Power prediction for a vessel without recorded data using data fusion from a fleet of vessels

A.I. Parkes[a,*], A.J. Sobey[a,1], D.A. Hudson[a], T.D. Savasta[a]

[a] *Maritime Engineering, University of Southampton, University Rd, Southampton SO16 7QF, England, UK*
[b] *Marine and Maritime Group, Data-centric Engineering, The Alan Turing Institute, The British Library, London, NW1 2DB, England, UK*

Abstract

Recent legislation in shipping applies additional pressure to reducing fuel consumption. However, this is impossible without accurate power prediction, as it is required to allow comparisons between novel efficiency improving advancements and to have confidence in route optimisation. This prediction is particularly difficult in rough weather, which the traditional prediction methods struggle to account for. Neural networks trained on an operational dataset from the vessel are a potential solution to this problem, as they have been shown to predict powering to a mean error of 2% across all weather conditions. However, the gathering of these data is expensive and time consuming. There is currently no literature looking at how data from one vessel can be used to make predictions about another, reducing the cost and allowing prediction of the performance of new vessels. This paper investigates the accuracy in predicting powering for an unseen vessel, using a neural network trained on a fusion of data, from a range of sensors located on other vessels in a fleet. It demonstrates the level of extrapolation that can be achieved from the use of multiple datasets on a real application and suggests that, for the fleet of vessels used, ship parameters are less important for accurate power prediction than having sufficient data across the desired prediction domain. It concludes that prediction of around 4% error can be achieved for most ships in the fleet and discusses the cause of the higher errors seen for a minority of other vessels.

Keywords:

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

Machine Learning, Shaft Power Prediction, Neural Networks, Ocean Engineering, Naval Architecture

1. Vessel Power Prediction in Waves

Reduction of emissions in the marine industry is important as it is estimated to be responsible for 2.5% of Global Greenhouse Emissions (International Maritime Organisation 2020). Therefore, the IMO's Marine Environment Protection Committee (MEPC) require a 50% reduction in annual shipping emissions by 2050 compared to 2008 (International Maritime Organization 2018). To meet this aim, accurate prediction of vessel power consumption will allow benchmarking of energy saving devices or new design concepts, allowing current and future vessels to be more efficient. In addition, weather routing optimisation can be used to ensure that the most efficient routes are taken to avoid bad weather or adverse currents. However, the prediction of emissions is difficult as the range of weather conditions encountered during operations rarely equates to the 'calm water' conditions traditional powering prediction methods are designed for (Holtrop 1984). Accurate prediction of power in weather is therefore essential to effectively benchmark power consumption and therefore to reduce it.

Predicting ship power requirements in weather is challenging. Traditional naval architecture techniques based on experimental towing tank data are expensive, as they require a number of tests and each new vessel will require these to be repeated. Due to this expense models are rarely tested in complex sea states, so these data are not applicable to realistic vessel operation. To reduce the expense of towing tank experiments, empirical formulae based on these tests have been developed for calm water conditions (Holtrop and Mennen 1982) (Holtrop 1984). These methods require multiple non-trivial vessel parameters which are known during the design phase of a vessel but may not be known for a ship in operation. Such parameters are used in neural networks to predict powering but with limited practical use (Ai-guo and Jia-wei 2009). Some extensions to these methods account for weather using Beaufort number, which is a coarse measure of wind strength, to infer typical wave height (Townsin et al. 1993). However, the use of coarse bins of 'sea state', instead of measured or hindcast wave height, reduces the accuracy of any predictions.

While simple empirical formulae are inaccurate in rough seas, it is possible to model vessel powering accurately in waves using Computational Fluid Dynamics simulations. However, these require prohibitively large run times as the air-sea interface is complex to model (Wackers et al. 2011). More recently continuous monitoring data, with frequencies around 30 seconds, provides an opportunity to produce fast and accurate powering estimates without the need for specific vessel parameters. However, it is difficult to analyse continuous monitoring performance data in waves using traditional regression approaches (Lakshmynarayanana and Hudson 2017), due to the heavy-tailed distributions and noise levels within the data. Therefore machine learning techniques are increasingly being used on large datasets to make these predictions, as they have the ability to approximate complex relationships, allowing them to model the vessel-weather relationships present in operational data (Parkes et al. 2018).

The literature shows that shaft powering of a vessel can be predicted with average accuracies of between 1.5-5% error with the use of a regression neural network trained with high frequency data from the vessel (Pedersen and Larsen 2009), (Radonjic and Vukadinovic 2015), (Bal Beşikçi et al. 2016), (Zissis et al. 2015). Interactions between weather, specifically wave height, and vessel propulsion can be approximated by neural networks to increase prediction accuracy (Petersen et al. 2012) (Hu et al. 2019), which is not possible with traditional techniques from naval architecture. The neural network is easier to implement than traditional methods, does not require any vessel-specific parameters, and can be implemented earlier in a vessel's operating life than other performance analysis techniques as it can use high frequency data without trimming or binning.

A number of ships now record continuous monitoring, or high frequency, data. This includes data from ships in all weather conditions encountered during operation. It is demonstrated that the power requirements for these vessels can be predicted accurately for any given vessel using neural networks and operational data (Petersen et al. 2012). However, the cost of gathering these data is high, in the region of £100,000s for each vessel. It is therefore impractical for every ship to be monitored purely for the determination of powering requirements. Hence, utilising these data across fleets, making predictions for different vessels from the data collected from another, is important to reduce the cost.

