Communications in Transportation Research 2 (2022) 100073

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Full Length Article

Data fusion and machine learning for ship fuel efficiency modeling: Part II – Voyage report data, AIS data and meteorological data

Yuquan Du[a,*], Yanyu Chen[a,b], Xiaohe Li[a,c], Alessandro Schönborn[d], Zhuo Sun[e]

[a] Centre for Maritime and Logistics Management, Australian Maritime College, University of Tasmania, Launceston, TAS 7250, Australia
[b] Institute for Marine and Antarctic Studies, College of Sciences and Engineering, University of Tasmania, Taroona, TAS 7053, Australia
[c] College of Power and Energy Engineering, Harbin Engineering University, Harbin, 150001, China
[d] Maritime Energy Management, World Maritime University, Fiskehamnsgatan 1, 201 24 Malmö, Sweden
[e] College of Transportation Engineering, Dalian Maritime University, Dalian, 116026, China

ARTICLE INFO

Keywords:
Ship fuel efficiency
Fuel consumption rate
Voyage report
AIS
Data fusion
Machine learning

ABSTRACT

When voyage report data is utilized as the main data source for ship fuel efficiency analysis, its information on weather and sea conditions is often regarded as unreliable. To solve this issue, this study approaches AIS data to obtain the ship's actual detailed geographical positions along its sailing trajectory and then further retrieve the weather and sea condition information from publicly accessible meteorological data sources. These more reliable data about weather and sea conditions the ship sails through is fused into voyage report data in order to improve the accuracy of ship fuel consumption rate models. Eight 8100-TEU to 14,000-TEU containerships from a global shipping company were used in experiments. For each ship, nine datasets were constructed based on data fusion and eleven widely-adopted machine learning models were tested. Experimental results revealed the benefits of fusing voyage report data, AIS data, and meteorological data in improving the fit performances of machine learning models of forecasting ship fuel consumption rate. Over the best datasets, the performances of several decision tree-based models are promising, including Extremely randomized trees (ET), AdaBoost (AB), Gradient Tree Boosting (GB) and XGBoost (XG). With the best datasets, their R^2 values over the training sets are all above 0.96 and mostly reach the level of 0.99–1.00, while their R^2 values over the test sets are in the range from 0.75 to 0.90. Fit errors of ET, AB, GB, and XG on daily bunker fuel consumption, measured by RMSE and MAE, are usually between 0.8 and 4.5 ton/day. These results are slightly better than our previous study, which confirms the benefits of adopting the actual geographical positions of the ship recorded by AIS data, compared with the estimated geographical positions derived from the great circle route, in retrieving weather and sea conditions the ship sails through.

1. Introduction

The International Maritime Organization (IMO, 2020) has been promoting energy-efficient operational measures in voyage management to save bunker fuel and mitigate ship emissions, including sailing speed optimization, trim optimization, weather routing, and virtual (just-in-time) arrival. These measures are preferred by shipping companies because of their cost effectiveness compared to technical solutions (Wan et al., 2018; Merkel et al., 2022). The shipping industry has been facing some frustrations during the process of implementing these energy-efficient measures owing to the inability to accurately quantify how much bunker fuel a ship will consume in one day or hour in different speed, displacement/draft, trim, weather, and sea conditions. Therefore, to advance these energy-efficient measures in the shipping industry, it is paramount to lay a solid theoretical foundation by developing models that can accurately quantify the relationship between fuel consumption rate (MT/day, or MT/h) and its determinants, including sailing speed, displacement/draft, trim, weather conditions, and sea conditions.

According to Yan et al. (2021), machine learning (ML) models are emerging in recent years as an effective approach for ship fuel efficiency analysis. Multiple data sources can assist shipping companies to develop ML models of ship fuel efficiency to be used in energy-efficient operational measures in voyage management, including voyage report data, sensor data, automatic identification system (AIS) data, and meteorological data. The systematic literature review of Yan et al. (2021) and a literature review conducted in our previous study ("*Data fusion and*

* Corresponding author.
E-mail address: yuquan.du@utas.edu.au (Y. Du).

https://doi.org/10.1016/j.commtr.2022.100073
Received 30 April 2022; Received in revised form 18 June 2022; Accepted 19 June 2022
Available online xxxx

machine learning for ship fuel efficiency modeling: Part I – voyage report data and meteorological data", referred to as "Li et al. (2022)" hereinafter) show that few existing studies explore the complementary advantages of different data sources and combine multiple data sources for the sake of improving data quality for ML ship fuel efficiency models. Therefore, Li et al. (2022) propose the following research questions regarding fusion of different data sources that interest both academics and industry professionals:

- **RQ1.** Is it possible to combine/fuse different but complementary data sources for the sake of modeling accuracy for ship fuel efficiency analysis? And how these data sources can be fused?
- **RQ2.** Compared to a single data source, what are the benefits of fusing different data sources in terms of modeling accuracy and generalization?
- **RQ3.** Selection of datasets and choice of models are two different decision dimensions but they rely on each other. When these two decisions are interwoven, how can we select the best datasets and best models?

In our previous study, Li et al. (2022) noticed the data quality issue of voyage report caused by the deck officers' practice of snapshotting and eye inspecting weather and sea conditions. To remedy this issue, Li et al. (2022) developed a solution of fusing voyage report data and publicly accessible meteorological data by replacing the information of snapshotted weather and sea conditions in voyage report with accurate hourly weather and sea conditions retrieved from meteorological data. Over the nine datasets from data fusion for eight 8,100-TEU to 14,000-TEU containerships, several ship-specific ML models of forecasting ship fuel consumption rate achieve high fit performances with R^2 values all above 0.96 and even reaching 0.99 to 1.00 for training sets, while their R^2 values for test sets are also promising between 0.74 and 0.90.

In Li et al. (2022), a key step before retrieving exact information of weather and sea conditions from meteorological data is calculating the ship's hourly geographical positions (< Timestamp, latitude, longitude >) along its sailing trajectory. Li et al. (2022) assumed the ship follows the great circle route approximated by the widely adopted rhumb line and adopted the rhumb line formulas (Bennett, 1996; Weintrit and Kopacz, 2011) to calculate the geographical locations the ship passes in a day. Several ship captains we consulted commented that the great circle route may not be followed in sailing for several reasons and using the geographical positions derived from the great circle route or the rhumb line may introduce inaccuracy when weather and sea conditions are retrieved from meteorological data. This is a prominent limitation of Li et al. (2022).

To address this limitation, we approached *MarineTraffic* headquartered in Greece and purchased the AIS data of the eight containerships experimented with by Li et al. (2022), because AIS data provides the detailed geographical positions of the ship forming its actual sailing trajectory. Meanwhile, AIS data also provides the information of the ship's heading at each geographical position, and this may make the calculation of the directions of wind/waves/sea currents relative to the ship's heading more reliable. The objective of this study is to investigate whether the introduction of actual geographical positions in AIS data will improve the information quality of weather and sea conditions retrieved from meteorological data and therefore further improve the fit performances of ML models when meteorological data and voyage report data are combined.

The rest of the paper is organized as follows. Section 2 describes AIS data, introduces the approach of fusing voyage report data, AIS data, and meteorological data, and constructs several datasets based on data fusion. Over these datasets of eight containerships, Section 3 experiments with eleven ML models and makes informed decisions about dataset choice and ML model selection. Section 4 draws concluding remarks.

2. Fusion of voyage report data, AIS data and meteorological data

The same eight mega containerships as those in Li et al. (2022) are considered in this study. Data sources of voyage report and meteorological data are also same to Li et al. (2022): a global shipping company provides the voyage report data of these eight ships and the sailing period recorded by the voyage report data spans from February 2014 to March 2016; Centre for Medium-range Weather Forecasts (ECMWF) (Hersbach et al., 2018) provides the data of wind, waves, and sea water temperature in the finest granularity of 0.25° (longitude) × 0.25° (latitude) × 1 h (time); Copernicus Marine Service (CMEMS, also a.k.a. "Copernicus") (Rio et al., 2014) provides the data of sea currents in the finest granularity of 0.25° (longitude) × 0.25° (latitude) × 3 h (time).

The purchased AIS data from *MarineTraffic* has 15 columns. Apart from the identification and particulars of the ship (*MMSI*, *Call Sign*, *Ship Name*, *Flag Country*, *Draft Designed*, *Length*) and the information about the voyage (*ETA*, *Destination Port*), AIS data contains the detailed navigation data including "*Timestamp (UTC)*", "*Navigation Status*", "*Longitude Position*", "*Latitude Position*", "*Ship Course*", "*Ship Heading*", and "*Sailing Speed*". There is a data entry every 3–5 min. "*Sailing Speed*" appears to be useful in our study. However, this study considers voyage report as the main data source of ship fuel consumption which records information on a daily basis. Therefore, "*Sailing Speed*" information in the time interval of 3–5 min from AIS data does not help in this study.

The information about "*Timestamp (UTC)*", "*Longitude Position*", and "*Latitude Position*" could be quite useful in that it helps us find the actual geographical positions of the ship in a day and recover its actual sailing trajectory on that day. Further, accurate detailed information of weather and sea conditions the ship sails through can be retrieved from meteorological data, according to the actual sailing trajectory. "*Ship Heading*" information could also be useful because it helps convert the (absolute) directions of wind, waves, and sea currents reported by meteorological data to relative directions of wind, waves, and sea currents against the ship's heading, which is desired in ship fuel efficiency modeling. Li et al. (2022) had to utilize "True Course" information in voyage report in the calculation of relative directions of wind, waves, and sea currents as a workaround because voyage report does not record the heading of the ship.

The approach of fusing voyage report data, AIS data, and meteorological data is illustrated in Fig. 1. First, for a given day recorded by voyage report, the ship's hourly geographical positions are retrieved from AIS data. Specifically, the ship's positions at the times of {00:00 h, 01:00 h, 02:00 h, …, 23:00 h} are selected. If a data entry does not exist corresponding to e.g. 05:00 h, the entry in AIS data whose timestamp is closest to this time point is used as the substitute. Second, according to these geographical positions, the hourly weather and sea condition information are queried and obtained from meteorological data including ECMWF (wind, waves, sea water temperature) and Copernicus (sea currents). Then the directions of wind, waves, and sea currents are converted to the relative directions to the ship's heading. Third, these hourly weather and sea conditions are aggregated and produce their daily averages. At last, daily average conditions of wind, waves, sea water temperature, and sea currents are used to replace the meteorological record in the voyage report.

In this data fusion approach, the noises of AIS data are not a concern because only hourly geographical positions of the ship are needed for the sake of retrieval of weather and sea conditions. Finer positions of the ship from AIS data are meaningless because our target in this study is the daily average weather and sea conditions the ship sails through for each day in the voyage report. Even if there were noises in sampling the ship's hourly geographical positions from AIS data, they would not cause a problem in calculating the daily average weather and sea conditions confronting the ship. Meanwhile, hourly positions of the ship already enable us to retrieve the weather and sea conditions in the resolution of 15–28 nautical miles (corresponding to the sailing speeds of 15–28 knots),

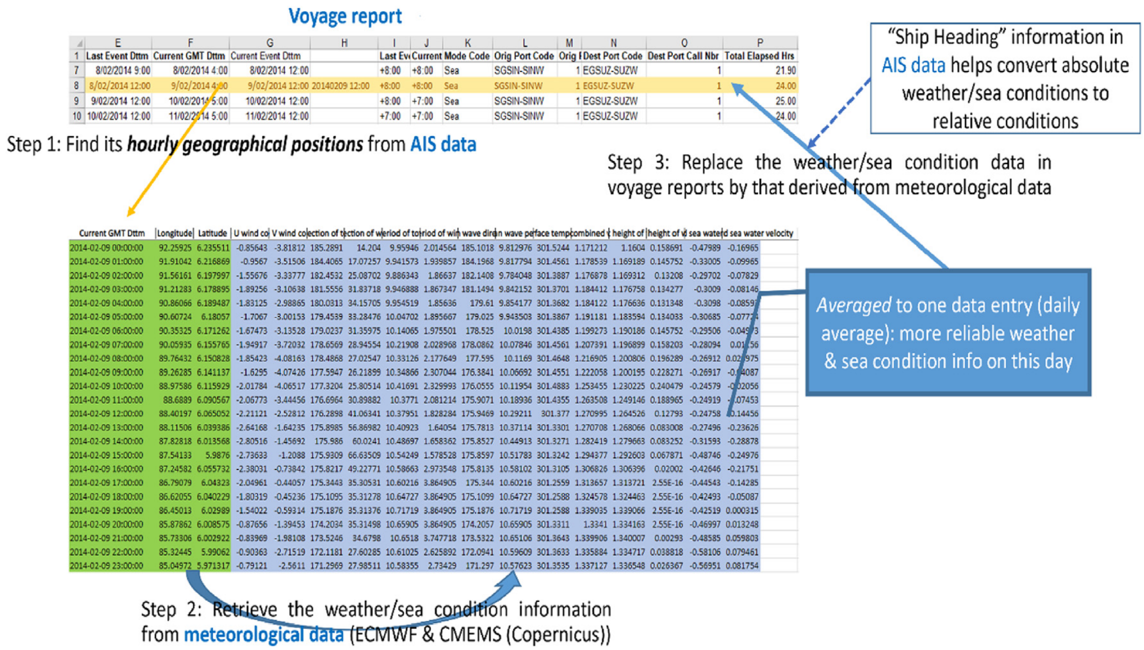


Fig. 1. Approach of fusing voyage report data, AIS data, and meteorological data.

which is sufficient for our research purpose. Finer positions of the ship, say half-hourly positions, will break the daily trajectory into too short segments (say each segment with 7–14 nautical miles) that do not make sense and might be even shorter than the geographical granularity of the finest meteorological dataset.

Similar to Li et al. (2022), this study also allows the conversion of the precise values representing wind speed and relative directions of wind, waves, and sea currents to fuzzy values. See Tables 2 and 3 and Fig. 1 of Li et al. (2022). This is because voyage reports usually adopt fuzzy values and our preliminary experiments show that fuzzy values sometimes overcome data noises/inaccuracy and improve fit performance of ML models. Overall, nine datasets are constructed from this data fusion approach, and the features of each dataset are listed in Table 1. "*Set1*" in Table 1 is exactly the same "*Set1*" in Li et al. (2022), which represents the voyage report.

3. Experimental results and discussion

The eight containerships, voyage report data and meteorological data, ML models, performance metrics of models, and experimental settings are all the same as our previous study Li et al. (2022), which allows experimental comparison with Li et al. (2022). Eleven ML models that are experimented with include artificial neural network (ANN) (Haykin, 2008), support vector machine (SVM) (Boser et al., 1992), ridge regression (Ridge) (Hoerl and Kennard, 1970), LASSO (Tibshirani, 1996), basic decision tree (DT) (Breiman et al., 1984), Extremely randomized trees (ET) (Geurts et al., 2006), random forest (RF) (Breiman, 2001), AdaBoost (AB) (Freund and Schapire, 1997; Drucker, 1997), gradient tree boosting (GB) (Friedman, 2001), XGBoost (XG) (Chen and Guestrin, 2016), and LightGBM (LB) (Ke et al., 2017).

3.1. Performances of eleven ML models over nine datasets and selection of the best datasets

Same as Li et al. (2022), for each dataset in Table 1, we randomly divided it to a training set (80% of data entries) and test set (20% of data entries), which results in a *split* of the dataset. For each split of the dataset, we experimented with a given ML model involving a process of *five-fold cross-validation* based hyperparameter optimization with the Bayesian Optimization method using the tree-structured Parzen Estimators of hyperopt 0.2.2 library (*Hyperopt*) (Bergstra et al., 2013), which is called a *run*. For each ML model over each dataset, we have 20 random splits of the dataset and thus 20 runs of experiments. Each performance metric (R^2, *MSE*, *RMSE*, *MAE* and *MAPE* for training set, R^2 *(test)* for test set, see definition in Li et al. (2022)) takes the average of 20 runs to overcome the influence of random splitting of the dataset. Experimental results of ship S1 are reported in Table 2, while the results of ships S2 to S8 can be found in Tables A1 to A7 in Appendix. Note that the performances over the best datasets found by Li et al. (2022), including *Set1* and $Set3_{precise}$, are also reported in Tables 2 and A1 to A7, for the convenience of comparison with Li et al. (2022).

When quality of datasets and performance of ML models are interwoven, shown in Tables 2 and A1 to A7, a voting scheme same as Li et al. (2022) is adopted. Each ML model acts as a voter and votes for the best datasets (candidates) by considering R^2 (with two decimal places) as the first priority and R^2 (test) (with two decimal places) as the secondary performance metric. The voting result is collated in Table 3 in which the last column is the votes of the corresponding ML models (voters). Fig. 2 is the Tally sheet that counts the votes received by each dataset: Fig. 2(a) considers all the models as voters; Fig. 2(b) does not consider DT, SVM, ANN, Ridge, and LASSO as voters, because their fit performances are significantly worse than ET, RF, AB, GB, XG and LB and thus they will not be preferred by industry applications; Fig. 2(c) further removes RF, GB and LB from the voter list because they are dominated by ET, AB, and XG against both R^2 and R^2 (test).

It can be seen from Fig. 2 that $AIS5_{precise}$ receives the largest number of votes, followed by $Set3_{precise}$ and *Set1*. $AIS5_{precise}$ receives 18 votes from ET, RF, AB, GB, XG and LB, according to Fig. 2(b), and 9 votes from ET, AB, and XG according to Fig. 2(c). This reveals that when AIS data is available for ship fuel efficiency analysis, $AIS5_{precise}$ is the best, and this dataset is better than *Set1* and $Set3_{precise}$ from Li et al. (2022) which is the Part I of this series of studies. This demonstrates the benefits of further fusing AIS data to voyage report data and meteorological data considered in Li et al. (2022). Therefore, we recommend using $AIS5_{precise}$ in practice by fusing voyage report data, AIS data and meteorological data. When AIS data is not available, we can combine voyage report data and meteorological data and utilize $Set3_{precise}$, or even adopt voyage report data *Set1* directly.

