Physics-based shaft power prediction for large merchant ships using Neural Networks

A.I. Parkes[a,*], A.J. Sobey[a], D.A.Hudson[a]

[a]*Fluid Structure Interactions, University of Southampton, University Rd, Southampton SO17 1BJ, England, UK*

Abstract

There are currently over 100,000 merchant ships operating globally. To reduce emissions requires predicting and benchmarking the power they use. This is relatively straightforward for calm conditions but becomes almost impossible in larger waves. Design power predictions for ships in weather are typically derived by applying a 'margin' onto a reference 'calm water power'. This is of questionable accuracy as the techniques available to estimate these 'margins' are inaccurate. To improve the accuracy and flexibility of such predictions this paper investigates the use of neural networks. For this, 27 months of continuous monitoring data are used from 3 vessels of the same design, sampled every five minutes. Multiple network sizes are considered and evaluated to determine the quantity and quality of data required for predictions. A key aspect is determining network architectures optimised not just for accuracy, but that give close relationships between the input variables and shaft power. Predictions are compared to the results of a regression, the conventional tool to determine shaft power from measured full-scale data from ships. The predictions from this network are similar in accuracy to those of standard practices, with an error less than 10%, but the scope for further improvements is large.

Keywords: Machine Learning, Shaft Power Prediction, Function approximation, Physics-based learning, Artificial Neural Networks

1. Introduction

It is estimated that 90% of the world's trade is seaborne, due to the efficiency of shipping as a mode of transport. Despite this efficiency, the sheer volume of trade means that global shipping is responsible for 3.1% of anthropogenic CO_2 emissions [20], equivalent to those of a major indus-

*Corresponding author
Email address: aiplu17@soton.ac.uk (A.I. Parkes)

trialised economy such as Germany or Japan. Despite this shipping is presently outside of the United Nations Framework Convention on Climate Change (UNFCCC [25]) commitments to reduce emissions. Implementation of an effective energy efficiency management (SEEMP [15]) plan for a vessel, as mandated by IMO, requires benchmarking of its fuel usage. Therefore prediction of fuel consumption, based on its power requirement, is extremely valuable. Accurate prediction of a ship's power requirement in different weather conditions is difficult using traditional methods based on model tests and/or numerical analysis [18]. Even more sophisticated methods, such as fitting high frequency operational data with regression curves [11] or comparing to design speeds to produce a weather margin [10] [16], struggle to give accurate predictions which would allow vessels to determine the penalties for travelling a given route.

Traditional techniques for power prediction at the design stage rely on computational analysis of the added resistance due to waves, or on towing tank tests at model scale, see Molland et al [18]. Much of the operational ship performance analysis is presently based on measured data, and focuses on trying to obtain an accurate regression curve fit to the power-speed relationship in calm water to provide a baseline performance [3]. It has advantages in its simplicity, but is time consuming and concentrates purely on the relationship between speed and power, ignoring fluctuations for weather. As it is the industry standard it is chosen as a means of comparison between the developed method and those in regular use for analysing ship performance. In order to derive such a regression fit, it is common to filter out performance data in waves above a certain, arbitrary, height. A choice must also be made on whether to derive the curve for the remaining data set, or whether to also filter for draught and vessel trim. It is extremely difficult to analyse ship performance data in waves using such methods shown by Lakshmynarnyanana[11], where the nature of the regression relationship is not known a priori. Th artificial neural network method (ANN) allows the possibility of deriving a method of predicting ship power based on all of the underlying physical parameters. This predictive model may be used in performance analysis as well as having the capability to be used for weather routing and deriving design margins for future ships.

Recorded data has historically been used for similar estimations, through use of the 'noon report' of the ship's position, wind speed, estimate of sea state and daily fuel consumption taken at noon each day during operation. This type of analysis requires a large number of voyages before the required quantity of data can be collated and also suffers from the averaging inherent in using one data point to represent the operation and weather conditions from a 24hr period. These reports are also reliant on observation and subject to human error. Despite these drawbacks noon reports are currently used to monitor vessel status and operational efficiency. Recent improvements in the ability to collect, store and transmit data allows for analysis of these different variables at a much higher frequency. The extra data can be combined with recent advances in forecasting environmental conditions using hindcast models to provide improved predictions. Dinham-Peren and Dand[4] highlight the potential benefits and some of the problems with using these data to derive performance benefits.

Beyond Dinham-Peren and Dand[4] there are a few other recent attempts to predict ship performance from data measured more frequently. These papers utilise a range of techniques, e.g. Trodden et al.[24], Aldous et al. [1] and Lu et al.[16]. Trodden et al.[24] investigated a method for associating segments of a data-stream with its corresponding ship activity to find the fuel efficiency; the method utilises a number of filtering techniques to determine the activity being performed. To validate the methodology, results from the data analysis of speed over ground are compared to fuel consumption data measured under sea-trial conditions and found to be in close agreement. The analysis of this paper utilises one month's worth of data, constituting 43,143 data-points. Aldous et al.[1] categorises the relevant sources of uncertainty in performance measurement. A sensitivity analysis conducted on the sources of uncertainty highlights the relative importance of each. The two major data acquisition strategies, continuous monitoring and noon reporting, are compared, using 9570 data points, after filtering, taken over 370 days. It was found that the number of observations in the data set has a significant effect on uncertainty, with more data reducing the uncertainty, with the observations taken at either 15 mins (continuous monitoring) or 24 hrs (noon reports). Lu et al.[16] looked at a semi-empirical addition to the method of Kwon[10] to estimate the ship's added resistance considering the specific ship type under varying draughts, speeds, encounter angles, sea states, fouling effect and engine degradation conditions.

Despite these attempts to utilise some of the available data there are limited attempts to apply soft computing or machine learning techniques on data from operational measurements. This is despite the use of Artificial Neural Networks (ANN) in a number of other marine applications, Jain and Deo [9] review the use of neural networks in ocean engineering. They show that the majority of applications of neural networks in ocean engineering are to predict natural variables in specific locations (wind speed and direction, wave height - Hu et al[6] and tide - Lee and Jeng[13]), but that there is some use for predicting non-natural variables like predicting ship parameters - Islam[8] and vessel location - Zissis et al [26]. The majority of papers reviewed by Jain and Deo [9] are simple supervised feed-forward networks with one or two hidden layers and a low number of inputs (with a few exceptions Makarynskyy [17] and Huang et al [7]).

Notable exceptions – applying soft computing to operational measurements - include Pedersen and Larsen [21], Besikci et al. [2] and Radonjic and Vukadinovic[23]. Pedersen and Larsen[21] also used an Artificial Neural Network approach to ship power prediction, looking at predictions over 10min periods, they used a Bayesian learning scheme. Four variables were investigated; ship speed, relative wind speed and direction, air temperature and sea water temperature. The sampling time was every 1 second, but these measurements were inconsistent, sometimes with gaps of more than 10 seconds; power and speed were updated at a different time period, every 13 seconds. Samples with excessive variance in the heading were excluded. The relative error of the predictions was less than 2.7% for the mean propulsive power over 10 min periods. This was seen to be significantly better than empirical or data-driven methods based on towing tank tests (e.g. Holtrop[5]). Besikci et al. [2] predict the fuel consumption of a vessel but use data from 'noon reports'. The parameters considered for fuel prediction are ship speed, revolutions per minute (RPM), mean draft, trim, cargo quantity on board, wind and sea effects, in which output from the ANN is fuel consumption. Only 233 points of data are used with the best prediction being reached with 12 neurons in one hidden layer which provides better performance than multiple regression analysis. Artificial Neural Networks have also been used to predict power for two boats by Radonjic and Vukadinovic[23] but the data used was from full scale trials, not measured from day to day use of a ship. Their results only concentrated on the ship speeds effect on power so predicting ship performance in weather is not possible from this model. The data used to train their networks includes vessel specifications such as length to beam ratio, this means a network trained on one vessels data will never be able to be used on another vessel, a vital application of this method. Importantly the focus of all of these approaches, marine or non-marine, is on the accuracy of the power prediction, but there is limited evidence of understanding how physics dependant these models are.