Data fusion is increasingly used in machine learning to improve the accuracy or pertinence of results (Elmas and Sonmez 2011) (Melendez-Pastor et al. 2017) and the potential of blending data from different instruments, time periods and applications is still under investigation. This dataset provides a particularly stochastic set of data, with continuous and discontinuous input distributions, as well as different operating profiles for each ship. The application of data fusion methods could allow power prediction for merchant vessels where there is no available data, such as those on charter agreements, and considerably reduce the cost for these methods.

This study provides insight into the potential of predicting powering for a vessel where no data exists by training networks on data from other vessels. Studies in similar areas, such as wind power prediction, show promising results (Tasnim et al. 2018). However, there are no known studies investigating this for vessel power prediction, which may be due to the comparative complexity of modelling vessel powering caused by the effect of second order variables; unmeasurable input variables such as piloting style or the interaction of draft and trim. Power prediction quality for networks trained on a sister ship is compared to networks trained on a fleet of non-sister vessels. Datasets from 21 Liquid Natural Gas carriers are analysed to understand the quantity of data that is required and the accuracy in prediction for vessels with different levels of similarity.

2. Fleet Data

Datasets of varying size from 21 Liquid Natural Gas carriers are analysed for their use for power prediction from neural networks. Some are sister ships, so have identical hull forms and similar machinery, where others are the only example of their vessel type. To preserve the anonymity of the data, the vessel types are denoted by letters and for each vessel type each sister ship is denoted by a number. There are 8 different classes of vessel A-H, with quantities of sister ships for each vessel type ranging from 1 to 7. The size of each dataset ranges from nearly 2,500,000 observations to fewer than 60,000, Figure 1, all datasets have a frequency of 30 seconds per observation.

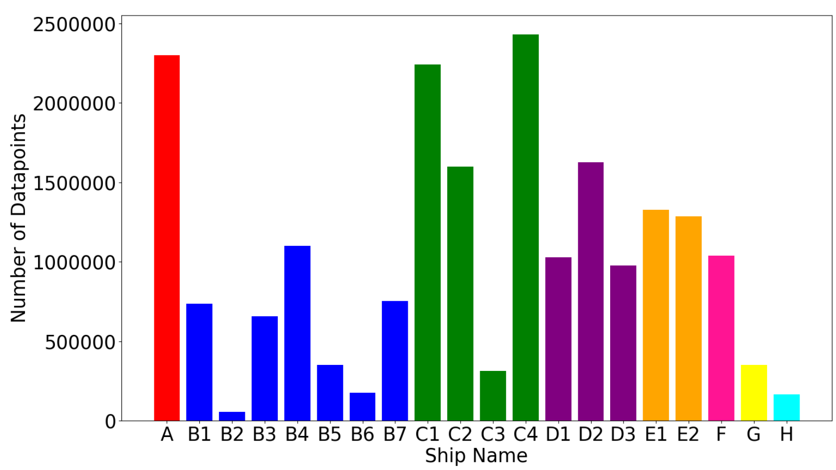


Figure 1: The size of each vessel's dataset, with each class of vessel signified by a different colour.

To ensure the validity of the results and to accurately measure the change in error, the ship datasets should be of comparable size. Therefore, only datasets with over 1,000,000 datapoints are used. The size is chosen to ensure the testing and validation sets[1] are over 150,000 datapoints, as statistical analysis showed

[1]The randomly selected subset of 15%, of the dataset which is used for testing a networks

that this size set approximates the distribution of whole dataset accurately for all variables used, and hence avoiding autocorrelation related bias in errors.

Out of the original 21 ships, 9 have more than the required 1,000,000 datapoints post trimming; with 2 sister ships in the D and E class, 3 sister ships in the C class and only 1 ship in each of the A and F classes. The size of the test set used for every vessel is 150,000, regardless of the overall size of the dataset. Although only 9 ships are used in this study, the methodology used is designed to be applicable to larger fleets, for example the data filtering applied is not tailored to the specific sensor errors found on each ship, instead only removing all points where vessel power is 0 or below.

The fleet of ships are all Liquid Natural Gas carriers with build dates spanning 12 years from 2003 to 2015. Amongst the 9 ships there are two different types of propulsion system: steam and diesel-electric. All ships are of a similar size, with a 17m range in length and 3m range in beam. For these large merchant vessels, 17m difference in size is not significant in relation to the overall size of the vessel, so powering relationships are expected to be similar. The C-class vessels have 2 propellers, while all other vessels have one, larger, propeller. Propeller number and size will not significantly affect powering but may change the routes operated by a vessel.

2.1. Input Variable Analysis

Over 100 variables are recorded on board each vessel including: engine, cargo, vessel condition and vessel movement data. A detailed study into variable selection for shaft power prediction has been performed, (Parkes et al. 2019), where data quantity was prioritised. Of the 100 input variables, nearly half of the variables did not measure correctly for over 30% of the time, leaving a set of 43 usable variables for prediction. In this study, 3 different sets of input variables are compared. The first set is 5 variables selected from a Naval Architecture perspective:

power prediction abilities.

(i) GPS ship speed (knots);

(ii) true wind speed (m/s);

(iii) apparent wind direction (degrees);

(iv) draught (m) - the distance from the base of the ship to the waterline; and

(v) trim (m) - the angle of the waterline on the vessel.

This set of 5 is compared to two other input variable selection methods: incrementally increasing the quantity of input variables ordered based on their correlation to shaft power and incrementally increasing the quantity of principal components from a Principal Component Analysis (PCA) of the full usable dataset. The study notes minimal difference in prediction accuracy between each of the approaches. This study uses the 5 Naval Architecture selected variables discussed in (Parkes et al. 2019). As expert opinion identifies that these 5 have a causal connection to the output; a change in any one of them causes an increase or decrease in required shaft power. The use of causally related input variables will allow comparison between vessels through analysing the relationship between inputs and shaft power for each, to identify differences in powering characteristics.