Table 1
Features contained in each dataset.

Original datasets	Data source	Features	Dataset								
			Set1	$AIS2_{precise}$ [b]	$AIS2_{fuzzy}$ [c]	$AIS3_{precise}$ [b]	$AIS3_{fuzzy}$ [c]	$AIS4_{precise}$ [b]	$AIS4_{fuzzy}$ [c]	$AIS5_{precise}$ [b]	$AIS5_{fuzzy}$ [c]
Voyage report data	Shipping company	Fuel consumption rate	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Sailing speed	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Displacement	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Trim	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Wind speed	✓								
		Wind direction (Rel.)	✓								
		Swell height	✓								
		Swell direction (Rel.)	✓								
		Sea currents speed	✓								
		Sea currents direction (Rel.)	✓								
		Sea water temperature	✓								
AIS + Meteorological data	AIS+ European Centre for Medium-range Weather Forecasts (ECMWF)	Wind speed		✓	✓	✓	✓	✓	✓	✓	✓
		Wind direction (Rel.)[a]		✓	✓	✓	✓	✓	✓	✓	✓
		Swell height		✓	✓	✓	✓	✓	✓		
		Swell direction (Rel.)[a]		✓	✓	✓	✓	✓	✓		
		Swell period									
		Wind wave height				✓	✓	✓	✓		
		Wind wave direction (Rel.)[a]				✓	✓	✓	✓		
		Wind wave period									
		Combined wave height				✓	✓			✓	✓
		Combined wave direction (Rel.)[a]				✓	✓			✓	✓
		Combined wave period									
		Sea water temperature		✓	✓	✓	✓	✓	✓	✓	✓
	AIS + Copernicus Marine Service	Sea current speed		✓	✓	✓	✓	✓	✓	✓	✓
		Sea current direction (Rel.)[a]		✓	✓	✓	✓	✓	✓	✓	✓

Note:

[a] Relative directions of wind/waves/sea currents are calculated based on ship's “heading” information from AIS data.
[b] The subscript “precise” means the directions of wind/waves/sea currents are calculated as the angles relative to ship's heading measured by degrees.
[c] The subscript “fuzzy” means the precise information of directions of wind/waves/sea currents is converted to fuzzy data as per Table 2 and Fig. 1 of Li et al. (2022), and wind speed is represented by Beaufort scale numbers as per Table 3 of Li et al. (2022).

Looking at the results here, one may ask why other datasets in Table 1 combine voyage report data, meteorological data, and AIS data but are not competitive with the original voyage report data *Set1*, and even the best dataset $AIS5_{precise}$ cannot always win the original voyage report dataset *Set1*. Similarly, regarding the results reported in our previous study (Li et al., 2022), one may ask why many datasets in Li et al. (2022) combine voyage report data and meteorological data but are not competitive with the original voyage report data *Set1*, and even the best dataset $Set3_{precise}$ in Li et al. (2022) cannot always win the original voyage report dataset *Set1*. Our deep investigation into the data reveals the following possible reasons. First, as reported by ECMWF and CMEMS in their websites, their meteorological data cannot avoid inaccuracy and errors, because these data relies on many types of collection equipment and the calculation of many models. As the evidence, we will see in Part III of this series of studies that the wind conditions contained in ECMWF are quite different from the actual wind conditions captured by the sensors on board the ships.

Second, the power of “average” calculation plays a critical role in reducing the quality of data used for model training. Specifically, the weather condition for a given day (corresponding to a voyage report data entry) is estimated by taking the average of the weather conditions at 24 waypoints (hourly waypoints) during the day. However, even accurate weather conditions at these waypoints cannot guarantee their daily average is closer to the actual weather condition (the reality). To provide an analogy for the sake of understanding, consider a situation in which we are estimating the actual average/mean value of a random variable through several observations. Assume the actual average value of this random variable is 10. Consider two different samples of observations: *Sample 1* = {9.5, 9.5, 9.5, 11, 12, 13} and *Sample 2* = {9, 7, 5, 11, 13, 15}. The deviation of data in *Sample 1* from the real average ({0.5, 0.5, 0.5, 1, 2, 3}) is much smaller than that of *Sample 2* ({1, 3, 5, 1, 3, 5}). However, the average value estimated through *Sample 1* is 10.75, which is worse than that estimated from *Sample 2* (i.e., 10, the same as the actual average).

Third, the quality of voyage report data might already be good enough. Specifically, when it turns to the snapshotted weather and sea condition data, a ship captain we consulted pointed out “***though the snapshotted weather and sea condition data is not desired, if you***

Table 2
The fit performance of eleven machine learning models for ship S1.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	*Set1*	0.846	0.643	81.022	8.934	6.851	7.995
	$AIS2_{precise}$	0.840	0.630	77.176	8.694	6.705	7.959
	$AIS2_{fuzzy}$	0.822	0.623	85.984	9.211	7.093	8.448
	$AIS3_{precise}$	0.827	0.624	83.223	9.057	6.908	8.190
	$AIS3_{fuzzy}$	0.837	0.630	78.690	8.719	6.714	7.970
	$AIS4_{precise}$	0.841	0.641	76.360	8.633	6.604	7.779
	$AIS4_{fuzzy}$	0.841	0.625	76.849	8.681	6.688	7.928
	$AIS5_{precise}$	0.838	0.618	78.187	8.788	6.765	7.982
	$AIS5_{fuzzy}$	0.835	0.635	79.655	8.869	6.857	8.166
	$Set3_{precise}$ [a]	0.847	0.617	73.848	8.532	6.522	7.697
ET	*Set1*	0.992	0.781	4.001	1.525	1.090	1.255
	$AIS2_{precise}$	0.958	0.773	20.546	4.176	3.152	3.765
	$AIS2_{fuzzy}$	0.955	0.766	21.556	4.174	3.164	3.797
	$AIS3_{precise}$	0.960	0.767	19.295	4.067	3.079	3.691
	$AIS3_{fuzzy}$	0.945	0.772	26.555	4.972	3.829	4.610
	$AIS4_{precise}$	0.959	0.768	19.646	4.095	3.091	3.719
	$AIS4_{fuzzy}$	0.966	0.769	16.417	3.578	2.716	3.266
	$AIS5_{precise}$	0.951	0.773	23.833	4.511	3.428	4.084
	$AIS5_{fuzzy}$	0.952	0.771	23.140	4.393	3.374	4.034
	$Set3_{precise}$ [a]	0.965	0.762	17.043	3.524	2.699	3.245
RF	*Set1*	0.964	0.761	18.837	4.321	3.194	3.721
	$AIS2_{precise}$	0.940	0.757	29.138	5.322	3.997	4.760
	$AIS2_{fuzzy}$	0.934	0.757	31.834	5.575	4.183	4.994
	$AIS3_{precise}$	0.932	0.753	32.837	5.657	4.221	5.028
	$AIS3_{fuzzy}$	0.943	0.756	27.491	5.186	3.895	4.635
	$AIS4_{precise}$	0.932	0.754	32.987	5.663	4.216	5.041
	$AIS4_{fuzzy}$	0.940	0.758	29.145	5.335	3.985	4.762
	$AIS5_{precise}$	0.938	0.751	30.021	5.416	4.040	4.798
	$AIS5_{fuzzy}$	0.949	0.766	24.816	4.914	3.678	4.368
	$Set3_{precise}$ [a]	0.936	0.756	30.736	5.506	4.112	4.911
AB	*Set1*	0.955	0.758	23.482	4.687	4.036	4.940
	$AIS2_{precise}$	0.956	0.762	21.288	4.333	3.661	4.495
	$AIS2_{fuzzy}$	0.947	0.759	25.519	4.648	3.801	4.620
	$AIS3_{precise}$	0.947	0.755	25.740	4.879	4.148	5.082
	$AIS3_{fuzzy}$	0.958	0.759	20.443	4.176	3.395	4.129
	$AIS4_{precise}$	0.946	0.761	26.012	4.816	4.091	5.023
	$AIS4_{fuzzy}$	0.951	0.751	23.966	4.460	3.733	4.568
	$AIS5_{precise}$	0.950	0.763	24.422	4.732	3.966	4.854
	$AIS5_{fuzzy}$	0.963	0.765	17.825	3.861	3.142	3.804
	$Set3_{precise}$ [a]	0.938	0.752	29.988	5.180	4.370	5.371
GB	*Set1*	0.987	0.764	6.570	2.238	1.633	1.893
	$AIS2_{precise}$	0.958	0.740	20.367	4.158	3.130	3.722
	$AIS2_{fuzzy}$	0.943	0.756	27.321	5.079	3.867	4.574
	$AIS3_{precise}$	0.961	0.749	19.024	3.972	2.993	3.552
	$AIS3_{fuzzy}$	0.955	0.759	21.837	4.113	3.167	3.757
	$AIS4_{precise}$	0.952	0.746	23.273	4.533	3.398	4.068
	$AIS4_{fuzzy}$	0.957	0.752	20.695	4.328	3.319	3.954
	$AIS5_{precise}$	0.941	0.738	28.137	4.950	3.745	4.445
	$AIS5_{fuzzy}$	0.955	0.754	21.469	4.385	3.360	3.967
	$Set3_{precise}$ [a]	0.962	0.743	18.367	3.776	2.825	3.330
XG	*Set1*	0.995	0.771	2.805	1.392	1.008	1.168
	$AIS2_{precise}$	0.964	0.755	17.318	3.687	2.753	3.217
	$AIS2_{fuzzy}$	0.951	0.759	23.599	4.375	3.297	3.850
	$AIS3_{precise}$	0.955	0.753	21.321	4.017	2.952	3.439
	$AIS3_{fuzzy}$	0.951	0.766	23.136	4.388	3.304	3.850
	$AIS4_{precise}$	0.959	0.757	19.655	4.101	3.045	3.564
	$AIS4_{fuzzy}$	0.945	0.756	26.670	4.796	3.626	4.259
	$AIS5_{precise}$	0.940	0.755	28.831	5.231	3.827	4.389
	$AIS5_{fuzzy}$	0.957	0.759	20.713	4.086	3.018	3.475
	$Set3_{precise}$ [a]	0.953	0.734	22.403	4.236	3.177	3.695
LB	*Set1*	0.989	0.755	5.857	2.183	1.652	1.924
	$AIS2_{precise}$	0.941	0.732	28.704	4.895	3.705	4.403
	$AIS2_{fuzzy}$	0.927	0.737	34.903	5.699	4.347	5.157
	$AIS3_{precise}$	0.931	0.742	33.072	5.560	4.152	4.962
	$AIS3_{fuzzy}$	0.922	0.739	37.613	5.982	4.551	5.421
	$AIS4_{precise}$	0.929	0.723	34.023	5.658	4.281	5.131
	$AIS4_{fuzzy}$	0.914	0.723	41.501	6.174	4.671	5.577
	$AIS5_{precise}$	0.927	0.731	34.888	5.579	4.095	4.907
	$AIS5_{fuzzy}$	0.938	0.736	29.667	5.060	3.760	4.486
	$Set3_{precise}$ [a]	0.943	0.723	27.467	4.806	3.609	4.272
SVM	*Set1*	0.861	0.784	73.082	8.540	6.365	7.156
	$AIS2_{precise}$	0.868	0.795	63.462	7.956	5.915	6.810
	$AIS2_{fuzzy}$	0.865	0.796	64.961	8.047	6.059	6.973
	$AIS3_{precise}$	0.864	0.794	65.598	8.088	6.076	7.012
	$AIS3_{fuzzy}$	0.876	0.796	59.629	7.692	5.732	6.604
	$AIS4_{precise}$	0.865	0.793	64.842	8.037	6.034	6.961
	$AIS4_{fuzzy}$	0.870	0.786	62.681	7.882	5.928	6.838
	$AIS5_{precise}$	0.863	0.799	65.964	8.114	6.080	6.999
	$AIS5_{fuzzy}$	0.867	0.798	64.071	7.994	5.977	6.866
	$Set3_{precise}$ [a]	0.858	0.786	68.382	8.263	6.143	7.059
ANN	*Set1*	0.869	0.781	68.911	8.290	6.391	7.296
	$AIS2_{precise}$	0.876	0.773	59.980	7.662	5.914	6.866
	$AIS2_{fuzzy}$	0.901	0.778	48.121	6.838	5.360	6.232
	$AIS3_{precise}$	0.865	0.784	65.208	8.036	6.231	7.285
	$AIS3_{fuzzy}$	0.864	0.781	66.001	8.030	6.270	7.379
	$AIS4_{precise}$	0.859	0.758	68.527	7.974	6.165	7.174
	$AIS4_{fuzzy}$	0.878	0.775	58.615	7.552	5.914	6.949
	$AIS5_{precise}$	0.871	0.780	62.329	7.848	6.054	7.041
	$AIS5_{fuzzy}$	0.868	0.773	63.823	7.814	6.053	7.071
	$Set3_{precise}$ [a]	0.854	0.778	70.184	8.366	6.437	7.518
Ridge	*Set1*	0.814	0.774	97.422	9.868	7.725	8.932
	$AIS2_{precise}$	0.826	0.786	83.624	9.143	7.097	8.337
	$AIS2_{fuzzy}$	0.823	0.783	85.424	9.241	7.252	8.528
	$AIS3_{precise}$	0.835	0.792	79.647	8.923	6.999	8.239
	$AIS3_{fuzzy}$	0.833	0.793	80.613	8.977	7.083	8.330
	$AIS4_{precise}$	0.832	0.790	80.693	8.981	7.014	8.239
	$AIS4_{fuzzy}$	0.828	0.788	82.842	9.100	7.205	8.459
	$AIS5_{precise}$	0.828	0.788	82.760	9.096	7.029	8.250
	$AIS5_{fuzzy}$	0.825	0.786	84.246	9.177	7.121	8.359
	$Set3_{precise}$ [a]	0.830	0.784	81.939	9.050	6.993	8.192
LASSO	*Set1*	0.814	0.773	97.552	9.875	7.711	8.917
	$AIS2_{precise}$	0.826	0.784	83.984	9.162	7.109	8.353
	$AIS2_{fuzzy}$	0.822	0.782	85.539	9.247	7.256	8.536
	$AIS3_{precise}$	0.834	0.793	79.815	8.932	6.999	8.238
	$AIS3_{fuzzy}$	0.832	0.791	81.130	9.005	7.087	8.327
	$AIS4_{precise}$	0.832	0.792	80.977	8.997	7.027	8.257
	$AIS4_{fuzzy}$	0.828	0.789	83.030	9.110	7.212	8.473
	$AIS5_{precise}$	0.828	0.788	82.784	9.097	7.041	8.271
	$AIS5_{fuzzy}$	0.825	0.784	84.263	9.178	7.124	8.369
	$Set3_{precise}$ [a]	0.829	0.786	82.204	9.064	6.997	8.191

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

snapshotted 8-m waves/swells, it is almost impossible that your ship sailed through good weather and sea conditions on average on that day". This comment indicates that the snapshotted weather and sea condition data might be representative, though to unknow degrees, for the actual weather and sea conditions the ship sails through in a day.

3.2. Performance comparison of ML models

While Table 3 reveals the performances of different ML models, we further report their performances over the best dataset $AIS5_{precise}$ of eight ships in Table 4. Tables 3 and 4 both confirm that ET, RF, AB, GB, XG and LB are good candidate models that can be adopted by the shipping industry. Their R^2 values over the best datasets are all above 0.95 and even reach the level of 0.99–1.00, while their R^2 performance over the test sets is in the range from 0.75 to 0.90. The remaining models, including DT, SVM, ANN, Ridge, and LASSO, are not recommended for industry applications because their R^2 values on the training sets are usually comparatively low, while the values of R^2 over the test sets have not shown any advantages compared to ET, RF, AB, GB, XG and LB.

Further, the fit performances of RF and LB are usually slightly dominated by ET, AB, GB, and XG, against both R^2 and R^2 (test), which confirms the sufficiency of only installing ET, AB, GB and XG in industry applications related to ship fuel efficiency analysis. GB can also be removed from industry installation once XG has already be installed

Table 3
Best performance of each machine learning model from ten datasets and the datasets that achieve the best performance. R^2 (with two decimal places) is considered as the first priority, and R^2 (test) (with two decimal places) is the secondary performance metric.

