*Manuscript
Click here to view linked References

Machine Learning for Propulsion Power Prediction and Analysis of Fouling Related Performance Deterioration

Anastasia Laurie[a*], Enrico Anderlini[a], Jesper Dietz[b], Giles Thomas[a]

[a]Department of Mechanical Engineering, University College London, United Kingdom

[b] A.P. Moller – Maersk, Copenhagen, Denmark

Abstract

Improving operational performance and reducing fuel consumption is of growing importance for shipping companies. Ship performance degrades over time due to hull and propeller fouling; therefore assessing when fouling effects are significant enough to warrant cleaning is critical. Advancements in onboard data logging systems, combined with machine learning techniques, unlock the potential to predict fouling effects accurately and determine when to clean. This study evaluates five models for propulsion power prediction: Multiple Linear Regression, Decision Tree (AdaBoost), K – Nearest Neighbours, Artificial Neural Network and Random Forest. The significance of environmental parameters was explored, and simulated power-speed curves created from predictions to identify performance deterioration due to fouling. The Random Forest model was most effective in predicting propulsion power, with an error of 1.17%. The addition of 'Days Since Clean' and 'Significant Wave Height' increased prediction accuracy by 0.07% and 0.12% respectively. Simulated power-speed curves revealed a 5.2% increase in effective power due to fouling. The findings highlight the potential of tree-based models for decision support systems while providing a method to determine when to conduct hull and propeller cleaning.

Keywords: Operational performance, Propulsion power, Biofouling, Machine learning, Random Forest, Hull cleaning.

*Corresponding author: zcakala@ucl.ac.uk, 26 Radley Mews London W8 6JP

1. Introduction

The shipping industry represents a growing proportion of global CO_2 release, with emissions from shipping having increased by 70% since 1990 (Cames et al., 2015). The International Maritime Organisation's (IMO) target to reduce 2008 Greenhouse Gas emission levels by 50% before 2050 (IMO, 2018) has put the shipping industry under pressure to improve performance and lower CO_2 outputs. Container ships transporting goods use 140 tonnes of fuel a day on average (Dagkinis and Nikitakos, 2015) and so there is an economic incentive to improve ship performance by reducing fuel consumption, with even small gains in performance yielding considerable financial savings (Terazono and Hume, 2017). Ship performance has traditionally been monitored using Noon Reports (NR), manually derived once per 24 hours, providing an average daily reading for each operational metric (Aldous et al., 2015). Several studies have implemented Machine Learning (ML) techniques to evaluate, model and predict the operational performance of ships from NR data. Besikci *et al.* (2016) created an Artificial Neural Network (ANN) to predict fuel consumption using NR, gaining a predictive performance far greater than a multiple regression model. Further studies by Pedersen and Larsen (2009) explored performance monitoring to predict propulsion power using NR data, again with an ANN, showing good accuracy with a predictive error of 7%.

With advances in ship monitoring, more vessels are now being equipped with high-frequency Continuous Monitoring (CM) systems. Variables are monitored using sensors every few seconds, yielding a greater number of samples during each voyage than NR data. Aldous (2015) found CM data to have less uncertainty than the lower frequency NR data. However, a challenge working with CM data is the large variability present due to the extensive range of operational and environmental situations experienced during a voyage. Nevertheless, operational performance models have yielded lower errors using CM data compared with NR data. Pedersen and Larsen built an ANN using CM data and achieved an absolute error of 1.65%. Subsequently, Petersen and Winther (2012) developed an improved ANN to predict fuel consumption using open-source ferry data, yielding a model with an error of only 1.50%. Comparing this to Gaussian Processes (GP) and Gaussian Mixture Models (GMM), the ANN was able to outperform both models. ML approaches with CM data have not been confined to using ANNs, however. Petursson (2009) applied the K-Nearest Neighbours (KNN) algorithm and Support Vector Regression (SVR) to Petersen's data set to predict shaft power, finding both algorithms exhibited high predictive accuracy. However, the work is difficult to compare to Petersen's due to the use of a different target variable. Chaal (2018) deployed the KNN algorithm along with Decision Tree regression on high-frequency data and compared both models' performances to an ANN, yielding very similar results. Tree-based methods were explored further by Soner *et al.* (2018) on Petersen's ferry data set, in the form of bagging, boosting, and Random Forest approaches. Direct comparison with Petersen's ANN showed the Random Forest obtained a reduced error of 43.5 litres/hour compared to the 47.2 litres/hour achieved by the ANN. No studies have since exploited the promise of tree-based algorithms, and few have compared an array of algorithms using the same data set.

There has been little investigation into the effects of specific environmental and operational factors on performance using ML. Parkes and Savasta (2019) analysed the effect of using different predictor variables within their ANN and advanced feature selection techniques before modelling. When introducing a wave height variable, an improvement in prediction accuracy of 0.5% was achieved, which was proven significant using hypothesis testing. However, no approaches since have explored data preparation or the addition of environmental factors in this way.

A key factor causing performance deterioration is biofouling; the additional surface roughness increases hull and propeller resistance and consequently propulsion power. Cleaning procedures, such as dry-docking and propeller cleaning, which remove fouling are both expensive and time-consuming (Akinfiev et al., 2007). Thus, it is desirable to determine when fouling effects are significant enough to warrant cleaning. Fouling is mainly influenced by environmental factors including water temperature, depth, and pH, while the consequent effects

on power output can vary by ship and increase with speed (Uzun et al., 2019). Current attempts to quantify fouling effects have been limited to empirical and simulation-based approaches. Empirical approaches are used predominantly, with experimental data used as a basis for mathematical fouling models. These studies have predicted an increase in effective power of 25% - 59% at a range of operating speeds (Schultz, 2007; Uzun et al., 2019), validated by NR data. CFD simulations have also been widely used, predicting an increase in effective power of 18.1% – 38% dependent on fouling severity (Demirel et al., 2017; Song et al., 2020).

Previous work with ML models has not yielded solutions with high enough accuracy to identify fouling or cleaning effects on performance. While real ship data has been used to validate empirical approaches, it is unclear how to exploit larger CM data sets to predict performance accurately enough to evaluate fouling effects and determine when to clean. Improved accuracy may lie in advanced preparation techniques and analysis into variable significance in modelling, which has received minimal investigation to date. There is also clear potential for data enrichment with weather and wave conditions to improve the prediction accuracy of current models if such data is not part of the original CM data set. To address the gaps in previous work, this study will use ML techniques to build a propulsion power model with sufficient accuracy to enable determination of fouling related performance deterioration. The study aims to provide the basis for operators to assess when to undertake hull and propeller cleaning. Feature selection techniques and statistical analysis of variable importance are explored to maximise model accuracy and investigate the effects of both cleaning and wave conditions on performance. Additionally, multiple models are created, allowing a direct comparison of ML techniques trained using the same CM data.

2. Data

2.1 Data Overview

Data was provided for five sister container ships (8700 TEU capacity) operating between Europe and South America, for twelve months from the 1st January 2018 to the 1st January 2019. This route, since it transects warm waters which encourages higher marine growth, should provide detectable fouling effects. The raw CM data was recorded at 10-second intervals; however, pre-processing steps averaged variables over 10-minute periods and ensured observations were time-aligned. Data from one ship, Ship A, was used in the modelling stage with the remaining data kept for later validation. The data from this ship contained 26 variables and 52548 observations.

Data cleaning, to eliminate erroneous and missing values from the data set, left 10571 observations remaining, a reduction of 41977 values. A running average was used to smooth the values of latitude and longitude to detect points far from the exact route of the ship, allowing removal of sporadic position readings. Boxplots were used to identify extreme outliers, which were removed conservatively to retain as much data as possible. To reduce variability which would affect model accuracy, observations recorded in shallow waters or when manoeuvring in port were removed. This was achieved by removing power values below a threshold of 7000 kW, to retain observations when sailing in open waters. The filtering, however, reduces the possible predictive range of any resulting models particularly in shallow water.