Although vessel speed is the most highly correlated variable to shaft power, only one measure of vessel speed is used as an input variable as the two variables contain significant redundancy between them. This would introduce unnecessary complexity for the network to model, given the addition of a highly correlated input variable would provide minimal additional information. If measured correctly, the speed through the water measurement is more hydrodynamically relevant to powering than the speed over ground. However, due to inaccuracies in the measuring equipment the speed through the water is often less reliable than speed over ground. Over all the observed datapoints the two speed measurements are within 1knot of each other over 80% of the time and as the speed through the water is deemed reliable it is used as the vessel speed input variable.

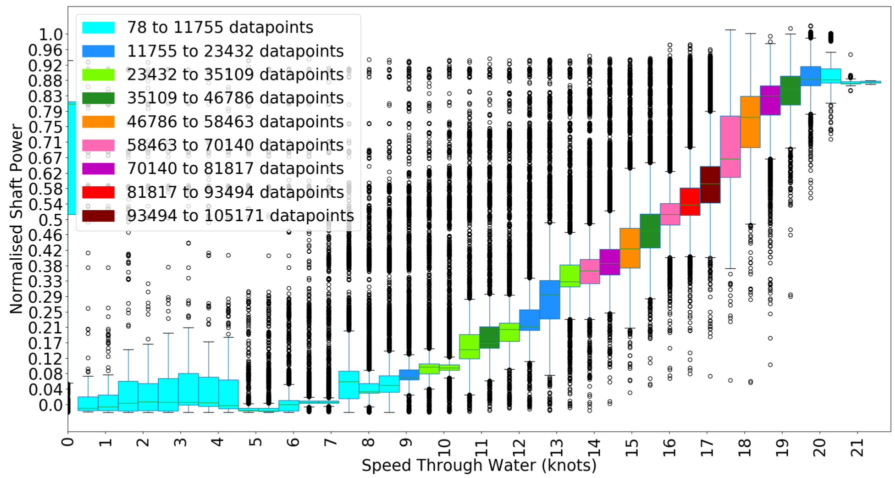
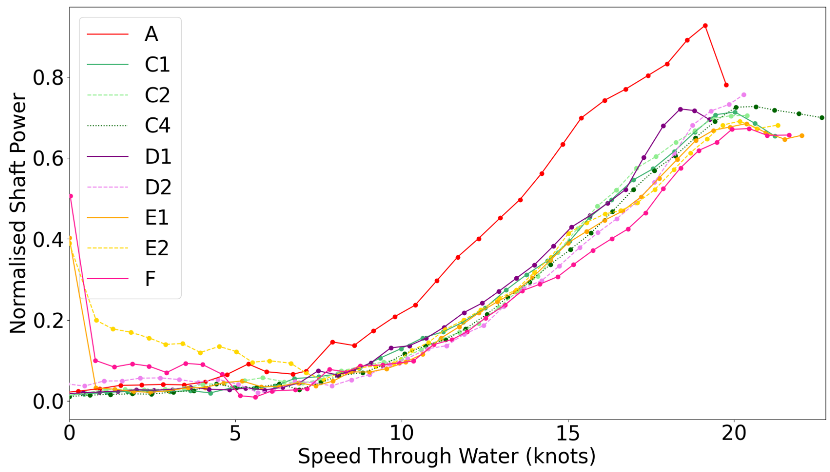
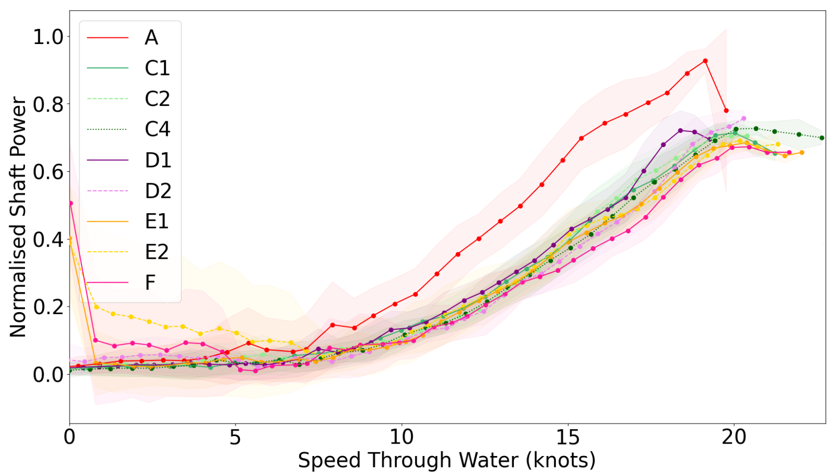


Figure 2: The distribution of the observed shaft powers for half knot bins of speed through the water for ship F. In the box and whisker plots the boxes contain 50% of the distribution and the whiskers extend to the datum which is at 1.5 times the interquartile range.

Power to speed through the water regression curves are often used in the Naval Architecture literature to define powering relationships for a specific vessel, as a vessel's speed is highly correlated to its power requirements. The distribution of power values observed at each vessel speed through the water illustrates this relationship, Figure 2. In the traditional regression method a single trend line describes the relationship between the two variables for the majority of the datapoints. However, the stochastic nature of the environment leads to a more complex patterns requiring other variables such as weather and vessel condition to make accurate predictions across the entire range. This is made more complex by variables like draft and trim which are multimodal and for certain vessels this approaches a discrete distribution.