There are currently limited efforts to use machine learning tools to predict ship power from real data, those that do use only a few input parameters. The focus for the available attempts is on the accuracy of prediction rather than the relationship between inputs and outputs, which will be vital to make the most of these tools. This paper presents an application of machine learning tools on measured ship data to predict shaft power in a range of ship and sea conditions. The focus will be on creating networks which approximate the relationships between inputs and outputs, physics-based, and not solely on the accuracy of the results, like much of the literature. Of the six well documented applications of neural networks to ship propulsion prediction, five ([2] [21] [14] [22] [23]) use one hidden layer with less than 50 neurons. A two layer neural network has also been applied [23], although the number of neurons in the layers is not specified. Previous studies refer to whether a function can be found that gives high accuracy, this does not necessarily imply the network will easily be able to find the real representation as many networks suffer from poor extrapolation [9], perhaps indicating that they have not found the real representation. Shallow networks can memorise data but are poor at generalisation, deeper networks are capable of learning features at various levels of abstraction [19] allowing explicit development of areas of the network to handle the weaker relationships between inputs and outputs. This can improve model generalisation [12] and so it is proposed that the use of larger networks will improve the ability to extrapolate beyond the available input data by becoming more physics-based. Guidance is given on the quantity of data required for this type of analysis and the type of architecture required to give a balance between accuracy while retaining a basis in the underlying physics of the ship's behaviour. The developed method is compared to a regression used on the same dataset, to highlight the differences in the machine learning methods and potential areas where current models might be improved. A method capable of determining the influence of weather on ship power performance allows its use in both weather routing and in providing a correction from measured data in a range of conditions back to a calm water, or reference, condition. The latter may provide more data for analysis of a range of ship operational and design effects, Dinham-

Peren and Dand[4].

2. Artificial Neural Networks

Artificial Neural Networks are collections of neurons that are grouped into layers with weighted connections, with a simple representation shown in Figure 1. Each neuron is connected to other neurons in the network and, each connection carries a unique weighting.A neuron has at least one input from other neurons and a single output which is multiplied by the weighting associated with a given connection, providing inputs to the next neurons in the chain. Explicitly, for one neuron with N inputs, let $\{x_i\}$ be the inputs into the neuron and $\{w_i\}$ the weights on the connections, where $i \in I = \{0, ..., N-1\}$. A bias input, x_0, is also included which is permanently set to either 1 or 0 with a corresponding variable weight value w_0 and an output y. As data is received a neuron receives input values from the neurons it is connected to and the activation of the neuron a is computed,

$$a = \sum_{i \in I} w_i x_i. \tag{1}$$

The output of the neuron is $y = f(a)$, where f is the activation function, which is selected by the user from a range of pre-determined functions, in this study a sigmoid activation function is used. The data was split with 70% used to train the network, 15% to test the final network with and the other 15% was used for 'validation', which is a standard split for these applications. There are a number of different types of neural network but here a back propagation neural network is used that may be divided into 4 broad steps and a brief summary is given:

Step 1: Hyperparameter Selection

The accuracy of the network is dependent on a number of key variables, the main ones are the number of hidden layers and neurons. As the number of units increases more complex relationships can be modelled by the network; the selected variables are shown in table 1 .

Step 2: Training

The training is performed using Backpropogation and scaled conjugate gradient, before which the weights are randomly initialised to between 0 and 1. These outputs are compared to their corresponding measured shaft power to calculate an error. Backpropogation then occurs to improve the accuracy of the net, where the weights connecting the neurons are changed based on the error between the estimate and actual result. This is repeated and as the weights are continually changed, scaled conjugate gradient descent is used to recognise that a set of weights, as close as possible to the optimal set, has been found. This is performed for all of the training data. The process is then repeated for the specified number of epochs, unless the error criterion is reached first.

Step 3: Validation

Validation is performed during the training process to see if it can terminate early as the network is already suitably trained - this saves time and prevents overfitting. The error is calculated at the end of each epoch for the testing and validation set, if the validation error increases for 5 consecutive epochs then training is prematurely stopped. A mean square error goal of 0.001 was set to stop training as an additional control to ensure there was a limit on the computational time but this goal was never reached.

Step 4: Verification – Testing

Once the training process has finished, testing of the net occurs. To do this the 'untouched' testing data is run though the network, no backpropogation occurs, the outputs are then compared to the measured power and the average absolute percentage error is calculated. Large networks with many layers and neurons can overfit a dataset by becoming too sensitive to specific datapoints, having a low training error but a high test error. There is also a danger of having too few layers or neurons in a network, underfitting can occur where both training and testing error are high as the network fails to model the basic relationships within the data. Before testing, a systematic hyperparameter tuning was performed showing that increasing the number of epochs provides limited improvements to the accuracy as the stopping criterion meant that the full number of epochs was seldom used . Automatic stopping after 5 epochs of increasing validation error seemed to provide limited improvements to the accuracy and the values used seem to provide a good balance between accuracy and computational cost.

3. Data Description

The data used for training the network in this study are from three large merchant vessels of the same design ('sister ships') from January 2014 to February 2016, with a varying number of months of data from each vessel.

Most of the vessel movement is repeated routes with occasional variations to the standard, thus covering a large range of geographic locations and recorded weather conditions. The total number of months worth of data is around 27, totalling 120,758 datapoints. Data is recorded once every 5 minutes on board for each variable apart from wave height which is hindcast weather data. Variables that were available for the training of the neural network are:

- GPS ship speed (knots)
- ship speed through the water (knots) (Speed Log)
- wave height (m)
- true wind speed (m/s)
- apparent wind direction (degrees)
- draught (m)
- trim (m)
- heading (degrees)

Training data was used to predict the shaft power (KW) of the vessels. Shaft power is the product of the shaft torque (T) and its angular velocity (ω) which can also be expressed in terms of the RPM of the engine (N).

$$\begin{aligned}\text{Shaft power} &= T\omega \\ &= \frac{2\pi NT}{60}\end{aligned}$$

Shaft power is the measure of how much power the engine transmits to the propeller via the shaft. Measuring this allows the calculation of the engine efficiency and is a more direct measure of power required by the vessel than the quantity of fuel used (where the engine performance affects the resulting power).

The data used shows high variance as the vessels operate in a range of conditions. This is substantially beyond the variation seen in many other power prediction applications such as for wind turbines or road vehicles. This can be seen in the coefficients of variation for the data; the wind speed, wave and shaft power all have coefficients of variation around 0.5 – 0.6 whereas GPS and Log speed have a coefficient of variation of 0.3.

The variation in the data is apparent in Figure 2 as a boxplot is drawn for all the shaft powers recorded at each speed. Figure 2 shows that there is a trend to the majority of the data with the mean power being approximately proportional to the cubic of the speed. The middle, darker, boxes are bounded by their upper and lower quartiles - meaning they contain half the datapoints for each plot. Most of the data is in the 14-20 knot range, with the higher values 16 and above having the highest densities. Lower or higher values than this have a lower density of data, and a lower variance. From around 11knots of ship speed upwards there is a variance of shaft power nearly half the range of observed shaft power. Ship speeds from 8-11knots show even higher variance of shaft power around two thirds the range. It demonstrates the difficulty in creating a relationship between the input variables, such as speed, and power. Much of the data fits these simplified relationships, as speed is so dominant, but will give inaccurate results at other points due to the variation caused by other factors.

Figure 3 shows the wind direction histogram and the average wind speed encountered. It shows that the majority of the time the vessel is facing the apparent wind nearly head on, meaning the vessel is traveling faster than any wind influence or that the vessel is traveling into a head wind. The average wind speed over each interval shows little variation.

The vessels operate in either a loaded or ballast condition. Therefore the draught tends to be around one of two values, figure 4 shows the split at around 0.45 of the maximum draught values. Trim, unlike draught, can be altered by the captain at any time, but it tends to be kept at one of 8 main values with a normalised value of 0.8 dominating.

3.1. Feature Selection

From the available 8 variables feature selection was performed to ensure the best results from training. If too many variables which are poorly correlated to the output variable are used a network can suffer from overfitting. However, not using a variable which includes relevant information not present in any other variable, will reduce the accuracy of the trained network. In addition, a focus of this research is ensuring a good correlation between input variables and outputs, so using only those with a strong influence on the power helps to provide a more direct correlation.

The Spearmans Rank correlation coefficient is used to evaluate correlation between the possible input variables. As figure 5 shows; speed from GPS and speed from the log have the highest correlation with shaft power, but as expected they also have high correlations with each other, showing a high level of redundancy between the variables. In this case speed from the GPS was chosen as the only speed variable as it has the higher correlation to shaft power. Wind speed, draught and wave height were all chosen as variables as they have between 0.5-0.1 correlation with shaft power, so have some effect on the shaft power. Trim shows little correlation with shaft power, but -0.42 correlation coefficient with draught. As the relationship between draught and trim is uncertain - both variables were included for training to see if any new information regarding this relationship could be found.

The final list of variables used for training after feature selection is:

- GPS ship speed (knots)
- wave height (m)
- true wind speed (m/s)
- apparent wind direction (degrees)
- draught (m)
- trim (m)

3.2. Data Processing

Filtering is applied to the data to remove wind speeds above 40knots, removing 74,775 data points (from 120,758 to 45,983). The effects of this filtering are shown in figure 4 where all the variables retain their distribution shape, just with reduced frequencies of data, apart from wind speed which loses its tail. It is interesting to note that there is more variation in the speeds the ship travels at than would be expected.