Ship	Model	Best R^2	Best R^2 (test)	Datasets
S1	DT	0.85	0.64	$Set1$
	ET	0.99	0.78	$Set1$
	RF	0.96	0.76	$Set1$
	AB	0.96	0.77	$AIS5_{fuzzy}$
	GB	0.99	0.76	$Set1$
	XG	1.00	0.77	$Set1$
	LB	0.99	0.76	$Set1$
	SVM	0.88	0.80	$AIS3_{fuzzy}$
	ANN	0.90	0.78	$AIS2_{fuzzy}$
	Ridge	0.84	0.79	$AIS3_{precise}$
	LASSO	0.83	0.79	$AIS3_{precise}$
S2	DT	0.85	0.65	$AIS2_{precise}$
	ET	0.98	0.78	$AIS5_{precise}$
	RF	0.96	0.77	$Set1$
	AB	0.98	0.75	$AIS2_{precise}$, $AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS4_{precise}$, $AIS5_{precise}$
	GB	0.99	0.77	$AIS4_{precise}$
	XG	0.99	0.77	$Set3_{precise}$
	LB	0.98	0.75	$Set3_{precise}$
	SVM	0.88	0.82	$AIS3_{precise}$, $AIS4_{precise}$
	ANN	0.91	0.79	$Set3_{precise}$
	Ridge	0.84	0.81	$AIS3_{precise}$, $AIS4_{precise}$
	LASSO	0.84	0.81	$AIS3_{precise}$, $AIS4_{precise}$
S3	DT	0.87	0.71	$AIS3_{precise}$
	ET	0.99	0.82	$AIS2_{precise}$, $Set3_{precise}$
	RF	0.96	0.81	$AIS2_{precise}$, $AIS5_{precise}$, $AIS5_{fuzzy}$
	AB	1.00	0.82	$AIS5_{precise}$
	GB	0.97	0.82	$AIS2_{precise}$, $AIS3_{precise}$, $AIS5_{precise}$, $AIS5_{fuzzy}$
	XG	0.98	0.82	$AIS5_{precise}$
	LB	0.95	0.80	$AIS2_{precise}$, $AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS5_{precise}$, $AIS5_{fuzzy}$, $Set3_{precise}$
	SVM	0.85	0.83	$AIS4_{precise}$
	ANN	0.87	0.80	$AIS3_{precise}$, $Set3_{precise}$
	Ridge	0.80	0.80	$AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS4_{precise}$, $AIS4_{fuzzy}$, $AIS5_{precise}$, $Set3_{precise}$
	LASSO	0.80	0.80	$AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS4_{precise}$, $AIS4_{fuzzy}$, $AIS5_{precise}$, $Set3_{precise}$
S4	DT	0.93	0.73	$AIS3_{fuzzy}$
	ET	1.00	0.87	$AIS2_{precise}$, $AIS3_{precise}$, $AIS4_{precise}$, $AIS5_{precise}$, $AIS5_{fuzzy}$, $Set3_{precise}$
	RF	0.98	0.86	$AIS2_{precise}$, $AIS5_{precise}$, $AIS5_{fuzzy}$
	AB	0.99	0.87	$AIS2_{precise}$, $AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS5_{precise}$, $AIS5_{fuzzy}$, $Set3_{precise}$
	GB	1.00	0.87	$AIS3_{precise}$
	XG	1.00	0.87	$AIS3_{fuzzy}$, $Set3_{precise}$
	LB	0.99	0.87	$AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS5_{precise}$, $AIS5_{fuzzy}$
	SVM	0.94	0.85	$AIS2_{fuzzy}$
	ANN	0.95	0.86	$Set3_{precise}$
	Ridge	0.83	0.82	$Set1$
	LASSO	0.83	0.81	$AIS3_{precise}$, $AIS4_{precise}$, $AIS5_{precise}$, $Set3_{precise}$
S5	DT	0.95	0.83	$AIS5_{fuzzy}$
	ET	1.00	0.90	$Set1$, $AIS2_{precise}$, $AIS3_{precise}$, $AIS4_{precise}$
	RF	0.98	0.89	$AIS3_{fuzzy}$, $AIS4_{fuzzy}$, $AIS5_{fuzzy}$
	AB	1.00	0.90	$AIS3_{fuzzy}$, $AIS5_{fuzzy}$
	GB	1.00	0.89	$AIS2_{precise}$, $AIS2_{fuzzy}$, $AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS4_{precise}$, $AIS5_{precise}$
	XG	1.00	0.89	$AIS3_{precise}$
	LB	0.99	0.88	$Set1$, $AIS2_{precise}$, $AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS5_{precise}$
	SVM	0.93	0.88	$Set1$
	ANN	0.94	0.89	$AIS2_{fuzzy}$
	Ridge	0.89	0.88	$AIS2_{precise}$, $AIS5_{fuzzy}$
	LASSO	0.89	0.88	$AIS2_{precise}$
S6	DT	0.86	0.57	$AIS4_{precise}$
	ET	0.99	0.77	$Set1$, $AIS2_{precise}$, $AIS2_{fuzzy}$, $AIS3_{fuzzy}$
	RF	0.96	0.77	$Set1$
	AB	0.99	0.75	$AIS3_{precise}$
	GB	0.97	0.79	$Set1$
	XG	0.97	0.79	$Set1$
	LB	0.97	0.77	$AIS2_{fuzzy}$
	SVM	0.86	0.77	$AIS2_{precise}$
	ANN	0.88	0.76	$AIS2_{precise}$
	Ridge	0.79	0.75	$AIS3_{precise}$, $AIS3_{fuzzy}$, $AIS4_{precise}$
	LASSO	0.79	0.75	$AIS3_{precise}$, $AIS4_{precise}$
S7	DT	0.88	0.68	$Set3_{precise}$
	ET	0.99	0.81	$Set3_{precise}$
	RF	0.97	0.82	$AIS5_{fuzzy}$
	AB	0.99	0.83	$AIS5_{precise}$
	GB	0.99	0.79	$Set3_{precise}$
	XG	0.99	0.78	$Set3_{precise}$
	LB	0.98	0.81	$AIS3_{precise}$, $AIS3_{fuzzy}$
	SVM	0.91	0.79	$Set1$
	ANN	0.90	0.82	$AIS4_{precise}$
	Ridge	0.82	0.76	$Set3_{precise}$
	LASSO	0.82	0.76	$Set3_{precise}$
S8	DT	0.93	0.78	$AIS5_{fuzzy}$
	ET	1.00	0.88	$Set1$, $AIS5_{precise}$, $Set3_{precise}$
	RF	0.98	0.86	$Set1$, $AIS5_{precise}$, $AIS5_{fuzzy}$, $Set3_{precise}$
	AB	1.00	0.87	$AIS5_{precise}$, $AIS5_{fuzzy}$
	GB	0.99	0.86	$AIS5_{precise}$, $AIS5_{fuzzy}$, $Set3_{precise}$
	XG	0.99	0.88	$Set1$
	LB	0.98	0.87	$Set1$
	SVM	0.91	0.87	$Set3_{precise}$
	ANN	0.92	0.86	$Set3_{precise}$
	Ridge	0.88	0.85	$Set3_{precise}$
	LASSO	0.88	0.85	$Set3_{precise}$

because GB and XG have close fit performances. Fit errors of ET, AB, GB, and XG on daily bunker fuel consumption, measured by RMSE and MAE, are usually between 0.8 and 4.5 ton/day, though fit errors might be over 4.5 ton/day occasionally if datasets are not carefully chosen.

The experimental results reported in Tables 3 and 4 also rank the performances of eleven ML models into the following four different tiers. The performances of the models in the same tier are quite close, while those of the models in different tiers are significantly different. All the experimental findings for fit performance of ML models are consistent with those from our Part I of this series of studies (Li et al., 2022).

- Tier 1: ET, AB, GB, and XG;
- Tier 2: RF, LB;
- Tier 3: DT, SVM, ANN; and
- Tier 4: Ridge, LASSO.

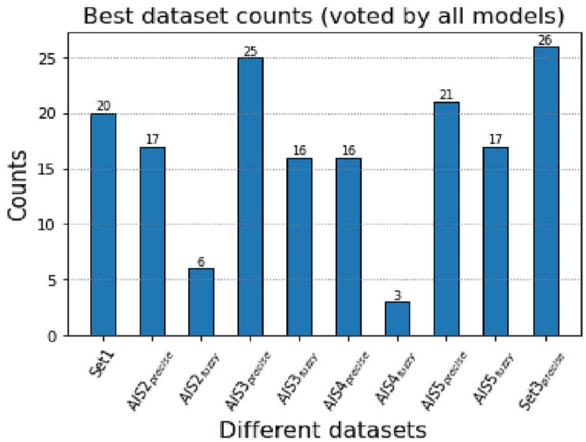
3.3. The impact of wave period

We further added "combined waves period" to the best dataset $AIS5_{precise}$ to see whether adding wave period information improves the experimental result. The experimental results of three best models (ET, AB, and XG) for ships S1, S3, S5, and S8 are shown in Fig. 3.

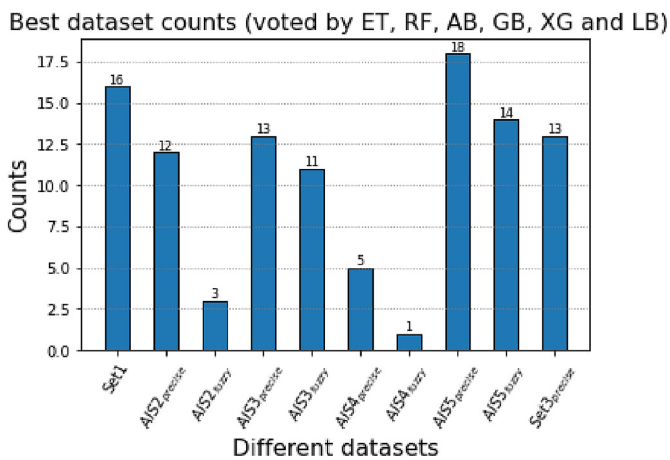
Fig. 3 reveals that including wave period information into models might improve the fit performance of models (Ships S1 and S5) but this improvement is often negligible. It might also slightly reduce the fit performance of models. This indicates that the influence of wave period on the fuel consumption rate of a mega conainership at sea is negligible and could be explained by the noises associated with the training data. By considering the consistent result in Part I of this series of studies (Li et al., 2022), we do not recommend including wave period into models, if voyage report data and meteriological data are combined, no matter whether AIS data is involved.

3.4. An experimental summary of this study and Li et al. (2022)

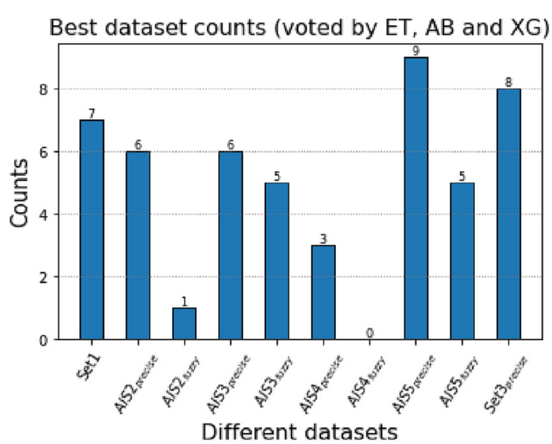
This section summarizes the experimental findings in Li et al. (2022) and this study. Fig. 4 illustrates the fit performances (R^2 and RMSE) of three best models (ET, AB and XG) over three best datasets: ***Set1*** is the original voyage report data, $\boldsymbol{Set3_{precise}}$ represents the best dataset by



(a) Best dataset counts (voted by all models)



(b) Best dataset counts (voted by ET, RF, AB, GB, XG and LB)



(c) Best dataset counts (voted by ET, AB, and XG)

Fig. 2. Best datasets voted by machine learning models.

fusing voyage report data and meteorological data, and $\boldsymbol{AIS5_{precise}}$ represents the best dataset by fusing voyage report data, meteorological data, and AIS data. Overall, as shown in the Tally sheet in Fig. 2, $\boldsymbol{AIS5_{precise}}$ is slightly better than $\boldsymbol{Set3_{precise}}$ which in turn is slightly better than ***Set1***. The fit errors of ET, AB and XG over these datasets are normally within 5 ton/day and can be as low as less than 1 ton/day.

Figs. 2 and 4 also reveal that the decision of selecting good ML models

Table 4
The fit performance of eleven machine learning models over dataset $AIS5_{precise}$.

Ship	Model	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
S1	DT	0.838	0.527	78.187	8.788	6.765	7.982
	ET	0.951	0.719	23.833	4.511	3.428	4.084
	RF	0.938	0.692	30.021	5.416	4.040	4.798
	AB	0.950	0.706	24.422	4.732	3.966	4.854
	GB	0.941	0.676	28.137	4.950	3.745	4.445
	XG	0.940	0.696	28.831	5.231	3.827	4.389
	LB	0.927	0.667	34.888	5.579	4.095	4.907
	SVM	0.863	0.752	65.964	8.114	6.080	6.999
	ANN	0.871	0.728	62.329	7.848	6.054	7.041
	Ridge	0.828	0.738	82.760	9.096	7.029	8.250
	LASSO	0.828	0.737	82.784	9.097	7.041	8.271
S2	DT	0.838	0.546	100.776	9.878	7.326	8.621
	ET	0.979	0.717	13.486	3.239	2.390	2.760
	RF	0.948	0.693	32.466	5.645	4.072	4.748
	AB	0.975	0.690	15.834	3.830	3.227	3.765
	GB	0.964	0.705	22.248	4.288	3.136	3.537
	XG	0.965	0.700	21.869	4.287	2.959	3.263
	LB	0.959	0.662	25.714	4.671	3.276	3.798
	SVM	0.880	0.758	74.547	8.599	6.191	6.738
	ANN	0.895	0.744	65.299	8.036	6.139	6.847
	Ridge	0.829	0.760	106.864	10.332	7.780	8.791
	LASSO	0.829	0.760	106.939	10.336	7.779	8.774
S3	DT	0.867	0.630	97.066	9.534	6.948	8.222
	ET	0.982	0.799	13.029	3.188	1.858	2.364
	RF	0.963	0.777	27.121	5.168	3.376	4.170
	AB	0.995	0.789	3.588	1.728	1.292	1.513
	GB	0.969	0.787	22.552	4.221	2.710	3.366
	XG	0.976	0.789	17.745	3.884	2.439	2.960
	LB	0.952	0.770	35.323	5.432	3.468	4.412
	SVM	0.840	0.795	115.821	10.727	6.746	8.262
	ANN	0.860	0.781	101.322	10.029	6.763	8.262
	Ridge	0.796	0.771	147.564	12.133	8.425	10.745
	LASSO	0.796	0.770	147.612	12.135	8.426	10.746
S4	DT	0.904	0.706	78.681	8.637	6.425	6.897
	ET	0.998	0.849	1.642	0.927	0.651	0.696
	RF	0.975	0.835	20.097	4.472	3.292	3.590
	AB	0.988	0.847	9.701	2.956	2.466	2.774
	GB	0.991	0.851	7.631	2.412	1.810	1.933
	XG	0.992	0.848	6.441	2.074	1.548	1.640
	LB	0.992	0.842	6.859	2.321	1.771	1.925
	SVM	0.927	0.836	59.875	7.686	5.642	6.018
	ANN	0.939	0.846	50.219	7.062	5.524	5.962
	Ridge	0.827	0.775	141.409	11.888	9.244	9.534
	LASSO	0.827	0.775	141.586	11.895	9.244	9.530
S5	DT	0.948	0.764	28.458	5.104	3.741	5.634
	ET	0.997	0.875	1.475	0.901	0.652	0.988
	RF	0.983	0.857	9.594	3.090	2.281	3.497
	AB	0.995	0.869	2.723	1.476	1.172	2.123
	GB	0.997	0.874	1.628	1.102	0.823	1.310
	XG	0.991	0.871	4.860	1.909	1.382	2.183
	LB	0.991	0.858	4.875	2.049	1.523	2.390
	SVM	0.918	0.856	45.274	6.711	4.874	7.385
	ANN	0.935	0.855	36.276	5.973	4.538	6.997
	Ridge	0.887	0.851	62.515	7.903	5.941	9.040
	LASSO	0.887	0.851	62.689	7.914	5.950	9.050
S6	DT	0.847	0.521	63.834	7.896	5.826	7.701
	ET	0.984	0.729	6.604	2.439	1.780	2.368
	RF	0.959	0.711	17.057	4.116	2.974	3.950
	AB	0.983	0.714	6.959	2.393	1.958	2.888
	GB	0.954	0.731	19.294	4.244	3.282	4.469
	XG	0.948	0.731	21.685	4.533	3.501	4.755
	LB	0.956	0.713	18.441	3.987	3.023	4.106
	SVM	0.846	0.738	64.572	8.017	5.703	7.479
	ANN	0.868	0.731	55.181	7.401	5.673	7.526
	Ridge	0.778	0.707	92.805	9.631	7.393	9.895
	LASSO	0.777	0.704	93.139	9.648	7.387	9.878
S7	DT	0.865	0.633	54.511	7.334	5.473	7.099
	ET	0.978	0.811	8.811	2.497	1.753	2.266
	RF	0.964	0.787	14.703	3.799	2.760	3.604
	AB	0.988	0.802	4.812	2.046	1.675	2.298
	GB	0.975	0.802	10.330	3.084	2.147	2.810
	XG	0.973	0.803	10.967	3.204	2.208	2.823

(continued on next page)

Table 4 (*continued*)

Ship	Model	R^2	R^2 (test)	MSE	RMSE (ton/ day)	MAE (ton/ day)	MAPE (%)
	LB	0.981	0.775	7.624	2.614	1.804	2.364
	SVM	0.854	0.789	58.893	7.641	5.411	6.831
	ANN	0.879	0.778	48.929	6.903	5.287	6.805
	Ridge	0.809	0.771	77.312	8.789	6.635	8.431
	LASSO	0.808	0.769	77.734	8.813	6.664	8.483
S8	DT	0.908	0.738	55.429	7.369	5.275	6.362
	ET	0.998	0.860	1.223	0.864	0.549	0.687
	RF	0.975	0.835	15.054	3.848	2.645	3.229
	AB	0.995	0.847	3.047	1.544	1.182	1.516
	GB	0.988	0.839	7.367	2.097	1.438	1.781
	XG	0.973	0.843	16.064	3.747	2.646	3.231
	LB	0.973	0.827	16.500	3.523	2.521	3.138
	SVM	0.897	0.838	62.015	7.865	5.503	6.604
	ANN	0.911	0.824	53.888	7.282	5.345	6.458
	Ridge	0.867	0.817	80.108	8.945	6.728	8.344
	LASSO	0.867	0.818	80.433	8.963	6.737	8.348

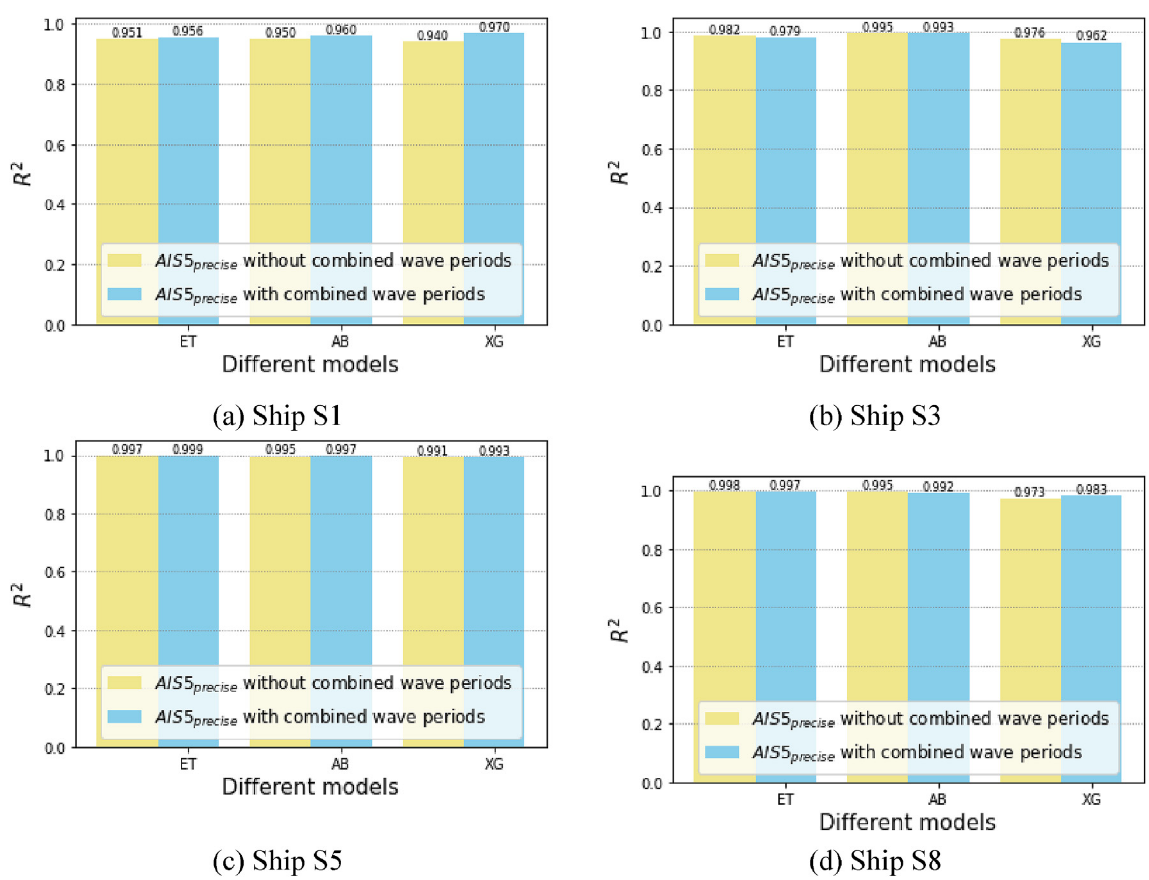
is interwoven with the decision of selecting good datasets. For instance, in Fig. 4, when the model AB is adopted, ***AIS5$_{precise}$*** demonstrates the quality of the best dataset. However, when ET or XG is adopted, ***Set1*** and ***Set3$_{precise}$*** have some chance to win.

