Review

Big Data Analytics and Machine Learning of Harbour Craft Vessels to Achieve Fuel Efficiency: A Review

Zhi Yung Tay [1,*], Januwar Hadi [1], Favian Chow [1], De Jin Loh [1] and Dimitrios Konovessis [2]

1 Engineering Cluster, Singapore Institute of Technology, 10 Dover Drive, Singapore 138683, Singapore; januwar.hadi@singaporetech.edu.sg (J.H.); 2001187@sit.singaporetech.edu.sg (F.C.); lohdejin@gmail.com (D.J.L.)
2 Department of Naval Architecture, Ocean & Marine Engineering, Strathclyde University, 100 Montrose Street, Glasgow G4 0LZ, UK; dimitrios.konovessis@strath.ed.uk
* Correspondence: zhiyung.tay@singaporetech.edu.sg; Tel.: +65-6592-1944

Abstract: The global greenhouse gas emitted from shipping activities is one of the factors contributing to global warming; thus, there is an urgent need to mitigate the adverse effect of climate change. One of the key strategies is to build a vibrant maritime industry with the use of innovation and digital technologies as well as intelligent systems. The digitization of the shipping industry not only provides a competitive edge to the shipping business model but also enhances ship operational and energy efficiency. This review paper focuses on the big data analytics and machine learning applied to harbour craft vessels with the aim to achieve fuel efficiency. The paper reviews the telemetry system requires for the digitalization of harbour craft vessels, its challenges in installation, the vessel monitoring and data transmission system. The commonly used methods for data cleaning are also presented. Last but not least, the paper considers two types of the machine learning systems, i.e., supervised and unsupervised machine learning systems. The multi-linear regression and hidden Markov model for supervised machine learning system and the artificial neural network, grey box model and long short-term memory model for unsupervised machine learning are discussed, and their pros and cons are presented.

1. Introduction

1.1. Impact of Industrial Revolution on Shipping Industry

The shipping industry has undergone a transformation from the use of steam engines in the First Industrial Revolution, followed by the widespread use of electrical and combustion engine-powered vessels in the Second Industrial Revolution. The shipping industry has also experienced a digital transformation in the Third Industrial Revolution, and is now entering the dawning of the Fourth Industrial Revolution that focuses on smart shipping based upon the integration of the internet of things (IoT), intelligent systems and innovative solutions. The transformation of the shipping industry from the First to the Third Industrial Revolutions has brought the maritime and shipping businesses a competitive edge. One of the digitalization technologies is the use of big data analytics (BDA) and machine learning (ML) to achieve fuel efficiency. The improvement in energy efficiency plays a role in reducing emissions intensity, which is part of the main goals in the 2018 Paris Climate Accord [1]. In addition, the United Nations Shipping Agency (UNSA) has reached an agreement to cut carbon emissions by at least 50 per cent by 2050, compared with the 2008 levels [2]. Evidenced from the significant benefits the First to Third Industrial Revolutions brought to the shipping industry, the Fourth Industrial Revolution that focuses on harnessing the infoCOMM technologies, networks and big data to create tech-enabled solutions has encouraged shipping companies to embrace innovative digital solutions.

1.2. Global Warming and Decarbonisation

The maritime sector is constantly searching for alternatives to fossil fuels with tighter environmental controls in place to aid in reducing climate change. Referring to Table 1, the total shipping activities account for about 3% of all man-made greenhouse emissions [3]. According to the report in [4], it is estimated that 16 ships create approximately the same amount of Sulphur pollution as all the cars in the world. The International Maritime Organization (IMO) has implemented the Initial IMO Strategy on Reduction of Greenhouse Gas (GHG) Emissions from Ships, to reduce gross annual GHG emissions from shipping by at least 50% by 2050, relative to the 2008 emission levels [5].

Table 1. Breakdown of GHG emissions [6].

	Third IMO GHG Study (Million Tonnes)						ICCT (Million Tonnes)		
	2007	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Global CO_2 Emissions	31,959	32,133	31,822	33,661	34,726	34,968	35,672	36,084	36,062
International Shipping	881	916	858	773	853	805	801	813	812
Domestic Shipping	133	139	75	83	110	87	73	78	78
Fishing	86	80	44	58	58	51	36	39	42
Total Shipping (% of global)	1100 (3.5%)	1135 (3.5%)	977 (3.1%)	914 (2.7%)	1021 (2.9%)	942 (2.6%)	910 (2.5%)	930 (2.6%)	932 (2.6%)

Several means in place to mitigate the carbon emission from shipping activities are:

- Energy Efficiency Measurement Index
  The IMO has implemented the Ship Energy Efficiency in Annex VI of the MARPOL [7] which include the Energy Efficiency Design Index (EEDI) in the ship design state, Ship Energy Efficiency Management Programme (SEEMP) in the ship operational planning stage and Energy Efficiency Operational Index (EEOI) in monitoring the energy efficiency and collection of data for continuous improvement in terms of carbon emission. The proposed ship energy efficiency concept aims to minimize GHG emissions by developing ways to lower fuel usage, more efficient ship design and switching to alternative fuels that emit lesser GHG [8].

- Alternative Marine Fuels
  To meet the increasingly strict emission regulation, alternative fuels such as liquified natural gas (LNG), ammonia, methanol and liquid hydrogen have become a more important part of the energy mix. LNG trade has expanded dramatically from 100 million tonnes in 2000 to approximately 300 million tonnes in 2017. However, LNG still generates carbon emissions but is significantly lesser than diesel. Ammonia and hydrogen could be produced from hydrocarbons, and green ammonia and green hydrogen which are produced from electrolysis powered by renewables or nuclear are excellent sources of zero-emission fuel [9]. Methanol on the other hand is easier to store and handle than LNG. However, ammonia, hydrogen and methanol have a lower energy content than conventional fuel.

- Electrification
  Electrification of marine vessels is becoming more commercially viable due to increasingly declining battery costs fueled by the growth of electric cars. Several commercial electric vessels have also been built. For example, the 4.3 MWh all-electric ferry, Ellen (Figure 1a), was built in the framework of the EU's Horizon 2020 program and is estimated to save 2000 tons of CO_2 per year in its operation. A small electric cruise ship, Brime Explorer (Figure 1b), and the Grimaldi GGSG ro-ro freighter (Figure 1c) were built to operate in Norway's fjords and the Mediterranean, respectively [10]. Nevertheless, there are several challenges in marine electrification, especially in the charging infrastructure, voyage distance and weight issues [11].
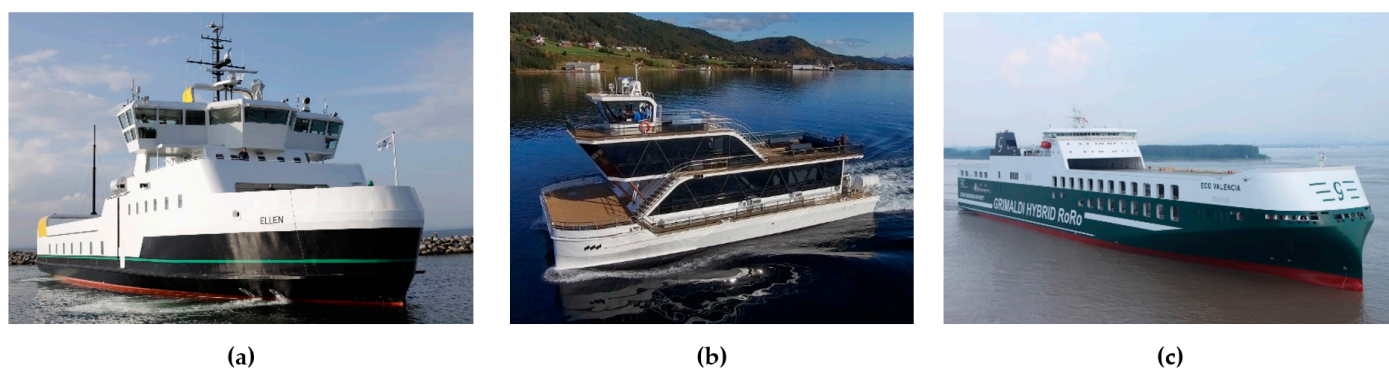
Figure 1. Examples of electric vessels: (**a**) Ellen Ferry; (**b**) Brim Explorer cruise ship; (**c**) Grimaldi GGSG ro-ro freighter.

There are also other means for mitigating carbon emissions such as wind-assisted propulsion ship [12], use of Flettner rotor [13], slow steaming [14] and many more. This review paper focuses on the digitalization of marine vessels, in particular, the harbour craft vessels (HCV) utilizing BDA and ML to achieve fuel efficiency. The paper is arranged as follows: Section 2 describes the various means of digitalization technologies applied in the maritime industry; Section 3 describes the general framework for BDA and ML systems applied to ship to achieve fuel efficiency; Section 4 covers the types of data acquisition systems; Section 5 discusses data filtering and preparation; Section 6 covers the BDA and ML models commonly used for HCV. Last but not least, the Conclusion is provided in Section 7.

2. Digitalization in Maritime Industry

Digitalization using BDA and ML has been utilized in the shipping industry to improve operational efficiency, productivity and to enhance fuel efficiency. BDA and ML are used to reorganize huge amounts of unstructured data and analyze these data to establish the correlations between diverse aspects that are difficult for human analysts to identify. ML helps to accelerate the process of BDA and is used to uncover trends and patterns of the data. Research conducted across industries has shown significant improvements in industries that are in pursuit of digitalization where these improvements have enabled industries to experience better economic performance [15]. Although digitalization in the shipping industry has been relatively slow compared with other industries, big players such as Rolls Royce and Wärtsilä have set up research and development centres to explore remote and autonomous shipping [16,17].

Machinery that is built for ships usually does not live up to its expected lifespan. This might be due to the lack of maintenance or the incapability of detecting faults sooner to prevent catastrophic damages. To reduce such risks, preventive maintenance that involves the evaluation of equipment condition via periodic check (i.e., BDA) and continuous equipment condition monitoring (i.e., IoT) enables the maintenance process to be much more efficient as remote diagnostics of ships' machinery will be made available [18]. When faults are discovered immediately, further engine damage can be prevented, thereby reducing the amount of fuel consumed. This in turn reduces the amount of GHG emissions and also results in 10–35 per cent more cost-effective operations [19]. According to a report in [20], predictive maintenance can reduce unexpected failures by 55% and maintenance costs are expected to reduce by an estimated 25% to 30%. Companies are also utilizing advanced technology such as weather routing, allowing ships adequate time to avoid bad weather [21,22]. In addition to that, the technology ensures that its seagoing assets' gas emissions and cargo temperatures are monitored from shore, thereby reducing maintenance costs and the risk of failure due to negligence [23]. Another type of technology that helps to reduce fuel consumption is the Marine Growth Prevention System (MGPS). The MGPS aims to combat marine organism growth and prevent the organism from depositing on the ship's systems, thus helping to eliminate corrosion. The MGPS aids in the efficient

operation of the seawater-supplied system and machinery [24], and reduces the ship's resistance. This in turns increases the energy savings and reduces the fuel consumption of a ship [24].

To stay ahead of an ever-evolving environment, there is a growing emphasis on the adoption of digital technologies in the maritime industry, employing BDA and ML [25–27]. It enables a large amount of data to be collected, stored and processed where many aspects of marine operations could be conducted via digital platforms, efficiently and effectively. BDA and ML to achieve fuel efficiency in ships are widely implemented in commercial vessels such as container ships, oil tankers and cruise liners. However, little work has been carried out on HCV such as tugboats, patrol vessels and ferries. The carbon emission from HCV should not be overlooked; therefore, countries such as Singapore have allocated a substantial amount of funding in encouraging the harbour craft sectors to invest in digital solutions [28]. In the next section, the BDA and ML frameworks applied to HCV are presented.

3. State-of-the-Art

Presented herein is state-of-the-art literature for BDA and ML in the recent two decades used for achieving fuel energy efficiency in a ship. The latest development of BDA and ML utilized in maritime related research is obtained from sources published in scientific journals and conferences, and the practical application of BDA and ML in the industry is also reviewed from information found in the public domain.

3.1. Big Data Analytics

The BDA could be used for identifying the pattern and correlation of fuel consumption with respect to the environment and ship data recorded, to improve the fuel efficiency through optimal vessel speed and voyage route. A research study on vessels fuel consumption by the use of BDA has been conducted in [29,30], where it is reported that the power, and thus the fuel required to propel the ship through water depends on the trim of the vessels. The optimum ship speed during the time of vessel delivery for fuel consumption changes over time due to a variety of factors such as engine wear, coating of vessels, etc. The use of BDA can help shipowners determine the optimum speed for fuel consumption, taking into consideration factors such as bunker cost, freight rates and schedules [31]. The use of BDA to calculate potential fuel savings can provide ship owners with detailed insight on all aspects of vessel operations impacting the fuel efficiency to inform on the return of investment (ROI) decision. Additionally, the Fujitsu Laboratory has developed technology that uses BDA for large ships to estimate fuel efficiency, speed and other performance in actual sea conditions in 2016 [32]. Other research on suggesting optimal vessel speed decisions in maritime logistics using weather big data has also been conducted [33,34]. It is also evidenced by a report in [35] that BDA has helped shipping lines such as Maersk in cutting its fuel consumption by 13%. Big data analytics are also used to investigate the speed optimization process for large container ships [36].