The final data selected includes all times when the ship is docked, manoeuvering or otherwise stationary. There were also periods where one variable would be constant for an unlikely length of time; it was decided these were due to malfunctions of the equipment. If the shaft power, ship speed, wind speed or wind direction stayed constant for more than a short period the whole section would be removed from the dataset, as it was assumed to be an error in the recording. Also, all of the time steps when the ship was not moving were removed - that is ship speed was equal to 0. This would leave all low, or manoeuvering, speeds, which were included to maximise the quantity of data available to train the network.

Finally, the apparent wind information was converted to true wind speed and true wind angle with the equations 2 and 3.

$$V_{true} = \sqrt{V_{app}^2 + V_{ship}^2 - 2V_{app}V_{ship}} \tag{2}$$

$$\alpha = \arccos \frac{V_{app} \cos \beta - V_{ship}}{\sqrt{V_{app}^2 + V_{ship}^2 - 2V_{app}V_{ship}}} \tag{3}$$

4. Neural Network Parameters

As powering requirements for ships in waves have not been investigated using machine learning techniques before, it is important to provide guidance on the quality and quantity of data required for this application and the optimal network architecture. This section aims to provide an idea for the number of hidden layers and neurons required to provide an accurate assessment in this application. Studies have suggested that 1 hidden layer is sufficient to produce accurate results [22] for this type of data. However, larger networks have been shown to have a number of advantages over shallow networks including feature abstraction, allowing explicit consideration of lower order terms which will allow better approximation of the relationships between variables with lower correlations in Figure 5, and better generalisation capabilities [12]. So, larger networks are explored to investigate how well differently sized trained networks reflect the input-output relationship. A range of 1 – 5 hidden layers are selected to test this hypothesis, each with between 1 and 100 neurons in each layer.

4.1. Numbers of Neurons

Figure 6 and table 2 show the percentage error in a net of varying numbers of layers for 1, 50 and 100 neurons on the test set, except where 5 hidden layers are used and 150 neurons are also investigated.

The results show that error decreases as both numbers of layers and numbers of neurons increase, this is not surprising as the larger complexity of the network allows for a higher-order fit to the data. In figure 6, 100 neurons and 50 neurons create approximately the same error for the numbers of layers tested. The computational time to train the network increases substantially with increasing neurons and layers meaning that if 50 and 100 neurons produce similar results then it is more efficient to use the lower number of neurons; 50 neurons are thus selected for further studies. It appears that an accuracy of 7-9% is easy to achieve using a basic network and limited treatment of the data. However, it is also important to understand the relationship between these results and the input variables.

4.2. Numbers of Layers

The dependence of the performance of the network on the number of hidden layers is investigated. To help with this the testing and training error for different numbers of layers is documented in table 2; it shows that 3 hidden layers create the lowest average error for both testing and training. The percentage error for 4 hidden layers is greater than that for 3, but the error reduces again for 5 hidden layers. For 2-4 hidden layers the simulations show a training error slightly lower than testing error but not significantly which implies that they are not starting to overfit the data. However, in the case where there are 5 hidden layers this difference is already starting to increase, suggesting that at this stage the network is starting to overfit but this does not appear to be statistically significant. For one hidden layer the error is larger for both the training and testing data and it is suggested this is because the data is underfitted. Larger quantities of data should allow a larger number of hidden layers to be used.

To look for signs of over or underfitting, and to determine whether the relationship between the input variables and the power are based on physicsal laws, or whether the neural network provides a function which produces a low error but gives a "black box" approximation which works but is not related to the real world relationship between inputs and outputs. To see the variation of the shaft power prediction for each variable the following process was followed:

1. Train the network,
2. set all but one value to be constant, the mode,
3. cycle the remaining variable from its minimum to maximum recorded values,
4. run the new dataset through the trained network.

Figures 7 - 11 are representations of the relationship between speed and draught corresponding to the networks in table 2. The draught and speed are selected as representative of the other variables, which show similar trends. The shaft power is normalised to preserve the confidentiality of the data.

At 1 hidden layer (figure 7), the speed-power graph shows a curve that is starting to represent the expected behaviour, which is that as ship speed increases shaft power output increases in a non-linear fashion, suggesting the relationships within this data are not too complex as a network with one layer can identify the relationship. The expectation would be that shaft power will increase approximately as the cube of the ship speed, but with variations about this trend caused by wave interference between bow and stern wave patterns, e.g. [18]. Variations in draught and trim will affect the bow and stern wave patterns which will have an effect on the powering of the vessel as well.

At 4 hidden layers (figure 10) the simulations show signs of overfitting, particularly in figure 10 b) where at around 11m draught a sharp change of gradient suggests erroneous extreme data points are skewing the trend, this is not shown in the accuracy of the training and testing data. 5 layers (figure 11) show similar signs of overfitting to 4 layers, and so larger numbers of hidden layers were not investigated. However, neither the percentage errors nor figures of the relationships between the input variables and power show definitive overfitting and due to the size of the data being used to train these networks a larger network would be needed to show this clearly.

Since a network with 1 layer appears to be under-fitting and a network with 4 or 5 layers show some signs of overfitting, the network that best captures the relationships in these data appear to be a 2 or 3 layered network which are used in the further investigations.

Simulations with multiple neurons, for the same number of hidden layers, are shown in figure 12 and figure 13. The aim is to select a number of layers that is robust to the number of neurons, giving similar trends at each simulation, as this implies the network is better at consistently identifying fundamental relationships in the data. It can be seen that the repeatability of the network is better for inputs that the power is more sensitive to, the speed correlates highly with the power and the resulting speed-power curve shows robustness to the network architecture. To develop physics-based networks it is the less sensitive variables that should be concentrated on.

Figure 12 a) and figure 13 a) show similar repeatability and distinct predictions although figure 13 a) does appear to be slightly more reliable. When looking at figure 12 b) and figure 13 b) it appears that the figure 12 b) is mapping the same curve each time, with some small difference, though there is some variance in how deep the trough at a draught of 4-8m is.

For figure 13 b) the results for 30 and 100 neurons per hidden layer have an extra peak at draughts of 7-9m which are not seen in any other simulation, this is a sign that the network is beginning to overfit the data or approximating an unrealistic function. This implies that 2 hidden layers produce more reliable results compared to 3 hidden layers. Even taking into account the fact that table 2 show a 3 layered network as having a lower error than a 2 layered network, figures 13 b) and 12 b) show that overfitting is already starting to occur at 3 layers for the relationship between individual variables and the output, therefore the network was chosen to have 2 hidden layers.

5. Accuracy and Reliability of the Net

5.1. Accuracy: Neural Net vs Regression

Having determined the accuracy and provided guidance on the neural network architecture using a "black box" assessment, the curves will be assessed to determine how much the trends in the variables are based on physical relationships, i.e. reflect the behaviour of the ship. Regression is currently the accepted practice in naval architecture to calculate ship speed-shaft power curves for calm waters from operational data, and prediction in weather is based on regression analysis of model testing [3].

Therefore, the relationships from regression are compared with simulations from the neural networks on the same dataset, shown in figure 14. The data are divided into wind 'bins' and a curve is created unique to that condition from a regression. When creating the simulation using the neural net the wind speed variable is altered to the relevant wind condition. For the three compared wind conditions the two methods, regression and neural networks, map similar relationships until 17 knots of ship speed.

For ship speeds below 17 knots there is a good correlation between the predictions from the neural network and the regression. At ship speeds higher than 17 knots the regression continues to show an increase in shaft power, whereas the neural network tends to flatten. This 'dropping off' occurs at the point at which less data are available and the neural network predicts shaft power based on the data it has seen, therefore predicting power can no go higher than what it has seen previously. The network predicts that the power or speed will not go much higher than the largest recorded value of power in the dataset. The ship does not often operate at speeds above 17knots, which stretches to beyond the designed maximum speed, and so it is less important for the analysis. The tailing off and unpredictability shown in figure 14 is explored further in the next sections.

5.2. Repeatability of analysis

Multiple identical simulations are replicated in order to check the reliability of the network, illustrated in figures 15 and 16. The only change between each simulation is the randomly selected 70% of the data that is used to train the network. A reliable network should map the same line each time, and this is a good indicator that it is the true relationship between the variable and the output. Figure 15 a) shows good repeatability for most ship speeds apart from the higher range from 18 knots and above. One of the simplest explanations for this is that these ships do not spend much time travelling at speeds above 18 knots, therefore less data for these regions leads to less reliable predictions. As well as the simulations giving different results, there is also a drop in power which suggests that the higher ship speeds require less shaft power, which does not reflect the behaviour seen by vessels. In figure 9 (a) the predicted powering is higher for 0 knots than 5 knots, this behaviour is hard to explain in terms of expected vessel behaviour, so is most likely an example of the poor extrapolation of the network. Figure 2 shows the low quantity of data in this region.