4. Conclusions and discussion

This study, as the Part II of this series of studies, was motivated by a limitation of our previous study (Part I) that weather and sea condition information derived from the great circle sailing route (suggested by industry professionals) might be inaccurate. In this study, AIS data is further fused to voyage report data and meteorological data in that AIS data provides actual geographical positions of the ship which further help to retrieve more accurate weather and sea condition information from meteorological data.

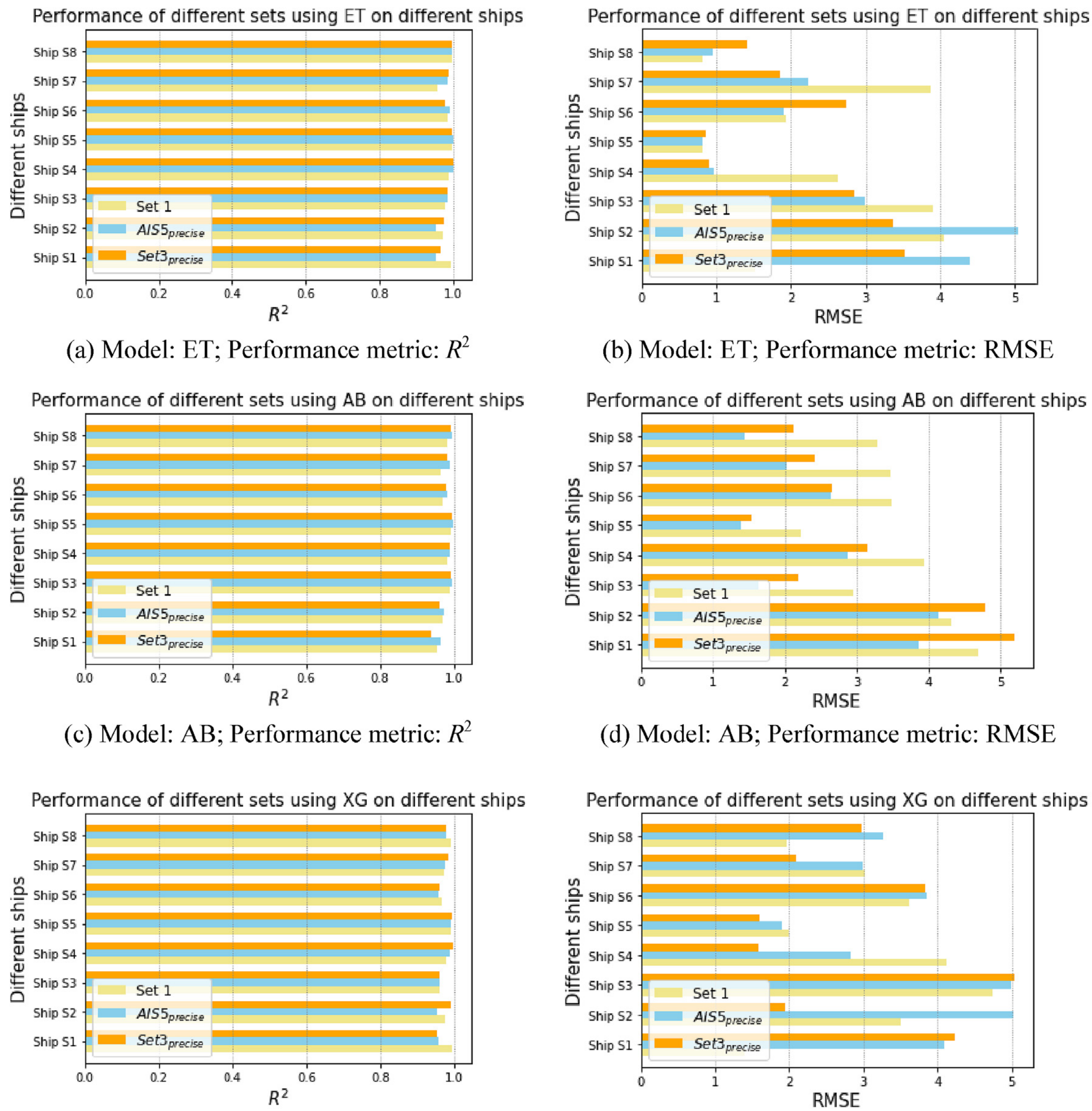
To summarize Part I (Li et al., 2022) and Part II (this study) of this series of studies, when dataset choice is considered, the original voyage report dataset ***Set1*** has a decent quality for ship fuel efficiency modeling; if more effort is paid to fuse voyage report data and meteorological data, data quality improves slightly and ***Set3$_{precise}$*** can be adopted. When AIS data is available, further including AIS data might also be beneficial, which suggests the adoption of the dataset ***AIS5$_{precise}$***. As far as ML model choice is concerned, we recommend the installation of four decision-tree based models including ET, AB, GB, and XG because they usually possess the highest fit performance and good generalization performance. Their performances are also quite robust against random splits of a dataset into training and test sets.

Overall, the best datasets found, including ***Set1***, ***Set3$_{precise}$***, and ***AIS5$_{precise}$***, ensure accurate fit performances of best ML models: R^2 on the training set is above 0.96 and even reaches 0.99 to 1.00, and R^2 on the test set is between 0.74 and 0.90; the fit errors measured by RMSE and MAE are between 0.5 and 4.5 ton/day. This accuracy is sufficient for many industry applications and energy-efficient operational measures for shipping companies, including sailing speed optimization, weather routing, and virtual arrivals. Therefore, apart from shipping companies, this research may also interest weather information service providers (WISPs) that are innovating weather routing, sailing speed optimization, and virtual (just-in-time) arrivals. International shipping associations that are pioneering in virtual arrival policy (BIMCO, 2021) may also find our studies useful, because a barrier to virtual arrival is the difficulty in quantifying the bunker fuel savings in different speed, draft, weather, and sea conditions (Merkel et al., 2022). Our studies also provide regulators such as IMO and EU with more quantitative evidence on how different data sources can be fully utilized for ship fuel efficiency analysis and what level of accuracy the state-of-the-art ML models can achieve to



(a) Ship S1 (b) Ship S3 (c) Ship S5 (d) Ship S8

Fig. 3. Fit performance of three best models (ET, AB, XG) over dataset *AIS5$_{precise}$*, with and without wave period information.



(a) Model: ET; Performance metric: R^2 (b) Model: ET; Performance metric: RMSE

(c) Model: AB; Performance metric: R^2 (d) Model: AB; Performance metric: RMSE

(e) Model: XG; Performance metric: R^2 (f) Model: XG; Performance metric: RMSE

Fig. 4. Fit performance (R^2 and RMSE) of three best models (ET, AB, XG) on three best datasets (*Set 1*, $Set3_{precise}$, $AIS5_{precise}$).

model a ship's bunker fuel consumption rate and the resultant emission rate.

Our industry collaborator, a global container shipping company, has been working with voyage report data as the main data source to clarify the bunker fuel efficiency issues they encountered. Though they do not have the research capacity of fusing different data sources and experimenting with state-of-the-art ML models, the suggestions of additionally using meteorological data and AIS data are all from them. Part I and Part II of this series of studies turned these basic ideas to solid research outcomes with intensive experiments with eight mega containerships. For the first time, we provide a clear answer to several questions a container shipping company may ask, including ‘does fusing voyage report data and other data sources improve the quality of data for ship fuel efficiency analysis?’, and ‘what level of accuracy can state-of-the-art ML models achieve over these fused datasets?’.

The reported fit and generalization performances of ET, AB, GB and XG are probably the highest level of accuracy we could achieve to model a mega containership's fuel consumption rate, if voyage report data is used as the main source of bunker fuel consumption. Section 3.1 discusses the main reasons why it is difficult, if not impossible, to further improve the modeling accuracy. These reasons boil down to the fact that voyage report data reports the “daily” bunker fuel consumption of a ship. Therefore, it is natural to further ask whether a data source with a finer granularity (sampling frequency) such as sensor data further improves the accuracy of ship fuel consumption rate modeling. This will be answered in a following study, referred to as “Part III” of this series of studies.

Replication and data sharing

Computer code in Python in this study is published in GitHub as a software infrastructure to reduce the exploration efforts of industry professionals. Best trained machine learning models are also published in GitHub, which enables maritime researchers to estimate the bunker fuel consumption rates of different sizes of mega containerships in different sailing speed, draft, trim and weather/sea conditions, though our raw data is confidential. The machine learning models published are completely black boxes, and one cannot conduct reverse engineering to access the original datasets. Readers can find the computer code and trained machine learning models in the URL: https://github.com/yuqua ndu/Data-driven-Ship-Fuel-Efficiency-Modeling.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is a part of the IAMU (International Association of Maritime Universities) research project titled "*Data fusion and machine learning for ship fuel efficiency analysis: a small but essential step towards green shipping through data analytics*" (Research Project No. 20210205_AMC). The materials and data in this publication have been obtained through the support of IAMU and The in Japan. Jean-Louis Bertholier developed the Python code of collecting meteorological data for ships during his Assistant Engineer internship at World Maritime University. This study has been conducted using E.U. Copernicus Marine Service Information; https://doi.org/10.48670/moi-00050. Hersbach et al. (2018) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results of this study and trained machine learning models published contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. We thank the two anonymous reviewers for their time spent on this series of three papers and their constructive comments.

Appendix A

Table A1
The fit performance of eleven machine learning models for ship S2.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	*Set1*	0.833	0.668	113.854	10.580	7.934	8.951
	$AIS2_{precise}$	0.848	0.647	95.131	9.497	7.026	8.169
	$AIS2_{fuzzy}$	0.850	0.635	94.066	9.585	7.155	8.275
	$AIS3_{precise}$	0.832	0.625	105.279	10.063	7.466	8.761
	$AIS3_{fuzzy}$	0.835	0.616	103.114	10.046	7.495	8.671
	$AIS4_{precise}$	0.854	0.634	91.071	9.389	6.915	8.058
	$AIS4_{fuzzy}$	0.832	0.638	104.706	10.121	7.549	8.763
	$AIS5_{precise}$	0.838	0.641	100.776	9.878	7.326	8.621
	$AIS5_{fuzzy}$	0.844	0.645	98.151	9.776	7.276	8.406
	$Set3_{precise}$ [a]	0.820	0.589	112.089	10.461	7.916	9.230
ET	*Set1*	0.971	0.786	19.857	4.055	2.986	3.306
	$AIS2_{precise}$	0.966	0.769	21.080	4.247	3.184	3.681
	$AIS2_{fuzzy}$	0.958	0.765	26.376	5.012	3.869	4.454
	$AIS3_{precise}$	0.973	0.775	16.741	3.697	2.653	3.048
	$AIS3_{fuzzy}$	0.968	0.769	19.826	4.047	2.961	3.398
	$AIS4_{precise}$	0.969	0.767	19.271	4.076	3.015	3.476
	$AIS4_{fuzzy}$	0.962	0.764	23.481	4.488	3.310	3.817
	$AIS5_{precise}$	0.979	0.776	13.486	3.239	2.390	2.760
	$AIS5_{fuzzy}$ [a]	0.952	0.766	29.638	5.037	3.783	4.339
	$Set3_{precise}$	0.974	0.765	15.842	3.377	2.445	2.780
RF	*Set1*	0.959	0.766	27.622	5.205	3.750	4.227
	$AIS2_{precise}$	0.947	0.753	32.873	5.683	4.092	4.765
	$AIS2_{fuzzy}$	0.955	0.751	27.856	5.237	3.790	4.396
	$AIS3_{precise}$	0.955	0.757	28.055	5.259	3.774	4.415
	$AIS3_{fuzzy}$	0.945	0.750	34.278	5.766	4.051	4.704
	$AIS4_{precise}$	0.951	0.753	30.772	5.477	3.879	4.491
	$AIS4_{fuzzy}$	0.946	0.748	33.751	5.721	4.078	4.720
	$AIS5_{precise}$	0.948	0.757	32.466	5.645	4.072	4.748
	$AIS5_{fuzzy}$	0.949	0.754	31.822	5.577	3.975	4.637
	$Set3_{precise}$ [a]	0.950	0.740	31.494	5.541	4.007	4.662
AB	*Set1*	0.968	0.762	21.779	4.305	3.609	4.143
	$AIS2_{precise}$	0.980	0.748	12.278	3.280	2.750	3.190
	$AIS2_{fuzzy}$	0.973	0.739	16.514	3.821	3.216	3.725
	$AIS3_{precise}$	0.980	0.749	12.642	3.279	2.739	3.205
	$AIS3_{fuzzy}$	0.977	0.746	14.184	3.472	2.922	3.409
	$AIS4_{precise}$	0.980	0.748	12.489	3.353	2.804	3.262
	$AIS4_{fuzzy}$	0.976	0.738	14.841	3.491	2.913	3.390
	$AIS5_{precise}$	0.975	0.754	15.834	3.830	3.227	3.765
	$AIS5_{fuzzy}$	0.971	0.754	18.197	4.128	3.549	4.162
	$Set3_{precise}$ [a]	0.961	0.743	24.755	4.778	4.073	4.729
GB	*Set1*	0.964	0.781	24.457	4.564	3.429	3.793
	$AIS2_{precise}$	0.980	0.764	12.651	3.166	2.496	2.842
	$AIS2_{fuzzy}$	0.964	0.749	22.561	4.354	3.304	3.719
	$AIS3_{precise}$	0.979	0.773	12.602	2.905	2.149	2.421
	$AIS3_{fuzzy}$	0.980	0.760	12.341	3.100	2.236	2.502
	$AIS4_{precise}$	0.988	0.772	7.589	2.274	1.677	1.884
	$AIS4_{fuzzy}$	0.973	0.763	16.682	3.628	2.707	3.057
	$AIS5_{precise}$	0.964	0.766	22.248	4.288	3.136	3.537
	$AIS5_{fuzzy}$	0.967	0.759	20.448	4.297	3.300	3.722
	$Set3_{precise}$ [a]	0.992	0.760	5.008	1.817	1.234	1.378

(*continued on next column*)