3.2. Machine Learning

Machine learning can be defined as the application of artificial intelligence and computer systems to learn from the environment, improve itself from experience without the need for any explicit programming and call for action that does not require human intervention. Machine learning focuses on enabling algorithms to learn from the data provided, gather insights and make predictions on previously unanalyzed data using the information gathered [37].

The shipping industry has been keeping an eye on the development of ML that can customize container freight and overcome tough operational problems encountered in everyday operations. For instance, ML can be used to forecast estimated travel time (ETA), even if there are congestions at transshipments points, weather-related difficulties, overbooking issues, and equipment paucity. The computation learns from the past data,

thus providing a much more definite forecast [38]. To improve fuel efficiency and vessel performance at sea, an algorithm was developed for an intelligent fuel oil consumption monitoring system that can propose an optimal trim condition to minimize the ship resistance during voyage [39]. Fujitsu Laboratories has developed a new artificial technology and teamed up with Mitsui O.S.K. Line, Ltd. (MOL) in improving fuel efficiency and reducing CO_2 emission of a ship through the use of operational big data [40]. ML is also used for fouling analysis in predicting a more accurate dry docking, cleaning and coating schedules as the vessel fuel consumptions are affected by the vessel speed and the fouling of the ship [41]. An ML approach was also developed by the Technical University of Denmark in predicting the main energy consumption under realistic operational conditions [42]. An artificial neural network (ANN)-based decision support system has also been developed for cargo vessel operation by employing a combination of traditional statistical analysis and ANNs [27].

3.3. Ship Energy Efficiency

The ship energy efficiency depends significantly on the ship resistance and propulsion. Theoretically, for a newly built ship, the total ship hull resistance is obtained from the bare hull resistance test, where the ship without propeller is run in calm water in the towing tank at a constant speed V and the ship hull resistance is measured by the computer in the towing carriage. The ship hull resistance R_T is then used to calculate the effective power, P_E, of the marine engine as follows [43]:

$$P_E = R_T \times V. \tag{1}$$

As the ship bare hull resistance test does not take into consideration the effect due to the propeller and wave as well as mechanical losses such as shaft and gears, the final ship engine power which is measured as the Total Engine Brake Power, P_{TEB}, has to take into account these losses coefficients, η_{Losses}, as shown in Equation (2) [44]

$$P_{TEB} = \frac{P_E}{\eta_{Losses}} \tag{2}$$

The fuel consumption of the ship depends significantly on P_{TEB}, where higher fuel consumption is required when a greater amount of P_{TEB} is needed, and vice versa. However, in practice when the ship is operating in the sea, the amount of fuel required to achieve the same V might differ from ship to ship depending on the performance of the engine and also due to the changes of the ship resistance affected by factors such as hull fouling, weather conditions and water depth. Thus, a more reliable method to measure the fuel consumption efficiency is by direct measurement from boats in the real sea and by the use of the Energy Efficiency Index (EEI), measured as [45]

$$EEI = \frac{\Delta^{\frac{2}{3}} \cdot V^3}{FC} \tag{3}$$

where Δ is the ship displacement, FC the fuel consumption and V the vessel speed. If Equations (1)–(3) are arranged into a single equation, the relationship between the EEI with the P_{TEB}, R_T, V, Δ, FC is

$$EEI \propto f\left(\frac{\Delta, P_{TEB}, V}{FC, R_T}\right) \tag{4}$$

Equation (4) shows that the vessel speed, mean draft and trim are the few parameters if properly adjusted, may reduce fuel consumption and carbon emissions, thereby increasing the energy efficiency. For HCV, the draft and trim do not significantly change due to their scale, thus one of the methods to improve the EEI is via the adjustment to the vessel speed. In addition, the speed of the vessel also depends on the shipping route and the

environmental conditions such as the wind and current velocity. Thus, the influence of the vessel speed, wind and current velocity on energy efficiency has to be taken into account.

4. Machine Learning with Big Data Analytics to Achieve Fuel Efficiency

4.1. Digitalisation Framework

The BDA and ML framework for a tugboat is described here. Although the framework described here is targeted for discovering the knowledge domain of the fuel consumption in tugboats to achieve fuel efficiency, it is applicable for other applications such as preventive maintenance, route optimization, etc., as described in Section 2. Figure 2 shows the schematic diagram of the digitalization process from the data collected in the ship and then transmitted to the land-based system via the network middleware. The information process via BDA is then transmitted back to the control bridge for ship route decision making. The data is continually fed to the ML system in improving the decision-making capability to improve energy efficiency. Machine learning with the BDA process involves several components, i.e.,:

- Telemetry: Sensors and data acquisition;
- Vessel monitoring system;
- Network middleware;
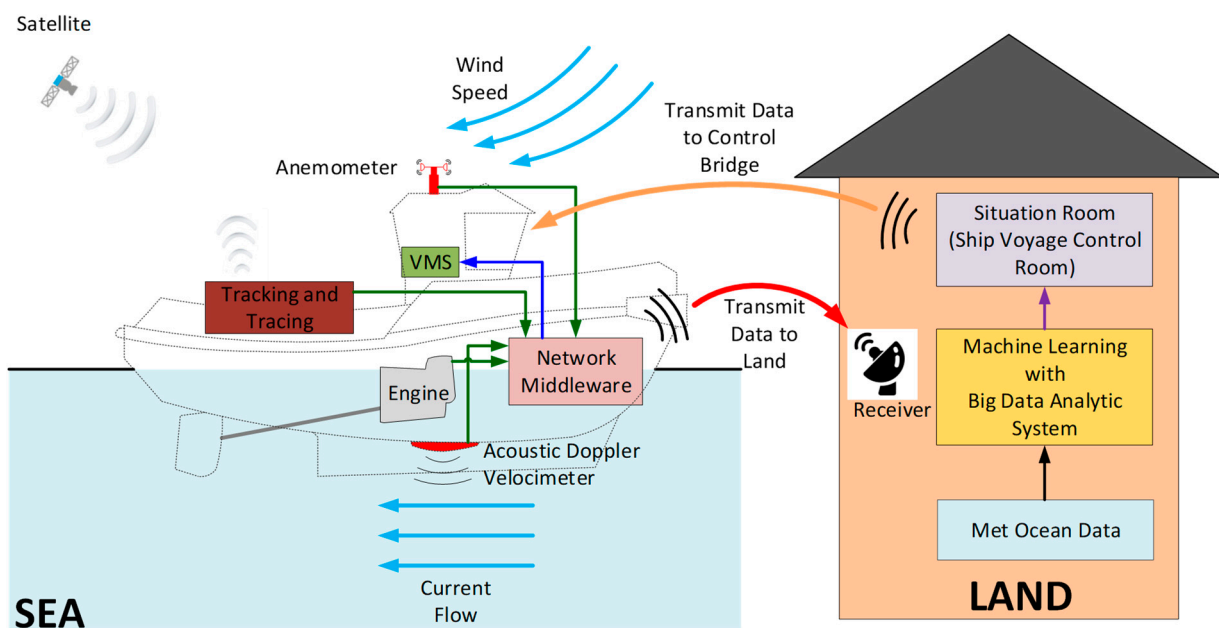- ML with BDA system.

Figure 2. Schematic diagram of digitalization process of tugboat involving data collection, transmission and ML with BDA.

The ML with BDA involves two states, i.e., STAGE 1 descriptive analytics; STAGE 2 predictive and prescriptive analytics (see Figure 3). The descriptive analytic is to find the patterns or correlation between the *EEI* with the various factors that have been identified such as the vessel speed *V*, vessel displacement Δ, vessel route, fuel consumption *FC*, wind velocity and current flow. Once the patterns and correlations have been identified, predictive analytics are used to predict the behavior of the vessel from the metocean data, when the vessel is travelling under different scenarios such as at a specified vessel route and vessel speed. These possible scenarios or solutions are then fed to the situation room for decision making.
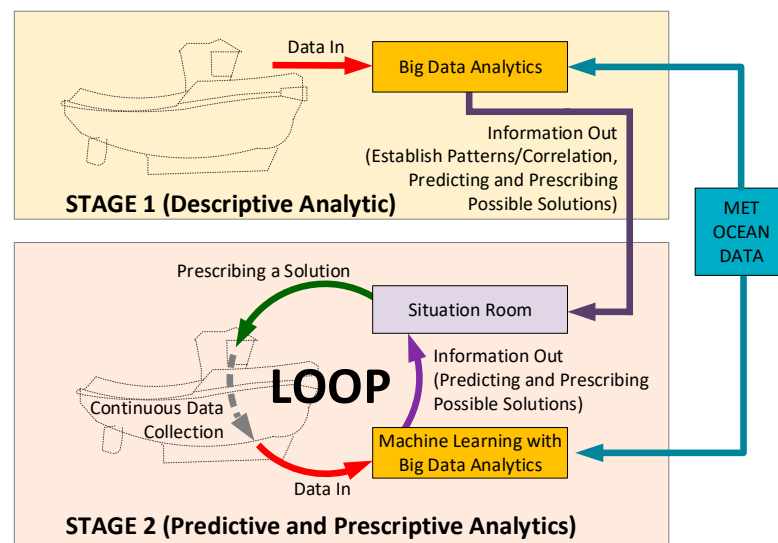
Figure 3. Machine learning with BDA process flow.

As presented in Figure 3, STAGE 1 only involves Descriptive Analytics where data are collected from the ship and fed into the BDA software to establish the patterns and correlations, as specified in Equations (1)–(3). In STAGE 2, the information is then transferred to the situation room to suggest the vessel on the optimized speed and route to be taken that could improve the *EEI*. This is a continuous loop process, where data will be collected continuously from the ship, and to be used in training the ML system with BDA, so that the decision-making process could be continuously improved.

4.2. Data Acquisition

4.2.1. Telemetry

Depending on the application of ML, various types of sensors must be installed on the ship to collect the necessary data.

Mass Flowmeter

For ship energy efficiency, one of the most important data is the fuel consumption, which could be collected by the flowmeter. There are several flowmeters in the market such as the volumetric flowmeter that measures the volume of the fuel consumed and the more accurate Coriolis mass flowmeter that measures the mass of the fuel consumed. The installation of Coriolis mass flowmeter was made mandatory by the Singapore Maritime Port Authority (MPA) from 1 July 2019 [46] to increase fuel quality and reliability, and also to prepare the sector for the rise in distillate bunker fuel deliveries after the IMO implemented a 0.5 per cent worldwide Sulphur cap on 1 January 2020 [47]. A Coriolis meter is based on motion mechanics principles as shown in Figure 4. As fluids enter the device, a driving coil induces the tubes to vibrate in opposition at their natural resonant frequency thereby creating sine waves. The fluid induces Coriolis force which causes the flow tubes to twist. The density of the fluid (mass) is measured by analyzing the frequency of the sine waves and the readings are highly accurate with typical measurement errors of ± 0.2 per cent. Coriolis mass flowmeter is equipped with built-in sensors for temperature, pressure, and density measurements with a display system [48]; therefore, the results could be stored and transmitted readily. Moreover, the Coriolis mass flowmeter requires low maintenance as there are no moving parts. The volumetric flowmeter on the other hand has moving parts that can be degraded over time which will result in inaccurate readings. Additionally, volumetric type meters contain separate temperature and pressure gauges that might be readily tampered with or gauges that are inaccurate.
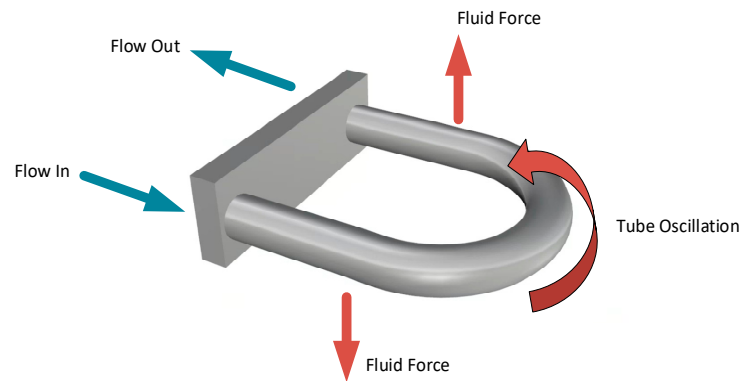
Figure 4. Schematic of a Coriolis meter.