To verify this hypothesis figure 15 b) is a histogram of the recorded ship speed data to give a reference of where the majority of the data lies. To quantify the reliability of the network at different ship speeds figure 15 c) shows the variation between the 5 simulations at any point i.e. the more variation between the simulations, the less reliable the prediction is. Figure 15 shows that the peaks in the histogram correspond directly to troughs in the 'range' line and the simulations in 15 a) are more repeatable . This shows that more data gives more accurate and reliable predictions, an unsurprising correlation, but more importantly that above 18 knots of ship speed the lack of data means less reliable and more importantly less accurate predictions. This means that with the current dataset, predictions for ship speed above 18 knots should not be considered, but since this is above standard operating conditions it is not a problem. However, the same issues occur at the lower end of the data set and predictions in these regions might be more valuable, and perhaps improvement in this region would reduce the average error of the prediction to below 8%.

To ensure this phenomenon was not purely observed for the ship speed variable, and therefore the "drop off" and difference in simulations was caused by lack of data at extremities rather than some other reason, the same plots were created for the wind speed, figure 16, where the results are consistent with the earlier analysis. However, there is a larger variation in these simulations, indicating that the power prediction is less sensitive to this variable. Whilst the prediction of the power appears to be good in areas, and there is a relationship between the simulation and the behaviour of the vessel, in more extreme conditions the neural network prediction does not appear to be accurate or repeatable. It is assumed that results skew the accuracy of the overall analysis and therefore the prediction accuracy of the vessel in waves can drop below 8% assuming a longer time period for data collection.

5.3. Quantity of data vs Error in the Net

With the current 2 layer, 50 neuron network an error of around 8% is common. Figure 17 shows that as the amount of data increases, which has been randomly sampled from across the 27 months worth of data provided, the error decreases.

This low error shows that the network is reliable and accurate to predict shaft power for the majority of weather conditions and also shows that the network parameters are correct for the nature of the data it is being trained on. This also implies that a ship only needs to be at sea for 2 months to produce accurate results. However, this is misleading because the data used to train the network is randomly sampled from the total 27 months, meaning this will most likely include a larger spread of conditions than if the months were sampled as consecutive whole months.

For figure 18, multiples of months are sampled whole and consecutively, i.e. the data used for the first point in figure 18 is the first 288 datapoints in the database, equating to the first days' worth of information. For figure 18, the third datapoint on the graph, which represents 3 days' worth of information, should be an average of 20 simulations. However, one of the simulations produced a highly erroneous error of over 2000%, although this does not change the conclusions drawn from this figure, including this number in the average meant that the value was around 1000% and made figure 18 impossible to read. For this reason the 2000% error was taken out of the average for this datapoint. For the same reason of making the figure more legible the scale is capped at 270% when the upper bounds for the error bars for the first 4 datapoints are in fact higher than this. This means the network (and ship) is exposed to a small range of conditions and therefore cannot predict accurately the shaft power for the conditions it has not yet been exposed to. This means the percentage error is initially large and although it does decrease quickly as more data is available and used to train the network, the network is not exposed to nearly as wide a range of conditions as through random sampling. As such, the error is significantly higher, taking more than 3 months to drop below 20%. Figure 17 and 18 only show up to 3 months of data, this is because after this point the decrease in error from 3 – 27 months is regular and slow, giving a final error of 7.8%. Including all 27 months obscures the initial curve so the values after 3 months have not been included to aid clarity.These two figures, 17 & 18, show that the quantity of data can be relatively low to achieve an acceptable accuracy, but that the important thing is that the data contains a good range of conditions to allow accurate prediction. In both figures 17 and 18 the points are averages over 20 simulations and the error bars show maximum and minimum error in the 20 simulations.

It is important to note that as the number of months in figure 18 increases beyond 3 the error continues to decrease and by the time nearly all of the data is used the error has dropped to around 8% - the same as in figure 17. The point at which the error converges for the method used in figure 18 is based purely on when in the 27 recorded months of data the ship encountered sufficient range of conditions. This is a function of the routes the ships operate and the time of year they are in different locations.

6. Discussion

Monitoring ship performance in order to improve efficiency and reduce fuel consumption is becoming more important to tackle emissions from shipping. Using data measured from vessels in operation to establish baseline performance, track changes over time, improve routing and improve design power margins is challenging. Noon report data is insufficient in quantity to allow this quickly and – given the 24 hour reporting frequency – does not readily allow performance in different weather conditions to be established accurately. Continuous monitoring data holds more promise, but most present methods of analysis rely on heavily filtering the data, typically resulting in only 10% of the data set being retained. This tendency towards calm water predictions result in requiring many months of data to establish performance baselines and changes and no insight into performance in waves.

Neural networks have rarely been used for predicting shaft power in marine vessels. In this case, the three regression-based speed curves normally used for this application can be replaced by one neural network trained off of the same training data. The neural network opens possibilities to easily create similar curves for analysis for these lesser understood variables.

Predictions from the neural network are consistent for ranges where there is a high quantity of data and it is shown that extreme ranges of ship speed and wind speed produce less consistent predictions. Collecting more data will ensure more data in the extreme ranges which would aid prediction. An alternative is to 'manufacture' more extreme data by looking at the small amount of extreme data already present and use machine learning tools or simulation tools, like computational fluid dynamics, to generate more datapoints.

The results from this study are compared with other similar attempts to predict vessel powering with neural networks in table 3. The other comparable papers use significantly less training data and smaller, shallower, networks. It is noted that both provide better accuracy but it is difficult to draw too many conclusions from this, as the measurement errors of each data set are unknown. Previous studies pre-filtered their data to remove transient periods where the ship is manoeuvring in or out of port and bucketed datasets to remove the effect of some variables which are assumed to be the reason for the lower variance in the data. Pedersen [21] split their dataset into 4 datasets which are used to train 4 models separately, further decreasing the number of datapoints for training and the range of conditions experienced within that dataset. The networks are only tested within the ranges of their training data so no extrapolation is observed. It may be that changing the frequency of data collection or filtering, like is seen in the other applications, would improve this prediction accuracy but in addition networks are selected to provide a physics-based assessment of the data and this may also reduce the accuracy. Error is being estimated for conditions that rarely occur and reducing the range should increase the accuracy of the prediction, essentially only allowing prediction in conditions where there is sufficient data to do so.

Including the rotational speed of the engine (RPM) as an additional variable in the network may further improve its accuracy. For a given vessel, propeller and operational condition (draught, trim, etc.) in calm water, there is a unique relationship between vessel speed and RPM. In waves, however, the varying torque on the propeller will change the RPM for vessel speed and power. This therefore provides an additional variable related to the influence of sea state on speed and power.

The extent to which this network represents the physics of ship performance is open to debate. As concluded, the network predictions are accurate and the figures shown in the paper follow patterns which can be considered logical. However, for some of the variables used to train the neural network there is little knowledge of how they affect ship powering, for example draught and trim. It is known that draught and trim have a significant influence on the ship powering but the effects of changes in draught and trim separately are still uncertain. Figure 5 shows that trim and draught are more correlated than any other two input variables although both have low correlation with shaft power, suggesting the underlying relationships between these variables are more complicated than standard regressions can identify. It can also be seen in figure 5 that the ship speed has the highest correlation with shaft power, this confirms what can be seen in figures 15 and 16 as speed has the best repeatability of all variables. This shows that the more correlated a variable is to the shaft power, that the model approximates a relationships between the variables and produces an output that is more physics-based. Those variables with a weaker correlation are more liable to be approximated in a manner that favours accuracy.

Ideally comparison to results produced through either scale-model testing in a towing tank, or physics-based computational models of performance such as computational fluid dynamics, would provide complete validation of the adopted approach. However, such results are not readily obtained without considerable expense, or in the case of simulation are of questionable accuracy in their own right.

7. Conclusions

This study applies artificial neural networks to continuous monitoring data for three sister merchant vessels, acquired during normal operational service. It is critical in a ship performance monitoring application to retain a representative relationship between input variables and outputs, whilst also achieving the highest possible prediction accuracy. Most applications of neural networks in this field to date lack attention to understanding the relationships between input and output. This study suggests an approach that is suitable for both of these aims; to 'fit' underlying relationships without prior knowledge of them and to attain a high prediction accuracy. In this case the network that is best for accuracy shows a similarly good response in terms of mapping the relationships for each variable and performing as well as equivalent regression methods. However, these relationships are strongest in relation to the variables that correlate best with the power, while those that the power is less sensitive to are not as well mapped. Further work is required to better define whether the functions mapped by the neural network echo the physical relationships between inputs and outputs. Selecting or training for these networks will improve their transfer learning abilities, including extrapolation, as the response will be physics-based, meaning that the input-output combination, or similar, need not have been available in training.

The results show that a simple backpropogation neural network with 2 layers and 50 neurons in each layer can predict the power with an accuracy of 8% and shows good repeatability of the relationships between the input variables and the measured shaft power. Such a network may be capable of providing a baseline for performance monitoring across a wide range of environmental conditions, thus allowing faster decision-making. The method may therefore also be of use in improving weather routing and establishing a power margin for newbuilds.