2.2 Feature Engineering

Additional features were derived to enrich the data set and combine features to reduce dimensionality. The features Trim, Draft, Froude Depth Number, True Wind Speed, True Wind Direction, and Days Since Clean were created. Additionally, wave information for Ship A from the Copernicus Marine Environment Monitoring Service (CMEMS) including Significant Wave Height (SWH), Dominant Wave Direction (DWD), and Dominant Wave Period (DWP) was appended to the data set. Ship A experienced significant wave heights in the range of 0.17 m to 6.21 m, dominant wave directions in all headings and dominant wave periods of range 2.12 s to 23.25 s. Parameters were derived using the Météo France WAve Model (MFWAM) (Lorente and

Piedracoba, 2017) which uses an irregular grid-based approach to obtain forecasted wave parameters with a temporal resolution of 3 hours and a spatial resolution of 0.083° x 0.083°. The model matched Ship A's position coordinates and time stamps to find the corresponding wave state. The parameters hold uncertainty as values are forecasted rather than real-time measurements; however, given the very short forecast period, they are deemed sufficiently close to actual conditions. The days since clean (DSC) parameter was created as an indicator of fouling. The little data available, only 12 months, meant a distinction between cleaning events could not be made. In reality, different cleaning events can have varying impacts on ship performance. However, an assumption was made that cleaning events were similar enough to be combined into a single variable, stating the days passed since a cleaning event:

$$\text{DSC} = \text{Date} - \text{Cleaning Date}. \tag{1}$$

2.3 Feature Selection

Feature selection methods were applied to reduce the dimensionality of the feature set and select the best features for the power model. A high dimensional feature set can cause data sparsity, leading to a model that overfits and is unable to generalise when making new predictions (de Mello and Ponti, 2018). Feature selection was undertaken to remove intercorrelated and redundant variables. This reduced the feature set's dimensionality while retaining features with the best predicting ability to improve predictive accuracy. Features were selected using Correlation analysis (Pearson coefficient), Recursive Feature Elimination (RFE) (Aurelien, 2017), and Extra Trees Regression (ETR) (James et al., 2013).

Fig. 1. provides a matrix of Pearson Correlation Coefficients between every variable in the data set, indicating the linear relationships present. Highly intercorrelated features, such as speed through water and speed over ground, were identified. Only one of these features is necessary and therefore speed over ground was removed from the data set. Shaft RPM was also removed as it was directly derived from power output. Redundant parameters with little to no correlation to propulsion power such as Number of Refrigeration units were further removed.

RFE and ETR methods were explored to select the appropriate features for modelling, as correlation analysis omits nonlinear relationships and is insufficient to select features alone. RFE was applied using a Decision Tree base model to determine the optimum number of features for modelling. RFE ranks features by recursively building models, in this case, Decision Trees, with different feature combinations to minimise a target error function. The method 'prunes' weaker features until the specified number of features remain. An initial RFE simulation, using a negative mean squared error target function, determined eight features would be optimum for power prediction. Fig 2 shows the mean error of multiple cross-validation simulations associated with different numbers of features included in the model, with the optimum indicated by the dotted line at eight features. The shaded area around the graph represents one standard deviation above and below the accuracy shown by the curve, to indicate the cross-validation variability, calculated as the deviation between cross-validation simulations. There is little difference in the error when choosing between 4 and 8 features, however selecting a larger number of features for RFE reduces computation time as fewer iterations are required to remove features. The RFE process was repeated to find the best combination of eight features to be used for modelling. Fig. 3. shows the features as ranked by RFE, where the most important features are given a rank of 1, and the remaining features are ranked such that a higher ranking denotes a lower importance in power prediction. These results were validated using ETR, which gives the importance of each feature in the model. Fig. 4. shows agreement with the RFE selection with a high importance denoting greater influence in the model. As a result of these methods, the following parameters were selected as input features for predictive modelling:

- Speed through water (knots)
- Relative wind speed (knots)
- Sea temperature (°C)
- Trim (m)
- Draft (m)
- FDN
- SWH (m)
- DSC (days)

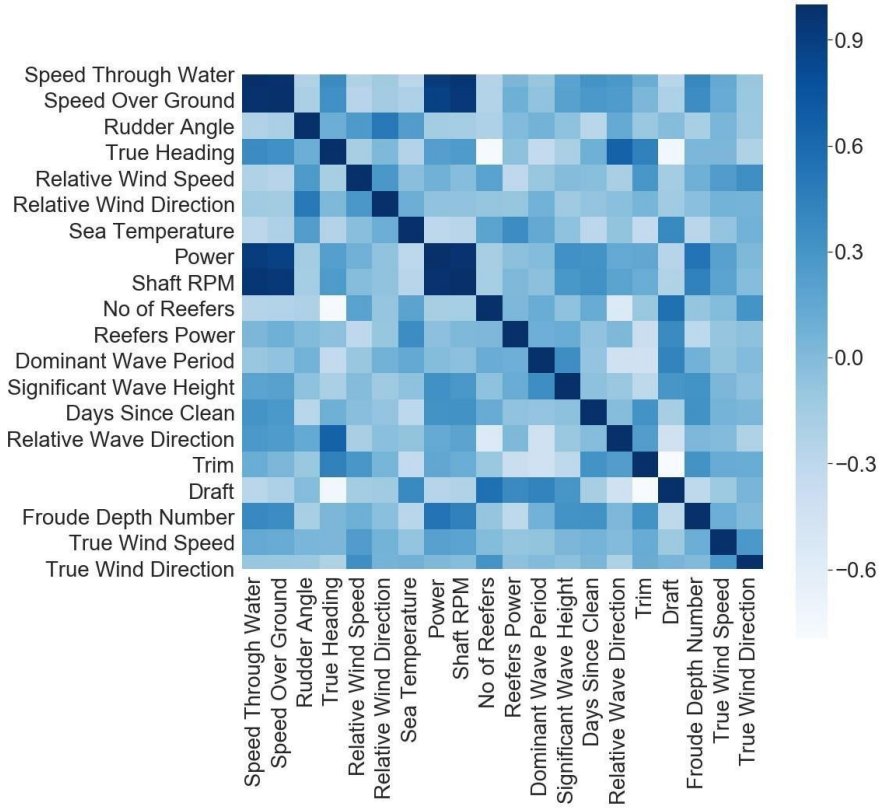

Fig. 1. Pearson correlation coefficient matrix for all variables.

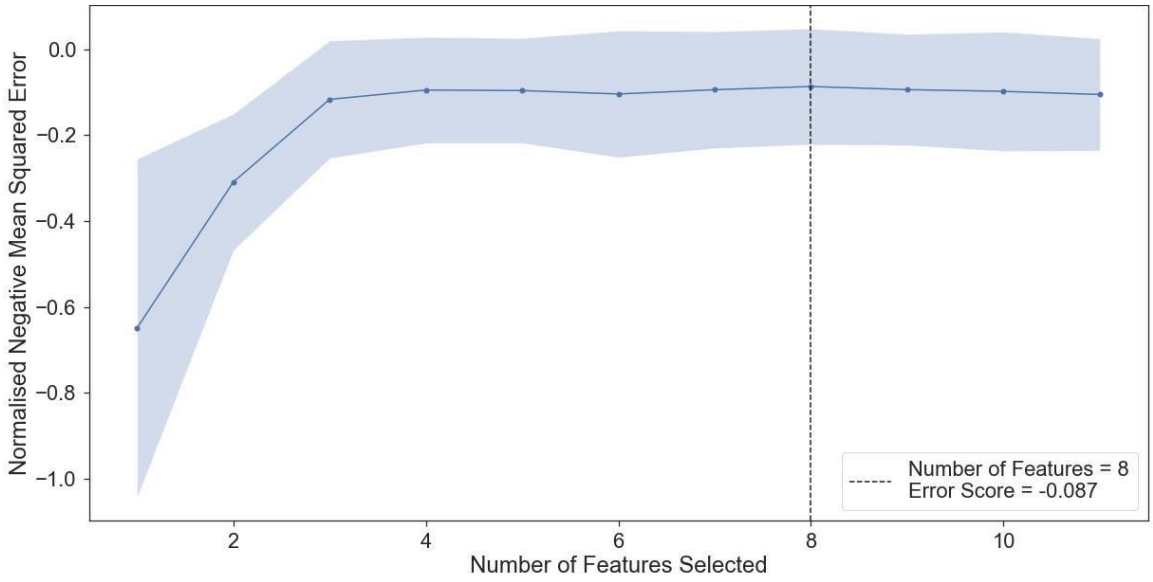

Fig. 2. Determination of the optimum number of input features using RFE. The variability of cross-validation is shown using a shaded section indicating a standard deviation about the mean score.