Table A1 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
XG	*Set1*	0.975	0.781	16.733	3.503	2.631	2.868
	$AIS2_{precise}$	0.959	0.757	25.467	4.457	3.004	3.294
	$AIS2_{fuzzy}$	0.962	0.754	23.570	4.567	3.274	3.615
	$AIS3_{precise}$	0.976	0.772	15.247	3.278	2.139	2.359
	$AIS3_{fuzzy}$	0.978	0.768	13.808	3.206	2.178	2.390
	$AIS4_{precise}$	0.966	0.765	20.767	4.157	2.743	3.008
	$AIS4_{fuzzy}$	0.960	0.762	24.749	4.576	3.229	3.544
	$AIS5_{precise}$	0.965	0.763	21.869	4.287	2.959	3.263
	$AIS5_{fuzzy}$	0.953	0.758	29.132	5.014	3.622	3.995
	$Set3_{precise}$ [a]	0.991	0.765	5.421	1.949	1.186	1.277
LB	*Set1*	0.946	0.761	36.850	5.784	4.429	4.834
	$AIS2_{precise}$	0.951	0.738	30.039	5.172	3.733	4.200
	$AIS2_{fuzzy}$	0.941	0.727	36.987	5.955	4.353	4.919
	$AIS3_{precise}$	0.972	0.745	17.363	3.757	2.709	3.079
	$AIS3_{fuzzy}$	0.961	0.730	24.200	4.627	3.333	3.789
	$AIS4_{precise}$	0.959	0.683	25.467	4.524	3.348	3.862
	$AIS4_{fuzzy}$	0.957	0.716	27.166	4.917	3.685	4.188
	$AIS5_{precise}$	0.959	0.733	25.714	4.671	3.276	3.798
	$AIS5_{fuzzy}$	0.941	0.733	36.457	5.684	4.108	4.702
	$Set3_{precise}$ [a]	0.980	0.748	12.589	3.053	2.179	2.442
SVM	*Set1*	0.848	0.797	103.306	10.147	7.260	7.779
	$AIS2_{precise}$	0.878	0.812	76.193	8.710	6.221	6.788
	$AIS2_{fuzzy}$	0.873	0.808	79.392	8.896	6.332	6.952
	$AIS3_{precise}$	0.875	0.816	78.296	8.829	6.359	6.937
	$AIS3_{fuzzy}$	0.871	0.810	80.452	8.950	6.470	7.131
	$AIS4_{precise}$	0.875	0.815	78.376	8.834	6.364	6.962
	$AIS4_{fuzzy}$	0.874	0.808	78.811	8.862	6.358	7.007
	$AIS5_{precise}$	0.880	0.809	74.547	8.599	6.191	6.738
	$AIS5_{fuzzy}$	0.869	0.804	81.630	9.012	6.541	7.180
	$Set3_{precise}$ [a]	0.864	0.812	84.860	9.176	6.608	7.210
ANN	*Set1*	0.876	0.787	84.367	9.093	6.935	7.682
	$AIS2_{precise}$	0.890	0.805	68.589	8.223	6.248	6.993
	$AIS2_{fuzzy}$	0.884	0.803	72.222	8.473	6.444	7.232
	$AIS3_{precise}$	0.893	0.815	66.528	8.111	6.172	6.926
	$AIS3_{fuzzy}$	0.891	0.811	67.646	8.181	6.182	6.952
	$AIS4_{precise}$	0.894	0.818	66.090	8.068	6.141	6.914
	$AIS4_{fuzzy}$	0.886	0.817	71.270	8.405	6.354	7.172
	$AIS5_{precise}$	0.895	0.797	65.299	8.036	6.139	6.847
	$AIS5_{fuzzy}$	0.887	0.802	70.576	8.362	6.365	7.100
	$Set3_{precise}$ [a]	0.908	0.791	56.693	7.365	5.581	6.171
Ridge	*Set1*	0.822	0.786	121.419	11.016	8.454	9.312
	$AIS2_{precise}$	0.829	0.810	107.128	10.345	7.775	8.763
	$AIS2_{fuzzy}$	0.822	0.803	110.983	10.529	7.876	8.873
	$AIS3_{precise}$	0.837	0.810	102.096	10.100	7.682	8.732
	$AIS3_{fuzzy}$	0.833	0.808	104.294	10.208	7.715	8.787
	$AIS4_{precise}$	0.836	0.813	102.598	10.125	7.695	8.759
	$AIS4_{fuzzy}$	0.833	0.809	104.538	10.220	7.723	8.798
	$AIS5_{precise}$	0.829	0.810	106.864	10.332	7.780	8.791
	$AIS5_{fuzzy}$	0.823	0.803	110.694	10.515	7.877	8.895
	$Set3_{precise}$ [a]	0.826	0.802	108.847	10.429	8.011	9.055
LASSO	*Set1*	0.822	0.785	121.508	11.020	8.471	9.331
	$AIS2_{precise}$	0.829	0.810	107.127	10.344	7.774	8.748
	$AIS2_{fuzzy}$	0.822	0.802	111.266	10.542	7.882	8.861
	$AIS3_{precise}$	0.836	0.809	102.493	10.119	7.676	8.700
	$AIS3_{fuzzy}$	0.831	0.809	105.332	10.259	7.719	8.779
	$AIS4_{precise}$	0.836	0.812	102.731	10.131	7.685	8.720
	$AIS4_{fuzzy}$	0.832	0.809	105.148	10.250	7.714	8.765
	$AIS5_{precise}$	0.829	0.810	106.939	10.336	7.779	8.774
	$AIS5_{fuzzy}$	0.823	0.803	110.874	10.524	7.884	8.887
	$Set3_{precise}$ [a]	0.824	0.796	110.162	10.492	8.034	9.042

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

Table A2
The fit performance of eleven machine learning models for ship S3.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	*Set1*	0.857	0.684	105.672	10.125	7.259	8.643
	$AIS2_{precise}$	0.854	0.699	106.123	10.087	7.292	8.759
	$AIS2_{fuzzy}$	0.850	0.695	108.384	10.209	7.427	8.876
	$AIS3_{precise}$	0.871	0.706	93.959	9.334	6.759	8.074
	$AIS3_{fuzzy}$	0.874	0.681	90.614	9.273	6.729	7.972
	$AIS4_{precise}$	0.867	0.692	97.233	9.628	6.941	8.311
	$AIS4_{fuzzy}$	0.869	0.687	94.625	9.621	7.005	8.343

(continued on next column)

Table A2 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS5_{precise}$	0.867	0.684	97.066	9.534	6.948	8.222
	$AIS5_{fuzzy}$	0.853	0.704	106.927	10.083	7.267	8.664
	$Set3_{precise}$ [a]	0.865	0.684	98.572	9.705	7.042	8.343
ET	*Set1*	0.977	0.800	17.021	3.911	2.462	2.964
	$AIS2_{precise}$	0.986	0.822	10.267	2.769	1.681	2.113
	$AIS2_{fuzzy}$	0.977	0.814	16.566	3.552	2.175	2.756
	$AIS3_{precise}$	0.982	0.818	12.791	3.030	1.702	2.189
	$AIS3_{fuzzy}$	0.988	0.813	8.588	2.438	1.478	1.867
	$AIS4_{precise}$	0.981	0.816	13.605	3.179	1.861	2.371
	$AIS4_{fuzzy}$	0.981	0.811	13.638	3.326	2.026	2.557
	$AIS5_{precise}$	0.982	0.828	13.029	3.188	1.858	2.364
	$AIS5_{fuzzy}$	0.983	0.826	12.304	2.986	1.751	2.218
	$Set3_{precise}$ [a]	0.985	0.821	10.758	2.846	1.716	2.181
RF	*Set1*	0.960	0.768	29.573	5.369	3.497	4.234
	$AIS2_{precise}$	0.963	0.807	26.977	5.153	3.378	4.167
	$AIS2_{fuzzy}$	0.959	0.803	29.977	5.426	3.478	4.315
	$AIS3_{precise}$	0.951	0.803	35.971	5.896	3.790	4.743
	$AIS3_{fuzzy}$	0.948	0.807	37.592	6.052	3.787	4.741
	$AIS4_{precise}$	0.959	0.797	29.386	5.381	3.477	4.317
	$AIS4_{fuzzy}$	0.954	0.802	33.398	5.680	3.620	4.511
	$AIS5_{precise}$	0.963	0.809	27.121	5.168	3.376	4.170
	$AIS5_{fuzzy}$	0.960	0.805	29.142	5.343	3.415	4.241
	$Set3_{precise}$ [a]	0.956	0.802	31.781	5.576	3.587	4.463
AB	*Set1*	0.988	0.798	9.177	2.942	2.371	2.718
	$AIS2_{precise}$	0.990	0.815	7.319	2.589	2.060	2.335
	$AIS2_{fuzzy}$	0.989	0.805	7.569	2.576	1.975	2.218
	$AIS3_{precise}$	0.995	0.813	3.799	1.682	1.270	1.480
	$AIS3_{fuzzy}$	0.994	0.803	4.303	1.839	1.356	1.568
	$AIS4_{precise}$	0.992	0.810	5.579	2.186	1.699	1.933
	$AIS4_{fuzzy}$	0.989	0.801	7.743	2.614	2.101	2.367
	$AIS5_{precise}$	0.995	0.820	3.588	1.728	1.292	1.513
	$AIS5_{fuzzy}$	0.995	0.802	3.489	1.623	1.154	1.330
	$Set3_{precise}$ [a]	0.991	0.812	6.328	2.183	1.712	1.998
GB	*Set1*	0.962	0.776	28.220	4.726	3.221	3.841
	$AIS2_{precise}$	0.966	0.815	25.195	4.467	2.872	3.617
	$AIS2_{fuzzy}$	0.962	0.798	27.998	4.806	3.066	3.870
	$AIS3_{precise}$	0.974	0.817	18.985	3.721	2.311	2.932
	$AIS3_{fuzzy}$	0.968	0.810	23.467	4.597	2.809	3.582
	$AIS4_{precise}$	0.960	0.813	29.026	4.863	3.092	3.881
	$AIS4_{fuzzy}$	0.966	0.811	25.102	4.373	2.668	3.431
	$AIS5_{precise}$	0.969	0.818	22.552	4.221	2.710	3.366
	$AIS5_{fuzzy}$	0.971	0.817	21.037	4.028	2.586	3.238
	$Set3_{precise}$ [a]	0.964	0.819	26.559	4.694	2.836	3.642
XG	*Set1*	0.959	0.778	30.013	4.738	3.214	3.744
	$AIS2_{precise}$	0.961	0.814	28.566	4.798	2.995	3.753
	$AIS2_{fuzzy}$	0.942	0.807	42.158	5.988	3.915	4.873
	$AIS3_{precise}$	0.964	0.809	26.403	4.516	2.727	3.416
	$AIS3_{fuzzy}$	0.969	0.806	22.552	4.331	2.511	3.237
	$AIS4_{precise}$	0.954	0.809	33.595	5.384	3.366	4.179
	$AIS4_{fuzzy}$	0.961	0.806	28.229	5.078	3.093	3.892
	$AIS5_{precise}$	0.976	0.819	17.745	3.884	2.439	2.960
	$AIS5_{fuzzy}$	0.959	0.808	30.356	4.995	3.159	3.921
	$Set3_{precise}$ [a]	0.961	0.810	28.714	5.030	3.052	3.828
LB	*Set1*	0.935	0.766	48.608	6.560	4.506	5.448
	$AIS2_{precise}$	0.945	0.802	40.487	5.912	3.868	4.848
	$AIS2_{fuzzy}$	0.925	0.802	53.776	7.068	4.742	5.928
	$AIS3_{precise}$	0.954	0.802	33.337	5.389	3.474	4.419
	$AIS3_{fuzzy}$	0.952	0.795	34.420	5.268	3.534	4.423
	$AIS4_{precise}$	0.938	0.801	44.899	6.546	4.381	5.462
	$AIS4_{fuzzy}$	0.947	0.784	38.969	5.866	3.946	4.969
	$AIS5_{precise}$	0.952	0.804	35.323	5.432	3.468	4.412
	$AIS5_{fuzzy}$	0.953	0.796	34.216	5.421	3.563	4.487
	$Set3_{precise}$ [a]	0.947	0.804	38.795	5.845	3.853	4.853
SVM	*Set1*	0.812	0.791	138.669	11.753	7.557	8.957
	$AIS2_{precise}$	0.842	0.825	114.633	10.675	6.588	8.092
	$AIS2_{fuzzy}$	0.829	0.817	123.943	11.116	6.904	8.485
	$AIS3_{precise}$	0.847	0.823	110.536	10.488	6.537	8.005
	$AIS3_{fuzzy}$	0.842	0.816	114.255	10.658	6.761	8.329
	$AIS4_{precise}$	0.845	0.828	112.062	10.563	6.588	8.077
	$AIS4_{fuzzy}$	0.842	0.822	114.254	10.668	6.734	8.244
	$AIS5_{precise}$	0.840	0.825	115.821	10.727	6.746	8.262
	$AIS5_{fuzzy}$	0.840	0.818	115.603	10.720	6.657	8.187
	$Set3_{precise}$ [a]	0.844	0.820	113.000	10.591	6.627	8.167
ANN	*Set1*	0.829	0.780	126.769	11.217	7.780	9.353
	$AIS2_{precise}$	0.849	0.812	109.169	10.413	7.022	8.680
	$AIS2_{fuzzy}$	0.849	0.813	109.194	10.415	6.917	8.561

(*continued on next column*)

Table A2 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS3_{precise}$	0.869	0.800	94.542	9.681	6.645	8.156
	$AIS3_{fuzzy}$	0.847	0.799	111.263	10.484	7.230	8.936
	$AIS4_{precise}$	0.860	0.799	101.451	10.008	6.840	8.418
	$AIS4_{fuzzy}$	0.847	0.800	110.709	10.470	7.192	8.881
	$AIS5_{precise}$	0.860	0.813	101.322	10.029	6.763	8.262
	$AIS5_{fuzzy}$	0.854	0.810	106.288	10.265	6.849	8.413
	$Set3_{precise}$ [a]	0.874	0.798	91.583	9.475	6.480	7.992
Ridge	$Set1$	0.780	0.778	162.676	12.739	9.007	11.114
	$AIS2_{precise}$	0.793	0.801	149.861	12.227	8.517	10.890
	$AIS2_{fuzzy}$	0.790	0.799	152.029	12.316	8.509	10.860
	$AIS3_{precise}$	0.801	0.797	143.924	11.981	8.319	10.627
	$AIS3_{fuzzy}$	0.802	0.798	143.536	11.966	8.330	10.651
	$AIS4_{precise}$	0.798	0.799	145.868	12.062	8.348	10.695
	$AIS4_{fuzzy}$	0.799	0.800	145.189	12.035	8.339	10.671
	$AIS5_{precise}$	0.796	0.804	147.564	12.133	8.425	10.745
	$AIS5_{fuzzy}$	0.793	0.802	149.516	12.214	8.459	10.779
	$Set3_{precise}$ [a]	0.801	0.796	144.061	11.987	8.329	10.615
LASSO	$Set1$	0.779	0.778	163.445	12.769	9.011	11.128
	$AIS2_{precise}$	0.793	0.800	149.869	12.227	8.508	10.879
	$AIS2_{fuzzy}$	0.790	0.799	152.048	12.317	8.502	10.852
	$AIS3_{precise}$	0.800	0.798	145.050	12.028	8.317	10.627
	$AIS3_{fuzzy}$	0.800	0.798	144.547	12.007	8.322	10.641
	$AIS4_{precise}$	0.798	0.799	146.172	12.075	8.346	10.694
	$AIS4_{fuzzy}$	0.799	0.799	145.246	12.037	8.329	10.657
	$AIS5_{precise}$	0.796	0.804	147.612	12.135	8.426	10.746
	$AIS5_{fuzzy}$	0.793	0.802	149.539	12.215	8.450	10.765
	$Set3_{precise}$ [a]	0.799	0.796	145.425	12.043	8.323	10.619

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

Table A3
The fit performance of eleven machine learning models for ship S4.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	$Set1$	0.906	0.758	81.312	8.851	6.363	6.681
	$AIS2_{precise}$	0.916	0.736	69.034	8.033	5.999	6.463
	$AIS2_{fuzzy}$	0.910	0.730	73.824	8.275	6.130	6.586
	$AIS3_{precise}$	0.916	0.741	68.065	8.086	6.006	6.432
	$AIS3_{fuzzy}$	0.926	0.725	60.675	7.518	5.564	5.919
	$AIS4_{precise}$	0.921	0.745	64.090	7.807	5.747	6.160
	$AIS4_{fuzzy}$	0.899	0.741	82.735	8.913	6.661	7.152
	$AIS5_{precise}$	0.904	0.749	78.681	8.637	6.425	6.897
	$AIS5_{fuzzy}$	0.912	0.758	72.400	8.366	6.185	6.648
	$Set3_{precise}$ [a]	0.916	0.746	68.063	8.094	6.036	6.523
ET	$Set1$	0.988	0.858	10.120	2.625	1.778	1.862
	$AIS2_{precise}$	0.999	0.867	0.705	0.629	0.394	0.422
	$AIS2_{fuzzy}$	0.999	0.863	0.811	0.714	0.462	0.495
	$AIS3_{precise}$	0.997	0.869	2.131	1.151	0.807	0.876
	$AIS3_{fuzzy}$	0.997	0.864	2.431	1.115	0.779	0.846
	$AIS4_{precise}$	0.997	0.870	2.521	1.191	0.845	0.915
	$AIS4_{fuzzy}$	0.998	0.864	1.826	1.115	0.768	0.830
	$AIS5_{precise}$	0.998	0.871	1.642	0.927	0.651	0.696
	$AIS5_{fuzzy}$	0.998	0.866	1.390	0.973	0.659	0.716
	$Set3_{precise}$ [a]	0.998	0.872	1.434	0.901	0.627	0.687
RF	$Set1$	0.974	0.848	22.794	4.752	3.335	3.501
	$AIS2_{precise}$	0.976	0.856	19.824	4.444	3.279	3.600
	$AIS2_{fuzzy}$	0.976	0.854	19.766	4.429	3.273	3.607
	$AIS3_{precise}$	0.972	0.854	22.477	4.723	3.464	3.773
	$AIS3_{fuzzy}$	0.973	0.855	21.949	4.659	3.424	3.751
	$AIS4_{precise}$	0.973	0.855	21.804	4.644	3.421	3.742
	$AIS4_{fuzzy}$	0.974	0.858	21.363	4.599	3.381	3.708
	$AIS5_{precise}$	0.975	0.859	20.097	4.472	3.292	3.590
	$AIS5_{fuzzy}$	0.975	0.858	20.309	4.491	3.298	3.602
	$Set3_{precise}$ [a]	0.975	0.853	20.349	4.497	3.331	3.618
AB	$Set1$	0.980	0.843	17.332	3.939	3.283	3.654
	$AIS2_{precise}$	0.987	0.865	10.568	3.102	2.595	2.916
	$AIS2_{fuzzy}$	0.985	0.862	11.992	3.358	2.790	3.103
	$AIS3_{precise}$	0.990	0.867	8.454	2.664	2.206	2.484
	$AIS3_{fuzzy}$	0.989	0.867	8.670	2.766	2.285	2.560
	$AIS4_{precise}$	0.988	0.864	9.894	2.953	2.477	2.782
	$AIS4_{fuzzy}$	0.987	0.864	10.576	3.117	2.607	2.916
	$AIS5_{precise}$	0.988	0.869	9.701	2.956	2.466	2.774
	$AIS5_{fuzzy}$	0.988	0.868	9.410	2.864	2.401	2.674
	$Set3_{precise}$ [a]	0.986	0.865	11.021	3.144	2.591	2.905