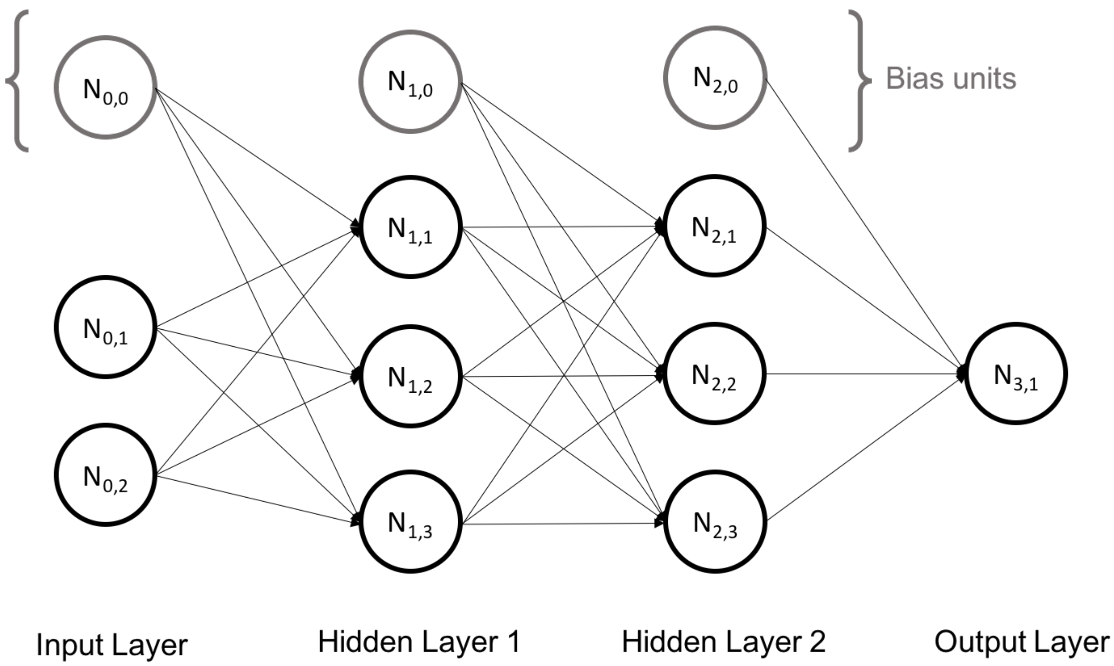


Figure 1: A neural network with 2 hidden layers and 3 neurons in each layer

Table 1: Selected hyperparameters

Hyperparameter	Value
Number of epochs	1000
Goal	0.001
Maximum number of validation failures	100
Performance function	mean squared error
Training algorithm	scaled conjugate gradient
Early stopping patience	5

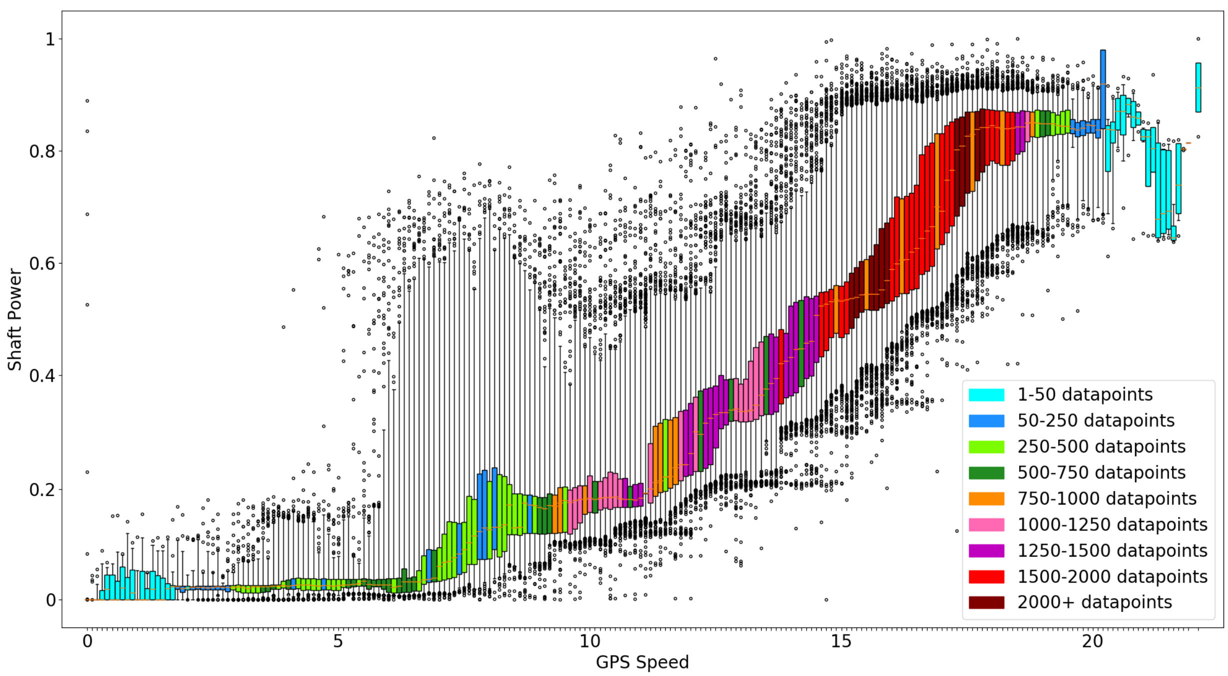


Figure 2: Box plots for every recorded value of shaft power at each recorded speed

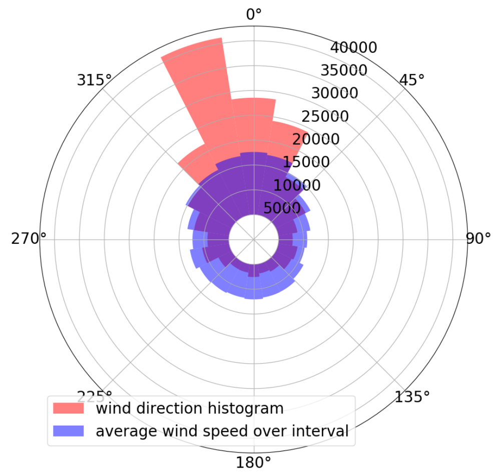


Figure 3: Wind direction histogram and average wind speed in each histogram bin

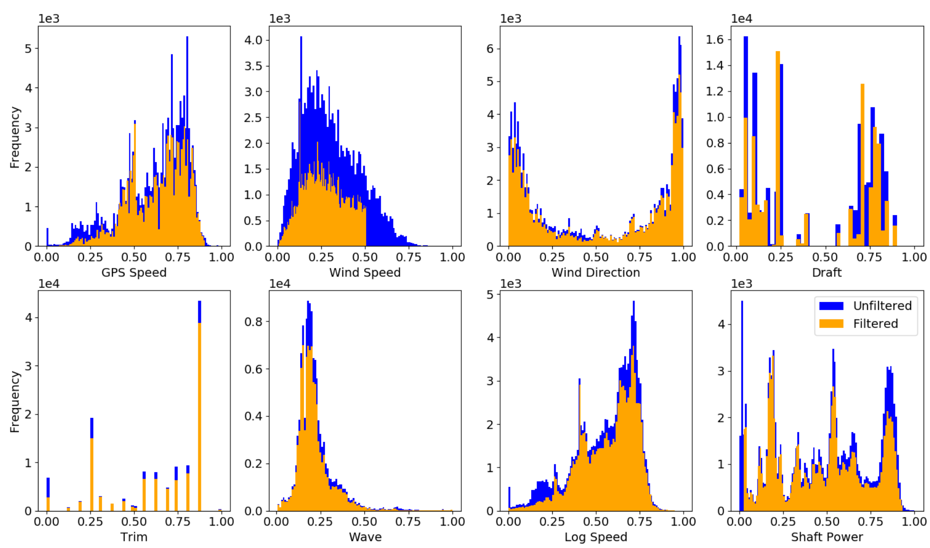


Figure 4: Histograms of all variables before and after filtering

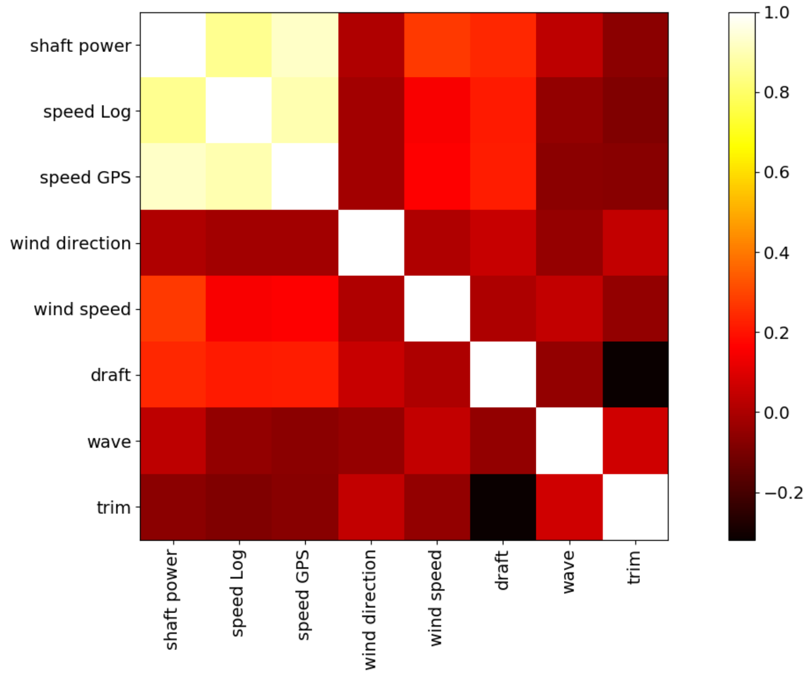


Figure 5: Heatmap of the Spearmans' rank correlation coefficient between all variables.