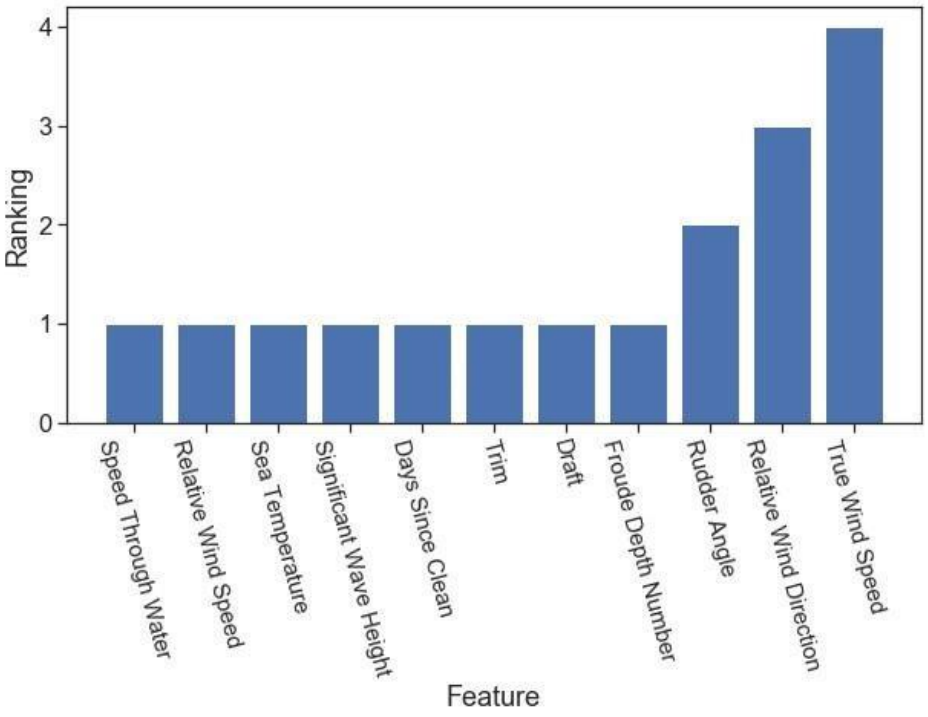

Fig. 3. Ranking of input features using RFE. Features given a rank of 1 are selected as inputs for the predictive model.

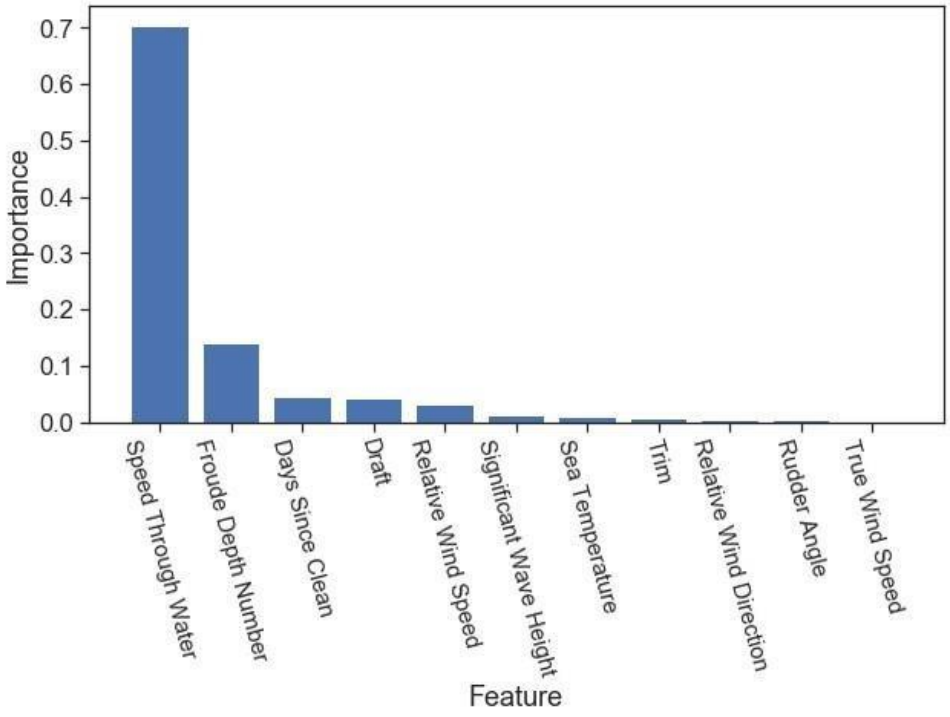

Fig. 4. Importance of input features in modelling using ETR. Features with a higher importance have a greater influence on prediction accuracy.

Fig. 5. shows histograms of all chosen features to show the operational profile of the ship and range of conditions experienced.