Wind Sensor

The ship energy efficiency depends on the ship resistance (Equation (4)) which in turn is affected by the environmental data such as wind, wave and current. Wind sensors, as shown in Figure 4, are used to collect data on the wind speed and direction the ship experienced. The mechanical wind sensor is shown in Figure 5a is also used to measure wind data; however, it operates with moving parts. The mechanical sensor operates by having a rotating cup and vane in measuring the wind speed and direction. The time it takes a mechanical sensor to physically start-up or record a change in wind direction causes observed variances in recorded wind speed [49]. For example, if a storm passes through a region and the wind abruptly changes direction, the sensor must slow down, stop, then resume to keep up with the shift. The inaccuracy in the mechanical wind sensor could be overcome by the ultrasonic wind sensor shown in Figure 5b. The Ultrasonic wind sensor [49], also known as a sonic anemometer, uses a microcontroller to measure the travelling time of the ultrasonic pulse in computing the wind speed and does not require any moving parts to operate, thus requires lesser maintenance and have a longer lifespan. Inertia does affect the ultrasonic sensor as it is capable of measuring changes of wind direction or high gust immediately and in real-time.
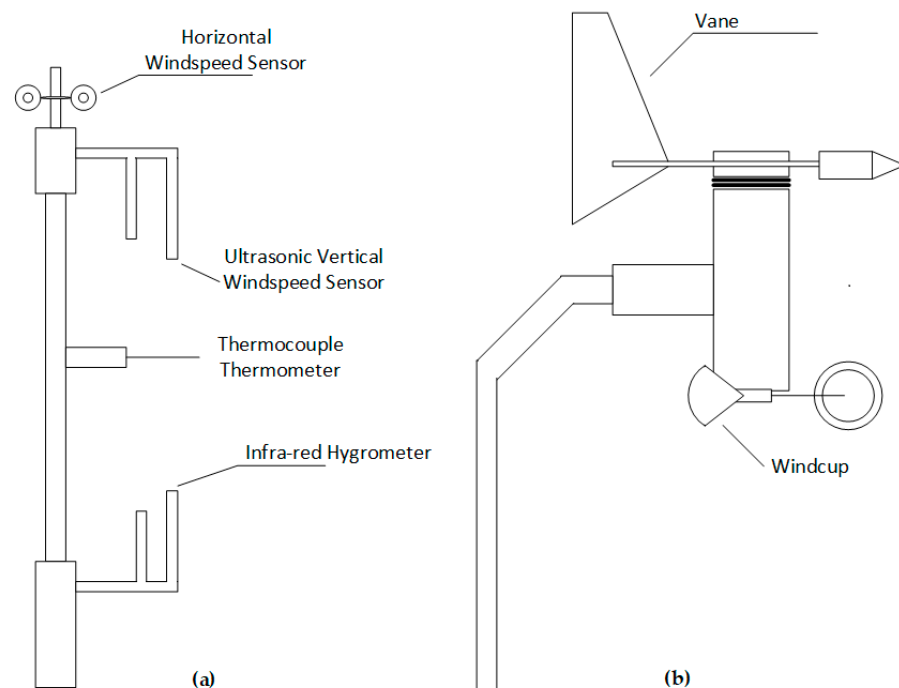

Figure 5. Schematic diagram of wind sensor. (**a**) Ultrasonic wind sensor; (**b**) mechanical wind Sensor.

Other Sensors

Other sensors to be installed onboard the ship for better accuracy in the prediction of fuel consumption are such as the acoustic Doppler profiler (ADP) to measure the current flow as the resistance of the ship is significantly affected by the current flow. However, the installation of the ADP is expensive and requires modification of ship structural configuration to fit the ADP. Tidal tables are thus used for predicting the average current flow when the real-time current profile is not available. The driveshaft RPM sensor is also used to collect the rotational speed of the propeller shaft. This RPM data could be used to categorize the operational activities of the HCV by comparison with the fuel consumptions and environmental data in the ML process.

4.2.2. Simulated/Online Data

Other than data accumulated from sensors, data could also be obtained from online platforms to aid in the analysis. Such examples include weather forecasts data and vessels' route data which could be used to plan voyages and avoid any routes that may include bad weather forecasts. Sea traffic such as shown in Figure 6 is available by providers such as MarineTraffic [50] and VesselFinder [51]. Data collected through simulations can also be of use to optimize the voyage and efficiency of the ship where multiple simulations can be run to determine the safest and fuel conserving route to take. Additionally, simulations can be run to test the optimal speed of the ship to reduce fuel consumption and carbon emissions. An example of ship resistance estimation is given in Figure 7.
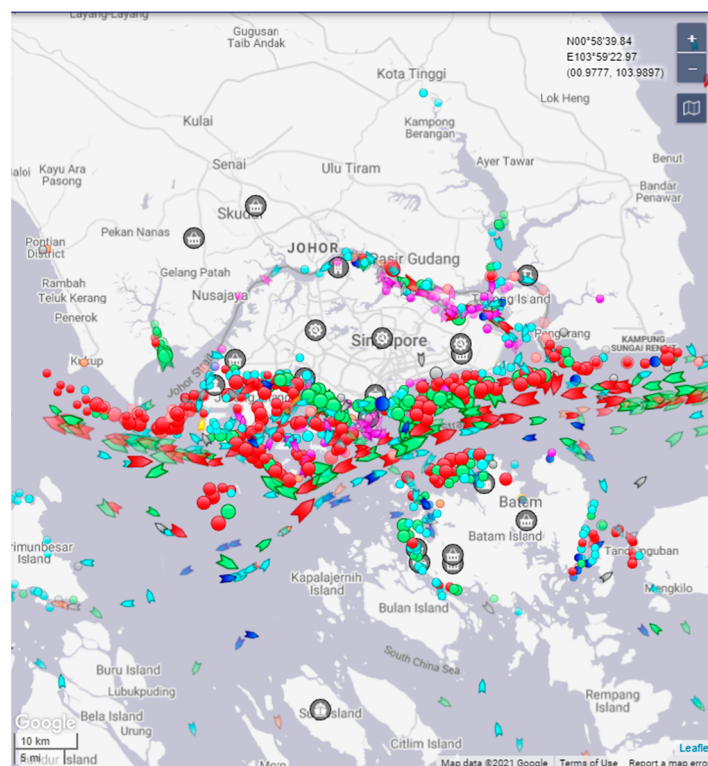

Figure 6. Screenshot from marinetraffic.com showing ships' locations surrounding Singapore (www.marinetraffic.com, accessed date: 22 October 2021).
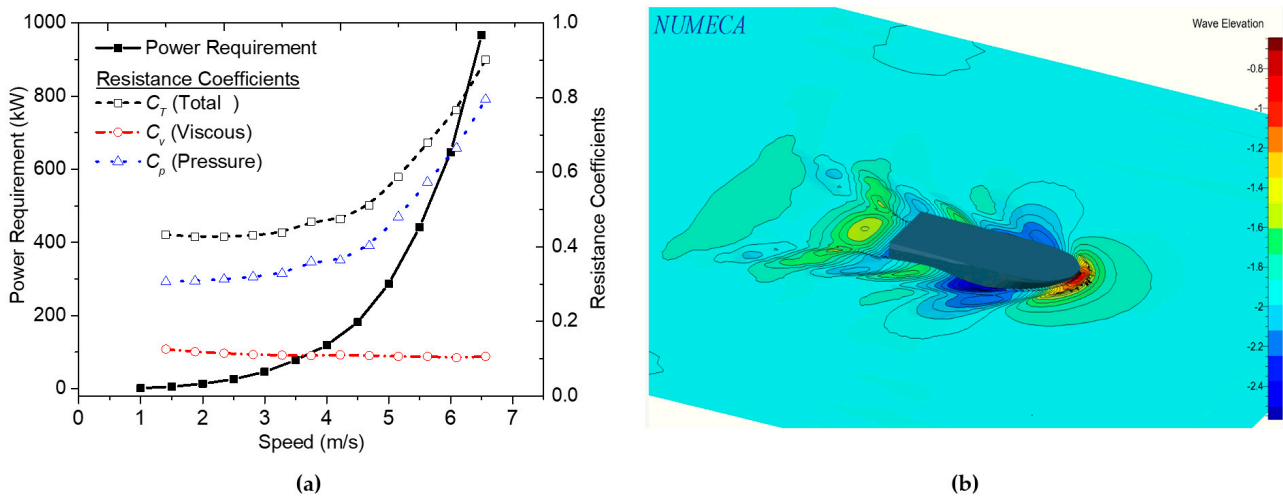
Figure 7. (**a**) Ship resistance/resistance coefficients vs. tugboat speed obtained from numerical simulation; (**b**) ship resistance model in FINE/Marine®.

4.2.3. Challenges in Sensors Installation

There are several challenges in the installation of sensors, such as the capital, operation, maintenance and troubleshooting costs. The installation and troubleshooting of the sensors must also accommodate the busy operational schedules of the HCV to minimize disruption. In some cases, the control system or sensor may or may not offer a warning if there is an intermittent malfunction. There may also be cases where the system is unaware of the program failure, thereby unable to offer feedback to the user. This is where troubleshooting comes in and performs the necessary procedures to figure out the faults to have the sensors operating again. The inclement weather makes the marine environment less appealing; therefore, any troubleshooting could only be carried out onshore. Additionally, conducting research experiments on board a ship may interfere with the ship operations, so may become prohibitively expensive due to the interference with the schedule of the ship operations [52]. Therefore, it is recommended that the ship crew should be educated and trained to troubleshoot and rectify the faults to minimize the downtime and potential inconvenience caused.

The conventional approach to ship cyber-security is based on the premise of keeping ship systems isolated from the Internet and ship/company intranets, which is still the standard for many safeties and security-sensitive ship owners. Even when the proprietary interface is accessible, data transmission is often serial and unidirectional [52]. With so many technologies onboard a ship, cyber security poses a significant threat. There are possibilities of cyber-attacks that may trigger ecological disasters such as oil spills by activating remotely-controlled or automated discharge valves, or by maliciously manipulating GPS signals and receivers to create groundings or accidents [23]. For prevention, companies have to invest heavily in cyber security as incidents of such magnitude could backfire on the efforts to reduce the effects of climate change.

4.3. Vessel Monitoring System and Data Transmission

Data collected from the ship, or the environment are recorded and stored in the data acquisition system where monitoring systems are usually installed onboard the ship for continuous monitoring of the operational conditions. The network middleware is used for communication and management of data where these data are transmitted to the shore for further analysis. There are several challenges in the transmission of data to shore as vessels are often operating at regions out of coverage of the shore station thereby having to deal with unstable connections. Sometimes, synchronizing big files between shore and the vessels becomes a significant issue due to the possibility of duplicate data and time lapse between machines. Some technologies can assist in the replication/synchronization of data

between the vessel and the land by establishing a private cloud between the firm and its vessels. The most commonly used method to transmit data is through electromagnetic wave transmission technology which is also known as very high frequency (VHF) radio transmission. Satellites are commonly used among ships as satellite communications (such as INMARSAT and COSPAS-SARSAT) are reliable for offshore operations and emergency communications. However, they tend to be costly; therefore, 4G is used as an alternative for the transmission of information that is not time sensitive.

5. Data Preparation and Filtering

Raw data collected from sensors tend to have errors and distortions. These data have to be pre-processed/filtered before they could be used for further analysis. The raw data has to be denoised and cleaned to transform them into useful information. Different types of errors may be accumulated throughout the process of data collection and this section aims to bring awareness to the most common errors retrieved from the data acquisition system and the data filtering methodology.