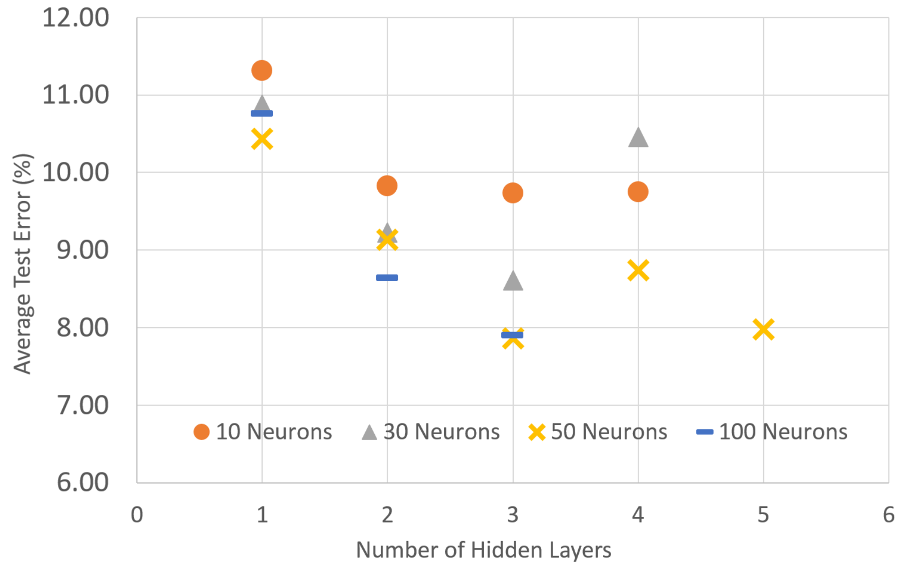
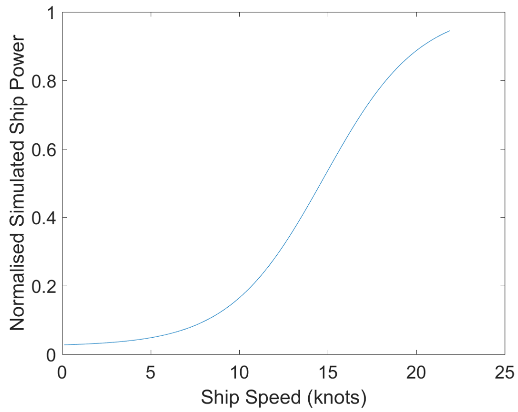


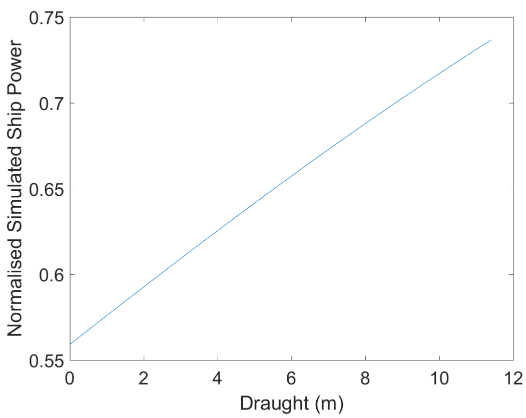
Figure 6: Plot showing number of hidden layers vs percentage testing error

Table 2: Table showing test and training error of various sized networks.

Layers	Neurons	Training Error(%)	Testing Error(%)
1	1	14.42	14.31
1	50	10.36	10.44
1	100	10.54	10.76
2	1	14.17	14.11
2	50	8.96	9.13
2	100	8.69	8.65
3	1	13.99	14.01
3	50	7.83	7.86
3	100	7.77	7.91
4	1	14.35	14.33
4	50	8.70	8.74
4	100	8.26	8.27
5	1	13.95	13.94
5	50	7.86	7.98
5	100	8.23	8.30
5	150	7.83	8.02

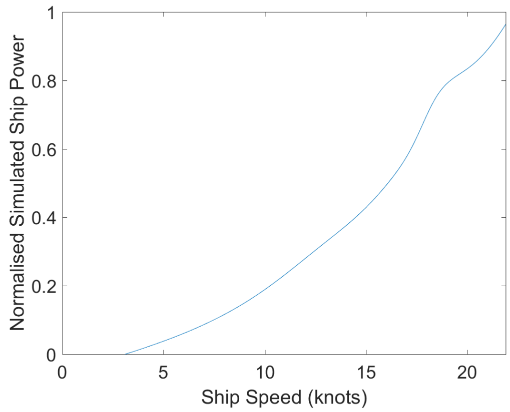


(a)

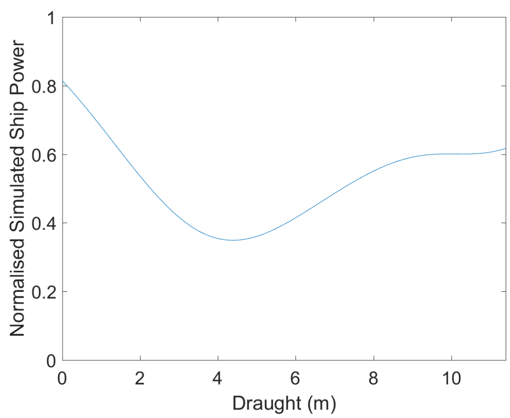


(b)

Figure 7: Simulation with 1 hidden layer and 1 neuron a) isolated speed and b) isolated draught

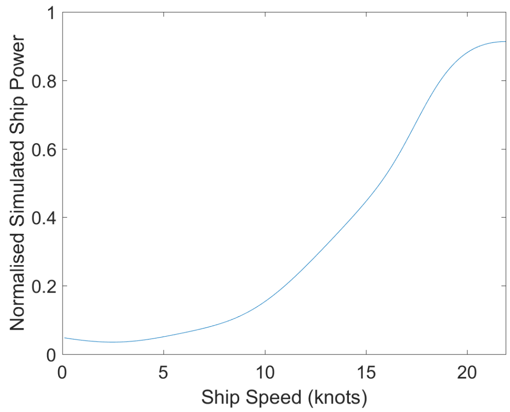


(a)

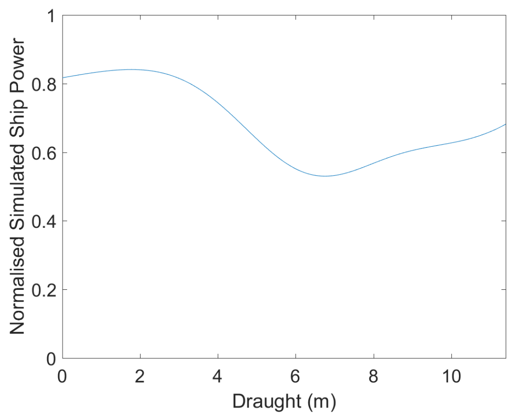


(b)

Figure 8: Simulation with 2 hidden layers and 50 neurons a) isolated speed and b) isolated draught

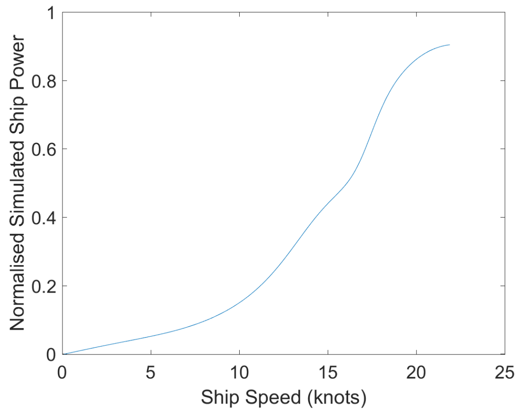


(a)