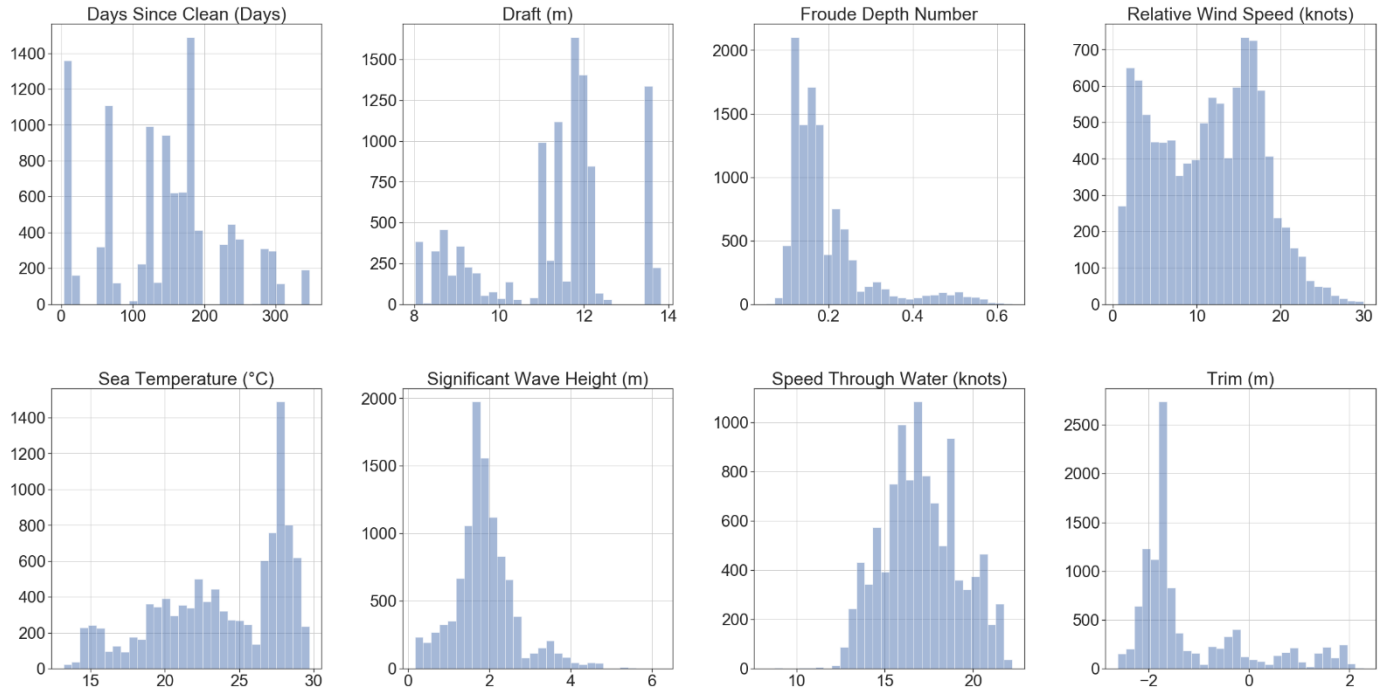

Fig. 5. Distributions of features selected for the modelling process, to indicate Ship A's operational profile with the y axis indicating the number of observations.

3. Regression Models

Four regression algorithms were evaluated for power prediction: KNN, Decision Tree (AdaBoost), Random Forest and ANN. Additionally, a multiple linear regression model was built as a baseline against which to compare the performance of each model. The best performing model was then used to investigate the effects of fouling and wave influences.

3.1 KNN

The KNN algorithm is an instance-based learner that makes predictions by observing the 'k' points closest to the input in the data space, using a specified distance metric (Aurelien, 2017). There are several methods of calculating the distance of a point to its 'k-nearest neighbours', with the Minkowski distance defined as:

$$d_M = \left(\sum_{i=1}^{n} (x_i - x^*{}_i)^P \right)^{\frac{1}{p}}. \quad (2)$$

When p = 2 this becomes the Euclidean distance:

$$d_E = \left(\sum_{i=1}^{n} (x_i - x^*{}_i)^2 \right)^{\frac{1}{2}}. \quad (3)$$

When the closest points are found, a weighted sum of the 'k neighbours' is taken to give the output prediction:

$$y = \frac{\sum_{i=1}^{k}(x_i)}{k}\,. \tag{4}$$

KNN has the potential to be used in problems with high nonlinearity as predictions are made based on similar known observations (Chaal, 2018). However, KNN can have issues with high dimensional data and requires features to be scaled to predict effectively.

3.2 Decision Tree

Decision Trees build a model by segmenting the feature space into regions (R, …, R_j) using recursive binary splitting to reduce the Residual Sum of Squares (RSS) (James et al., 2013), given as:

$$RSS = \sum_{j=1}^{j}\sum_{i\,\in R}\left(y_i - \bar{y}_{Rj}\right)^2\,. \tag{5}$$

The process of recursive binary splitting selects the appropriate predictor X_j and cut point s at each node, from all possible predictors and cut point values, to reduce the RSS (James et al., 2013). Mathematically, this defines a pair of half-planes:

$$R_1(j,s) = \{X|X_j < s\} \;\; and \; R_2(j,s) = \{X|X_j \geq s\}, \tag{6}$$

Where j and s are found to minimise the following:

$$\sum_{i:\, x_i \,\in R_1(j,s)}\left(y_i - \bar{y}_{R_1}\right)^2 \quad + \sum_{i:\, x_i \,\in R_2(j,s)}\left(y_i - \bar{y}_{R_2}\right)^2. \tag{7}$$

Splitting continues at each decision node along the branches until a termination criterion is reached, and the regions (R, .., R_j) are created. Predictions are made using the mean value of training observations in the region in which the test value belongs. Decision Trees alone are considered weak predictors and therefore, often combined via an ensemble method (Breiman et al., 2017). Combining trees with a boosting method, such as Adaptive Boosting (AdaBoost), trains a set of Decision Trees sequentially such that each tree is grown using information from trees grown previously (James et al., 2013).

$$\hat{f} = \sum_{b=1}^{B}\lambda\hat{f}_b(x)\,. \tag{8}$$

3.3 Random Forest

Random Forests use a bagging method to train trees in parallel. Random samples with replacement are selected from the training set, and trees are fitted to these samples, the results of which are pooled to create an average value (James et al., 2013):

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x). \tag{9}$$

At each split in a Random Forest tree, only a subset of predictor variables are considered. This decorrelates the Decision Trees, giving a robust model undominated by a single predictor variable (Breiman et al., 2017).

3.4 ANN

ANNs are black-box models which aim to emulate neurons in the brain. Multi-layer perceptrons are a class of ANN which have three layers of nodes as a minimum: an input, hidden, and output layers (Aurelien, 2017). Each neuron connects to every neuron in the next layer. Each connection between neurons has an associated weight, which determines its activation, creating a weighted sum to which a bias term is added. The output of a given neuron can be described as a function of the activation, with the sigmoid function often used to 'squash' the weighted sum (Dreyfus, 2005):

$$y = sig\left(\sum_{i \in I} w_i x_i\right), \tag{10}$$

Where w_i are the weights of each connection, x_i the neuron inputs, f is the activation function and i∈

I, where I = {0, …, N-1}.

Multi-layer perceptrons are feedforward networks, which learn through the alteration of their weights via the back-propagation algorithm (Annema, 1995). ANNs can exploit both linear and nonlinear relationships between predictor variables and outputs in prediction. However, they usually require a large amount of data to provide high prediction accuracy (Aurelien, 2017).

4. Modelling Approach

4.1 Power Prediction Modelling

The modelling approach used for this work is given in Fig. 6.

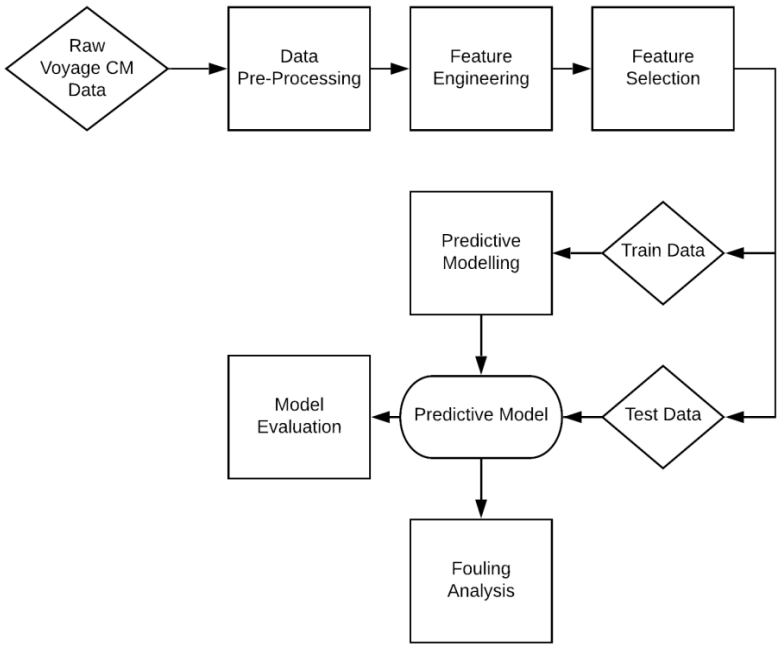

Fig. 6. Flow diagram outlining the data mining and modelling approach taken in this work.

