integrating FCR over the sailing period as equation (2.5) gives the accumulative fuel oil consumption (FOC), which is another output $Oil_{WBM}$ of the sub-model WBM_Voyage.

$$ FCR = \frac{SFOC \cdot P_B}{\eta_s} = \frac{Q \cdot 2\pi \cdot n}{\eta_s} \tag{2.4} $$

$$ FOC = \sum_{i=1}^{n} FCR_i \cdot \Delta t_i \tag{2.5} $$

2.3 BBM_Speed Sub-model

The sub-model BBM_Speed predicts the wave-caused speed loss. The multilayer perceptron (MLP) is adopted here for BBM modelling. There are three layers of nodes as illustrated in figure 2. The output values are expressed in equations (2.7) and (2.8). $w_{ij}$ is the connection weight from the $i$th neuron in input layer to the $j$th neuron in hidden layer. $v_{jt}$ is the connection weight from the $j$th neuron in hidden layer to the $t$th neuron in output layer. $\theta_j$ is the bias of the $j$th neuron in hidden layer. $\gamma_t$ is the bias of the $t$th neuron in output layer. $f$ is the activation function of hidden layer neurons. $g$ is the activation function of output layer neurons. The activation function of the hidden layer is commonly selected as a continuous, smooth, and monotone increasing function with upper and lower bounds.

$$ b_j = f(\sum_{i=1}^{n} w_{ij} a_i + \theta_j) \tag{2.6} $$

$$ c_t = g(\sum_{j=1}^{p} v_{jt} b_j + \gamma_t) \tag{2.7} $$

The output variable of the BBM_Speed is the wave-caused speed loss $Speed_{BBM}$. The input variables are related to wave information, propeller revolution rate, and WBM output $Speed_{WBM}$.

2.4 BBM_Oil Sub-model

The sub-model BBM_Oil predicts the wave-caused fuel consumption increase. The polynomial regression is adopted here for BBM modeling[9]. It is a form of regression analysis in which the
relationship between the output variable and the input variable is modeled as an $n$th degree polynomial. The output variable of the BBM_Oil is the wave-caused fuel consumption increase $Oil_{BBM}$. The input variables are related to wave information, propeller revolution rate, and WBM output $Oil_{WBM}$.

2.5 Data Enrichment

The NR dataset records the noontime weather information, noontime vessel motion, noontime operation state such as propeller revolution rate, and daily fuel consumption for the past 24 hours. The AIS dataset supplements relatively high-frequency motion state at each moment of signal transmission. The typical NR and AIS datasets are plotted together in figure 3. Between two adjacent NR data points, the ship speed fluctuated with the time, which means the weather conditions could not be constant within that period. Since the daily fuel consumption represents the accumulative amount for the past 24 hours, it must be associated with the weather change process and ship sailing process within the same period, which could not be simply represented by the noontime instant information in NR dataset. Obviously, only NR and AIS datasets could not serve well in revealing the correlation among fuel consumption, weather condition, and ship operation states.

The weather hindcast data are introduced here to enrich the NR and AIS datasets. For a certain AIS data point, the wind and current information are determined by querying from the weather hindcast dataset with the timestamp, the latitude, and the longitude of the AIS data point. The propeller revolution rate is retrieved from the neighbouring NR data point. Through the sub-model WBM_Voyage, the steady speed without wave effect $Speed_{WBM}$ is determined and assumed to be constant till the next AIS data point. The fuel consumption over this period without wave effect $Oil_{WBM}$ is also determined. Repeating the above process could supplement every AIS data point with weather information, steady speed value $Speed_{WBM}$, and fuel consumption value $Oil_{WBM}$. The wave-caused speed loss $Speed_{BBM}$ and wave-caused fuel consumption increase $Oil_{BBM}$ on each NR data point could be derived from AIS data points.

The weather change process and ship sailing process between two adjacent NR data points are recovered mostly from the enriched NR and AIS datasets. The characteristics of those stochastic variables are better represented by the statistics such as mean value and variance value over the same period. Therefore, the input variables of the sub-model BBMs are set to be statistics of weather information, propeller revolution rate, and WBM outputs. The output variables of the sub-model BBMs are statistics of the wave-caused speed loss and wave-caused fuel consumption increase, respectively. The training data and validation data for the sub-model BBMs could be obtained from the enriched NR and AIS datasets.

![Figure 3. Typical Example of NR and AIS data.](image-url)

3. Case Study

3.1 Datasets
The NR and AIS data collected from a 57000DWT bulk carrier are used to demonstrate the fidelity and capability of GBM model. The vessel particulars are listed in Table 1. The weather hindcast data are retrieved from NOAA and ECMWF[16]. The variables available in the datasets are listed in Table 2. The sampling period for the AIS dataset is 1 hour, and updating period of the weather data is 3 hours.

The datasets cover the period from August 15, 2018 to May 10, 2019, which contains 138 days in sailing. The 138 data points in NR dataset are crosschecked with AIS data and weather hindcast data to remove the incorrect NR data caused by human errors. After data pre-processing, 51 effective data points are left in the NR dataset for further training and validation.