5.1. Types of Errors

5.1.1. Measurement Error

The measurement error is the error that occurs in the data gathering chain. This could be due to the flaws in the measuring instruments which result in the difference between the real value and the actual data recorded. An example of different measurement errors is shown in Figure 8. Figure 8 shows an example of different RPM values with respect to time for the shaft of a propeller. The significance of the measurement error could be quantified by computing the mean squared error (MSE) given as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{5}$$

where n is the number of data points, Y_i the observed values and $\hat{Y}_i$ the predicted/recorded values. The MSE indicates the distance a regression line (black lines given in Figure 8) is to a set of points (observed values) where the square of the bracket terms in Equation (1) is to eliminate any negative signs and to give more weight to larger differences. Therefore, data with smaller MSE imply a higher accuracy in the recorded/predicted data. It is to note that the lower the MSE, the closer it is to determine the optimum fit line.
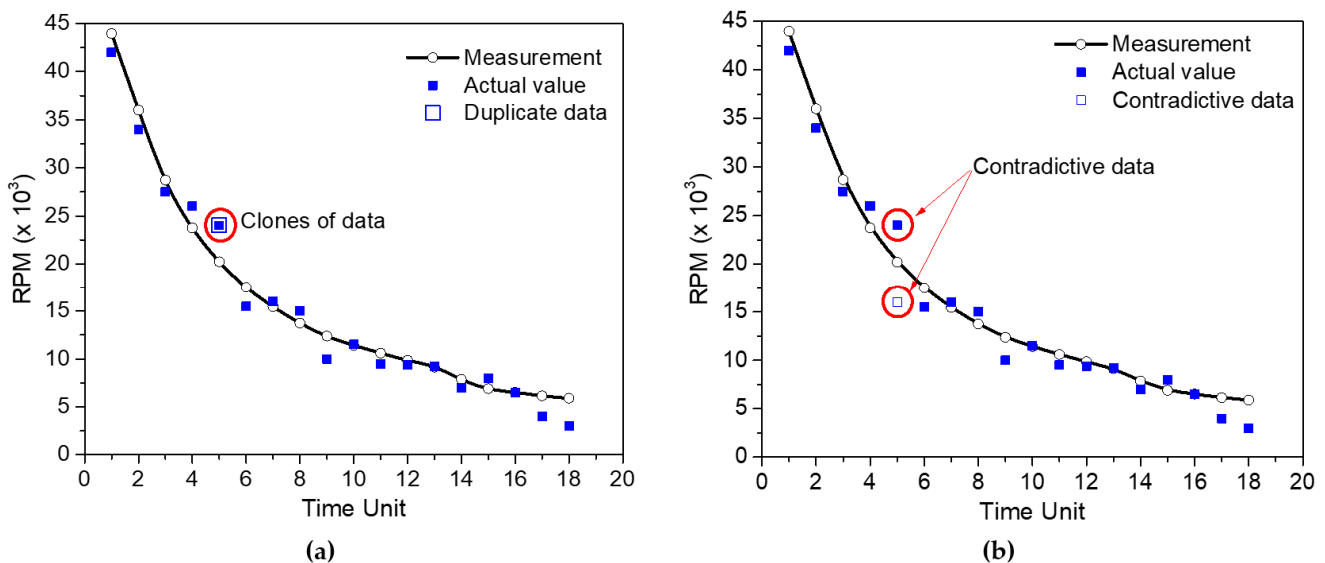

Figure 8. Example of deviation between real and measured value for RPM data, showing (**a**) duplicate data time unit 5 and 5.5; (**b**) contradictive data at time unit 5.

5.1.2. Inconsistent Data

In some cases, the data recorded by the sensors may be inconsistent where duplicate data, contradictive data and outliers may be recorded.

Duplicate Data

Duplicates of records may arise due to error in the handling of data while moving data between systems. In such cases, the duplicate data are easy to spot and appear as clones or duplicates as shown in Figure 8a. The two units of duplicated data by the sensor are denoted by the open and close square symbols in the red circle. The inclusion of duplicate data in the data analytic may indicate inaccurate or stale data. Fortunately, the duplicate data could be removed by simply deleting these clones, but one has to be careful to not deduplicate the original data.

Contradictive Data

Contradictive data occurs when multiple data are being recorded at a certain time. An example of contradictive data is shown in Figure 8b where there are two RPM values recorded at unit time 5. There are several methodologies proposed to remove contradictory data [53,54], but the removal of contradictory data contributes to the incompleteness in the dataset, thereby reducing the soundness of any information from such set of data [55]. Nwagwu et al. [55] proposed a novel approach for visually identifying contradictory data in a large and noisy dataset by applying a mutual exclusion rule in identifying contradictory data.

Outliers

Outliers are values in a random population sample that deviates abnormally from other values. An example of outliers is shown in Figure 9 presented by a spike in the temperature between 50 and 51 days. Outliers could be due to anomalies resulting from a faulty instrument. These outliers should be removed from the dataset but if a gathered value is extremely rare, it might cause the mean or standard deviation to drift drastically. As a result, removing such values is a crucial component of the data filtering process [56]. In some ways, this definition of outliers is delegated to the analyst in deciding the data point that constitutes abnormality. It is important to describe normal observations before aberrant observations may be identified. One way to do that is to examine the graphed data's overall form for significant aspects such as symmetry and deviations from assumptions. Another way is to examine the data for outlier findings that are not found in the rest of the data. Scatter plots and box plots are two graphical approaches for finding outliers, as well as an analytical procedure for detecting outliers when the distribution is normal.
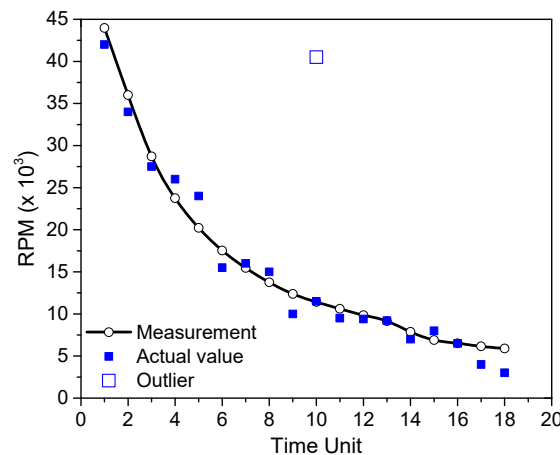

Figure 9. Example of an outlier recorded by RPM sensor.

5.2. Filtering of Raw Operational Data Techniques

Several methods could be utilized to remove the errors as presented in the previous sections. Some of the commonly used filtering techniques are given in the following sections. The pros and cons for each technique are summarized at the end of this section in Table 2.

Table 2. Advantages and disadvantages of different filtering methodologies.

Filtering Methods	Advantages	Disadvantages
CCT	Intuitive, straightforward, time information content is preserved	Applicable only to data that is normally distributed and without noise
HWT	Simple, computationally efficient, suitable for signal with sudden transition (not continuous), time information content is preserved	Shift sensitivity, poor directionality, lack of phase information [54],
FFT	Able to filter varying frequencies signal, able to convert discrete data into continuous data, maintain information on amplitudes, harmonics and phase	Time information content of signal is lost, relatively computationally expensive, sensitive to the length of Fourier transformation used
KF	Time information content is preserved, ideal for signals that are continuously changing and uncertain, light on memory, fast	Initial state probability density function has to be known, sensitive to initial estimate of state

5.2.1. Control Chart Techniques

The control chart technique (CCT) is a statistic mathematic formulated algorithm implemented in time series data to detect irregularities by sliding a predefined window along a stream of data points. This technique acts as a condition statement, i.e., accepting values within a threshold boundary and considers the data point as an outlier beyond this threshold boundary. Through this condition, the outliers could be eliminated as presented in Equation (6). The threshold boundary is taken as 3σ exclusively for data that follows a Gaussian distribution.

$$\chi_j = \begin{cases} \text{outlier} & \text{if } |\chi_j - \mu| > 3\sigma \\ \text{normal} & \text{otherwise} \end{cases} \tag{6}$$

where χ_j represents the data point for the *j*th number of observations, μ refers to the average of the observation sum data and σ is the standard deviation.

This technique is applied in [57,58] to filter out the outliers in the fuel oil consumption (FOC) time series. The detected outliers would be either replaced with previous data values or removal depending on the consistency of the data's resolution to be retained. This approach is intuitive and straightforward; however, the downside of the method is that it is only applicable to normally distributed data and does not comprise noisy data captured initially in the raw data as presented in Figures 10 and 11.
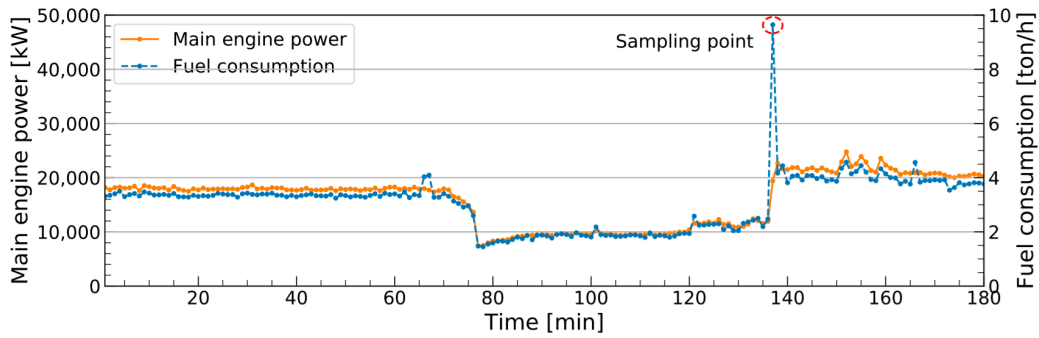
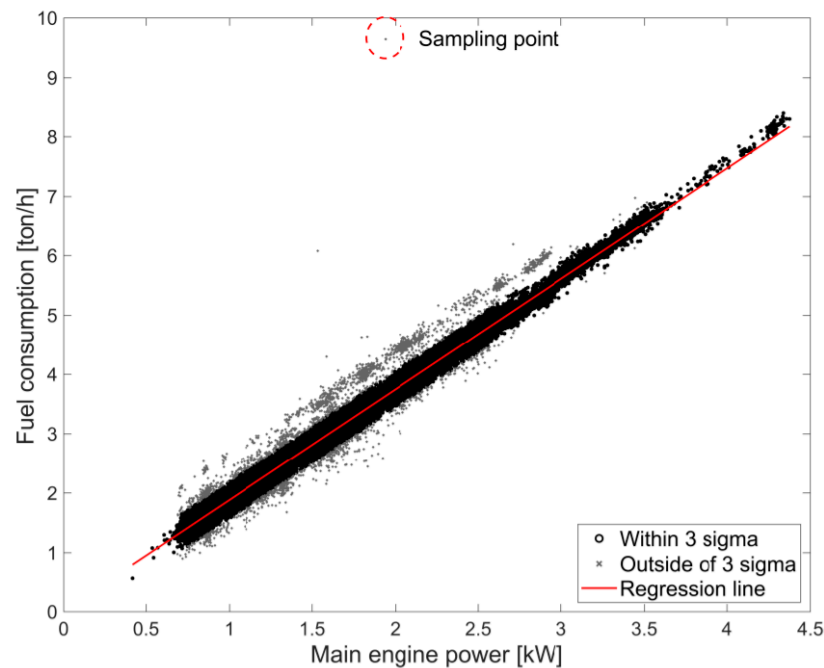**Figure 10.** Outlier detected around sampling point for main engine power and fuel consumption [59].

Figure 11. Outlier detection using 3σ rule [59].

5.2.2. Haar Wavelet Transformation

The Haar wavelet transformation (HWT) is a discrete wavelet transform system used to denoise and detects outliers. Subasi [60] suggested the use of HWT to decompose signals into several levels of signal components that yield a better result in training the ANN for accurate classification and diagnosis [60]. The HWT follows a straightforward technique in breaking down the signal into coefficients by sliding a fixed duration size of a window along the signal represented as

$$\psi(t) = \begin{cases} 1 & \text{for} & 0 \le t < \frac{t}{2} \\ -1 & \text{for} & \frac{1}{2} \le t < \frac{t}{2} \\ 0 & \text{for} & \text{otherwise} \end{cases} \tag{7}$$

where $\psi(t)$ is the Haar wavelet and t the time.

By sliding the window through the signal, it decomposes the signal into multiple levels to produce details and approximation coefficients equation, as shown in Figure 12. The decomposition levels depend on the degree of signal refinement to be accomplished by incrementing the level accordingly. The higher the level introduced, the more sensitive the fluctuation of data points will be filtered, contributing to the likelihood of relevant data points being filtered, resulting in an overall smoother signal. Therefore, the selection of the

level is an important factor while modelling the HWT algorithm. Figures 12 and 13 show the decomposition of the signal data points in many subsets of increments of the levels. The newly form collection of wavelet coefficients functions as a threshold in evaluating the signal data points' acceptance or rejection. Tay et al. [61] applied the HWT wavelet decomposition to the fuel consumption data collected from an HCV. An example of the filtering data is given in Figure 14.
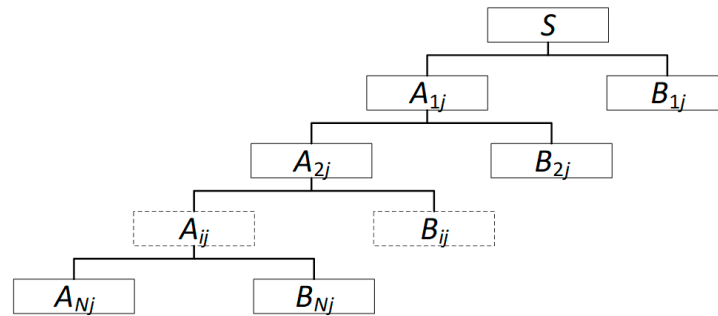