The Ship A data was divided randomly, without replacement, into a training set and an unseen testing set, with 70% and 30% split between the two sets, respectively. The propulsion power was given as the target output for modelling, and the features selected using RFE and ETR taken as inputs to each model. Standardisation was applied after splitting to ensure that no prior bias was introduced to the test set. Before training, the hyperparameters for each model were tuned using a grid search algorithm. Grid search iteratively determines

the best combination of model parameters to give optimal model performance. The hyperparameters selected for each model are given in Tables 1-4. Due to the low amount of training data available after pre-processing, all models were trained using ten-fold cross-validation to ensure good generalisation and maximise the efficiency of data usage.

Table 1. Hyperparameters for Random Forest.

Random Forest	
Maximum Features	Square Root N Features
Number of Estimators	1000
Maximum Depth	50
Minimum Samples Split	2
Minimum Leaf Samples	1
Bootstrap	False

Table 2. Hyperparameters for Decision Tree.

Decision Tree with Adaboost	
Learning Rate	1
Number of Estimators	10
Loss	Linear

Table 3. Hyperparameters for KNN.

KNN	
Number of Neighbours	7
Weight	Distance
Distance Metric	Minkowski

Table 4. Hyperparameters for ANN.

ANN	
Activation Function	ReLU
Hidden Layer Size	100
Number of Epochs	1000
Learning Rate	Constant
Solver	Limited-memory Broyden Fletcher Goldfarb Shanno (LBFGS)

The Root Mean Square Percentage Error (RMSPE) and Mean Absolute Percentage Error (MAPE) were chosen as evaluation metrics to indicate model performance on the testing set, enabling direct comparison with previous work. The RMSPE and MAPE can be calculated as:

$$RMSPE = \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{(\hat{y}_i - y_i)^2}}{y_i} \times 100, \tag{11}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)}{y_i} \times 100. \tag{12}$$

Where n denotes the number of observations, y denotes the actual value of propulsion power and $\hat{y}$ denotes the predicted value of propulsion power.

The lower the RMSPE and MAPE values, the more accurately the model can predict propulsion power. The model with the lowest errors was taken forward to investigate fouling and wave effects.

4.2 Cleaning and Environmental Effects

The influence of the parameters DSC and SWH on power prediction accuracy was investigated by incorporating each feature as an additional input to the predictive model. Statistical testing determined whether the mean of the predictive errors of each model was lower than the mean of the predictive errors of a base model without either parameter. The non-parametric Wilcoxon Signed Rank test was selected to test significance as the distribution of model errors would be paired but non-gaussian as indicated by a Shapiro-Wilks test for normality. The Wilcoxon Signed-Rank test determines whether there is a difference in the means of two samples (Wilcoxon, 1945). The individual differences, D_i, of the samples are taken:

$$D_i = |x_{2,i} - x_{1,i}|. \tag{13}$$

The differences, D_i, are then ranked in order of smallest absolute difference to largest absolute difference, and the rank of a pair denoted as R_i, such that:

$$R_i = rank(D_i). \tag{14}$$

Cases of $D_i = 0$ were excluded in accordance with the original Wilcoxon method (Wilcoxon, 1945), resulting in a reduced sample size denoted by n_r. An alternative approach by Pratt (Pratt, 1959) is more conservative and includes these differences in the ranking process. The Wilcoxon test statistic W^+ is then calculated:

$$W^+ = \sum_{i=1}^{n_r} R_i \, sgn(x_{2,i} - x_{1,i}). \tag{15}$$

The W^+ value is compared to a critical value given by $W_{critical,\,nr}$ which can be found using the Wilcoxon Signed-Rank test table of critical values. H_0 is then rejected if $W^+ > W_{critical,\,nr}$. An associated p-value can also be calculated, and if this is smaller than the given significance level, then the null hypothesis can be rejected.

The errors of each model, including DSC and SWH, were compared to the errors of a base model, testing at a significance level $\alpha = 0.05$ the following hypotheses:

$$H_0: \mu_P = \mu_X\,, \tag{16}$$

$$H_1: \mu_P > \mu_X\,, \tag{17}$$

Where μ_P is the mean of the base model errors and $\mu_X = \mu_{DSC}$ or μ_{SWH}, the mean of the models including DSC or

SWH, respectively.

4.3 Fouling Analysis

Fouling related performance deterioration was determined in terms of effective power increase by making predictions on a synthetic test set given in Table 5 with increasing DSC. All parameters were held constant except DSC, which was adjusted between 0 and 360 days. Predictions were made for a range of speeds between 14 and 22 knots to build a set of simulated power-speed curves at different DSC values.

Table 5. Prediction Parameters for Fouling Analysis.

Prediction Set – Fouling	
Speed through water (knots)	$14 \leq v \leq 22$
DSC (days)	$0 \leq DSC \leq 360$
Draft (m)	9
Trim (m)	-1
FDN	0.33
Sea Temperature (°C)	23
Significant Wave Height (m)	1.8
Relative Wind Speed (m/s)	10

These curves were created by applying a cubic fit to the prediction points at a given DSC to emulate the empirical power-speed relationship:

$$P = \frac{1}{2}\rho S v^3 C_T, \tag{18}$$

Where P is the propulsion power, ρ the water density, S the submerged area of the ship, v the ship velocity through the water and C_T the total resistance coefficient.

The effective power increase due to fouling was measured as the offset between curves at differing DSC. The offset could be equated to an excess fuel usage per day to find the consequent additional fuel cost associated with this increase:

$$\Delta C = C_{fuel} \times SFOC \times \Delta P \times t, \tag{19}$$

Where ΔC is the cost increase due to fouling effects, C_{fuel} is the fuel cost, t is time in operation, ΔP is the increase in power after a given number of DSC and SFOC is the specific fuel oil consumption of the ship's main engine.

5. Results and Discussion

5.1 Power Prediction Modelling

The results obtained for each model's predictions on the unseen test set are given in Table 6. The consistently low errors obtained indicate that all models were able to generalise well. The best performing model was the Random Forest, achieving the lowest errors of all models with an RMSPE of 0.0264% and a MAPE of 1.17%. However, both the KNN and ANN also achieved low errors, with RMSPE of 0.0302% and 0.0317% respectively.

Table 6. RMSPE and MAPE for All Models.

Model	RMSPE (%)	MAPE (%)
Random Forest	0.0264	1.171
KNN	0.0302	1.245
ANN	0.0317	1.893
Linear Regression	0.0930	6.453
Decision Tree	0.0932	6.987

The results imply that nonlinear relations are present between the input features and power output, enhancing the predictive ability of these models compared to linear models such as Linear Regression, which is unable to pick up such trends and therefore displays a lower predictive accuracy. The low KNN errors are likely due to similar operating conditions giving similar power outputs, and so a point will be analogous with neighbouring points in the feature space. The Decision Tree (AdaBoost) has significantly higher errors than the Random Forest, which are as high as the 7% obtained by Pedersen and Larsen's ANN using NR data (Pedersen and Larsen, 2009). The performance difference between the two tree-based models can be explained by how each model is trained. The Decision Tree is likely dominated by a single input feature at split points and uses this variable predominantly to give an output prediction. However, such domination by one feature is not possible in the Random Forest model as it uses random combinations of input features to make decisions at split points (James et al., 2013). Therefore, while the Random Forest model uses this dominating feature, it is enhanced by other features in the predictor set, leading to a lower predictive error. While ANNs have shown superior performance to other models in past studies (Parkes et al., 2018), the ANN performance was affected by the limited amount of data available for training, due to the extensive data cleaning undertaken. ANNs require many more observations to produce accurate models than available in this study (Parkes et al., 2019). Fig. 7. shows the true and predicted values on the unseen test set of power for each model. The straight lines shown by the Random Forest, KNN and ANN show low deviation between the predicted and true values. In contrast, the Linear Regression model appears to over-predict power values as indicated by the arching of the graph, and the Decision Tree discretises power levels.