(*continued on next column*)

Table A3 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
GB	$Set1$	0.977	0.851	19.591	4.196	3.176	3.352
	$AIS2_{precise}$	0.989	0.868	9.077	2.587	1.960	2.099
	$AIS2_{fuzzy}$	0.988	0.868	9.667	2.653	2.046	2.189
	$AIS3_{precise}$	0.995	0.868	4.407	1.836	1.343	1.452
	$AIS3_{fuzzy}$	0.991	0.871	7.288	2.277	1.677	1.808
	$AIS4_{precise}$	0.993	0.866	5.462	1.884	1.438	1.531
	$AIS4_{fuzzy}$	0.991	0.863	6.981	2.207	1.677	1.803
	$AIS5_{precise}$	0.991	0.873	7.631	2.412	1.810	1.933
	$AIS5_{fuzzy}$	0.990	0.875	8.437	2.566	1.943	2.087
	$Set3_{precise}$ [a]	0.989	0.866	8.845	2.500	1.838	1.957
XG	$Set1$	0.977	0.858	19.657	4.126	3.068	3.185
	$AIS2_{precise}$	0.991	0.864	7.508	2.341	1.721	1.808
	$AIS2_{fuzzy}$	0.992	0.870	6.889	2.321	1.715	1.816
	$AIS3_{precise}$	0.994	0.868	4.689	1.591	1.126	1.196
	$AIS3_{fuzzy}$	0.995	0.873	4.260	1.605	1.117	1.188
	$AIS4_{precise}$	0.993	0.861	5.846	1.821	1.344	1.426
	$AIS4_{fuzzy}$	0.989	0.865	8.965	2.311	1.733	1.832
	$AIS5_{precise}$	0.992	0.870	6.441	2.074	1.548	1.640
	$AIS5_{fuzzy}$	0.987	0.878	11.059	2.835	2.157	2.299
	$Set3_{precise}$ [a]	0.995	0.869	3.758	1.585	1.140	1.201
LB	$Set1$	0.968	0.844	28.153	5.010	3.861	4.044
	$AIS2_{precise}$	0.979	0.851	17.028	3.711	2.865	3.064
	$AIS2_{fuzzy}$	0.988	0.854	9.860	2.778	2.152	2.313
	$AIS3_{precise}$	0.989	0.866	9.039	2.625	2.009	2.161
	$AIS3_{fuzzy}$	0.986	0.865	11.633	3.097	2.388	2.564
	$AIS4_{precise}$	0.978	0.855	18.105	3.969	3.112	3.336
	$AIS4_{fuzzy}$	0.981	0.846	15.360	3.672	2.862	3.083
	$AIS5_{precise}$	0.992	0.865	6.859	2.321	1.771	1.925
	$AIS5_{fuzzy}$	0.987	0.873	10.530	2.901	2.250	2.440
	$Set3_{precise}$ [a]	0.987	0.855	10.943	2.871	2.200	2.340
SVM	$Set1$	0.906	0.842	81.874	9.015	6.318	6.374
	$AIS2_{precise}$	0.930	0.848	56.850	7.333	5.462	5.814
	$AIS2_{fuzzy}$	0.936	0.845	51.762	6.893	5.122	5.485
	$AIS3_{precise}$	0.929	0.846	57.822	7.548	5.488	5.811
	$AIS3_{fuzzy}$	0.917	0.842	68.045	8.187	5.956	6.291
	$AIS4_{precise}$	0.926	0.850	60.077	7.704	5.660	5.977
	$AIS4_{fuzzy}$	0.917	0.846	67.802	8.170	5.974	6.346
	$AIS5_{precise}$	0.927	0.860	59.875	7.686	5.642	6.018
	$AIS5_{fuzzy}$	0.919	0.854	66.517	8.094	5.967	6.387
	$Set3_{precise}$ [a]	0.921	0.857	63.718	7.972	5.848	6.146
ANN	$Set1$	0.925	0.845	65.521	8.076	6.102	6.390
	$AIS2_{precise}$	0.936	0.862	52.288	7.217	5.653	6.104
	$AIS2_{fuzzy}$	0.936	0.842	52.055	7.184	5.607	6.059
	$AIS3_{precise}$	0.944	0.859	46.191	6.779	5.331	5.782
	$AIS3_{fuzzy}$	0.941	0.848	48.047	6.911	5.366	5.842
	$AIS4_{precise}$	0.943	0.856	46.674	6.812	5.354	5.803
	$AIS4_{fuzzy}$	0.943	0.852	46.629	6.809	5.277	5.720
	$AIS5_{precise}$	0.939	0.868	50.219	7.062	5.524	5.962
	$AIS5_{fuzzy}$	0.930	0.858	57.295	7.541	5.889	6.371
	$Set3_{precise}$ [a]	0.947	0.856	42.555	6.513	5.034	5.502
Ridge	$Set1$	0.825	0.821	152.631	12.351	9.343	9.548
	$AIS2_{precise}$	0.822	0.802	145.669	12.066	9.341	9.625
	$AIS2_{fuzzy}$	0.816	0.796	150.467	12.263	9.500	9.837
	$AIS3_{precise}$	0.832	0.807	137.392	11.717	9.182	9.497
	$AIS3_{fuzzy}$	0.826	0.802	141.931	11.909	9.280	9.647
	$AIS4_{precise}$	0.831	0.806	138.025	11.744	9.187	9.487
	$AIS4_{fuzzy}$	0.826	0.802	142.348	11.926	9.279	9.628
	$AIS5_{precise}$	0.827	0.808	141.409	11.888	9.244	9.534
	$AIS5_{fuzzy}$	0.822	0.803	145.201	12.046	9.355	9.690
	$Set3_{precise}$ [a]	0.833	0.811	135.334	11.629	9.033	9.406
LASSO	$Set1$	0.824	0.823	153.402	12.382	9.347	9.537
	$AIS2_{precise}$	0.822	0.803	145.771	12.070	9.340	9.620
	$AIS2_{fuzzy}$	0.815	0.797	150.852	12.279	9.508	9.835
	$AIS3_{precise}$	0.831	0.806	137.956	11.741	9.191	9.499
	$AIS3_{fuzzy}$	0.824	0.804	143.555	11.977	9.303	9.641
	$AIS4_{precise}$	0.831	0.806	138.364	11.759	9.193	9.489
	$AIS4_{fuzzy}$	0.825	0.802	143.263	11.964	9.292	9.626
	$AIS5_{precise}$	0.827	0.808	141.586	11.895	9.244	9.530
	$AIS5_{fuzzy}$	0.822	0.805	145.687	12.067	9.364	9.692
	$Set3_{precise}$ [a]	0.832	0.809	135.961	11.656	9.053	9.417

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

Table A4
The fit performance of eleven machine learning models for ship S5.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	*Set1*	0.939	0.821	33.699	5.588	4.144	6.259
	$AIS2_{precise}$	0.947	0.795	29.259	5.181	3.809	5.761
	$AIS2_{fuzzy}$	0.946	0.799	30.200	5.379	3.954	5.947
	$AIS3_{precise}$	0.942	0.793	31.964	5.502	4.072	6.137
	$AIS3_{fuzzy}$	0.941	0.812	32.694	5.643	4.181	6.328
	$AIS4_{precise}$	0.940	0.793	33.054	5.628	4.130	6.221
	$AIS4_{fuzzy}$	0.941	0.800	32.814	5.558	4.121	6.247
	$AIS5_{precise}$	0.948	0.800	28.458	5.104	3.741	5.634
	$AIS5_{fuzzy}$	0.945	0.826	30.556	5.217	3.848	5.813
	$Set3_{precise}$ [a]	0.947	0.785	29.488	5.182	3.764	5.625
ET	*Set1*	0.998	0.895	1.057	0.805	0.569	0.857
	$AIS2_{precise}$	0.999	0.896	0.493	0.529	0.371	0.559
	$AIS2_{fuzzy}$	0.998	0.893	1.165	0.835	0.604	0.914
	$AIS3_{precise}$	0.998	0.895	0.890	0.707	0.495	0.748
	$AIS3_{fuzzy}$	0.997	0.893	1.963	1.134	0.815	1.234
	$AIS4_{precise}$	0.998	0.895	1.105	0.768	0.550	0.842
	$AIS4_{fuzzy}$	0.998	0.893	1.248	0.820	0.579	0.879
	$AIS5_{precise}$	0.997	0.894	1.475	0.901	0.652	0.988
	$AIS5_{fuzzy}$	0.998	0.891	1.055	0.808	0.584	0.883
	$Set3_{precise}$ [a]	0.997	0.892	1.413	0.854	0.619	0.935
RF	*Set1*	0.982	0.884	9.951	3.140	2.354	3.594
	$AIS2_{precise}$	0.982	0.878	9.753	3.116	2.290	3.509
	$AIS2_{fuzzy}$	0.983	0.883	9.496	3.072	2.249	3.461
	$AIS3_{precise}$	0.982	0.879	10.152	3.168	2.328	3.562
	$AIS3_{fuzzy}$	0.983	0.887	9.681	3.103	2.298	3.538
	$AIS4_{precise}$	0.981	0.875	10.565	3.238	2.356	3.594
	$AIS4_{fuzzy}$	0.983	0.885	9.200	3.026	2.224	3.413
	$AIS5_{precise}$	0.983	0.879	9.594	3.090	2.281	3.497
	$AIS5_{fuzzy}$	0.984	0.887	9.003	2.996	2.238	3.464
	$Set3_{precise}$ [a]	0.981	0.874	10.498	3.225	2.390	3.663
AB	*Set1*	0.990	0.895	5.408	2.213	1.830	3.156
	$AIS2_{precise}$	0.994	0.886	3.187	1.671	1.336	2.365
	$AIS2_{fuzzy}$	0.995	0.893	2.917	1.577	1.231	2.199
	$AIS3_{precise}$	0.996	0.892	2.337	1.372	1.073	1.942
	$AIS3_{fuzzy}$	0.997	0.897	1.684	1.136	0.841	1.554
	$AIS4_{precise}$	0.995	0.888	2.854	1.584	1.262	2.249
	$AIS4_{fuzzy}$	0.995	0.894	2.937	1.618	1.277	2.286
	$AIS5_{precise}$	0.995	0.889	2.723	1.476	1.172	2.123
	$AIS5_{fuzzy}$	0.996	0.895	2.299	1.392	1.081	1.951
	$Set3_{precise}$ [a]	0.995	0.886	2.543	1.525	1.209	2.217
GB	*Set1*	0.993	0.895	3.885	1.743	1.360	2.158
	$AIS2_{precise}$	0.996	0.892	2.273	1.265	0.946	1.496
	$AIS2_{fuzzy}$	0.997	0.893	1.801	1.015	0.782	1.257
	$AIS3_{precise}$	0.995	0.890	2.674	1.204	0.879	1.397
	$AIS3_{fuzzy}$	0.996	0.894	2.307	1.143	0.839	1.335
	$AIS4_{precise}$	0.997	0.888	1.936	1.010	0.720	1.132
	$AIS4_{fuzzy}$	0.994	0.893	3.367	1.426	1.047	1.656
	$AIS5_{precise}$	0.997	0.893	1.628	1.102	0.823	1.310
	$AIS5_{fuzzy}$	0.993	0.892	3.733	1.665	1.282	2.046
	$Set3_{precise}$ [a]	0.993	0.887	3.519	1.359	1.021	1.610
XG	*Set1*	0.990	0.892	5.361	1.995	1.520	2.370
	$AIS2_{precise}$	0.997	0.884	1.919	1.190	0.830	1.304
	$AIS2_{fuzzy}$	0.993	0.886	3.701	1.539	1.118	1.750
	$AIS3_{precise}$	0.996	0.885	2.079	1.206	0.857	1.347
	$AIS3_{fuzzy}$	0.991	0.893	4.859	1.785	1.314	2.077
	$AIS4_{precise}$	0.990	0.884	5.595	2.020	1.420	2.218
	$AIS4_{fuzzy}$	0.992	0.888	4.440	1.860	1.368	2.169
	$AIS5_{precise}$	0.991	0.891	4.860	1.909	1.382	2.183
	$AIS5_{fuzzy}$	0.992	0.887	4.417	1.897	1.403	2.206
	$Set3_{precise}$ [a]	0.993	0.878	3.601	1.605	1.133	1.749
LB	*Set1*	0.986	0.879	7.810	2.636	2.028	3.173
	$AIS2_{precise}$	0.987	0.879	7.311	2.490	1.883	2.957
	$AIS2_{fuzzy}$	0.984	0.869	9.040	2.831	2.115	3.308
	$AIS3_{precise}$	0.991	0.882	5.103	2.066	1.528	2.380
	$AIS3_{fuzzy}$	0.985	0.884	8.471	2.658	1.990	3.143
	$AIS4_{precise}$	0.987	0.846	7.256	2.420	1.848	2.954
	$AIS4_{fuzzy}$	0.984	0.878	9.032	2.837	2.146	3.365
	$AIS5_{precise}$	0.991	0.880	4.875	2.049	1.523	2.390
	$AIS5_{fuzzy}$	0.983	0.879	9.198	2.795	2.104	3.278
	$Set3_{precise}$ [a]	0.987	0.873	7.382	2.350	1.758	2.725
SVM	*Set1*	0.931	0.884	38.408	6.173	4.382	6.630
	$AIS2_{precise}$	0.916	0.881	46.567	6.810	4.916	7.414
	$AIS2_{fuzzy}$	0.912	0.886	48.618	6.946	4.984	7.542
	$AIS3_{precise}$	0.912	0.880	48.897	6.986	5.055	7.619

(*continued on next column*)

Table A4 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS3_{fuzzy}$	0.913	0.882	48.270	6.942	5.017	7.597
	$AIS4_{precise}$	0.913	0.878	47.993	6.918	5.019	7.549
	$AIS4_{fuzzy}$	0.915	0.881	47.104	6.851	4.980	7.544
	$AIS5_{precise}$	0.918	0.878	45.274	6.711	4.874	7.385
	$AIS5_{fuzzy}$	0.920	0.880	44.342	6.608	4.801	7.317
	$Set3_{precise}$ [a]	0.916	0.873	46.421	6.785	4.917	7.472
ANN	*Set1*	0.926	0.886	40.737	6.373	4.900	7.545
	$AIS2_{precise}$	0.938	0.876	34.278	5.776	4.398	6.845
	$AIS2_{fuzzy}$	0.943	0.885	31.933	5.592	4.281	6.685
	$AIS3_{precise}$	0.942	0.881	32.259	5.619	4.257	6.634
	$AIS3_{fuzzy}$	0.935	0.881	35.988	5.948	4.524	7.035
	$AIS4_{precise}$	0.940	0.880	33.504	5.733	4.354	6.741
	$AIS4_{fuzzy}$	0.938	0.878	34.189	5.806	4.435	6.882
	$AIS5_{precise}$	0.935	0.878	36.276	5.973	4.538	6.997
	$AIS5_{fuzzy}$	0.939	0.881	33.836	5.787	4.393	6.840
	$Set3_{precise}$ [a]	0.935	0.879	36.157	5.956	4.544	7.075
Ridge	*Set1*	0.875	0.868	69.368	8.325	6.341	9.937
	$AIS2_{precise}$	0.886	0.875	63.247	7.949	5.972	9.145
	$AIS2_{fuzzy}$	0.884	0.874	64.133	8.004	5.987	9.116
	$AIS3_{precise}$	0.892	0.868	60.011	7.743	5.826	8.883
	$AIS3_{fuzzy}$	0.892	0.871	59.890	7.735	5.845	8.893
	$AIS4_{precise}$	0.890	0.872	60.879	7.799	5.867	8.914
	$AIS4_{fuzzy}$	0.890	0.873	60.924	7.802	5.897	8.940
	$AIS5_{precise}$	0.887	0.874	62.515	7.903	5.941	9.040
	$AIS5_{fuzzy}$	0.887	0.875	62.587	7.907	5.952	9.027
	$Set3_{precise}$ [a]	0.889	0.868	61.610	7.846	5.934	9.109
LASSO	*Set1*	0.874	0.868	69.799	8.351	6.357	9.948
	$AIS2_{precise}$	0.886	0.875	63.278	7.951	5.976	9.156
	$AIS2_{fuzzy}$	0.884	0.874	64.249	8.012	5.995	9.128
	$AIS3_{precise}$	0.891	0.869	60.385	7.766	5.839	8.900
	$AIS3_{fuzzy}$	0.891	0.870	60.464	7.771	5.876	8.925
	$AIS4_{precise}$	0.889	0.869	61.278	7.824	5.879	8.934
	$AIS4_{fuzzy}$	0.890	0.873	61.092	7.813	5.909	8.957
	$AIS5_{precise}$	0.887	0.874	62.689	7.914	5.950	9.050
	$AIS5_{fuzzy}$	0.887	0.874	62.829	7.923	5.967	9.032
	$Set3_{precise}$ [a]	0.888	0.868	61.988	7.870	5.953	9.129