Figure 12. Multi-level decomposition by HWT showing the wavelet coefficients A_{ij} and B_{ij}. N is the total level of decomposition considered in HWT.
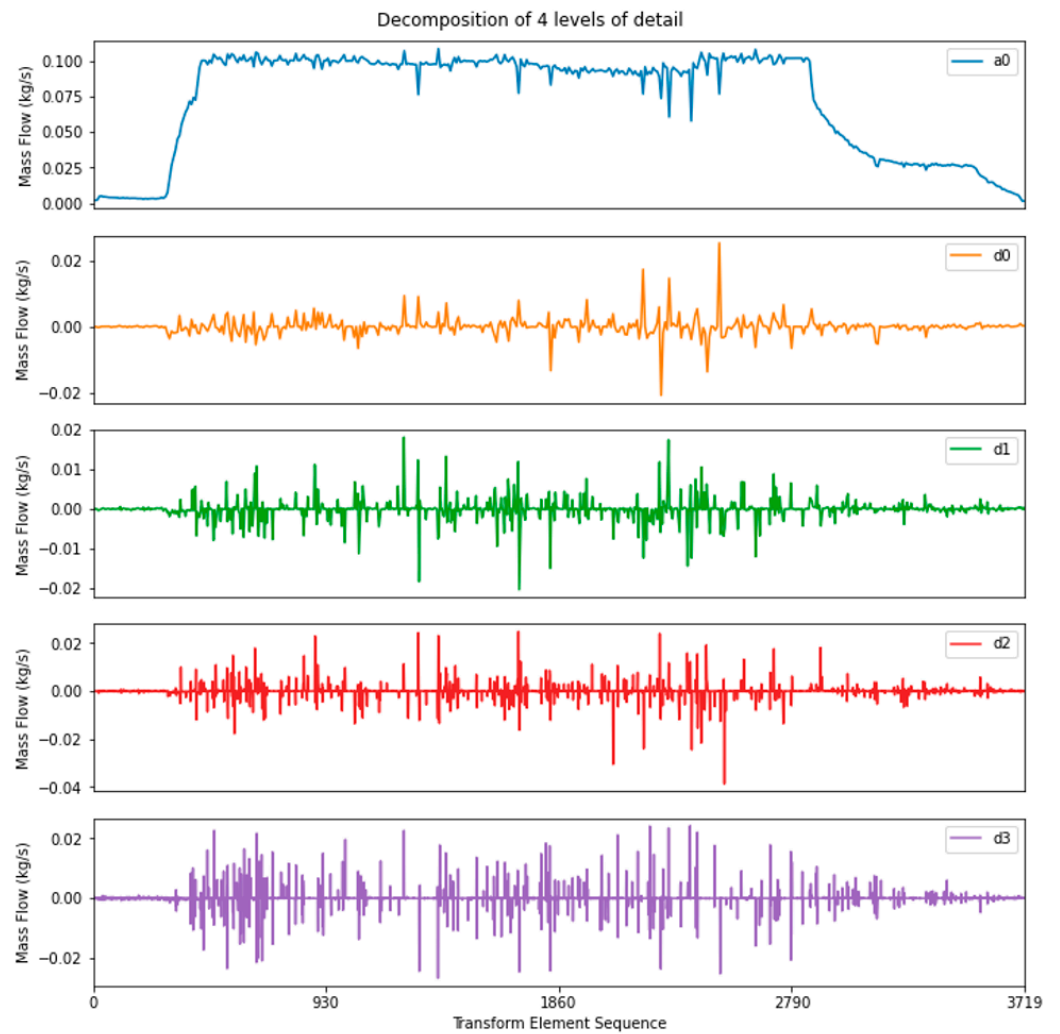

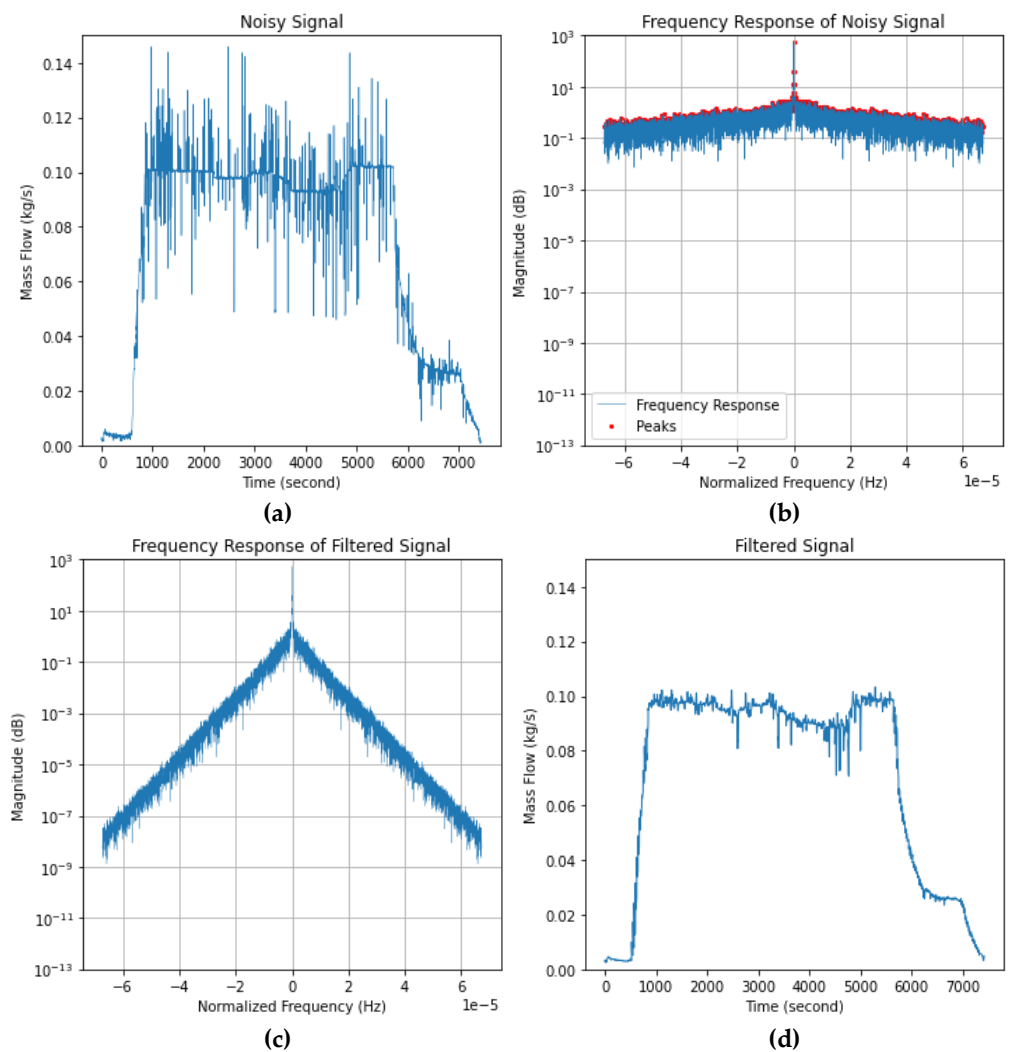**Figure 13.** Visualization of different decomposition levels by HWT.

Figure 14. Fast Fourier transform to detect and filter out the noise. (**a**) Noisy signal. (**b**) Frequency response of noisy signal. (**c**) Frequency response of filtered signal. (**d**) Filtered signal.

5.2.3. Fast Fourier Transform

The Fast Fourier Transform (FFT) technique is commonly used in a wide variety of applications, including audio and image compression formats, among others. Most signals are highly compressible in the FFT domain, which represents transform scales as a function of frequency in the detection of noise to be removed. Furthermore, FFTs can accelerate the detection and filtering processes, making them useful in digital signal processing.

In FFT, the signal was transformed from time-series to frequency domain represented as power spectral density (PSD) [62]. This PSD represents the signal's intensity magnitude where the PSD's peak represents the noise to be filtered out as depicted in Figure 14. To denoise, a threshold is set based on the optimal response to be retained. The PSD could be applied to ship operational data such as the fuel consumption recorded by mass flowmeters as the sensor measurement noise may vary irregularly, making denoising challenging, as simple filtering techniques do not perform well.

5.2.4. Kalman Filter

The Kalman filter (KF) is a low pass filter that acts as an optimal estimator to minimize the MSE based on the measurement and estimated data. The filtering behavior depends on the hyperparameter known as the covariance matrix. If the covariance matrix is set to a low value, it will obtain a smooth function by removing a sudden spike in values; otherwise, it will trust the measurement data input if the covariance values are set as a high value. This

approach was used in [63] to denoise all the ship operational data provided that the trend remains close to speed over ground (SOG) data shown in Figure 15.
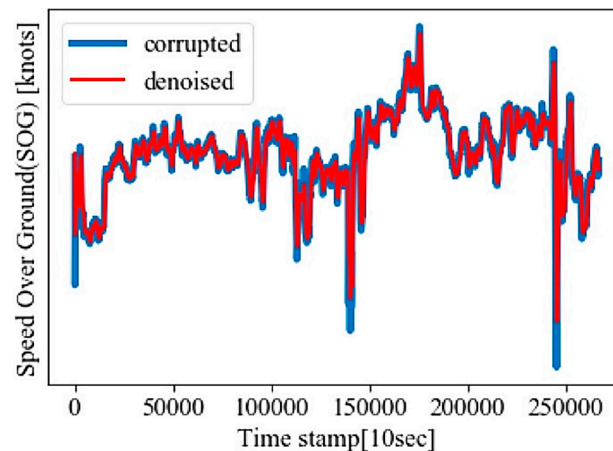

Figure 15. Denoising of SOG using KF [63].

6. Supervised and Unsupervised Machine Learning Model

A substantial research effort was carried out on various statistical models in forecasting fuel consumption in the past few years. Most of the statistical models widely utilized in the recent literature papers are generalized into two major categories, regression models (RM) and ML models.

As for the development of RM in predicting fuel consumption, various operational and environmental factors of the ship have been taken into accounts to improve the polynomial regression's accuracy to predict fuel consumption with varying speed conditions [64]. Kee et al. [57] proposed a multilinear regression model to analyze the tugboats' service performance to ensure optimum fuel efficiency is met. However, RM has some disadvantages as there is quite a fair bit of an inference made due to ambiguity in collected data and exposure to the effect of sudden spikes and noisy data signals. Moreover, the operational data's noon reports are based on the operator's findings, leading to immense errors in developing the RM.

As data acquisition (DAQ) is well established in collecting real-time operational data rapidly, this eventually led to extensive research exploring ML models' capability for fuel consumption prediction. The ML model's key factors are its distinct benefit of generalizing the relationship between multiple dimensional operational data obtained from DAQ and allowing more accurate predictions than the RM. Literature papers on ML models shown in [65] showed that the ANN ML model is proven to outperform RM. However, achieving a stable ML model takes a substantial amount of time delegated to pre-process raw data to ensure no noise and outliers are captured during the models' training. Therefore, the following sections will study the various machine learning methods in enhancing the prediction of the machine. The flow chart describing the methodology for supervised and unsupervised ML is shown in Figure 16. The major procedures involving the filtering of raw data, creation of score dataset, K-mean clustering and activity labels are described in detail in [61]. The details for unsupervised and supervised ML are in the following sections.

The comparison of the different machine learning techniques is given in Table 6 at the end of this section.
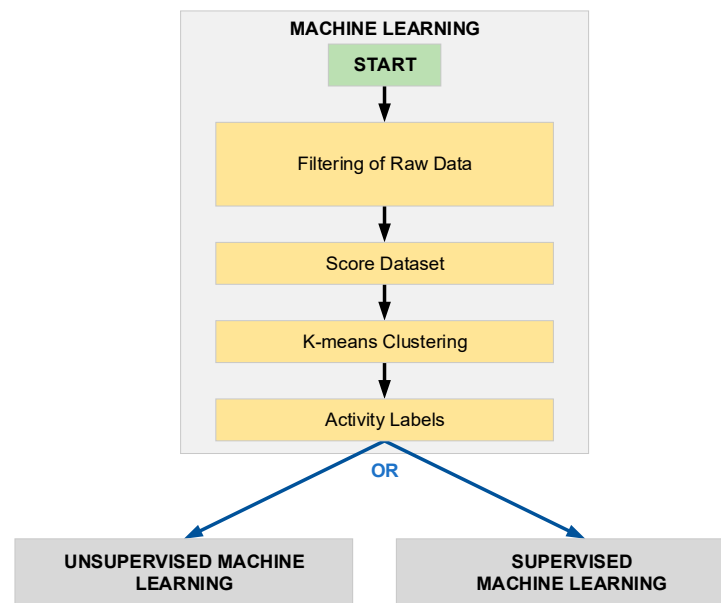
Figure 16. Flow chart for machine learning process [61].

6.1. Supervised Machine Learning

6.1.1. Multi Linear Regression Model (MLR)

The MLR models are simple to understand and incorporate in applications. It necessitates awareness of the input variables that are positively associated with the target variable, which is FOC. It employs the least-squares method to construct a regression line for predictions based on the relationship between the input and output variables. To achieve quality prediction, specific requirements such as multicollinearity and autocorrelation relationships between variables must be avoided when modelling the MLR. Kee et al. [57] suggested using MLR to estimate the fuel consumption of towing tugboats operating between laden and ballast conditions along the Malacca Straits. The primary goal was to build an MLR model capable of predicting recommended vessel speed, thereby enabling the operator to maximize fuel performance. The input factors, i.e., travelled distance, travelled hours, vessel speed, vessel deadweight and wind speed, that influence the fuel consumption are well-established; therefore, the MLR model was able to achieve an R2 score of as high as 0.91, indicating that the output has a major impact on the variables. This model was validated using the fuel consumption analysis method, which offers a ground truth to support the model's prediction capability.