**Table 1. Vessel particulars.**

| Vessel type         | Bulk carrier                  |
|---------------------|-------------------------------|
| Year built          | 2012                          |
| Gross Tonnage       | 33044                         |
| summer DWT          | 56741t                        |
| Length overall      | 189.99m                       |
| Breadth moulded     | 32.26m                        |
| Depth moulded       | 18.0m                         |
| Draft Scantling moulded | 12.8m                       |
| Main Engine         | MAN B&W. 6S50MCC              |
| M.C.R               | $9480\text{kW} \times 127\text{r/min}$ |

**Table 2. Variable List in Dataset.**

| Variable                  | Source | Variable                  | Source  |
|---------------------------|--------|---------------------------|---------|
| NR timestamp              | NR     | Draft                     | NR      |
| NR ship location          | NR     | Trim                      | NR      |
| Propeller speed           | NR     | AIS timestamp             | AIS     |
| Daily fuel consumption    | NR     | AIS location              | AIS     |
| Daily traveled distance   | NR     | AIS speed                 | AIS     |
| Daily steamed time        | NR     | AIS heading               | AIS     |
| Daily average speed       | NR     | Hindcast Wind speed       | ECMWF   |
| Daily wind direction      | NR     | Hindcast Wind direction   | ECMWF   |
| Daily wind grade          | NR     | Hindcast Current speed    | NOAA    |
| Daily current direction   | NR     | Hindcast Current direction| NOAA    |
| Daily current grade       | NR     |                           |         |

3.2 Parameters in WBM

The open water characteristics of the propeller, as shown in figure 4, are obtained from CFD simulation. The values of the propeller-hull interaction coefficients $\omega_p$ and $t_p$ are set to 0.2 and 0.1, respectively. The SFOC for the engine is shown in figure 5.
3.3 Features and Hyperparameters of BBMs

The multilayer perceptron (MLP) is adopted in sub-model BBM_Speed to predict the wave-caused speed loss. The polynomial regression is applied in sub-model BBM_Oil to predict the wave-caused fuel consumption increase. The features of the BBMs are determined by the correlation analysis and presented in table 3. The hyperparameters for the BBMs are listed in table 4.

### Table 3. Correlation analysis.

| Parameters                               | Corr. with speed | Corr. with fuel consumption |
|------------------------------------------|------------------|-----------------------------|
| longitudinal relative current speed      | 0.903            | 0.612                       |
| longitudinal relative wind speed         | 0.416            | 0.351                       |
| transverse relative current speed        | 0.099            | 0.137                       |
| transverse relative wind speed           | 0.150            | 0.093                       |
| propeller revolution rate                | 0.053            | 0.496                       |

### Table 4. Hyperparameters of BBMs.

| Model            | Hyperparameter      | Value               |
|------------------|---------------------|---------------------|
| BBM_Oil          | Polynomial Order    | 6                   |
|                  | Number of hidden layers | 1                   |
|                  | Hidden layer size   | 8                   |
|                  | Activation function | hyperbolic tan function |
|                  | Solver              | lbfgs               |
|                  | L2 penalty          | 1                   |
|                  | Tolerance           | $10^{-3}$           |
| BBM_Speed        | Number of hidden layers | 1                   |
|                  | Hidden layer size   | 8                   |
|                  | Activation function | hyperbolic tan function |
|                  | Solver              | lbfgs               |
|                  | L2 penalty          | 1                   |
|                  | Tolerance           | $10^{-3}$           |

4. Results

4.1 Validation

The enriched NR and AIS datasets are split into training set and validation set by K-Folds cross validator. The total number of the folds is set to 4. The training set is used to train the GBM and the pure BBMs, respectively. A pure WBM is also adopted for comparison [1]. The validation set is used to test the performance of above three models. From the validation results presented in table 5, the GBM could predict the fuel consumption with the relative error 4.8%, which is much better than 8.5% of the BBM and 28.5% of WBM.
Table 5. Validation results.

|                      | GBM | BBMs | WBM |
|----------------------|-----|------|-----|
| Speed prediction     |     |      |     |
| Error mean\(^a\)     | 3.6%| 4.0% | 2.9%|
| Error standard deviation\(^a\) | 4.9%| 5.2% | 2.2%|
| Fuel consumption prediction |     |      |     |
| Error mean           | 4.8%| 8.5% | 28.5%|
| Error standard deviation | 17.9%| 18.8% | 5.0%|

\(^a\) An error of a data point is calculated by \(error = \frac{|estimation - true| \cdot (true)^{-1} \cdot 100\%\). “Mean” represents the average of all errors, and “standard deviation” represents the standard deviation of all errors. The true value of speed is from the AIS dataset and the true value of fuel consumption is from the NR dataset.

4.2 Application
The established GBM could be applied to predict the estimated time arrival (ETA) and the total fuel consumption of any voyage. The prediction results for a past voyage are presented in table 6. The case vessel departed from Port Gibraltar at 20:53UTC, Oct. 10, 2018, and arrived at Port Lagos at 01:29UTC, Oct. 22, 2018. According to the NR dataset, the actual fuel consumption of this voyage is 339.4MT, while the GBM prediction is 337.5MT. The relative error of prediction is only 0.6%.