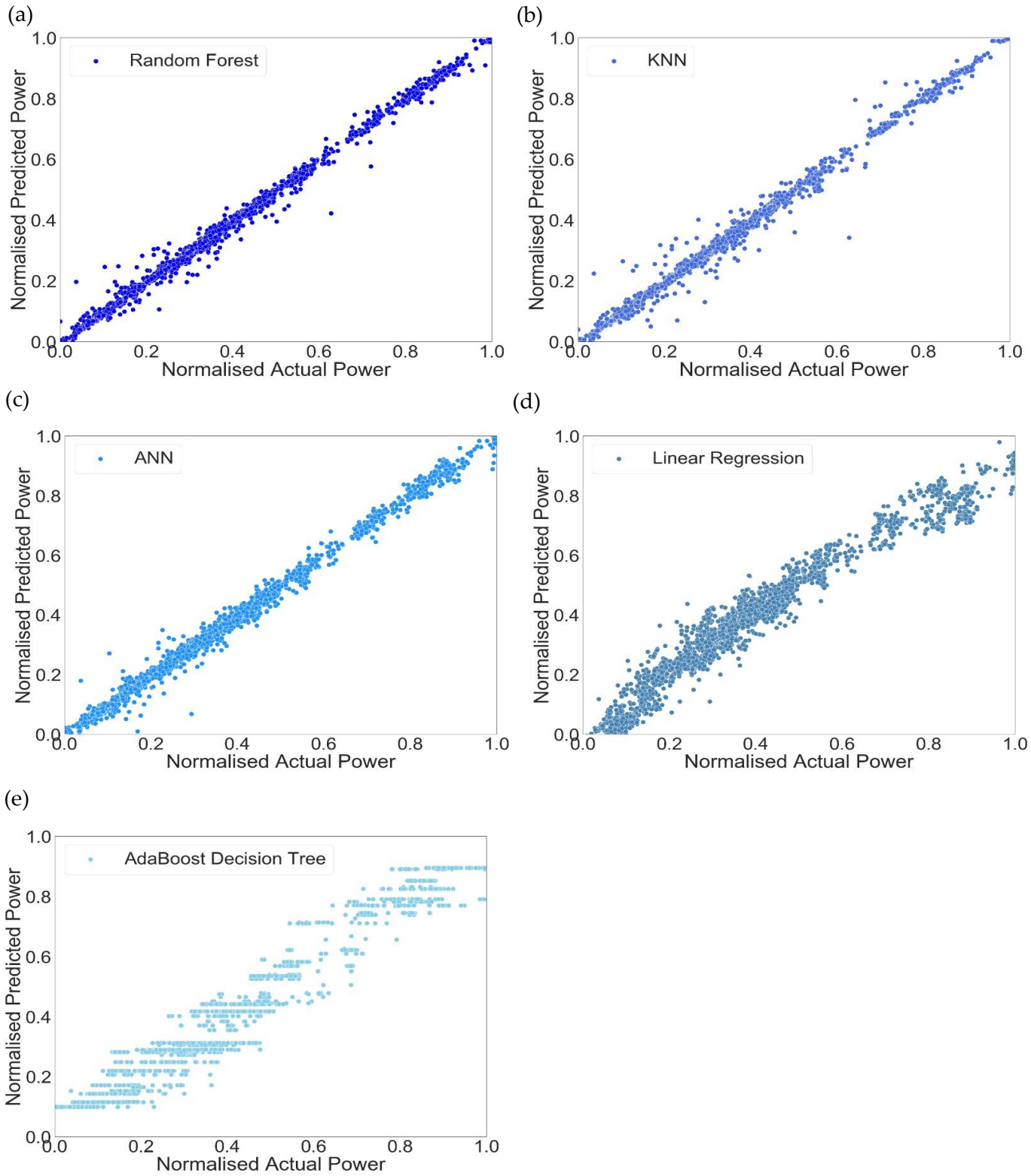

Fig. 7. True power against predicted power for: (a) Random Forest (b) KNN (c) ANN (d) Linear Regression (e) AdaBoost Decision Tree.

Care must be taken when directly comparing these results to previous studies as a different data set was used. Nevertheless, the Random Forest model achieved higher predictive accuracy than comparable studies shown in Table 7. The improved performance over Petersen's ferry study (Petersen et al., 2012) is particularly surprising

as a ferry follows a repeated route and has much lower operational variability than a container ship. However, the high accuracy achieved in this study can largely be attributed to filtering power values over a 7000 kW threshold while using RFE and ETR to select input features with the most predictive potential alongside hyperparameter optimisation. This highlights the importance of pre-processing before modelling.

Table 7. Comparison of Model Performance to Previous Studies.

Study	Model	Error (%)
Current Study: Ship A Data set	Random Forest	1.17
Parkes (Parkes et al., 2019)	ANN	1.70
Pedersen (Pedersen and Larsen, 2009)	ANN	1.65
Petersen (Petersen et al., 2012)	ANN	1.50

The lower error obtained than in previous studies demonstrates the potential of the Random Forest model to be used within industry tools, such as voyage planning and decision support systems. The model could provide more accurate predictions than the naval architecture models currently used, which have errors of around 5% (Lu et al., 2015). However, it should be noted that the model is limited to predicting travel in open waters due to the filtering threshold applied, rendering it unsuitable for prediction of conditions such as shallow water, docking and manoeuvring unless a significant amount of data is available for these conditions.

5.2 Cleaning and Environmental Effects

The results obtained using the Random Forest model with DSC and SWH included independently are shown in Table 8. Including DSC led to reductions in the RMSPE and MAPE of 0.0005% and 0.07% respectively. Including SWH reduced the RMSPE and MAPE by 0.0008% and 0.12%.

Table 8. RMSPE and MAPE for the DSC and SWH Models.

Model	RMSPE (%)	MAPE (%)
Base Model	0.0264	1.171
DSC Model	0.0259	1.101
SWH Model	0.0256	1.048

For the DSC variable, a p-value of 4.24×10^{-10} was obtained from the Wilcoxon significance test, which was far smaller than the specified significance level. Therefore, the null hypothesis could be rejected, and the addition of the DSC variable proved to be a statistically significant feature in the power prediction model. However, the improvement when incorporating DSC, while significant, is small and requires obtaining potentially confidential cleaning information. Despite this, the result shows that DSC does improve the predictive ability of a model, indicating that the model can distinguish a change in power output due to cleaning, which further implies that fouling has a notable influence on ship performance.

The addition of SWH was also statistically significant, with a p-value of 4.26×10^{-46} obtained. The improvement in predictive accuracy of 0.12% is much smaller than the 0.5% found by Parkes *et al.* when adding a similar wave height metric (Parkes et al., 2019), likely because of the difference in modelling approaches used and limited data size. Nevertheless, the addition of SWH improved model accuracy more than DSC, indicating that environmental effects have a greater impact on power prediction, and therefore ship performance, than cleaning.

However, unlike cleaning, shipping companies cannot control wave effects. Avoiding adverse wave conditions would require advanced voyage planning and could result in disruptions to shipping schedules.

The results obtained by the model, including DSC, were shown to be reproducible by applying the same process on sister ship data (Table 9). Wave data was only available for Ship A, which eliminated potential to compare all sister ships using the SWH model. The slight discrepancies between the sister ship results may be due to varying amounts of data available and some ships experiencing a more extensive range of operating and environmental conditions. The higher errors obtained for Ship D and E suggest slight overfitting; however, these errors are still low. These results increase confidence in the method for predicting propulsion power across multiple ships.

Table 9. Model Results on Sister Ships.