(a)

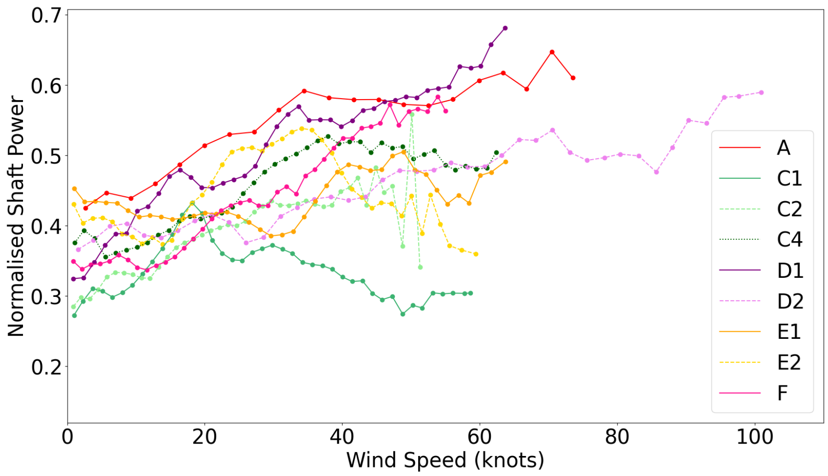


(b)

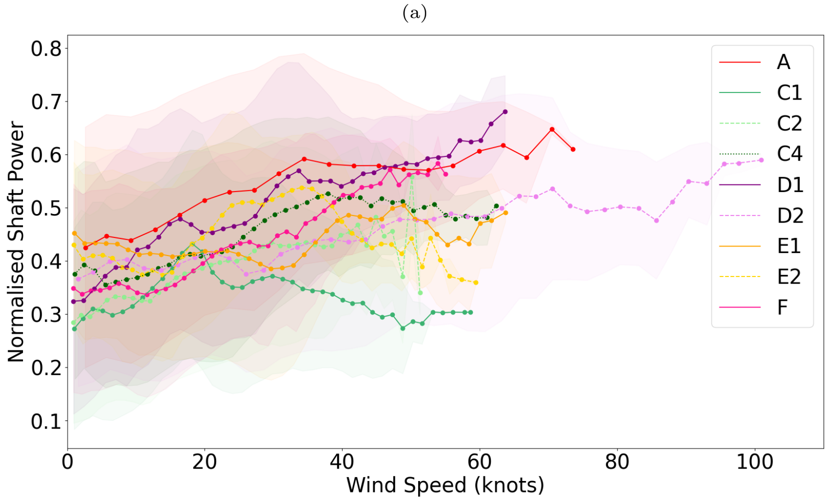
Figure 3: Comparison of the average observed power at half knot intervals of speed through the water for all vessels in the fleet (A), with shaded regions ± one standard deviation for each interval (B).

To compare the power profiles of the vessels across the fleet, the average power observed at each half knot interval of speed is plot for all of the vessels, Figure 3a. No noise or secondary relationships are captured by these speed-power curves, however a difference in propulsion relationship is clear for ship A. The required power for ship A is around 20% of the maximum power higher than for all of the other vessels, for all speeds over 7knots. This vessel is the only steam powered ship, as well as the oldest vessel in the fleet by 7 years, so a difference in propulsion characteristics is expected. This difference in performance may cause problems if attempting to predict powering of ship A by training on vessel data from the other ships.

There is an increase in average shaft power for near zero speeds for ships E1, E2 and F, Figure 3a. This is unexpected but may be caused by a sensor error, as minimal data filtering procedures are used on the datasets, or heavy weather, as the measurement used is the speed through the water resistance from waves and wind may affect the speed measurement.



(a)

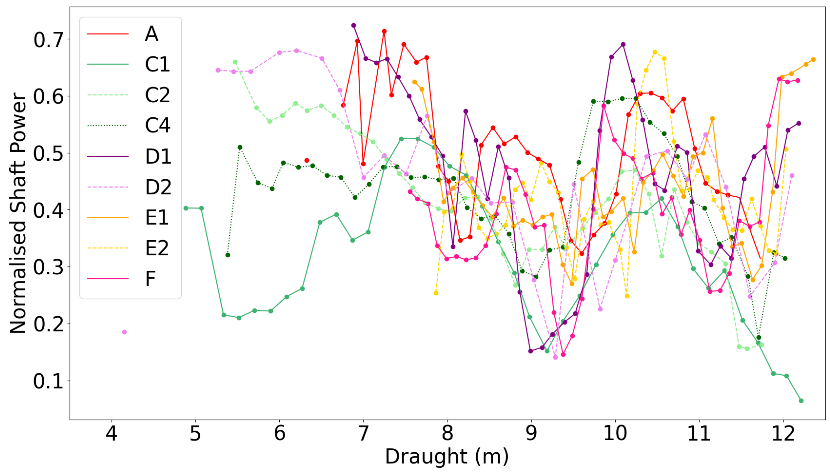


(b)

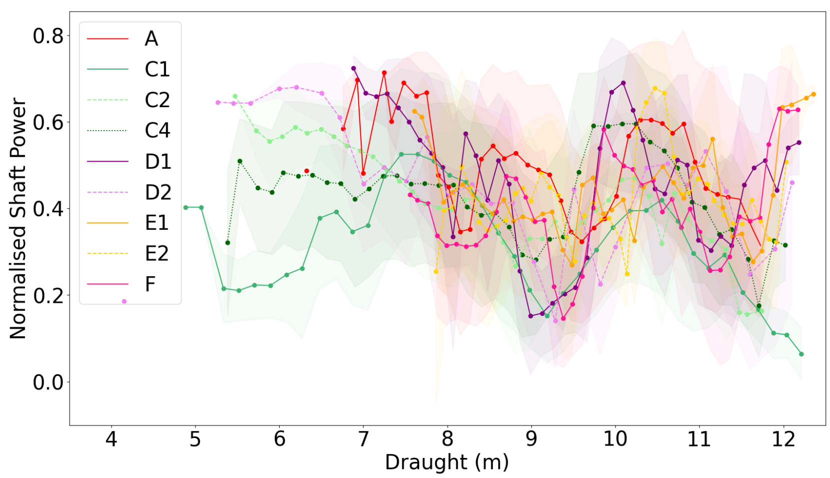
Figure 4: A comparison of the average observed power against wind speed for all of the vessels in the fleet (A), with shaded regions ± one standard deviation for each interval (B).