Table 6. GBM application results.

| Voyage          | Source | Departure time  | Arrival time  | Voyage time (h) | Distance traveled (nm) | Total oil (MT) |
|-----------------|--------|-----------------|---------------|-----------------|------------------------|----------------|
| Gibraltar to Lagos | GBM    | 10/10/2018 20:53| 10/22/2018 01:13 | 290.1           | 3368                   | 337.5          |
|                 | NR     | 10/10/2018 22:30| 10/22/2018 04:00 | 295.3           | 3419                   | 339.4          |
|                 | AIS    | 10/10/2018 20:53| 10/22/2018 01:29 | 292.8           | 3379                   | NA             |

Another important application of the proposed GBM is weather routing. Figure 6 presents a 12-day voyage optimization result from Port Louis, Mauritius to Port Singapore, Singapore. The total fuel consumption of the actual voyage is 382.3MT, while the optimized route in red line could save 28MT fuel oil (7.3%) with the same arrival time.

Figure 6. Weather Routing Optimization.
5. Conclusion

In this paper, a novel GBM approach is proposed to establish a digital twin model for voyage performance prediction. The NR and AIS datasets collected from a 57000DWT bulk carrier are used to demonstrate the fidelity and capability of the proposed GBM. The weather hindcast data are introduced to enrich the NR and AIS datasets, which are split into training and validation sets to develop GBM. The voyage performance prediction from the GBM shows better accuracy than those from pure WBM or pure BBMs. An arrival time forecast and a weather routing showcase are also presented to demonstrate the application effects of GBM.

The proposed GBM provides a satisfying prediction of ship speed and fuel consumption without mandatory sensor-collected data, thus applicable for a variety of vessels. In those cases where more sensors are available onboard, the proposed approach can incorporate sensor data to improve the model accuracy further.

6. Reference

[1] Haranen M, Pakkanen P, Kariranta R and Salo J 2016 White, Grey and Black-Box Modelling in Ship Performance Evaluation 1st Hull Performance Insight Conf. pp 115–27
[2] Moreno-Gutiérrez J, Pájaro-Velázquez E, Amado-Sánchez Y, Rodríguez-Moreno R, Calderay-Cayetano F and Durán-Grados V 2019 Comparative analysis between different methods for calculating on-board ship’s emissions and energy consumption based on operational data Sci. Total Environ. 650 pp 575–84
[3] Nielsen J B, Sandvik E, Pedersen E, Asbjørnslett B E and Fagerholt K 2019 Impact of simulation model fidelity and simulation method on ship operational performance evaluation in sea passage scenarios Ocean Eng. 188 p 106268
[4] Bialystocki N and Konovessis D 2016 On the estimation of ship's fuel consumption and speed curve: A statistical approach J. Ocean Eng. Sci. 1 pp 157–66
[5] Safaei A A, Ghassemi H and Ghiasi M 2019 VLCC’s fuel consumption prediction modeling based on noon report and automatic identification system Cogent Eng. 6 p 1595292
[6] Gkerekos C, Lazakis I and Theotokatos G 2019 Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study Ocean Eng. 188 p 106282
[7] Pedersen B P and Larsen J 2009 Modeling of Ship Propulsion Performance World Marit. Technol. Conf. pp 1–10
[8] Liang Q, Tvete H A and Brinks H W 2019 Prediction of vessel propulsion power using machine learning on AIS data, ship performance measurements and weather data J. Phys. Conf. Ser. 1357 p 012038
[9] Farag Y B A and Ölçer A I 2020 The development of a ship performance model in varying operating conditions based on ANN and regression techniques Ocean Eng. 198 p 106972
[10] Coraddu A, Oneto L, Baldi F, Cipollini F, Atlar M and Savio S 2019 Data-driven ship digital twin for estimating the speed loss caused by the marine fouling Ocean Eng. 186 p 106063
[11] Leifsson L T, Sevvarsdóttir H, Sigurdsson S T and Vésteinsson A 2008 Grey-box modeling of an ocean vessel for operational optimization Simul. Model. Pract. Theory 16 pp 923–32
[12] Safaei A A, Ghassemi H and Ghiasi M 2018 Correcting and enriching vessel's noon report data using statistical and data mining methods Eur. Transp. - Trasp. Eur. pp 1–13
[13] Lewis E V 1988 Principles of naval architecture: Volume II: Resistance, propulsion and vibration
[14] Holtrop J and Memmen G G J 1982 An Approximate Power Prediction Method Netherlands Sh. Model Basin, NSMB, Wageningen, Publ. No. 689, Publ. Int. Shipbuild. Progress, ISP, Vol. 29, Nr 335 pp 166-70
[15] Blendermann W 1994 Parameter identification of wind loads on ships J. Wind Eng. Ind. Aerodyn. 51 pp 339–51
[16] Zyczkowski M, Szłapczyńska J and Szłapczyński R 2020 Review of weather forecast services for ship routing purposes Polish Marit. Res. 26 pp 80–9