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

Table A5
The fit performance of eleven machine learning models for ship S6.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	*Set1*	0.837	0.636	67.292	8.143	5.917	7.777
	$AIS2_{precise}$	0.823	0.572	73.999	8.507	6.292	8.362
	$AIS2_{fuzzy}$	0.839	0.584	67.435	8.118	5.944	7.841
	$AIS3_{precise}$	0.847	0.568	63.921	7.913	5.799	7.668
	$AIS3_{fuzzy}$	0.841	0.581	66.494	8.095	5.969	7.923
	$AIS4_{precise}$	0.855	0.571	60.340	7.645	5.580	7.407
	$AIS4_{fuzzy}$	0.833	0.594	69.924	8.261	6.086	8.043
	$AIS5_{precise}$	0.847	0.581	63.834	7.896	5.826	7.701
	$AIS5_{fuzzy}$	0.832	0.591	70.398	8.318	6.145	8.148
	$Set3_{precise}$ [a]	0.832	0.576	69.684	8.275	6.119	8.113
ET	*Set1*	0.985	0.765	6.050	1.928	1.359	1.796
	$AIS2_{precise}$	0.991	0.767	3.917	1.699	1.229	1.633
	$AIS2_{fuzzy}$	0.992	0.765	3.275	1.519	1.124	1.491
	$AIS3_{precise}$	0.988	0.764	4.917	1.561	1.136	1.516
	$AIS3_{fuzzy}$	0.986	0.765	5.657	1.944	1.414	1.886
	$AIS4_{precise}$	0.988	0.762	4.791	1.801	1.318	1.760
	$AIS4_{fuzzy}$	0.987	0.763	5.498	1.871	1.381	1.838
	$AIS5_{precise}$	0.984	0.763	6.604	2.439	1.780	2.368
	$AIS5_{fuzzy}$	0.989	0.759	4.566	1.904	1.413	1.874
	$Set3_{precise}$ [a]	0.979	0.752	8.706	2.743	2.010	2.678
RF	*Set1*	0.956	0.766	18.155	4.225	3.016	4.012
	$AIS2_{precise}$	0.956	0.747	18.527	4.279	3.106	4.133
	$AIS2_{fuzzy}$	0.961	0.753	16.394	4.028	2.959	3.942
	$AIS3_{precise}$	0.954	0.746	19.175	4.366	3.148	4.187
	$AIS3_{fuzzy}$	0.954	0.752	19.080	4.337	3.155	4.211
	$AIS4_{precise}$	0.956	0.747	18.285	4.259	3.068	4.089
	$AIS4_{fuzzy}$	0.958	0.755	17.786	4.194	3.069	4.097
	$AIS5_{precise}$	0.959	0.747	17.057	4.116	2.974	3.950
	$AIS5_{fuzzy}$	0.962	0.751	16.017	3.985	2.922	3.886
	$Set3_{precise}$ [a]	0.953	0.740	19.498	4.382	3.173	4.211
AB	*Set1*	0.969	0.770	12.857	3.481	2.871	4.105

(*continued on next column*)

Table A5 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS2_{precise}$	0.984	0.752	6.749	2.491	2.043	2.988
	$AIS2_{fuzzy}$	0.980	0.759	8.310	2.799	2.346	3.410
	$AIS3_{precise}$	0.985	0.752	6.184	2.309	1.883	2.810
	$AIS3_{fuzzy}$	0.983	0.766	7.124	2.537	2.098	3.091
	$AIS4_{precise}$	0.983	0.755	6.924	2.537	2.113	3.111
	$AIS4_{fuzzy}$	0.975	0.756	10.579	3.112	2.619	3.768
	$AIS5_{precise}$	0.983	0.750	6.959	2.393	1.958	2.888
	$AIS5_{fuzzy}$	0.980	0.759	8.222	2.643	2.186	3.192
	$Set3_{precise}$ [a]	0.980	0.755	8.175	2.647	2.186	3.210
GB	*Set1*	0.965	0.786	14.509	3.538	2.597	3.507
	$AIS2_{precise}$	0.974	0.774	10.728	2.915	2.250	3.046
	$AIS2_{fuzzy}$	0.971	0.778	12.005	3.246	2.535	3.453
	$AIS3_{precise}$	0.968	0.768	13.413	3.271	2.531	3.437
	$AIS3_{fuzzy}$	0.974	0.766	10.865	2.713	2.076	2.823
	$AIS4_{precise}$	0.974	0.768	10.657	3.097	2.384	3.237
	$AIS4_{fuzzy}$	0.970	0.775	12.628	3.386	2.650	3.600
	$AIS5_{precise}$	0.954	0.765	19.294	4.244	3.282	4.469
	$AIS5_{fuzzy}$	0.960	0.765	16.628	3.852	2.995	4.059
	$Set3_{precise}$ [a]	0.971	0.770	11.917	3.111	2.384	3.226
XG	*Set1*	0.966	0.786	14.223	3.620	2.692	3.641
	$AIS2_{precise}$	0.971	0.773	11.988	3.174	2.438	3.306
	$AIS2_{fuzzy}$	0.964	0.777	15.126	3.599	2.766	3.757
	$AIS3_{precise}$	0.964	0.762	15.014	3.485	2.673	3.639
	$AIS3_{fuzzy}$	0.965	0.774	14.662	3.570	2.728	3.697
	$AIS4_{precise}$	0.971	0.768	12.290	3.278	2.516	3.416
	$AIS4_{fuzzy}$	0.972	0.779	11.539	3.130	2.381	3.228
	$AIS5_{precise}$	0.948	0.765	21.685	4.533	3.501	4.755
	$AIS5_{fuzzy}$	0.957	0.758	17.993	3.859	2.929	3.955
	$Set3_{precise}$ [a]	0.959	0.771	17.299	3.835	2.890	3.902
LB	*Set1*	0.951	0.773	20.401	4.334	3.285	4.472
	$AIS2_{precise}$	0.957	0.769	18.000	4.086	3.154	4.284
	$AIS2_{fuzzy}$	0.965	0.774	14.771	3.662	2.829	3.877
	$AIS3_{precise}$	0.965	0.751	14.716	3.549	2.709	3.676
	$AIS3_{fuzzy}$	0.961	0.765	16.089	3.729	2.875	3.906
	$AIS4_{precise}$	0.961	0.756	16.273	3.800	2.915	3.961
	$AIS4_{fuzzy}$	0.962	0.764	15.650	3.683	2.832	3.853
	$AIS5_{precise}$	0.956	0.749	18.441	3.987	3.023	4.106
	$AIS5_{fuzzy}$	0.963	0.757	15.608	3.608	2.754	3.761
	$Set3_{precise}$ [a]	0.963	0.754	15.520	3.514	2.682	3.646
SVM	*Set1*	0.838	0.748	67.236	8.175	5.625	7.308
	$AIS2_{precise}$	0.859	0.768	58.991	7.655	5.491	7.220
	$AIS2_{fuzzy}$	0.840	0.762	66.704	8.154	5.866	7.724
	$AIS3_{precise}$	0.849	0.766	63.237	7.922	5.687	7.503
	$AIS3_{fuzzy}$	0.835	0.768	68.880	8.281	5.903	7.754
	$AIS4_{precise}$	0.840	0.769	66.795	8.150	5.817	7.656
	$AIS4_{fuzzy}$	0.842	0.766	66.097	8.114	5.750	7.527
	$AIS5_{precise}$	0.846	0.770	64.572	8.017	5.703	7.479
	$AIS5_{fuzzy}$	0.826	0.774	72.707	8.522	6.072	7.993
	$Set3_{precise}$ [a]	0.843	0.767	65.144	8.045	5.755	7.629
ANN	*Set1*	0.851	0.740	61.550	7.798	5.849	7.715
	$AIS2_{precise}$	0.875	0.763	52.267	7.189	5.479	7.270
	$AIS2_{fuzzy}$	0.855	0.767	60.467	7.758	5.885	7.788
	$AIS3_{precise}$	0.865	0.774	56.395	7.477	5.698	7.549
	$AIS3_{fuzzy}$	0.854	0.776	61.192	7.813	5.943	7.845
	$AIS4_{precise}$	0.868	0.775	55.403	7.419	5.662	7.506
	$AIS4_{fuzzy}$	0.859	0.769	58.976	7.652	5.795	7.653
	$AIS5_{precise}$	0.868	0.765	55.181	7.401	5.673	7.526
	$AIS5_{fuzzy}$	0.854	0.765	60.861	7.781	5.944	7.870
	$Set3_{precise}$ [a]	0.859	0.772	58.184	7.599	5.750	7.603
Ridge	*Set1*	0.758	0.729	100.434	10.018	7.588	10.192
	$AIS2_{precise}$	0.773	0.740	94.924	9.740	7.360	9.841
	$AIS2_{fuzzy}$	0.771	0.738	95.615	9.775	7.376	9.863
	$AIS3_{precise}$	0.787	0.749	89.046	9.434	7.255	9.691
	$AIS3_{fuzzy}$	0.785	0.747	89.799	9.473	7.247	9.679
	$AIS4_{precise}$	0.786	0.750	89.615	9.464	7.237	9.663
	$AIS4_{fuzzy}$	0.784	0.748	90.306	9.500	7.233	9.661
	$AIS5_{precise}$	0.778	0.743	92.805	9.631	7.393	9.895
	$AIS5_{fuzzy}$	0.777	0.741	93.183	9.651	7.389	9.899
	$Set3_{precise}$ [a]	0.775	0.745	93.218	9.652	7.454	9.977
LASSO	*Set1*	0.753	0.724	102.272	10.109	7.629	10.199
	$AIS2_{precise}$	0.772	0.738	95.454	9.767	7.342	9.801
	$AIS2_{fuzzy}$	0.770	0.735	96.087	9.799	7.368	9.835
	$AIS3_{precise}$	0.786	0.747	89.527	9.459	7.238	9.656
	$AIS3_{fuzzy}$	0.784	0.749	90.435	9.507	7.231	9.643
	$AIS4_{precise}$	0.785	0.747	89.820	9.475	7.231	9.647
	$AIS4_{fuzzy}$	0.783	0.749	90.743	9.523	7.206	9.600

(*continued on next column*)

Table A5 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS5_{precise}$	0.777	0.741	93.139	9.648	7.387	9.878
	$AIS5_{fuzzy}$	0.777	0.741	93.454	9.665	7.379	9.870
	$Set3_{precise}$ [a]	0.774	0.744	93.502	9.667	7.443	9.960

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

Table A6
The fit performance of eleven machine learning models for ship S7.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	$Set1$	0.828	0.680	69.472	8.260	6.302	8.155
	$AIS2_{precise}$	0.863	0.680	55.328	7.389	5.573	7.264
	$AIS2_{fuzzy}$	0.853	0.670	59.404	7.669	5.797	7.542
	$AIS3_{precise}$	0.856	0.670	58.237	7.557	5.694	7.432
	$AIS3_{fuzzy}$	0.850	0.671	60.824	7.759	5.896	7.686
	$AIS4_{precise}$	0.855	0.667	58.606	7.606	5.720	7.432
	$AIS4_{fuzzy}$	0.856	0.671	58.321	7.622	5.775	7.511
	$AIS5_{precise}$	0.865	0.678	54.511	7.334	5.473	7.099
	$AIS5_{fuzzy}$	0.867	0.669	53.839	7.272	5.433	7.071
	$Set3_{precise}$ [a]	0.880	0.683	48.319	6.903	5.173	6.749
ET	$Set1$	0.956	0.806	17.780	3.880	2.884	3.713
	$AIS2_{precise}$	0.983	0.830	6.818	2.117	1.487	1.925
	$AIS2_{fuzzy}$	0.971	0.819	11.791	3.070	2.178	2.810
	$AIS3_{precise}$	0.979	0.830	8.690	2.410	1.676	2.170
	$AIS3_{fuzzy}$	0.973	0.826	10.900	2.886	2.007	2.593
	$AIS4_{precise}$	0.983	0.829	6.702	2.040	1.405	1.820
	$AIS4_{fuzzy}$	0.976	0.820	9.693	2.688	1.894	2.454
	$AIS5_{precise}$	0.978	0.834	8.811	2.497	1.753	2.266
	$AIS5_{fuzzy}$	0.983	0.828	6.851	2.236	1.591	2.050
	$Set3_{precise}$ [a]	0.987	0.805	5.176	1.848	1.259	1.639
RF	$Set1$	0.964	0.793	14.369	3.774	2.813	3.649
	$AIS2_{precise}$	0.962	0.808	15.226	3.874	2.845	3.726
	$AIS2_{fuzzy}$	0.959	0.806	16.605	4.019	2.964	3.884
	$AIS3_{precise}$	0.958	0.810	16.981	4.069	2.963	3.883
	$AIS3_{fuzzy}$	0.958	0.811	17.018	4.095	2.979	3.891
	$AIS4_{precise}$	0.957	0.809	17.317	4.118	3.023	3.958
	$AIS4_{fuzzy}$	0.957	0.806	17.549	4.147	3.037	3.972
	$AIS5_{precise}$	0.964	0.813	14.703	3.799	2.760	3.604
	$AIS5_{fuzzy}$	0.966	0.817	13.809	3.684	2.693	3.513
	$Set3_{precise}$ [a]	0.961	0.794	15.501	3.920	2.867	3.740
AB	$Set1$	0.964	0.790	14.672	3.464	2.781	3.712
	$AIS2_{precise}$	0.984	0.813	6.382	2.399	1.970	2.684
	$AIS2_{fuzzy}$	0.986	0.812	5.520	2.083	1.663	2.269
	$AIS3_{precise}$	0.991	0.813	3.818	1.747	1.405	1.959
	$AIS3_{fuzzy}$	0.988	0.816	4.669	1.917	1.519	2.098
	$AIS4_{precise}$	0.987	0.810	5.060	2.008	1.642	2.260
	$AIS4_{fuzzy}$	0.984	0.812	6.476	2.401	2.023	2.782
	$AIS5_{precise}$	0.988	0.826	4.812	2.046	1.675	2.298
	$AIS5_{fuzzy}$	0.987	0.820	5.055	2.014	1.624	2.245
	$Set3_{precise}$ [a]	0.982	0.777	7.272	2.415	1.888	2.558
GB	$Set1$	0.962	0.803	15.408	3.756	2.777	3.605
	$AIS2_{precise}$	0.971	0.815	11.694	3.275	2.287	2.984
	$AIS2_{fuzzy}$	0.963	0.811	14.984	3.741	2.665	3.503
	$AIS3_{precise}$	0.978	0.818	8.854	2.705	1.774	2.312
	$AIS3_{fuzzy}$	0.973	0.821	10.691	3.106	2.141	2.809
	$AIS4_{precise}$	0.973	0.811	10.692	3.036	2.085	2.729
	$AIS4_{fuzzy}$	0.975	0.814	9.991	2.994	2.068	2.705
	$AIS5_{precise}$	0.975	0.826	10.330	3.084	2.147	2.810
	$AIS5_{fuzzy}$	0.977	0.822	9.116	2.812	1.991	2.610
	$Set3_{precise}$ [a]	0.986	0.785	5.466	2.156	1.442	1.880
XG	$Set1$	0.972	0.813	11.021	3.022	2.222	2.865
	$AIS2_{precise}$	0.972	0.810	11.116	3.156	2.168	2.779
	$AIS2_{fuzzy}$	0.968	0.814	12.937	3.415	2.437	3.161
	$AIS3_{precise}$	0.981	0.816	7.733	2.613	1.674	2.122
	$AIS3_{fuzzy}$	0.977	0.822	9.345	2.859	1.912	2.453
	$AIS4_{precise}$	0.973	0.813	10.730	3.019	2.111	2.714
	$AIS4_{fuzzy}$	0.972	0.821	11.364	3.182	2.236	2.883
	$AIS5_{precise}$	0.973	0.827	10.967	3.204	2.208	2.823
	$AIS5_{fuzzy}$	0.976	0.831	9.854	2.981	2.094	2.702
	$Set3_{precise}$ [a]	0.986	0.784	5.731	2.093	1.424	1.808
LB	$Set1$	0.957	0.789	17.547	4.044	3.053	3.968
	$AIS2_{precise}$	0.963	0.796	14.730	3.673	2.655	3.489
	$AIS2_{fuzzy}$	0.959	0.794	16.325	3.769	2.812	3.691
	$AIS3_{precise}$	0.976	0.809	9.713	2.740	1.932	2.536

(*continued on next column*)

Table A6 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS3_{fuzzy}$	0.975	0.805	10.022	3.034	2.163	2.840
	$AIS4_{precise}$	0.963	0.790	15.034	3.546	2.604	3.424
	$AIS4_{fuzzy}$	0.976	0.788	9.920	2.910	2.171	2.846
	$AIS5_{precise}$	0.981	0.803	7.624	2.614	1.804	2.364
	$AIS5_{fuzzy}$	0.973	0.813	10.759	2.940	2.095	2.762
	$Set3_{precise}$ [a]	0.982	0.785	7.152	2.366	1.742	2.283
SVM	$Set1$	0.906	0.786	38.185	6.078	4.323	5.574
	$AIS2_{precise}$	0.867	0.819	53.666	7.308	5.142	6.525
	$AIS2_{fuzzy}$	0.855	0.812	58.832	7.646	5.481	6.978
	$AIS3_{precise}$	0.861	0.820	56.159	7.463	5.273	6.678
	$AIS3_{fuzzy}$	0.860	0.820	56.543	7.490	5.326	6.747
	$AIS4_{precise}$	0.860	0.821	56.451	7.482	5.268	6.680
	$AIS4_{fuzzy}$	0.868	0.816	53.522	7.291	5.227	6.663
	$AIS5_{precise}$	0.854	0.815	58.893	7.641	5.411	6.831
	$AIS5_{fuzzy}$	0.855	0.811	58.735	7.637	5.447	6.893
	$Set3_{precise}$ [a]	0.871	0.748	51.533	7.113	5.173	6.591
ANN	$Set1$	0.863	0.786	55.639	7.392	5.651	7.274
	$AIS2_{precise}$	0.877	0.822	49.757	7.015	5.388	6.972
	$AIS2_{fuzzy}$	0.880	0.815	48.398	6.909	5.348	6.960
	$AIS3_{precise}$	0.886	0.818	46.000	6.707	5.154	6.669
	$AIS3_{fuzzy}$	0.891	0.816	44.189	6.596	5.071	6.574
	$AIS4_{precise}$	0.895	0.820	42.238	6.442	4.942	6.414
	$AIS4_{fuzzy}$	0.886	0.819	46.178	6.751	5.216	6.783
	$AIS5_{precise}$	0.879	0.805	48.929	6.903	5.287	6.805
	$AIS5_{fuzzy}$	0.869	0.806	53.158	7.238	5.571	7.180
	$Set3_{precise}$ [a]	0.892	0.771	43.321	6.515	5.071	6.587
Ridge	$Set1$	0.790	0.781	85.163	9.224	6.955	8.817
	$AIS2_{precise}$	0.807	0.797	77.989	8.828	6.677	8.485
	$AIS2_{fuzzy}$	0.806	0.793	78.557	8.860	6.742	8.596
	$AIS3_{precise}$	0.813	0.798	75.843	8.705	6.551	8.321
	$AIS3_{fuzzy}$	0.813	0.795	75.785	8.701	6.614	8.420
	$AIS4_{precise}$	0.811	0.800	76.446	8.740	6.580	8.334
	$AIS4_{fuzzy}$	0.811	0.797	76.688	8.754	6.640	8.432
	$AIS5_{precise}$	0.809	0.799	77.312	8.789	6.635	8.431
	$AIS5_{fuzzy}$	0.809	0.796	77.344	8.791	6.707	8.543
	$Set3_{precise}$ [a]	0.820	0.758	72.381	8.498	6.520	8.315
LASSO	$Set1$	0.789	0.781	85.405	9.238	6.961	8.819
	$AIS2_{precise}$	0.807	0.796	78.356	8.848	6.703	8.536
	$AIS2_{fuzzy}$	0.806	0.792	78.750	8.871	6.753	8.625
	$AIS3_{precise}$	0.811	0.796	76.696	8.753	6.608	8.410
	$AIS3_{fuzzy}$	0.811	0.796	76.732	8.756	6.644	8.459
	$AIS4_{precise}$	0.809	0.798	77.292	8.787	6.634	8.425
	$AIS4_{fuzzy}$	0.809	0.797	77.165	8.781	6.659	8.477
	$AIS5_{precise}$	0.808	0.798	77.734	8.813	6.664	8.483
	$AIS5_{fuzzy}$	0.808	0.796	77.682	8.811	6.719	8.568
	$Set3_{precise}$ [a]	0.819	0.758	72.827	8.524	6.550	8.374

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

Table A7
The fit performance of eleven machine learning models for ship S8.