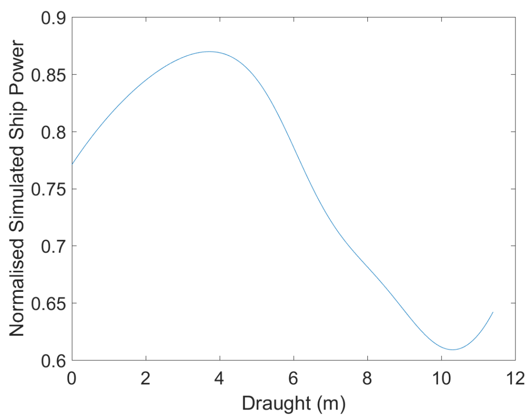


(b)

Figure 9: Simulation with 3 hidden layers and 50 neurons a) isolated speed and b) isolated draught

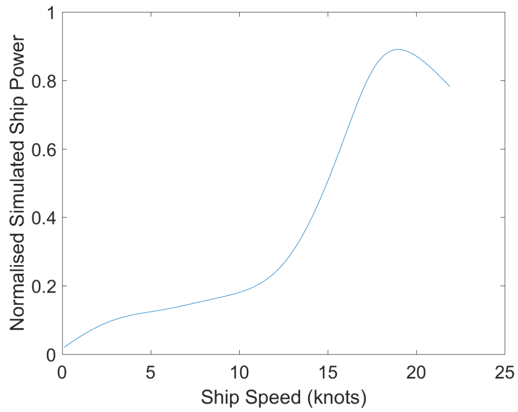


(a)

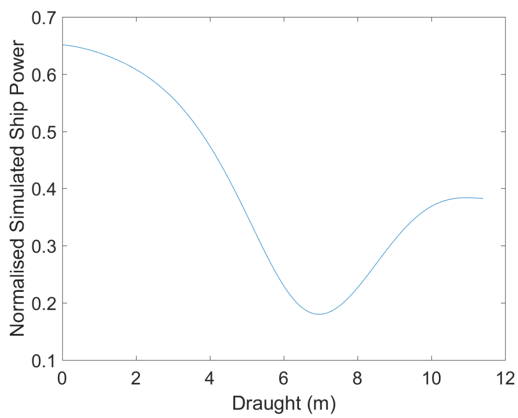


(b)

Figure 10: Simulation with 4 hidden layers and 50 neurons a) isolated speed and b) isolated draught

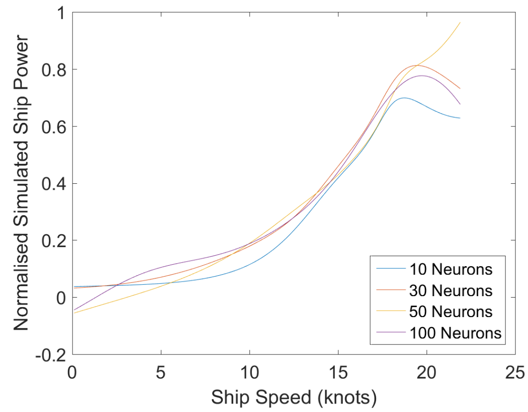


(a)

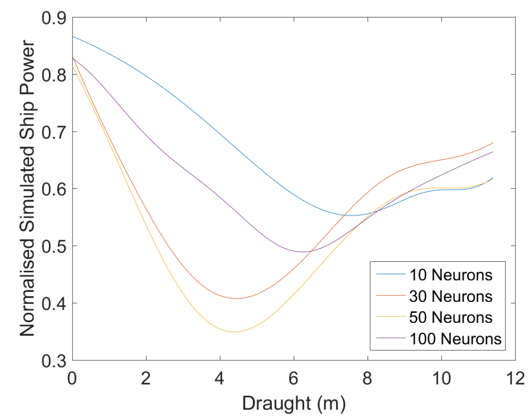


(b)

Figure 11: Simulation with 5 hidden layers and 150 neurons a) isolated speed and b) isolated draught

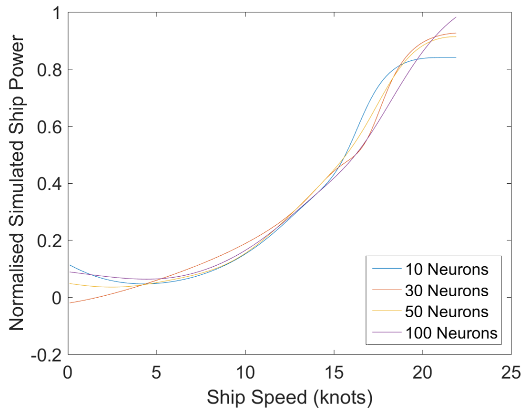


(a)

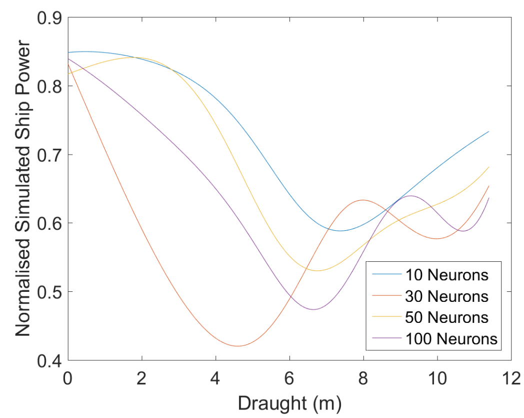


(b)

Figure 12: Simulations with 2 hidden layers and a range of neurons, a) speed, b) draught

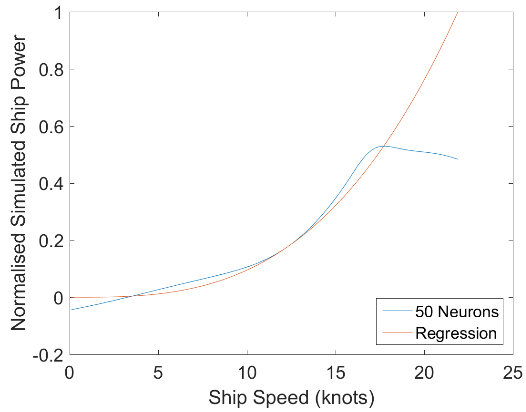


(a)

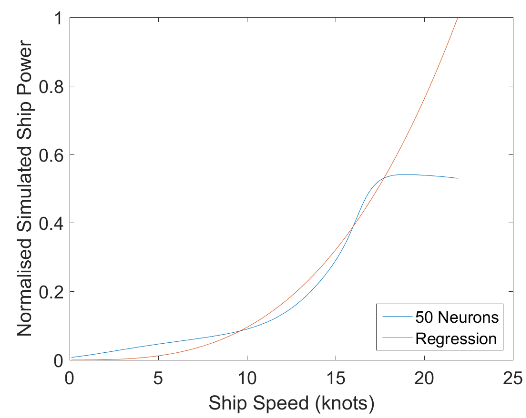


(b)

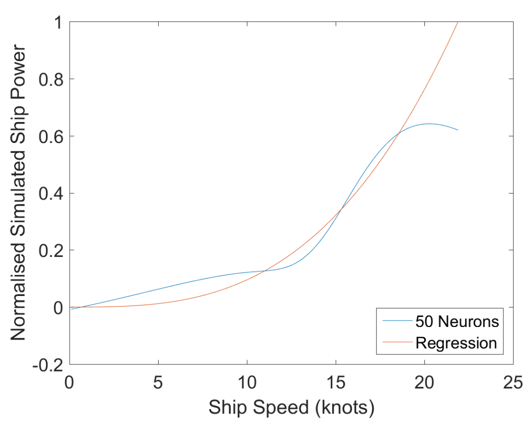
Figure 13: Simulations with 3 hidden layers and a range of neurons, a) speed, b) draught



(a)

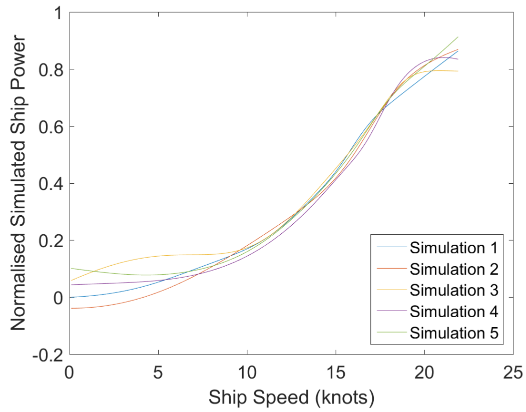


(b)