To compare vessel datasets the power-variable curves are analysed and the power to wind direction curves show the same trend across the fleet for the wind direction domain from 0° to 360°. This means wind direction relationships learnt for one vessel should transfer to another vessel well. However, the wind speed curves show a less even distribution of data across the range of observed wind speeds. Ship D2 is the only vessel to experience the highest speeds of 60-

100mph, Figure 4a, which is extreme, equivalent to Beaufort 12 or 'hurricane force'. Although only 0.7% of the dataset records wind speed values above 60mph, when plotted temporally, the set of datapoints containing these high wind readings create a smooth curve which suggests that these data is not anomalous. This 60+knot area in the wind speed domain is sparsely populated, both in terms of the total number of datapoints and in the variety of vessels with datapoints populating it, making it difficult to predict behaviour in this region as the distributions are not representative of the behaviour.



(a)



(b)

Figure 5: Comparison of the average observed power against draft for all vessels in the fleet (A), with shaded regions ± one standard deviation for each interval (B).

A similar area of sparse data can be observed in the draft variable domain, Figure 5a. The power-draft curves show that only 4 vessels; D2, C1, C2 and C4, operate at drafts below 7m. The quantity of datapoints below 7m of draft is 0.21% of the combined 4 vessel's data. The sparseness of this section means the data in it may not have representative distributions, which may be the cause of the separation between all four lines below 7m of draft, Figure 5a. The power-trim curve is also analysed, but due to the coupling of draft and trim, does not provide any further insight.

The power prediction is shown to be most sensitive to changes in the speed of the vessel. However, a significant spread of power can be observed based on the additional noise provided from the other variables. Due to the complexity of these relationships, machine learning techniques are required to accurately model the powering of a vessel.

3. Artificial Neural Networks

Artificial neural networks are made up of layers of interconnected neurons, with size denoted as **(Layers, Neurons)** in this study. They approximate relationships between pre-assigned inputs and outputs by optimising the weightings of the connections between neurons. The more layers and neurons a network contains, the more complex the relationships within a dataset it can model.

The networks in this study are written using the Keras 2.3.1 libraries (Chollet et al. 2015) with Tensorflow 2.0.0 (Abadi et al. 2015) backend. Feed forward networks are used exclusively. From domain knowledge there should exist temporal effects within the dataset, such as breaking waves on the ship or manoeuvring (Simsir and Ertugrul 2009). But from time series analysis and trialling the use of recurrent networks, it is discovered that the data frequency of 30 seconds is too coarse to identify any temporal correlations. The datasets are therefore treated as time invariant and no networks involve recursive elements and all network outputs are predicted independently. The inputs and targets are scaled between 0-1, as both normalisation and scaling are trialled and scaled inputs produced improved prediction performance.

Alongside an early stopping procedure, the maximum epoch limit is 1000, although in practice this limit is never reached, as the early stopping terminates training after 10 epochs of unimproved error values. A relatively small batch size of 50 is chosen to balance training time and accuracy. As training the networks in this study is in the order of hours, a small batch size is possible. The learning rule used is AdaMax (Kingma and Ba 2014) with ReLU activation functions throughout the network. The use of a technique to regularise training such as L1 and L2 regularisation and dropout are trialled and are shown to decrease accuracy of predictions, so no regulariser is used. The error function used is mean absolute, as it is commonly used in other ship power prediction literature (Grabowska and Szczuko 2015) this will allow cross study comparisons. All selected hyperparameters for this study are listed in Table 1, these parameters are selected from a small parametric search as well as experience predicting powering from ship datasets..

The number of layers and neurons and the configuration of the neurons in the networks used is briefly investigated with a parametric study of prediction accuracies for sizes in the range **(1, 50)** to **(3, 400)**, to identify if the network parameters used in ship power prediction literature produce results with similar accuracy to previous applications. The maximum number of layers investigated is 3 as there are only five input variables, and the dataset is not evenly distributed enough for meaningful feature extraction using more layers.

From the initial parametric study using the entire dataset as training data, the median error decreases for increasing size of network from **(1, 50) - (3, 300)**, with **(3,300)** networks producing a median testing error of 1.98% and median training error of 1.95% suggesting that this size network does not overfit the dataset. No difference in error is noted between networks where the number of neurons decreases for each layer and networks with the same number of neurons in each layer. Networks of size **(3, 400)** have a similar distribution of errors but a median error 0.10% higher. This suggests that the **(3, 400)** network does not capture any additional relationships within the dataset to reduce error, compared to the **(3, 300)** network, but is instead beginning to overfit the dataset. For these reasons networks of size **(3, 300)** are used for the rest of this study.

Table 1: Selected hyperparameters

Hyperparameter	Value
Number of hidden layers	3
Number of neurons in each hidden layer	300
Number of epochs	1000
Batch Size	50
Early Stopping Patience	10
Error function	Mean Absolute Error
Learning rule	AdaMax (Kingma and Ba 2014)
Activation Function	ReLU
Regularising technique	None
Initialiser	Random Normal ($\mu = 0, \sigma = 0.1$)

4. Verify Datasets for Power Prediction

For each ship, 10 networks with the parameters specified in Section 3, are trained and tested on their operational dataset[2]. The only difference for each of the 10 network runs is the randomly sampled training, testing and validation sets, the random initialiser and any stochastic elements of the optimiser. This acts as a benchmarking of the method and datasets against results in the literature and allows an initial comparison of the prediction accuracies of the ships.