Ship	RMSPE (%)	MAPE (%)
Ship A	0.026	1.10
Ship B	0.050	1.72
Ship C	0.035	1.38
Ship D	0.066	2.74
Ship E	0.072	2.37

5.3 Fouling Analysis

Using a refined model incorporating both DSC and SWH with a MAPE of 1.02%, fouling effects were explored with results given as power-speed curves in Fig. 8. Diving reports from Ship A's cleaning events confirmed there was light hull fouling and 60% fouling coverage on the propeller; therefore some performance deterioration, as seen, was expected. There is a clear distinction between the curves for the clean (60 DSC) and fouled ship (360 DSC), showing an average increase in power requirement of ~5.2% across all speeds for the given conditions. As all other parameters were held constant, including the sea state, the change between the curves was a result of a change in the DSC parameter. The increase in power due to lack of cleaning, and therefore fouling, aligned with expectation (Rawson and Tupper, 2001) and was verified by calculating the corresponding power increase across the ship's full operational profile. An average power increase due to fouling of ~3% over a year is expected by shipping companies; therefore, the 5.2% average power increase observed appears reasonable.

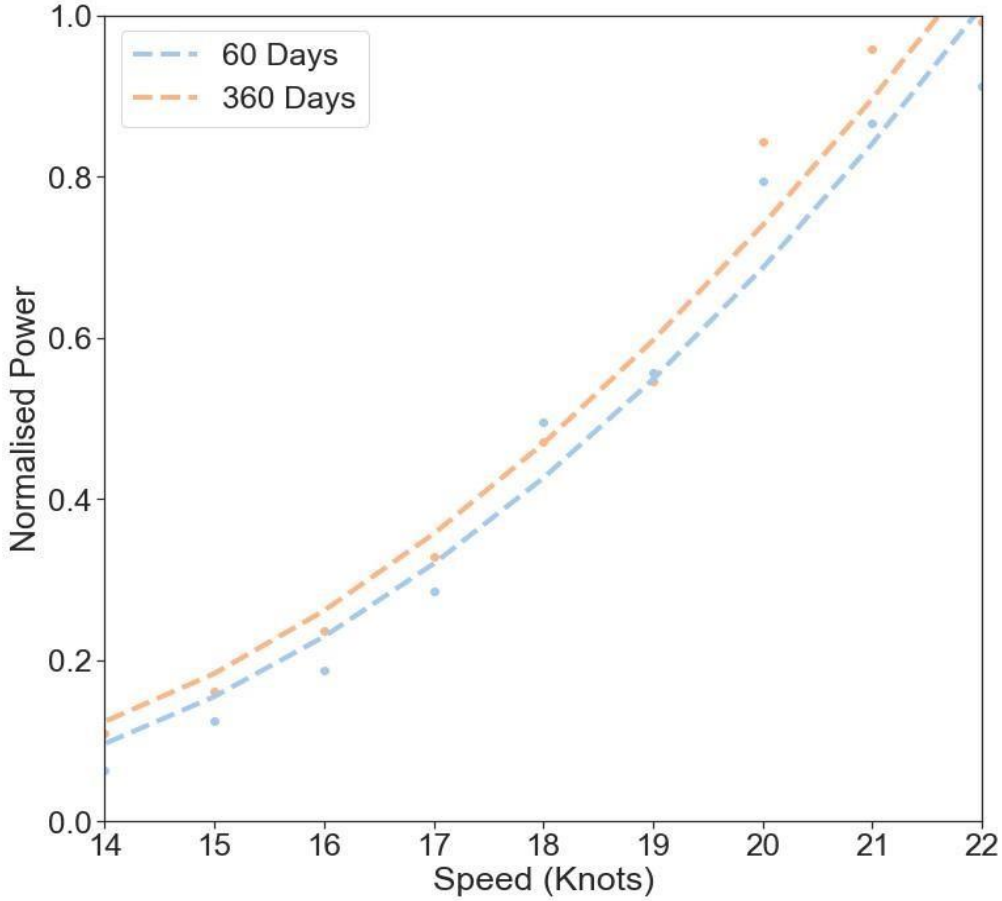

Fig. 8. Speed-power curves for 60 and 360 DSC, showing an offset of ~5.2%.

Previous empirical methods by Uzun *et al.* (2019) found an effective power increase of 25% over three years due to fouling, while Schultz (2007) predicted a range of 4-59% power increase at 30 knots depending on fouling severity. Demirel *et al's* (2017) simulations found an 18.1% power increase at 24 knots for a lightly slime fouled ship. Predictions have been made at a lower speed range compared to previous studies, and so the average increase of ~5.2% across speeds from 14 knots to 22 knots appears to fit in with the lower-end estimates. Therefore, this method provides an attractive approach for identifying performance deterioration due to fouling, using real-world data as opposed to simulation or empirical models.

While the magnitude of the power increase due to fouling appears relatively small, it could have significant financial ramifications for a shipping company. Ship A has a MAN-B&W 9S90ME-C8 main engine, with an SFOC of 0.184 kg/kW when operating at 80% MCR. Assuming the ship operates 24 hours a day (Armstrong, 2013), using Heavy Fuel Oil (HFO) IFO 380 costing 0.42 £/kg (April 2020) (Ship and Bunker, 2020.), the additional fuel used per day due to the fouling related power increase is ~5868.86 kg. This equates to an added daily cost of approximately £2500. Shipping companies can monitor the cumulative additional fuel cost, comparing it to the cost of undertaking a cleaning procedure, to determine when it is financially efficient to clean their ships. The cost of halting shipping operations during cleaning must also be considered (Armstrong, 2013). Therefore, shipping companies would likely wait until the performance deterioration due to fouling gives a considerable cost rise due to the additional power requirements before undertaking a cleaning procedure.

The evolution of power increase due to fouling can be tracked as more data is recorded. Fig. 9. was created by training models with 10, 11 and 12 months of data after a cleaning event and using each model to make predictions at a speed of 22 knots with increasing DSC. Linear regression was applied to each set of predictions, assuming a simple linear relationship between DSC and power, giving gradients that indicate power increase

due to fouling as a percentage above a base operating power. The gradients are built up from 10 months of data onwards as this is when the error is small enough to yield reliable predictions. These gradients evolve, with 12 months of data after a cleaning event giving a steeper gradient and therefore higher power increase due to fouling than prior months. An operator could monitor the gradient evolution, which would be expected to settle as more data is added, using it to continuously evaluate fouling severity, reassess the associated power increase cost and determine whether to carry out cleaning. This regression approach gives a clear profile of fouling effects on power, however, could be considered oversimplified, assuming a constant relationship between fouling and power. Marine growth and its resistive effects are unlikely to vary linearly and so a more sophisticated approach may improve the treatment of fouling influence on performance.