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
DT	$Set1$	0.916	0.774	54.181	7.305	5.213	6.441
	$AIS2_{precise}$	0.910	0.734	54.413	7.305	5.175	6.255
	$AIS2_{fuzzy}$	0.909	0.751	55.016	7.391	5.225	6.303
	$AIS3_{precise}$	0.904	0.764	58.172	7.506	5.430	6.557
	$AIS3_{fuzzy}$	0.910	0.757	54.043	7.255	5.193	6.272
	$AIS4_{precise}$	0.910	0.746	54.031	7.280	5.178	6.240
	$AIS4_{fuzzy}$	0.907	0.747	56.211	7.451	5.288	6.384
	$AIS5_{precise}$	0.908	0.771	55.429	7.369	5.275	6.362
	$AIS5_{fuzzy}$	0.925	0.777	45.620	6.622	4.665	5.687
	$Set3_{precise}$ [a]	0.916	0.769	50.649	6.985	4.922	5.949
ET	$Set1$	0.998	0.882	1.556	0.811	0.551	0.679
	$AIS2_{precise}$	0.998	0.866	0.927	0.774	0.471	0.588
	$AIS2_{fuzzy}$	0.995	0.862	2.789	1.167	0.761	0.941
	$AIS3_{precise}$	0.997	0.870	2.074	1.178	0.744	0.925
	$AIS3_{fuzzy}$	0.997	0.870	1.736	0.987	0.628	0.776
	$AIS4_{precise}$	0.996	0.865	2.660	1.228	0.795	0.986
	$AIS4_{fuzzy}$	0.997	0.863	1.868	1.000	0.639	0.791
	$AIS5_{precise}$	0.998	0.877	1.223	0.864	0.549	0.687
	$AIS5_{fuzzy}$	0.998	0.874	1.459	0.949	0.616	0.766
	$Set3_{precise}$ [a]	0.995	0.876	2.783	1.404	0.907	1.120
RF	$Set1$	0.978	0.859	13.895	3.707	2.535	3.124

(continued on next column)

Table A7 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS2_{precise}$	0.974	0.840	15.552	3.918	2.680	3.274
	$AIS2_{fuzzy}$	0.976	0.840	14.417	3.782	2.623	3.196
	$AIS3_{precise}$	0.970	0.848	17.980	4.202	2.858	3.508
	$AIS3_{fuzzy}$	0.975	0.848	15.355	3.895	2.688	3.294
	$AIS4_{precise}$	0.974	0.839	15.743	3.948	2.718	3.318
	$AIS4_{fuzzy}$	0.974	0.840	15.800	3.958	2.747	3.356
	$AIS5_{precise}$	0.975	0.856	15.054	3.848	2.645	3.229
	$AIS5_{fuzzy}$	0.976	0.860	14.471	3.781	2.619	3.204
	$Set3_{precise}$ [a]	0.976	0.855	14.566	3.798	2.624	3.187
AB	*Set1*	0.982	0.870	11.601	3.288	2.747	3.479
	$AIS2_{precise}$	0.993	0.852	4.466	2.032	1.640	2.081
	$AIS2_{fuzzy}$	0.990	0.860	6.002	2.344	1.935	2.443
	$AIS3_{precise}$	0.993	0.857	4.187	1.850	1.461	1.869
	$AIS3_{fuzzy}$	0.993	0.859	4.258	1.782	1.406	1.799
	$AIS4_{precise}$	0.992	0.853	4.887	2.044	1.651	2.115
	$AIS4_{fuzzy}$	0.994	0.856	3.765	1.844	1.442	1.826
	$AIS5_{precise}$	0.995	0.866	3.047	1.544	1.182	1.516
	$AIS5_{fuzzy}$ [a]	0.996	0.874	2.605	1.432	1.084	1.388
	$Set3_{precise}$	0.991	0.863	5.365	2.114	1.693	2.148
GB	*Set1*	0.983	0.875	10.771	3.062	2.188	2.750
	$AIS2_{precise}$	0.990	0.845	6.089	1.946	1.312	1.638
	$AIS2_{fuzzy}$	0.991	0.842	5.472	1.961	1.317	1.636
	$AIS3_{precise}$	0.989	0.851	6.797	2.127	1.416	1.777
	$AIS3_{fuzzy}$	0.985	0.850	9.273	2.484	1.694	2.119
	$AIS4_{precise}$	0.991	0.849	5.801	1.838	1.227	1.539
	$AIS4_{fuzzy}$	0.992	0.846	4.749	1.968	1.355	1.695
	$AIS5_{precise}$	0.988	0.859	7.367	2.097	1.438	1.781
	$AIS5_{fuzzy}$	0.986	0.863	8.510	2.330	1.622	2.011
	$Set3_{precise}$ [a]	0.985	0.860	9.102	2.427	1.670	2.075
XG	*Set1*	0.991	0.877	5.538	1.956	1.429	1.791
	$AIS2_{precise}$	0.979	0.841	12.914	3.086	2.202	2.713
	$AIS2_{fuzzy}$	0.989	0.850	6.647	2.089	1.460	1.815
	$AIS3_{precise}$	0.985	0.842	9.459	2.492	1.759	2.163
	$AIS3_{fuzzy}$	0.989	0.850	6.678	2.185	1.473	1.824
	$AIS4_{precise}$	0.982	0.839	11.047	2.796	1.970	2.439
	$AIS4_{fuzzy}$	0.987	0.847	7.669	2.271	1.552	1.925
	$AIS5_{precise}$	0.973	0.863	16.064	3.747	2.646	3.231
	$AIS5_{fuzzy}$	0.978	0.861	13.557	3.275	2.288	2.797
	$Set3_{precise}$ [a]	0.979	0.856	12.821	2.974	2.114	2.589
LB	*Set1*	0.979	0.871	13.718	3.540	2.601	3.309
	$AIS2_{precise}$	0.973	0.832	16.182	3.417	2.447	3.043
	$AIS2_{fuzzy}$	0.983	0.837	9.937	2.763	1.977	2.484
	$AIS3_{precise}$	0.982	0.844	10.828	2.743	1.936	2.441
	$AIS3_{fuzzy}$	0.979	0.848	12.946	3.069	2.203	2.759
	$AIS4_{precise}$	0.969	0.824	18.763	4.017	2.929	3.677
	$AIS4_{fuzzy}$	0.973	0.835	16.409	3.765	2.725	3.415
	$AIS5_{precise}$	0.973	0.849	16.500	3.523	2.521	3.138
	$AIS5_{fuzzy}$	0.978	0.860	13.784	3.213	2.262	2.836
	$Set3_{precise}$ [a]	0.976	0.852	14.749	3.261	2.338	2.882
SVM	*Set1*	0.900	0.862	64.371	8.014	5.742	6.905
	$AIS2_{precise}$	0.892	0.846	64.923	8.048	5.555	6.624
	$AIS2_{fuzzy}$	0.886	0.841	68.842	8.289	5.746	6.819
	$AIS3_{precise}$	0.906	0.861	56.735	7.523	5.223	6.371
	$AIS3_{fuzzy}$	0.901	0.856	59.608	7.712	5.314	6.426
	$AIS4_{precise}$	0.906	0.862	56.668	7.517	5.201	6.343
	$AIS4_{fuzzy}$	0.900	0.856	60.112	7.746	5.316	6.423
	$AIS5_{precise}$	0.897	0.858	62.015	7.865	5.503	6.604
	$AIS5_{fuzzy}$	0.892	0.854	65.006	8.054	5.607	6.685
	$Set3_{precise}$ [a]	0.910	0.869	54.154	7.349	5.117	6.123
ANN	*Set1*	0.914	0.857	55.217	7.398	5.605	6.809
	$AIS2_{precise}$	0.903	0.834	58.524	7.590	5.568	6.678
	$AIS2_{fuzzy}$	0.896	0.830	62.810	7.876	5.865	7.024
	$AIS3_{precise}$	0.913	0.854	52.371	7.201	5.280	6.420
	$AIS3_{fuzzy}$	0.910	0.850	54.338	7.351	5.383	6.496
	$AIS4_{precise}$	0.917	0.849	50.268	7.033	5.138	6.252
	$AIS4_{fuzzy}$	0.913	0.847	52.479	7.207	5.293	6.397
	$AIS5_{precise}$	0.911	0.846	53.888	7.282	5.345	6.458
	$AIS5_{fuzzy}$	0.912	0.848	53.388	7.276	5.335	6.431
	$Set3_{precise}$ [a]	0.924	0.862	46.222	6.733	4.964	5.959
Ridge	*Set1*	0.866	0.842	86.315	9.288	7.004	8.561
	$AIS2_{precise}$	0.858	0.829	85.639	9.248	7.008	8.573
	$AIS2_{fuzzy}$	0.856	0.826	87.056	9.324	7.077	8.631
	$AIS3_{precise}$	0.868	0.839	79.392	8.905	6.681	8.300
	$AIS3_{fuzzy}$	0.866	0.836	81.009	8.996	6.756	8.358
	$AIS4_{precise}$	0.867	0.838	79.957	8.936	6.693	8.299
	$AIS4_{fuzzy}$	0.864	0.835	81.764	9.037	6.780	8.364

(*continued on next column*)

Table A7 (*continued*)

Model	Dataset	R^2	R^2 (test)	MSE	RMSE (ton/day)	MAE (ton/day)	MAPE (%)
	$AIS5_{precise}$	0.867	0.840	80.108	8.945	6.728	8.344
	$AIS5_{fuzzy}$	0.865	0.838	81.568	9.026	6.771	8.361
	$Set3_{precise}$ [a]	0.879	0.853	72.818	8.529	6.512	7.959
LASSO	*Set1*	0.865	0.842	87.140	9.332	7.023	8.576
	$AIS2_{precise}$	0.857	0.828	85.994	9.267	7.011	8.567
	$AIS2_{fuzzy}$	0.855	0.826	87.359	9.340	7.076	8.621
	$AIS3_{precise}$	0.867	0.839	80.087	8.944	6.729	8.334
	$AIS3_{fuzzy}$	0.864	0.837	81.870	9.043	6.794	8.365
	$AIS4_{precise}$	0.867	0.838	80.157	8.948	6.700	8.306
	$AIS4_{fuzzy}$	0.864	0.834	81.947	9.047	6.787	8.369
	$AIS5_{precise}$	0.867	0.840	80.433	8.963	6.737	8.348
	$AIS5_{fuzzy}$	0.864	0.838	82.001	9.050	6.782	8.359
	$Set3_{precise}$ [a]	0.878	0.852	73.581	8.573	6.525	7.966

Note:

[a] $Set3_{precise}$ is the best dataset in Li et al. (2022) which is the Part I of this series of studies.

References

Bennett, G.G., 1996. Practical rhumb line calculations on the spheroid. J. Navig. 49 (1), 112–119.

Bergstra, J., Yamins, D., Cox, D.D., 2013. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the 30th International Conference on Machine Learning (ICML 2013), pp. 115–123.

BIMCO, 2021. Just in Time Arrival Clause for Voyage Charter Parties 2021. https://www.bimco.org/contracts-and-clauses/bimco-clauses/current/just-in-time-arrival-clause-for-voyage-charter-parties-2021. (Accessed 5 April 2022).

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Haussler, D. (Ed.), Proceedings of the Annual Conference on Computational Learning Theory. ACM Press, Pittsburgh, PA, pp. 144–152.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Drucker, H., 1997. Improving regressors using boosting techniques. ICML 97, 107–115.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42.

Haykin, S.O., 2008. Neural Networks and Learning Machines, the third ed. Prentice Hall, New Jersey.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., 2018. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47. (Accessed 10 September 2021).

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.

IMO, 2020. Fourth IMO GHG study 2020. https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Fourth%20IMO%20GHG%20Study%202020%20Executive-Summary.pdf. (Accessed 19 March 2022).

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Liu, T.Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 3146–3154.

Li, X., Du, Y., Chen, Y., Nguyen, S., Zhang, W., Schönborn, A., Sun, Z., 2022. Data fusion and machine learning for ship fuel efficiency modeling: Part I – Voyage report data and meteorological data. Commun. Transport. Res. 2, 100074. https://doi.org/10.1016/j.commtr.2022.100074.

Merkel, A., Kalantari, J., Mubder, A., 2022. Port call optimization and CO2-emissions savings–Estimating feasible potential in tramp shipping. Marit. Transport Res. 3, 100054.

Rio, M.H., Mulet, S., Picot, N., 2014. Beyond GOCE for the ocean circulation estimate: synergetic use of altimetry, gravimetry, and in situ data provides new insight into geostrophic and Ekman currents. Geophys. Res. Lett. 41 (24), 8918–8925.

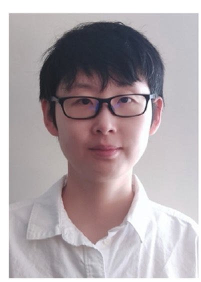
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B 58 (1), 267–288.

Wan, Z., El Makhloufi, A., Chen, Y., Tang, J., 2018. Decarbonizing the international shipping industry: solutions and policy recommendations. Mar. Pollut. Bull. 126, 428–435.

Weintrit, A., Kopacz, P., 2011. A novel approach to Loxodrome (Rhumb-Line), orthodrome (great circle) and geodesic line in ECDIS and navigation in general. Methods and Algorithms in Navigation, Marine Navigation and Safety of Sea Transportation 123–132.

Yan, R., Wang, S., Psaraftis, H.N., 2021. Data analytics for fuel consumption management in maritime transportation: Status and perspectives. Transport. Res. E Logist. Transport. Rev. 155, 102489.

Yuquan Du is a Senior Lecturer in the Centre for Maritime and Logistics Management at Australian Maritime College (AMC), University of Tasmania, Australia. He obtained his PhD in Control Theory and Control Engineering in June 2012 from Nankai University, and then worked for 3 years as a Research Fellow at the Centre for Maritime Studies, National University of Singapore. His current research concentrates on machine learning models, optimization models, and solution algorithms for management science problems in transportation, logistics, and supply chain management. He is a member of the Australian Maritime Logistics Research Network.



Yanyu Chen is a PhD candidate in the Institute for Marine and Antarctic Studies (IMAS), University of Tasmania (UTAS), Australia. She obtained her master's degree in Information and Communication Technology in July 2021 from UTAS, and then worked as a Research Assistant in the Centre for Maritime and Logistics Management at Australian Maritime College, UTAS. Her current research concentrates on machine learning models, deep learning models, optimization models, and solution algorithms for research problems in marine biology, fishery management, and supply chain management.

Xiaohe Li graduated from Harbin Engineering University. He obtained his PhD in Power Engineering and Engineering Thermophysics from Harbin Engineering University in March 2022. His current research focuses on ship fuel consumption prediction models, machine learning models, energy efficiency optimization models, solution algorithms, and energy efficiency management software development for the shipping industry.

Alessandro Schönborn is an Assistant Professor at World Maritime University (WMU) in the Maritime Energy Management (MEM) specialization. His research interests focus on fuel and propulsion technology, and on marine renewable energy from wave, tidal and ocean currents, but he has further interests in solar and wind power for marine applications. He graduated from University College London (UCL) with an M.Eng. degree in Mechanical Engineering and holds a PhD in combustion (UCL) for a thesis on renewable fuels for diesel engines. Before joining WMU, he worked at MAN Energy Solutions, SINTEF Energy Research, and Lund University.

Zhuo Sun is a Professor at the Department of Logistics in Dalian Maritime University. He received his dual bachelor's degree in Civil Engineering and Computer Science from Dalian University of Technology in China in 2003, master's degree in Hydraulics from Dong-A University in South Korea in 2005, and PhD in Urban Planning from Nagoya University in Japan in 2008. He conducted his postdoctoral research in the Centre for Maritime Studies of National University of Singapore from 2009 to 2012. His research interests are in the areas of shipping network design and port planning. He won the Gold Prize for the "*Next Generation Container Port Challenge*" in the Singapore in 2013 and the third-place winner of the "*The Last Mile Delivery" Algorithm Competition* organized by Alibaba, Hong Kong University of Science and Technology, and INFORMS in 2016. He developed a spatial planning tool named MicroCity (https://microcity.github.io).