[2]For all future results in this study 10 repetitions are performed to increase validity.

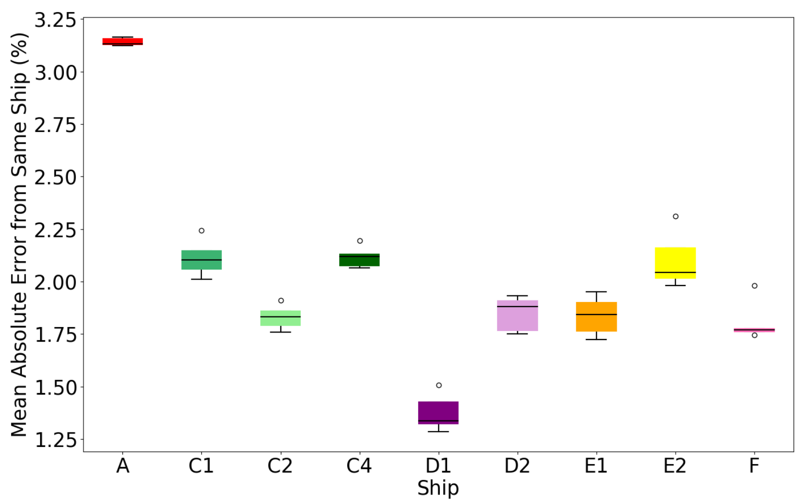


Figure 6: The distribution of mean absolute error from 10 networks of size **(3, 300)** for individual vessels, showing consistent predictions for individual ships and a maximum difference of 1% between the mean error for different ships. In the box and whisker plots the boxes contain 50% of the distribution and the whiskers contain 90% of the distribution and the circles show outliers. Where no upper whisker is visible, the 75th and 95th percentiles coincide.

For most of the ships, the distribution of the mean prediction errors from the different vessels ranges between 1.78-2.13%. Ship D1 can be predicted the most accurately, and exhibits a lower average error of 1.35%, and ship A is the most difficult to predict, with all of the errors above 3% Figure 6. These accuracies are similar to those in the literature. Powering is predicted to within a 5% error (Pedersen and Larsen 2009) with a best result of 1.5% from (Petersen et al. 2012), where the wave height is used as an additional input variable which has been shown to give an increase in accuracy of 0.5% (Parkes et al. 2019). Each ship shows consistent predictions, with low standard deviations, around 0.25%. Networks with the parameters in Table 1 produce errors inline with the power prediction literature for every ship, therefore this size of network is used in the following sections.

Within these predictions there is no relationship between error value and ship class, as error values within ship classes vary just as much as between ship classes. This suggests that the variation between ships is likely to be due to factors not relating to hull form or ship parameters. These factors may be a specific sensor error, the vessel conditions experienced, or differences in piloting and operation. The next two sections document the use of a network trained on one vessel to predict powering for a different vessel using two different approaches. The first trains networks on data from one ship and tests on all other ships in the fleet separately, the second trains networks on a fusion of data from all of the ships in the fleet apart from one, which the network is tested on.

5. Prediction when trained on a different ship

Networks are trained with data from one ship and tested on all other ships separately to evaluate the prediction accuracy from networks trained on data from one ship.

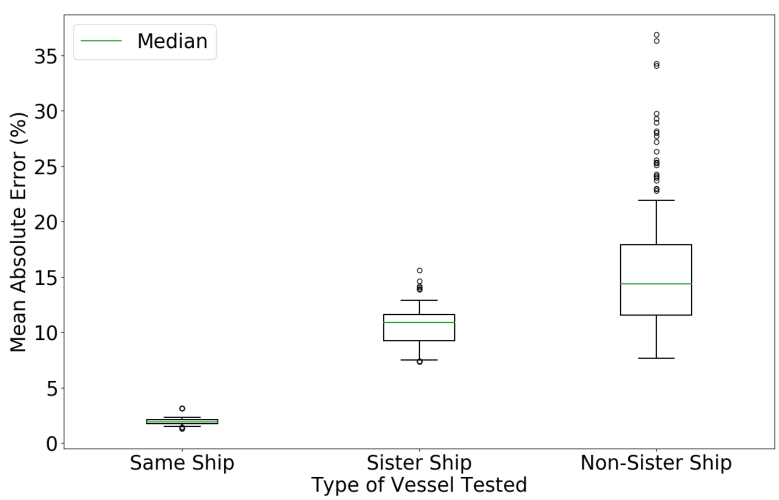


Figure 7: Box and whisker plots to illustrate the distribution of prediction accuracy from multiple networks when tested on: the ship the network was trained on; a sister ship; and a ship of a different class.

The error in prediction for networks trained on the data from a sister ship is between 7.34-15.63% with a median of 10.89%, Figure 7. This error is a significant increase compared to the error from networks trained on the same ship. The propulsion systems and hull forms on sister ships should be identical, hence theoretical powering relationships should be the same: which implies errors in line with those from a network trained on the same ship, Section 4. However, in operation the differences in routes, conditions, and operators increases the difference in the relationships within each dataset, which means that error in power prediction increases from 2% to 11%. An error in prediction of around 11% is large compared to the power reduction produced by energy saving devices like air lubrication, which is around 5%, making networks trained on a single sister ship unusable for many practical applications.