6.1.2. Hidden Markov Model

The hidden Markov model (HMM) uses a Markov Chain in which a certain set of states could be partially observable (hidden) or observable. An example of the Markov Chain with five hidden states (HS), denoted by s_1, s_2, s_3, s_4 and s_5, is shown in Figure 17 [58]. The transition information is quantified in terms of livelihood or the transition probability value, denoted as a_{ij}, where i is the original state and j the subsequent state. The state information could be deduced within the state, i.e., a_{ii} or between two different states, i.e., a_{ij}. The transition probability may be arranged in a Stochastic Matrix known as the Transition Matrix (TM). Similarly, the probability of the Markov Chain for Observable States (OS) could also be obtained, where its Stochastic Matrix is known as the Emission Matrix (EM). The HMM is applicable for a time-series dataset where the HMM utilizes Markov Chain to create a probabilistic correlation between states. By using the HMM, the ML can predict the fuel consumption based on the environmental condition given. A comparison of the prediction data (PR) for fuel consumption of a tugboat with the ground truth (FR) is shown in Figure 18. Note that vs. represents the vessel speed. Figure 18 indicates the correlation between the vessel speed and the fuel consumption, thereby could

be used for assisted decision making on the optimal vs. to achieve fuel efficiency. The fuel consumption profile for a tugboat differs from other harbour craft vessels such as the ferries or patrol vessels, where the latter have their fuel consumption influenced significantly by the vessel speed, i.e., an increase in fuel consumption increases with an increase in vessel speed. On the other hand, the tugboat working operations involve tugging and cruising in which both operations could lead to high fuel consumption. During the tugging operation, there is a significant drop in the vessel speed as shown in Figure 18, but the fuel consumption (FR) remains high. One way to deduce the correlation between the FR and vs. is by taking into consideration the vessel shaft RPM. However, if the vessel shaft RPM is not available, the classification model based on the K-mean clustering method [61] could be used to classify the different operational activities of the tugboat based on their FR and VS. This classification model can then be utilized to train the machine learning model, as described in Section 6.2.2.
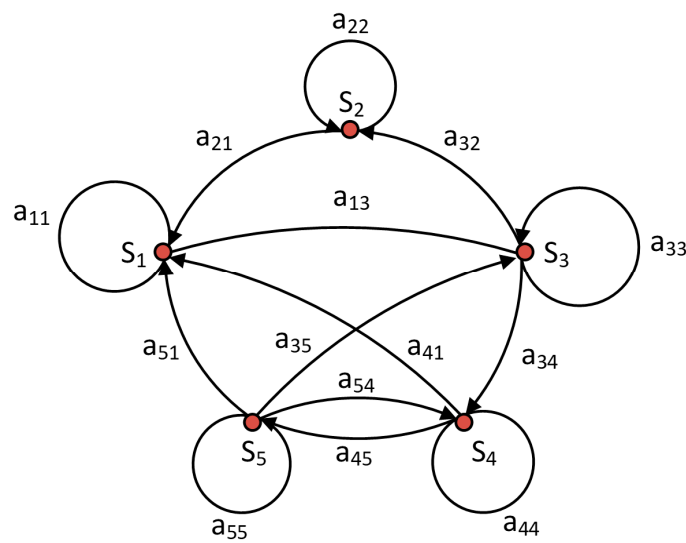

Figure 17. A Markov chain with 5 states (labelled S_1 to S_5) with selected state transitions.
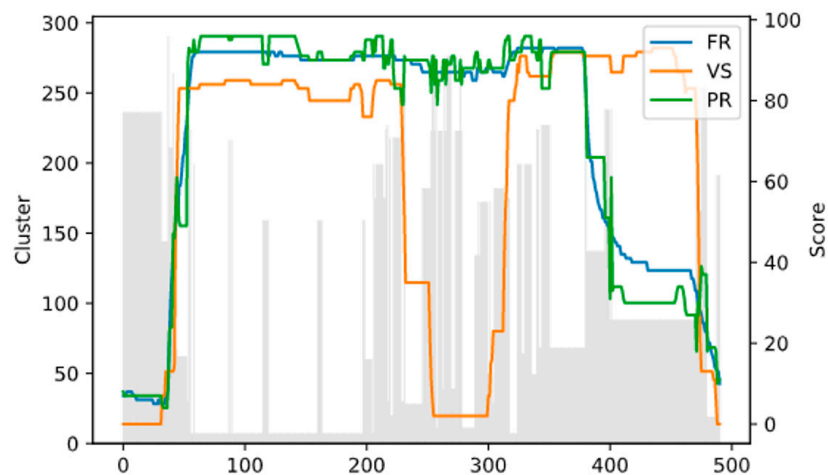
Figure 18. A segment of score dataset, ground truth (FR) vs. prediction (PR).

6.2. Unsupervised Machine Learning

6.2.1. Artificial Neural Network Model (Black Box Model)

The ANN is an ML model based on the black box model (BBM) that works similarly to human neurons in neural network architecture to recognize and decode the dataset's underlying relationship. Furthermore, ANN can learn nonlinear functions that are appropriate

for high fluctuation datasets such as those used in onboard ship measurements to predict the desired outcome. As a result, it is commonly used to improve ship powering performance by forecasting FOC based on a variety of operational data factors. Jeon et al. [25] proposed using the ANN model to analyze and forecast ship fuel consumption based on the table's dataset in [66]. Pre-processing raw operational data is more important than tuning hyperparameters in the neural network to improve prediction accuracy to model the neural network accurately. Furthermore, tuning neural network hyperparameters does not outperform the data pre-processing in terms of enhancing the robustness of the model learning capability. Several works in [25] focused on pre-processing the raw input data. Such pre-processing involves smoothing spline filtering algorithms applied to the entire signal in denoising outliers as shown in Figure 19, and Gaussian mixture model (GMM) clustering techniques shown in Figure 17, with the data parameters listed in Table 3.
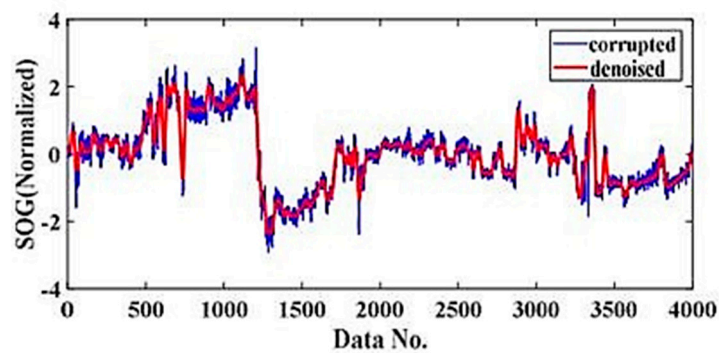

Figure 19. Denoised data for SOG using spline filtering [25].

Table 3. Data information used for ship prediction [25].

Data	Parameter	Remarks
Input	Average draft (m) Trim (m)	Ship State
	ME Power (kW) Shaft Speed (RPM)	Engine operation
	STW (knots) SOG (knots)	Navigation speed
	Relative Wind Speed (m/s)	Weather condition
Output	ME Fuel Consumption (tonnes/day)	Fuel consumption

The clustering technique classifies the operation with a high-frequency signal as a single cluster to indicate that the data cluster had somewhat similar operation states, which might aid the neural network's learning by reducing computation time and focusing the cluster's effect on FOC. The analysis also emphasized the significance of clustering data parameters that significantly influenced the FOC reported in [63] study. A clustering methodology applied to the engine power of a vessel is shown in Figure 20.

The ANN model was trained on post-processed datasets by varying hyperparameters such as the predefined range of hidden layers and neurons within the selected activation feature. The post-processed data is divided into three sets: a training set for learning, a validating set to prevent overfitting, and a testing set to validate model functionality. Through this method, the model will arrive at a converging solution in which raising the hyperparameter hidden layers and neurons represented as configuration have no significant impact on model accuracy, as shown in Table 4, and *R* values do not vary much with node increment.
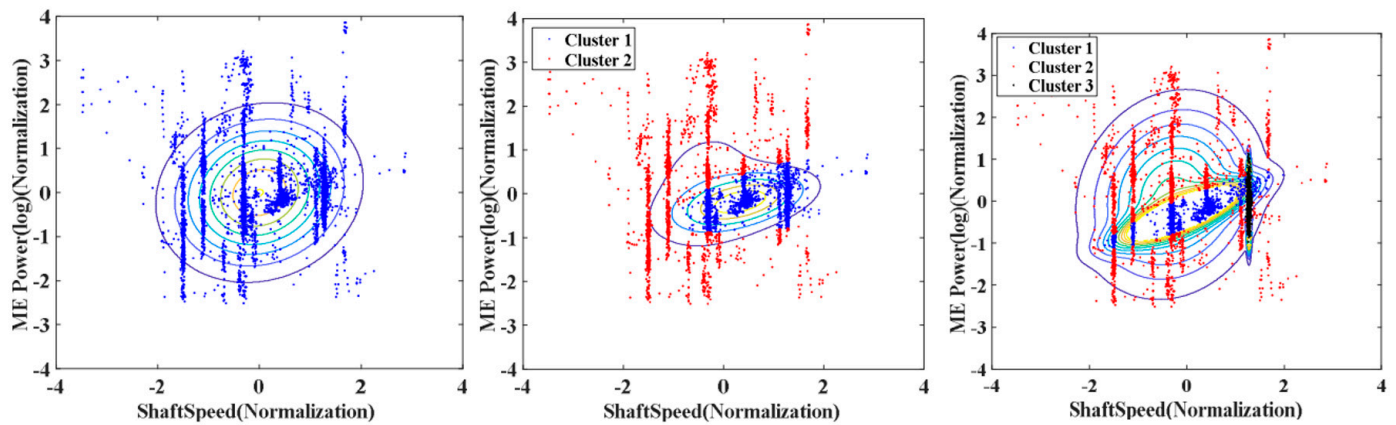
Figure 20. Clustering of data using GMM [25] 6.2.2. Grey Box Model (Hybrid Black Box–White Box Model).

Table 4. Variation in configuration affecting *R* values [25].

No.	MSE	Iteration	Configuration	*R* Values
S10	0.1963	29	7-4-3-1	0.9118
S11	0.1711	30	7-4-4-1	0.9235
S12	0.1648	33	7-4-5-1	0.9309
S13	0.1799	30	7-4-6-1	0.9227
S14	0.1808	32	7-4-7-1	0.9317

Table 5 shows the comparison of the *R* values and computation time for three different ML techniques, i.e., ANN, regression method (RM) and support vector (SV) summarized in [25]. Table 5 shows that in general, the ANN outperformed the RM and SV with the exponential sigmoid function having the highest prediction accuracy (*R* value). In comparison, the SV is able to predict relatively higher accuracy than the RM method with the quadratic SV method having the higher accuracy. However, the *R* values for RM and SV are very small, i.e., lesser than 0.5, and the computational time for the quadratic SV method is significantly high compared with other methods compared in the table.

Table 5. Models Performance [25].

Models	Customized Function	*R* Values	Time (s)
ANN	Exponential sigmoid	0.9636	0.5200
	Tangent sigmoid	0.9383	0.4585
	ReLU	0.8790	0.5439
Regression Method	Linear	0.0449	0.1167
	Interaction	0.2862	0.1892
	Pure Quadratic	0.1526	0.1425
	Full Quadratic	0.3582	0.2406
Support Vector	Gaussian RBF	0.0813	0.4109
	Linear	0.0071	6.5650
	Quadratic	0.4810	380.9915

In some cases, where there is a lack of data collected from the ships and environment due to difficulty in data collection or incomplete dataset, the grey box model (GBM), which is a hybrid of a white-box model (WBM) and BBM, could be utilized in the data prediction. To predict the fuel consumption of a tugboat, for instance, the BBM is trained by using historical operational data to forecast the ship FOC. To improve the accuracy in the data prediction, one of the possibilities is to include the ship resistance as part of the data training. However, as the ship resistance is not easy to obtain (recorded) by sensors, this could be carried out by using a WBM where the ship resistance under various sea

conditions could be obtained via numerical simulation or experimental test. Thus, the inclusion of operational data and simulated ship resistance used in the ML produce the GBM that may increase the accuracy of the predicted data.