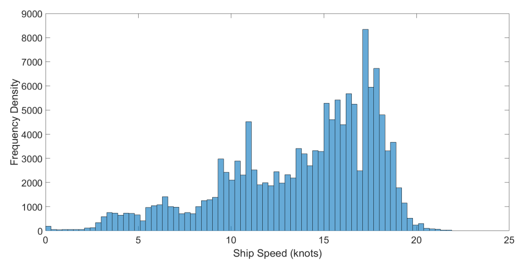


(c)

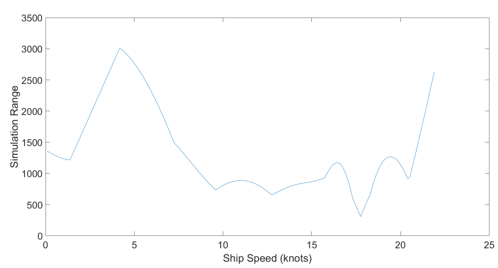
Figure 14: Neural network compared to a regression predicting ship speed for a) high wind speeds (20-30 knots), b) mid wind (10-20 knots) speeds and c) calm water (0-10 knots)



(a)

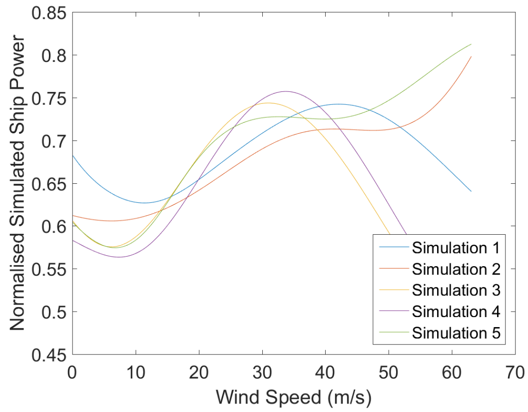


(b)

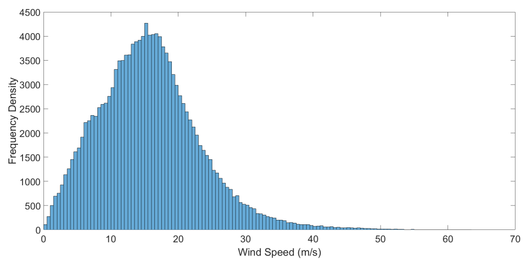


(c)

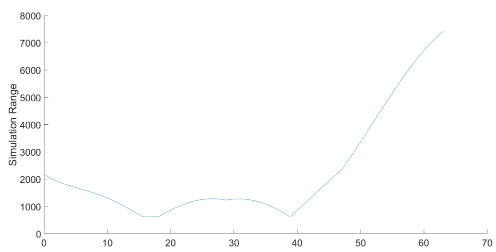
Figure 15: Repeatability of power prediction from ship speed for a) prediction, b) histogram of ship speeds and c) variation between simulations



(a)



(b)



(c)

Figure 16: Repeatability of power prediction from wind speed for a) prediction, b) histogram of wind speeds and c) variation between simulations

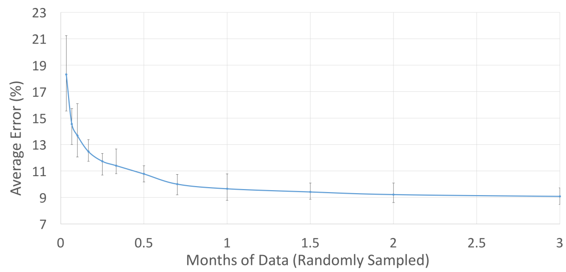


Figure 17: Randomly sampled months of data vs percentage error.

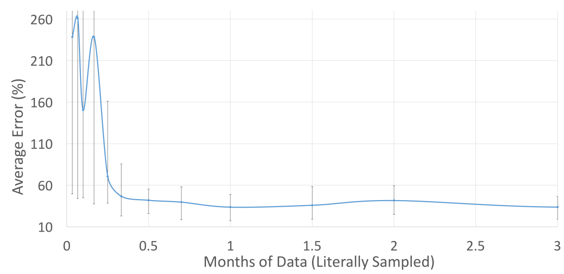


Figure 18: Consecutively recorded months of data vs percentage error.

	Coefficient of Variation		Frequency	Quantity	Network	Relative
	Ship speed	Powering		of Data	Size	Error(%)
LNG dataset	**0.3**	**0.59**	**5 mins**	**45,983**	**3L 50N**	**7.8**
Pedersen [21]	0.006	0.001	10 mins	679	1L 5-10N	2.7
Bal Besikci [2]		0.26	24 hours	233	1L 12N	6

Table 3: The coefficients of variation for ship speed and powering and quantity of data in the Shell dataset compared to other datasets used for neural network applications

[1] Aldous, L., Smith, T., Bucknall, R. and Thompson, P. [2015], 'Uncertainty analysis in ship performance monitoring', *Ocean Engineering* **110**, 29–38.

[2] Bal Beşikçi, E., Arslan, O., Turan, O. and Ölçer, A. [2016], 'An artificial neural network based decision support system for energy efficient ship operations', *Computers & Operations Research* **66**, 393–401.

[3] Carlton, J. [2012], *Marine Propellers and Propulsion*.

[4] Dinham-Peren, T. and Dand, I. [2010], 'the Need for Full Scale Measurements', *RINA Conference, William Froude* pp. 1–13.

[5] Holtrop, J. [1984], 'A statistical re-analysis of resistance and propulsion data'.

[6] Hu, Q., Zhang, R. and Zhou, Y. [2016], 'Transfer learning for short-term wind speed prediction with deep neural networks', *Renewable Energy* **85**.

[7] Huang, W., Murray, C., Kraus, N., Rosati, J. and Huang, W. [2003], 'Development of a regional neural network for coastal water level predictions', *Ocean Engineering* **30**, 2275–2295.

[8] Islam, M. [2001], 'Application of Artificial Intelligence Techniques in automatic hull form generation', *Ocean Engineering* **28**, 1–6.

[9] Jain, P. and Deo, M. C. [2006], 'Neural networks in ocean engineering', *Ships and Offshore Structures* **1**(1), 25–35.

[10] Kwon, Y. [2008], 'Speed loss due to added resistance in wind and waves', *Nav. Archit.* **3**, 14–16.

[11] Lakshmynarayanana, P.A. Hudson, D. [2017], 'Estimating added power in waves for ships through analysis of operational data', *HullPIC* .

[12] Lawrence, S., Giles, C. L. and Tsoi, A. C. [1996], 'What Size Neural Network Gives Optimal Generalization Convergence Properties of Backpropagation', *Networks* (UMIACS-TR-96-22 and CS-TR-3617), 1–37.

[13] Lee, T. and Jeng, D. [2002], 'Application of artificial neural networks in tide-forecasting', *Ocean Engineering* **29**(9), 1003–1022.

[14] Leifsson, L. T., Sævarsdóttir, H., Sigurdsson, S. T. and Vésteinsson, A. [2008], 'Grey-box modeling of an ocean vessel for operational optimization', *Simulation Modelling Practice and Theory* **16**(8), 923–932.

[15] Lloyd's Register [2012], 'Implementing a Ship Energy Efficiency Management Plan (SEEMP) - Guidance for shipowners and operators', pp. 1–4.

[16] Lu, R., Turan, O., Boulougouris, E., Banks, C. and Incecik, A. [2015], 'A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping', *Ocean Engineering* **110**(July 2014), 18–28.

[17] Makarynskyy, O. [2004], 'Improving wave predictions with artificial neural networks', *Ocean Engineering* **31**(5-6), 709–724.

[18] Molland, A. F., Turnock, S. R. and Hudson, D. A. [2011], *Ship Resistance and Propulsion*.

[19] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R. and Muharemagic, E. [2015], 'Deep learning applications and challenges in big data analytics', *Journal of Big Data* **2**(1), 1–21.

[20] *Note by International Maritime Organization (IMO). 2012 Guidelines for the development of a SEEMP* [n.d.], *MEPC 63/23, Annex 9* .

[21] Pedersen, B. P. and Larsen, J. [2009], 'Modeling of Ship Propulsion Performance', *World Maritime Technology Conference* (2008), 1–10.

[22] Petersen, J. P., Jacobsen, D. J. and Winther, O. [2012], 'Statistical modelling for ship propulsion efficiency', *Journal of Marine Science and Technology* **17**(1), 30–39.

[23] Radonjic, A. and Vukadinovic, K. [2015], 'Application of ensemble neural networks to prediction of towboat shaft power'.

[24] Trodden, D. G., Murphy, A. J., Pazouki, K. and Sargeant, J. [2015], 'Fuel usage data analysis for efficient shipping operations', *Ocean Engineering* **110**, 75–84.

[25] *United Nations Framework Convention on Climate Change* [1994], *UN General Assembly* .

[26] Zissis, D., Xidias, E. K. and Lekkas, D. [2015], 'A cloud based architecture capable of perceiving and predicting multiple vessel behaviour', *Applied Soft Computing Journal* **35**, 652–661.