When networks are tested on non-sister ships the minimum mean error is the same, around 7.5%. However, the mean, mean error increases by 3.5%, and the maximum mean error increases to 36.93%. Therefore, the distribution of non-sister ship errors has a similar distribution to the sister ship errors but with a longer tail. The distribution of errors for non-sister ships demonstrates that the same error in prediction for sister ships can be obtained by a non-sister ship. This means that the operation and experienced conditions of a vessel affect error of prediction more than vessel proportions. The following section evaluates the errors from networks trained on data from more than one ship to test an unseen ship.

6. Prediction when trained on a fused dataset of all other ships

To emulate a more realistic situation, where data is available from some but not all vessels in a fleet; this section uses a form of k-fold cross validation where data from all of the vessels except for one and uses this network to predict powering for the unused vessel. This increases the size of dataset used to train the networks from around 700,000 to around 5,600,000.

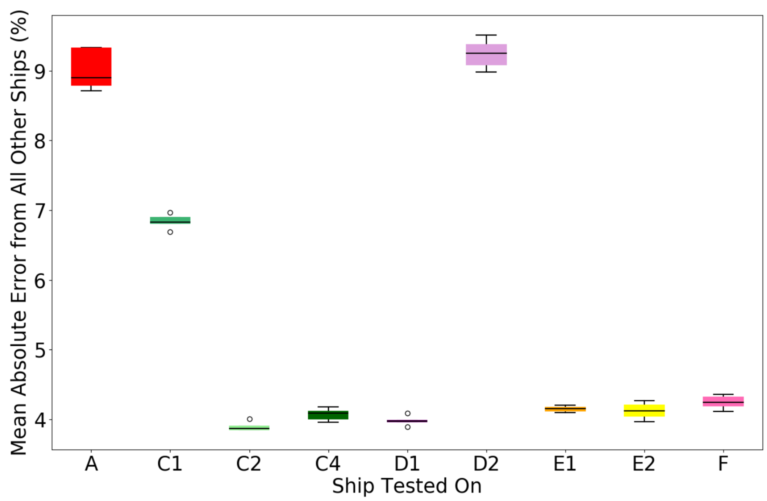


Figure 8: The distribution of mean absolute errors from multiple networks trained on all of the vessels apart from the ship tested on.

Errors in the range $(4 \pm 0.25)\%$ are observed for 6 of the 9 vessels tested, Figure 8. The three vessels which have a higher error are ships C1 with $(6.83 \pm 0.14)\%$ error in prediction and ships A and D2 with $(8.89 \pm 0.44)\%$ error and $(9.26 \pm 0.27)\%$. Ships C1 and D2 have sister ships with errors in line with $(4 \pm 0.25)\%$, which shows that the dimensions of the vessel are not causing the high errors for these ships. Therefore, including parameters like vessel length and hull form as an input would not improve prediction accuracies for this fleet.

It is suggested that the difference in power-speed curves, in Section 2.1, explains why the errors in prediction for ship A are high, as the speed-power relationship in the other 8 datasets are 20% lower than ship A, Figure 3a. This is due to the difference in propulsion systems and age of vessel as ship A is the only steam powered vessel in the fleet.

Ship D2 has errors over double that of its sister ship D1, Figure 8. This suggests that the area of the input variable space for this vessel is not covered by the training dataset. The operating conditions experienced by the vessels differ from the rest and this is confirmed by the wind speed curves, Figure 4a.

This region of high winds explains why the errors for ship D2 are high. No other ship experiences the same extreme conditions so a network trained on all other ships cannot predict accurately for high winds. It is confirmed that when the area of extreme weather (60+knots) and unusually low drafts were removed from the dataset, the errors for ship D2 reduce to 3.6%, which is in line with the other vessels.

Ship C1 shows an error 2% lower than the other anomalous ships, A and D2. Ship C1 is one of the four ships to operate at low drafts, Figure 5a. Although all of the power-draft relationships in this region show different relationships, due to the low quantity of data, ship C1 is the only vessel to show an increasing trend in this region, with an increase in power for an increase in draft. When the region of draft below 7m is removed from all of the datasets, the powering for ship C1 is predicted with a 4.6% error from a network trained on all other vessels. This is within 0.5% of all other ships when appropriate input variable ranges are selected, demonstrating that this difference in low draft behaviour is likely to be the reason for the difference in error.

7. Discussion

It is shown that networks trained on data from a single sister ship can predict powering with an error of around 11%, compared to 4% from a fleet of non-sister vessels. The decrease in error is unlikely to be caused solely by the increase in number of datapoints in the training set; as the range of operational conditions experienced is shown to be more important for accurate prediction than absolute number of datapoints (Parkes et al. 2018). This suggests that, for the fleet used in this study, vessel parameters are less important for accurate power prediction than sufficient data across the desired prediction domain.

Error in power prediction from a network, trained on data from 8 ships, for an unseen ship is $(4 \pm 0.25)\%$ for most vessels. For the 3 ships with a higher error in prediction, a significant difference in propulsion system, experienced conditions or behaviour is observed through analysis of the vessel data. If suitable discretion is used in choosing appropriate vessels and regions of input data for prediction, accurate power prediction from a fusion of data is possible using neural networks, without operational data from that specific vessel data or a sister ship.

The ships used in this study are all liquid natural gas carriers of similar size and proportions. Investigations into the effect caused by utilising a more varied fleet is of interest. This will allow the assessment of how relevant the powering relationships within a dataset from one vessel type is to a different type of vessel, by utilising techniques such as multi source domain adaptation. The use of a larger fleet may also allow the effects of latent variables such as piloting style or specific sensor characteristics to be analysed more completely, creating the potential to adapt the method to account for these.