Coraddu et al. [29] conducted a study to examine the capability of the GBM for a Handymax product tanker. The research considered the parametric calculation to forecast ship resistance in calm water for the WBM. Although the calculated outcome is marginally off compared with its counterpart predicted by the computational fluid dynamic (CFD) model, the slightly inaccurate resistance data complement the historical operational data to optimize the model performance to a greater extent. Due to the domain knowledge of the vessel characteristics preserved by the WBM, the GBM can achieve an equally high precision as the BBM in predicting FOC with limited historical operational data available. Figure 21 shows the comparison of the least mean per centage absolute error (MAPE) between the GBM and BBM. It can be seen that the MAPE reduces significantly for the Naïve (N)-GBM and Advanced (A)-GBM, thus implying that the WBM helps in increasing the accuracy of the predicted data.
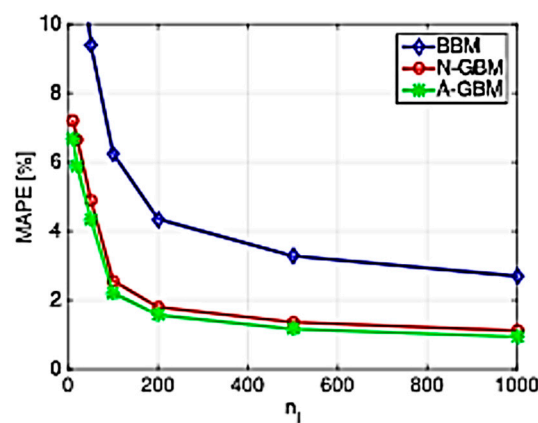

Figure 21. Comparison of performance between GBM and BBM [29]. **Note:** 1. N-GBM: output of WBM is used as a new feature that the BBM can use for training the model. 2. A-GBM: regularization process is changed to include some a priori information.

6.2.2. Long Short-Term Memory Model

The long short-term memory (LSTM) model networks (see Figure 22) are well-suited to classifying, processing and making predictions based on time-series data that have lags of unknown duration between important events in a time series [67]. The LSTM architecture comprises three distinct gates, namely the forget gate (A), input gate (B) and output gate (C). Unlike the ANN model, the LSTM model has its specified embedded activation neural network layers where the number of neural network layers could be added repeatedly. The LSTM model is effective when knowledge about the previous values has a substantial effect on the present values [68].
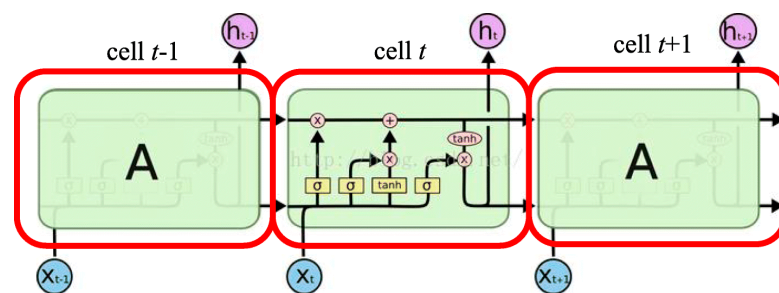

Figure 22. LSTM network architecture model [69].

For datasets with missing values or near zero data such as those shown in Figure 18, the ensembled method such as the LSTM is useful in regenerating the input data. The authors have considered the LSTM model that combines the time-series data collected from tugboat with classification data obtained from a K-mean clustering process. Their research outcome found that the combined model has a better capability in forecasting fuel consumption by using lesser historical data with a faster convergence duration, Table 6.

Table 6. Advantages and disadvantages of different machine learning techniques [70–73].

Filtering Methods	Advantages	Disadvantages
MLR	• Simple to implement • Dimensionality reduction • Regularization and cross-validation could be used to increase accuracy	• Outliers can have huge effect on the regression • Assumes linear relationship between attributes
HMM	• Efficient learning algorithm • Able to handle inputs of variable length	• Contain large number of unstructured parameters • Unable to express dependencies between hidden states • Unable to capture higher order correlation • Viterbi algorithm is computationally expensive
ANN	• Ability to predict with incomplete knowledge • Less sensitive to incomplete (corrupted) data • Capable for parallel processing	• Performance depending on computer processing power • Black box nature (hard to interpret) • Heavy depending on trial and error • Computational time is hard to estimate
GBM	• Considers the prior knowledge of the target system • Applies intelligent statistical technique • Complement the limitation of both WBM and BBM	• Inherit drawback from both WBM and BBM • Relatively more complex as it takes data from the WBM and BBM
LSTM	• Learning from stored information from previous time-step • Able to deal from problems with missing data • Less sensitive to tuning parameters such as learning rate • Able to handle the vanishing gradient problem	• Requires high memory bandwidth • Inaccurate with large weight initialization • Prone to overfitting

An example of the utilization of LSTM model in predicting the fuel consumption is given in Figure 23. Figure 23 compares the accuracy of the LSTM method in terms of $R2$ score and shows that an $R2$ of up to 0.94 could be achieved when a four-month fuel data is used in training the ensembled system. It is to note here that the accuracy in predicting the fuel consumption drops when it is used to predict the consumption for larger future values. This accuracy could be further improved when a combined LSTM model that takes input from the time-series operational data, i.e., fuel rate, vessel speed and wind effect are used with the classification model (see Figure 24). The classification model is obtained from the K-mean clustering method [60] where the operational activity is clustered based on its various states. The utilization of the time-series operational data and classification model in the LSTM model allows a more accurate prediction of the fuel consumption where an $R2$ close to 1.0 could be achieved as shown in Figure 24. The LSTM model, therefore,

outperforms the HMM and ANN in the fuel prediction when there are missing data in the dataset collected from the sensors.
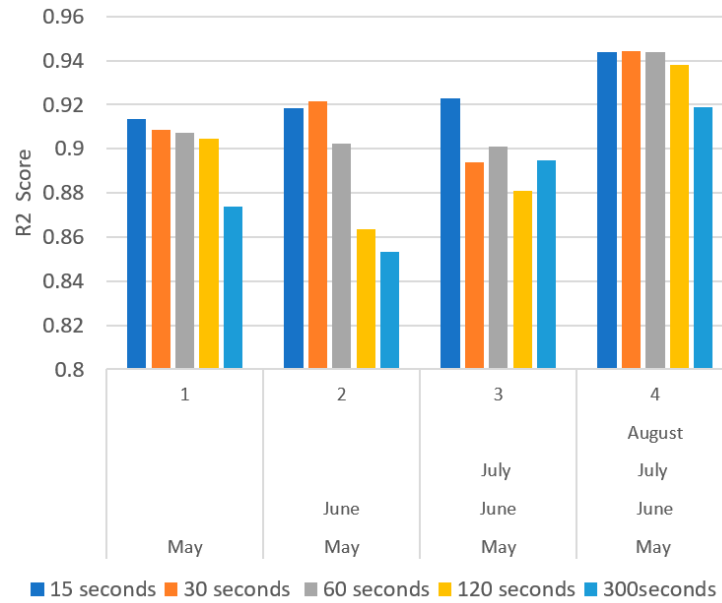

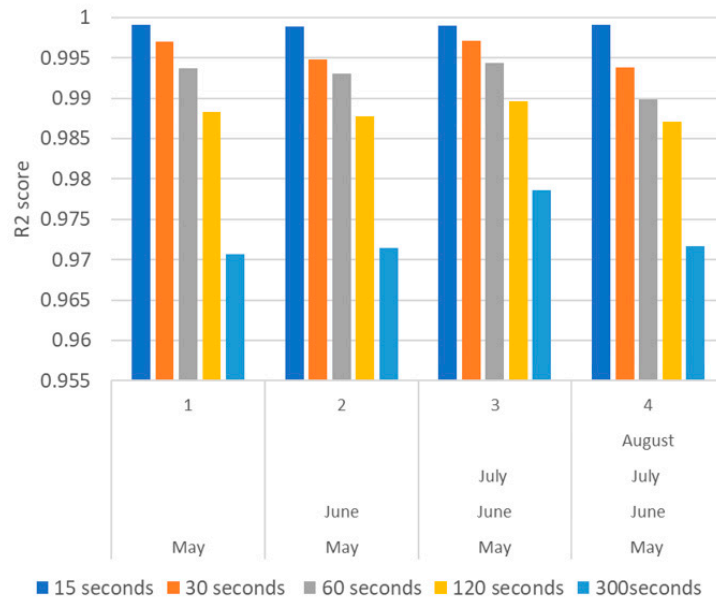**Figure 23.** Non-combined LSTM model prediction accuracy comparison.

Figure 24. Combined LSTM model prediction accuracy comparison.

7. Conclusions

This paper reviewed the big data analytics and machine learning techniques applied to harbour craft vessels with the aim to achieve ship energy efficiency. The numerous filtering techniques, i.e., CCT, HWT, FFT and KF, used in filtering data collected from HCV are presented where simple filtering technique such as the HWT is suitable for filtering fewer complex data whereas data with varying signal frequencies could be effectively filtered out by the FFT. The machine learning technique classified into the supervised and unsupervised techniques, were also presented where their pros and cons are compared. These machine learning techniques could be used in predicting fuel consumption given the environmental loadings. The supervised technique considers the MLG model and the

HMM where the former is simple to use for data with fewer variables, whereas the latter uses the probabilistic correlation between different states to predict the fuel consumption. In unsupervised machine learning, the ANN, GBM and LSTM models are considered. The GBM utilized simulated data to achieve higher accuracy in the prediction compared with the ANN. The LSTM is well-suited for classifying, processing and making predictions based on time-series data with an unknown duration between important events in a time series where the deep learning LSTM combined with autoencoder ensemble learning outperforms the conventional machine learning methods such as ANN. The autoencoder ANN is useful for analyzing fuel consumption data obtained from tugboats, which are prone to missing data and a drop in vessel speed during tugging operation, in which the autoencoder is able to regenerate the input data by encoding sets of data collected.

Author Contributions: Conceptualization, Z.Y.T.; methodology, J.H. and D.J.L.; validation, J.H. and D.J.L.; writing—original draft preparation, F.C.; writing—review and editing, Z.Y.T.; visualization, Z.Y.T. and F.C.; supervision, Z.Y.T. and D.K.; project administration, Z.Y.T.; funding acquisition, Z.Y.T. and D.K. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. United Nations Framework Convention on Climate Change (UNFCC). *Adoption of the Paris Agreement—Paris Agreement*; United Nations: New York, NY, USA, 2015.
2. Saul, J.; Chestney, N. UN Shipping Agency Reaches Deal to Cut CO_2 Emissions | Reuters. Available online: https://www.reuters.com/article/us-imo-emissions-idUSKBN1HK20F (accessed on 24 August 2021).
3. KPMG International. *The Pathway to Green Shipping*; KPMG International: Berlin, Germany, 2021.
4. Pearce, F. How 16 Ships Create as Much Pollution as All the Cars in the World. Available online: http://www.dailymail.co.uk/sciencetech/article-1229857/How-16-ships-create-pollution-cars-world.html (accessed on 21 October 2021).
5. International Maritime Organisation (IMO). *Adoption of the Initial IMO Strategy on Reduction of GHG Emissions from Ships and Existing IMO Activity Related to Reducing GHG Emissions in the Shipping Sector*; International Maritime Organisation (IMO): London, UK, 2018.
6. Florian Frese. Shipping Emissions and 6 Strategies to Avoid Maritime Pollution. Available online: https://container-xchange.com/blog/shipping-emissions/ (accessed on 24 August 2021).
7. International Maritime Organisation (IMO). Energy Efficiency Measures. Available online: https://www.imo.org/en/OurWork/Environment/Pages/Technical-and-Operational-Measures.aspx (accessed on 21 October 2021).
8. International Maritime Organisation (IMO). *Guidelines for Voluntary Use of the Ship Energy Efficiency Operational Indicator (EEOI)*; International Maritime Organisation (IMO): London, UK, 2009.
9. DNV GL. *DNV Group Technology & Research, White Paper 2020: Ammonia as a Marine Fuel DNV*; DNV GL: Bærum, Norway, 2020.
10. ECSA. The Race to Zero Emission | ECSA. Available online: https://www.ecsa.eu/resources/race-zero-emission (accessed on 24 August 2021).
11. Anwar, S.; Zia, M.Y.I.; Rashid, M.; de Rubens, G.Z.; Enevoldsen, P. Towards ferry electrification in the maritime sector. *Energies* **2020**, *13*, 6506. [CrossRef]
12. Viola, I.M.; Sacher, M.; Xu, J.; Wang, F. A numerical method for the design of ships with wind-assisted propulsion. *Ocean Eng.* **2015**, *105*, 33–42. [CrossRef]
13. Traut, M.; Gilbert, P.; Walsh, C.; Bows, A.; Filippone, A.; Stansby, P.; Wood, R. Propulsive power contribution of a kite and a Flettner rotor on selected shipping routes. *Appl. Energy* **2014**, *113*, 362–372. [CrossRef]
14. Cariou, P. Is slow steaming a sustainable means of reducing CO_2 emissions from container shipping? *Transp. Res. Part D Transp. Environ.* **2011**, *16*, 260–264. [CrossRef]
15. Sima, V.; Gheorghe, I.G.; Subic, J.; Nancu, D. Influences of the industry 4.0 revolution on the human capital develpment and consumer behavior: A systematic review. *Sustainability* **2020**, *12*, 4035. [CrossRef]
16. Rolls-Royce Press Release. Press Releases-Rolls-Royce Opens Autonomous Ship Research and Development Centre in Finland –Rolls-Royce. Available online: https://www.rolls-royce.com/media/press-releases/2018/25-01-2018-rr-opens-autonomous-ship-research-and-development-centre-in-finland.aspx (accessed on 26 October 2021).
17. Wärtsilä Successfully Tests Remote Control Ship Operating Capability. Available online: https://www.wartsila.com/media/news/01-09-2017-wartsila-successfully-tests-remote-control-ship-operating-capability (accessed on 26 October 2021).
18. Mishra, B. The Significance of Machine Learning in Shipping & Maritime. Available online: https://seanews.co.uk/features/the-significance-of-machine-learning-in-shipping-maritime/ (accessed on 24 August 2021).