Although meaningful feature extraction has not been noted during this study, a full investigation into whether it is possible is of interest. Methods to encourage feature extraction include utilising cascade networks or training each layer of a network on a different ship. The latter would provide flexibility for more diverse fleets; where a modular approach to network layers could allow only the most relevant vessels to be used for prediction.

8. Conclusion

Power prediction is a difficult task for ships in waves through traditional methods. To update these approaches, modern machine learning based methods demonstrate high accuracy but require vessel specific operational data for training. However, many vessels do not collect operational data, as it is expensive or the operator does not own the vessel, but these vessels still require accurate power prediction. Therefore, this study investigates the ability to make predictions for ships without data, from a fusion of data from all other ships in the fleet. First, the accuracy of using a neural network trained on operational data from a specific vessel is verified to be possible to with 2%, but extends this to include no manual trimming of the dataset. When using a network trained on 8 similar vessels to predict the powering, errors of $(4 \pm 0.25)\%$ are observed for most vessels. This error is less than half that from a network trained on a single sister ship. The ships with higher errors can be shown to have experienced different environmental conditions or have different propulsion systems, visible from a preliminary statistical analysis. It is therefore possible to extend the range of extrapolation using this fleets' data and means it is possible to accurately predict the behaviour of a new vessel, or one where data are not being collected, with a sensible selection of fused dataset comprised of similar vessels.

Acknowledgements

This work was kindly funded by Shell Shipping and Maritime. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

References

Abadi, M. et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. https://tensorflow.org.

Bal Beşikçi, E. et al. (2016). "An artificial neural network based decision support system for energy efficient ship operations". In: *Computers & Operations Research*, pp. 393–401. ISSN: 03050548. DOI: 10.1016/j.cor.2015.04.004.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

Elmas, C. and Y. Sonmez (2011). "A data fusion framework with novel hybrid algorithm for multi-agent decision support system for forest fire". In: *Expert Systems with Applications* 38, pp. 9225–9236.

Grabowska, K. and P. Szczuko (2015). "Ship resistance prediction with artificial neural networks". In: *2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 168–173.

Ai-guo, C. and Y. Jia-wei (2009). "Research on the genetic neural network for the computation of ship resistance". In: *2009 International Conference on Computational Intelligence and Natural Computing*. Vol. 1, pp. 366–369.

Holtrop, J. (1984). "A statistical re-analysis of resistance and propulsion data". In: *International shipbuilding progress* 31.363, pp. 272–276. ISSN: 0020-868X. DOI: 10.1007/s12011-015-0572-4.

Holtrop, J. and G. G. Mennen (1982). "An approximate power prediction method". In: *Int. Shipbuilding Progress* 29.

Hu, Z. et al. (2019). "Prediction of fuel consumption for enroute ship based on machine learning". In: *IEEE Access* 7, pp. 119497–119505.

International Maritime Organisation (2020). *Fourth IMO Greenhouse Gas Study*.

International Maritime Organization (2018). *Marine Environment Protection Committee (MEPC), 72st session*.

Kingma, D. P. and J. Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Lakshmynarayanana, P. and D. A. Hudson (2017). "Estimating added power in waves for ships through analysis of operational data". In: *2nd Hull Performance and Insight Conference: HullPIC'17*.

Melendez-Pastor, C., R. Ruis-Gonzalez, and J. Gomez-Gil (2017). "A data fusion system of GNSS data and on-vehicle sensors data for improving car positioning precision in urban environments". In: *Expert Systems with Applications* 80, pp. 28–38.

Parkes, A. I., A. J. Sobey, and D. A. Hudson (2018). "Physics-based shaft power prediction for large merchant ships using neural networks". In: *Ocean Engineering* 166, pp. 92–104.

Parkes, A. I. et al. (2019). "Efficient vessel power prediction in operational conditions using machine learning." In: *Practical Design of Ships and Other Floating Structures(PRADS), September 2019, Yokohama, Japan.*

Pedersen, B. P. and J. Larsen (2009). "Prediction of full-scale propulsion power using artificial neural networks". In: *Proceedings of the 8th international con-*

ference on computer and IT applications in the maritime industries (COMPIT'09), Budapest, Hungary May, pp. 10–12.

Petersen, J. P., D. J. Jacobsen, and O. Winther (2012). "Statistical modelling for ship propulsion efficiency". In: *Journal of marine science and technology* 17.1, pp. 30–39.

Radonjic, A. and K. Vukadinovic (2015). "Application of ensemble neural networks to prediction of towboat shaft power". In: *Journal of Marine Science and Technology (Japan)* 20.1, pp. 64–80. ISSN: 09484280. DOI: 10.1007/s00773-014-0273-2.

Simsir, U. and S. Ertugrul (2009). "Prediction of manually controlled vessels' position and course navigating in narrow waterways using Artificial Neural Networks". In: *Applied Soft Computing* 9.4, pp. 1217–1224.

Tasnim, S. et al. (2018). "Wind power prediction in new stations based on knowledge of existing Stations: A cluster based multi source domain adaptation approach". In: *Knowledge-Based Systems* 145, pp. 15–24.

Townsin, R. L. et al. (1993). "Estimating the influence of weather on ship performance." In: *Transaction of the Royal Institution of Naval Architects* 135, pp. 191–209.

Wackers, J. et al. (2011). "Free-surface viscous flow solution methods for ship hydrodynamics". In: *Archives of Computational Methods in Engineering* 18.1, pp. 1–41.

Zissis, D., E. K. Xidias, and D. Lekkas (2015). "A cloud based architecture capable of perceiving and predicting multiple vessel behaviour". In: *Applied Soft Computing* 35, pp. 652–661.