19. Cheliotis, M.; Lazakis, I.; Theotokatos, G. Machine learning and data-driven fault detection for ship systems operations. *Ocean Eng.* **2020**, *216*, 107968. [CrossRef]

20. Sullivan, G.P.; Pugh, R.; Melendez, A.P.; Hunt, W.D. *Operations & Maintenance Best Practices A Guide to Achieving Operational Efficiency*; US Department of Energy: Energy Efficiency & Renewable Energy: Washington, DC, USA, 2010.

21. Chen, H. Weather Routing: A New Approach. 2009. Available online: http://ww1.jeppesen.com/documents/marine/commercial/Safety_at_Sea.pdf (accessed on 16 November 2021).

22. Gershanik, V. Weather routing optimisation-challenges and rewards. *J. Mar. Eng. Technol.* **2011**, *10*, 29–40. [CrossRef]

23. Barthwal, N.; Agarwala, N. Industry 4.0 in the Shipping Industry: Challenges and Preparedness—The Prevailing Scenario the Digital Revolution in the Shipping Industry. 2019, pp. 1–12. Available online: https://maritimeindia.org/industry-4-0-in-the-shipping-industry-challenges-and-preparedness-the-prevailing-scenario/ (accessed on 16 November 2021).

24. Anish. 17 Pro Tips For Efficient Marine Growth Prevention System (MGPS) On Ships. Available online: https://www.marineinsight.com/tech/marine-growth-prevention-system/ (accessed on 21 October 2021).

25. Jeon, M.; Noh, Y.; Shin, Y.; Lim, O.K.; Lee, I.; Cho, D. Prediction of ship fuel consumption by using an artificial neural network. *J. Mech. Sci. Technol.* **2018**, *32*, 5785–5796. [CrossRef]

26. Munim, Z.H.; Dushenko, M.; Jimenez, V.J.; Shakil, M.H.; Imset, M. Big data and artificial intelligence in the maritime industry: A bibliometric review and future research directions. *Marit. Policy Manag.* **2020**, *47*, 577–597. [CrossRef]

27. Fam, M.L.; Tay, Z.Y.; Konovessis, D. An Artificial Neural Network Based Decision Support System for Cargo Vessel Operations. In Proceedings of the 31st European Safety and Reliability Conference, Angers, France, 19–23 September 2021; pp. 3391–3398.

28. Wei, T.T. Ship Agency and Harbour Craft SMEs Can Tap $3.7m Digitalisation Fund. Available online: https://www.straitstimes.com/singapore/transport/ship-agency-and-harbour-craft-smes-can-tap-37m-digitalisation-fund (accessed on 21 October 2021).

29. Coraddu, A.; Oneto, L.; Baldi, F.; Anguita, D. Vessels fuel consumption forecast and trim optimisation: A data analytics perspective. *Ocean Eng.* **2017**, *130*, 351–370. [CrossRef]

30. Coraddu, A.; Oneto, L.; Baldi, F.; Anguita, D. Vessels fuel consumption: A data analytics perspective to sustainability. In *Soft Computing for Sustainability Science*; Springer: Berlin, Germany, 2018; pp. 11–48.

31. Trelleborg Marine Systems. *Use of Big Data in the Maritime Industry*; Trelleborg Marine Systems: London, UK, 2018.

32. Fujitsu Laboratories Ltd. Fujitsu Develops High-Accuracy Fuel Efficiency Estimates Through a Ship's Operational Data—Fujitsu Global. Available online: https://www.fujitsu.com/global/about/resources/news/press-releases/2016/0510-03.html (accessed on 24 August 2021).

33. Lee, H.; Aydin, N.; Choi, Y.; Lekhavat, S.; Irani, Z. A decision support system for vessel speed decision in maritime logistics using weather archive big data. *Comput. Oper. Res.* **2018**, *98*, 330–342. [CrossRef]

34. Perera, L.P. Handling Big Data in Ship Performance and Navigation Monitoring. In Proceedings of the Smart Ship Technology Conference, London, UK, 24–25 January 2017; pp. 89–97.

35. Oyku. Maersk and Digital Revolution in Shipping Industry—Digital Innovation and Transformation. Available online: https://digital.hbs.edu/platform-digit/submission/maersk-and-digital-revolution-in-shipping-industry/ (accessed on 24 August 2021).

36. Khor, Y.S.; Døhlie, K.A.; Konovessis, D. Optimum speed analysis for large containerships. *J. Ship Prod. Des.* **2013**, *29*, 93–104. [CrossRef]

37. What Is the Definition of Machine Learning? | Expert.Ai | Expert.Ai. Available online: https://www.expert.ai/blog/machine-learning-definition/ (accessed on 26 October 2021).

38. Ozkan, U.C. Machine Learning: How Will It Integrate Into the Shipping Industry? Available online: https://www.morethanshipping.com/machine-learning-will-integrate-shipping-industry/ (accessed on 24 August 2021).

39. Ramesh, K.; Konovessis, D.; Thong, S.K.; You, X. Development of Intelligent Ship Fuel Consumption Algorithms. In Proceedings of the Maritime Technology and Engineering III: Proceedings of the 3rd International Conference on Maritime Technology and Engineering (MARTECH 2016), Lisbon, Portugal, 4–6 July 2016; CRC Press: Boca Raton, FL, USA, 2016; p. 161.

40. Anan, T.; Higuchi, H.; Hamada, N. New artificial intelligence technology improving fuel efficiency and reducing CO_2 emissions of ships through use of operational big data. *Fujitsu Sci. Tech. J.* **2017**, *53*, 23–28.

41. Green Steam. How Machine Learning Tackles Hull Fouling. Available online: https://greensteam.com/articles/machine-learning-and-hull-fouling/ (accessed on 23 May 2018).

42. Petersen, J.P.; Winther, O.; Jacobsen, D.J. A machine-learning approach to predict main energy consumption under realistic operational conditions. *Ship Technol. Res.* **2012**, *59*, 64–72. [CrossRef]

43. Kristensen, H.O.; Lützen, M. Prediction of Resistance and Propulsion Power of Ships. *Clean Shipp. Curr.* **2012**, *1*, 1–52.

44. Chakraborty, S. How The Power Requirement of a Ship Is Estimated? Available online: https://www.marineinsight.com/naval-architecture/power-requirement-ship-estimated/ (accessed on 24 August 2021).

45. Górski, W.; Abramowicz-Gerigk, T.; Burciu, Z. The influence of ship operational parameters on fuel consumption. *Zesz. Nauk. Morska Szczec.* **2013**, 49–54.

46. Circulars, P.M. Mandatory Adoption of Mass Flow Metering System for Distillates Delivery in the Port of Singapore from 1 July 2019. Available online: https://www.mpa.gov.sg/web/portal/home/port-of-singapore/circulars-and-notices/detail/befec4bc-c64e-44ff-b24d-9e330978ea0e (accessed on 24 August 2021).

47. MPA. MPA Extends Mandatory Use of Mass Flow Meters to Distillates Delivery. Available online: https://www.mpa.gov.sg/web/portal/home/media-centre/news-releases/detail/beae3877-f7c6-48f4-9996-9b90928bb331 (accessed on 21 October 2021).

48. MI News Network. Can Coriolis Flow Meters Reduce Bunker Quantity Disputes On Ships? Available online: https://www.marineinsight.com/maritime-law/can-coriolis-flow-meters-reduce-bunker-quantity-disputes-on-ships/ (accessed on 21 October 2021).

49. KRemington. Ultrasonic Wind Sensors or Cup Anemometers? That Is the Question. Available online: https://www.windpowerengineering.com/ultrasonic-wind-sensors-or-cup-anemometers/ (accessed on 24 August 2021).

50. MarineTraffic: Global Ship Tracking Intelligence | AIS Marine Traffic. Available online: https://www.marinetraffic.com/en/ais/home/centerx:104.1/centery:1.2/zoom:11 (accessed on 26 October 2021).

51. Free AIS Ship Tracking of Marine Traffic-VesselFinder. Available online: https://www.vesselfinder.com/ (accessed on 26 October 2021).

52. Thombre, S.; Zhao, Z.; Ramm-Schmidt, H.; Garcia, J.M.V.; Malkamaki, T.; Nikolskiy, S.; Hammarberg, T.; Nuortie, H.; Bhuiyan, M.Z.H.; Sarkka, S.; et al. Sensors and AI Techniques for Situational Awareness in Autonomous Ships: A Review. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–20. [CrossRef]

53. Hodge, V.; Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [CrossRef]

54. Fernandas, F.C.A.; Van Spaendonck, R.L.C.; Burrus, C.S. A new framework for complex wavelet transforms. *IEEE Trans. Signal Process.* **2003**, *51*, 1825–1837. [CrossRef]

55. Nwagwu, H.C.; Okereke, G.; Nwobodo, C. Mining and visualising contradictory data. *J. Big Data* **2017**, *4*, 1–11. [CrossRef]

56. Wedin, O.; Bogren, J.; Igor Grabec. Data filtering methods. *RoadIdea* **2013**, *3*, 45.

57. Kee, K.K.; Lau Simon, B.Y.; Yong Renco, K.H. Artificial neural network back-propagation based decision support system for ship fuel consumption prediction. *IET Conf. Publ.* **2018**, *2018*, 1306. [CrossRef]

58. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [CrossRef]

59. Kim, Y.R.; Jung, M.; Park, J.B. Development of a fuel consumption prediction model based on machine learning using ship in-service data. *J. Mar. Sci. Eng.* **2021**, *9*, 137. [CrossRef]

60. Subasi, A. Epileptic seizure detection using dynamic wavelet network. *Expert Syst. Appl.* **2005**, *29*, 343–355. [CrossRef]

61. Tay, Z.Y.; Hadi, J.; Konovessis, D.; Loh, D.J.; Tan, D.K.H.; Chen, X. Efficient Harbour Craft Monitoring System: Time-Series Data Analytics and Machine Learning Tools to Achieve Fuel Efficiency by Operational Scoring System. In Proceedings of the ASME 2021 40th International Conference on Ocean, Offhsore and Arctic Engineering OMAE 2021, Online, 21–30 June 2021. OMAE2021-62658.

62. Brunton, S.L.; Kutz, J.N. *Data Driven Science & Engineering-Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2019.

63. Jeon, M.; Noh, Y.; Jeon, K.; Lee, S.; Lee, I. Data gap analysis of ship and maritime data using meta learning. *Appl. Soft Comput.* **2021**, *101*, 107048. [CrossRef]

64. Bialystocki, N.; Konovessis, D. On the estimation of ship's fuel consumption and speed curve: A statistical approach. *J. Ocean Eng. Sci.* **2016**, *1*, 157–166. [CrossRef]

65. Gkerekos, C.; Lazakis, I.; Theotokatos, G. Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study. *Ocean Eng.* **2019**, *188*, 106282. [CrossRef]

66. Perera, L.P.; Mo, B. Machine intelligence based data handling framework for ship energy efficiency. *IEEE Trans. Veh. Technol.* **2017**, *66*, 8659–8666. [CrossRef]

67. Lianne and Justin. 3 Steps to Forecast Time Series: LSTM with TensorFlow Keras | Towards Data Science. Available online: https://towardsdatascience.com/3-steps-to-forecast-time-series-lstm-with-tensorflow-keras-ba88c6f05237 (accessed on 10 September 2021).

68. Zhu, Y.; Zuo, Y.; Li, T. Predicting Ship Fuel Consumption Based on LSTM Neural Network. In Proceedings of the 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), Guangzhou, China, 13–15 November 2020; pp. 310–313.

69. Yu, E.; Wei, H.; Han, Y.; Hu, P.; Xu, G. Application of time series prediction techniques for coastal bridge engineering. *Adv. Bridg. Eng.* **2021**, *2*, 1–18. [CrossRef]

70. ML-Advantages and Disadvantages of Linear Regression-GeeksforGeeks. Available online: https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/ (accessed on 25 October 2021).

71. Understanding of LSTM Networks-GeeksforGeeks. Available online: https://www.geeksforgeeks.org/understanding-of-lstm-networks/ (accessed on 25 October 2021).

72. Advantages and Disadvantages of Hidden Markov Model. Available online: https://www.slideshare.net/joshiblog/advantages-and-disadvantages-of-hidden-markov-model (accessed on 25 October 2021).

73. Strengths and Weaknesses of Hidden Markov Models. Available online: https://compbio.soe.ucsc.edu/html_format_papers/tr-94-24/node11.html (accessed on 25 October 2021).