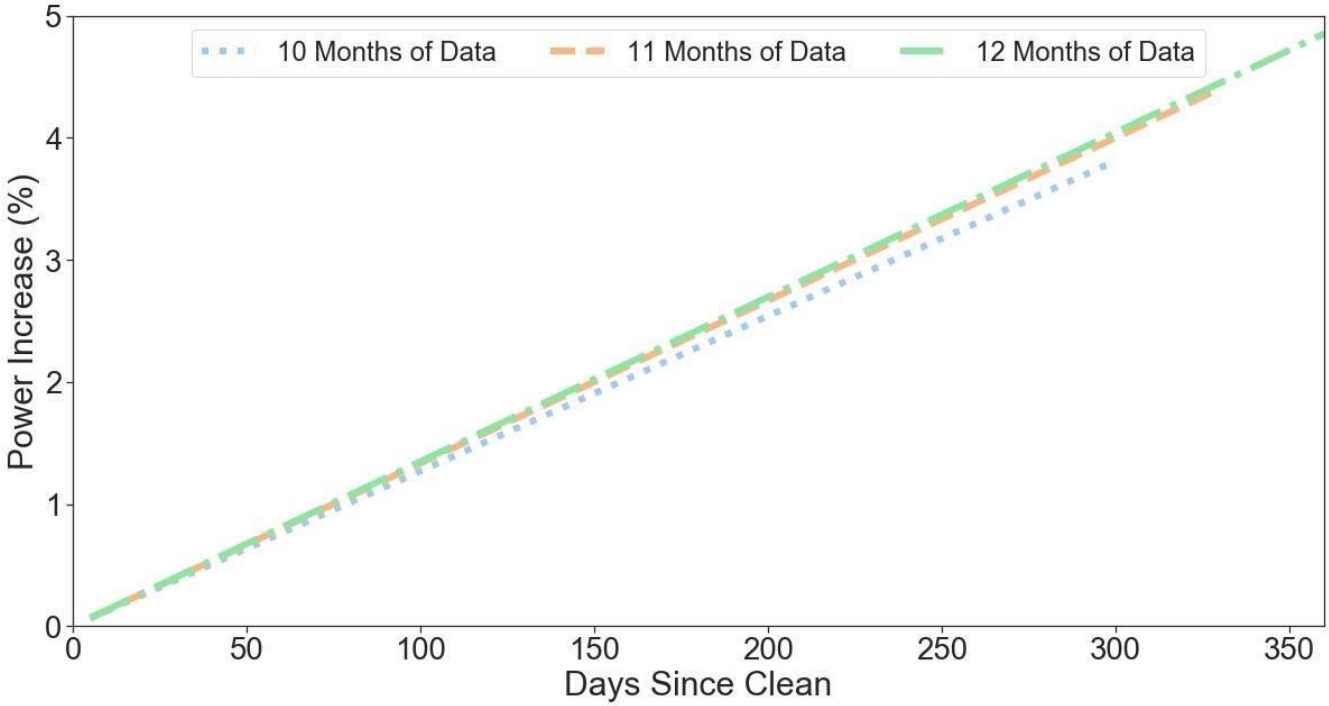

Fig. 9. Gradients to show percentage power increase due to fouling above a base operating power predicted with 10, 11 and 12 months of data.

6. Conclusions

This study used ML techniques to accurately predict power and performance degradation due to fouling, to assess when hull and propeller cleaning should be undertaken. A set of prospective ML models were evaluated for power prediction, with the Random Forest achieving the lowest error of all models investigated at 1.17%. A low error was achieved due to in-depth feature selection using techniques including RFE and intensive filtering, demonstrating the importance of pre-processing in predictive modelling. The statistical significance of the parameters DSC and SWH were demonstrated to improve model accuracy by 0.07% and 0.12% respectively, combining to produce a model with an error of 1.02%. Simulated speed trials showed an effective power increase on average of 5.2% after a year since cleaning, which equates to ~£2500 in excess fuel per day.

A higher accuracy was achieved compared to previous approaches, with sufficiently low error to assess fouling effects, demonstrating the potential of tree-based methods in performance prediction. The developed model has the potential for integration into voyage planning and operational guidance tools to reduce fuel consumption through parameter optimisation, replacing current naval architecture models. The subsequent fouling approach provided an attractive alternative to current empirical and CFD methods, determining the power increase due to

fouling using real-world data. The power increase can be continuously assessed and equated to an excess fuel cost, allowing shipping companies to determine when it is financially viable to undertake cleaning. Using ML techniques to monitor fouling related performance deterioration provides a means to aid shipping companies in deciding when to clean container ships; helping to reduce power requirements, fuel usage and cost.

There is scope for future work to further develop the approach and directly benefit shipping companies. The addition of wave enrichment data could be explored further to exploit improved predictive accuracy using environmental data and identify conditions with severely detrimental effects on performance that could be avoided in voyage planning. An analysis of which features cause increased fouling levels could be explored further, including factors such as sea temperature and sea state. Obtaining data over a larger time frame would allow distinction between cleaning events, enabling direct comparison of pre- and post-cleaning performance to better quantify fouling effects. More data may also enable the inclusion of shallow water observations to widen the prediction field. Further work could implement this model into voyage planning tools and operational guidance methods in place of dynamic models.

References

Akinfiev T., Janushevskis A., Lavendelis E. 2007. A Brief Survey of Ship Hull Cleaning Devices. Transport and Engineering Mechanics, 1407, pp. 133–146.

Aldous L., Smith T., Bucknall R., Thompson P. 2015. Uncertainty analysis in ship performance monitoring. Ocean Engineering, 110, pp. 29–38.

Annema, A.J. 1995. Feed-Forward Neural Networks, Vector Decomposition Analysis, Modelling and Analog Implementation. Norwell: Kluwer Academic Publishers.

Armstrong, V.N. 2013. Vessel optimisation for low carbon shipping. Ocean Engineering, 73, pp. 195– 207.

Aurelien, G. 2017. Hands-on Machine Learning with Scikit-Learn & TensorFlow. Sebastopol: O'Reilly. pp. 33-79.

Bal Beşikçi E., Arslan O., Turan O., Ölçer A.I. 2016. An artificial neural network-based decision support system for energy efficient ship operations. Computer Operational Research, 66, pp. 393–401.

Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 1984. Classification and Regression trees. New York: Chapman & Hall.

Cames M., Graichen J., Siemons A., Cook V. 2015. Directorate General for Internal Policies Policy Department A: Economic and Scientific Policy Emission Reduction Targets for International Aviation and Shipping. [online]. Available from: http://www.europarl.europa.eu/RegData/etudes/STUD/2015/569964/IPOL_STU(2015)569964_E N.pdf. [Accessed 27 January 2019].

Chaal, M. 2018. Ship Operational Performance Modelling for Voyage Optimization through Fuel Consumption Minimization. Thesis (Masters), World Maritime University.

Dagkinis I., Nikitakos N. 2015. Slow Steaming Options Investigation Using Multi Criteria Decision Analysis Method. Econship, pp 1–15.

Demirel Y.K., Turan O., Incecik A. 2019. Predicting the effect of biofouling on ship resistance using CFD. Applied Ocean Research, 62, pp. 100–118.

Demirel Y.K., Uzun D., Zhang Y., Fang H.C., Day A.H., Turan O. 2017. Effect of barnacle fouling on ship resistance and powering. Biofouling, 33, pp. 819–834.

Dreyfus, G. 2005. Neural Networks Methodology and Applications. 2nd ed. Heidelberg: SpringerVerlag.

Fernandes de Mello R., Antonelli Ponti M. 2018. Machine Learning A Practical Approach on the Statistical Learning Theory. Cham: Springer International Publishing.

IMO. 2018. Initial IMO Strategy on Reduction of GHG Emissions from Ships, Resolution MEPC. 304(72).

James G., Witten D., Hastie T., Tibshirani R. 2013. An introduction to statistical learning : with applications in R. London: Springer. pp. 303-330.

Lorente P, Piedracoba S. 2017. The new CMEMS IBI-WAV forecasting system : skill assessment using in situ and HF radar data. 8th EuroGOOS Conference, 3-5 October 2017, Bergen.

Lu R., Turan O., Boulougouris E., Banks C., Incecik A. 2015. A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. Ocean Engineering, 110, pp. 18–28.

Parkes A.I., Savasta T.D., Sobey A.J., Hudson D.A. 2019. Efficient vessel power prediction in operational conditions using machine learning. The 14th International Symposium on Practical Design of Ships and Other Floating Structures, 22 - 26 September, Yokohama: Springer. pp. 1-18.

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

Pedersen B.P., Larsen J. 2009. Prediction of Full-Scale Propulsion Power using Artificial Neural. 8th International Conference on Computer and IT Applications in the Maritime Industries, 10-12 May 2009. Budapest: TuTech. Pp.537-550.

Petersen J.P., Jacobsen D.J., Winther O. 2012. Statistical modelling for ship propulsion efficiency. Journal of Marine Science and Technology, 17, pp. 30-39.

Pétursson, S. 2009. Predicting Optimal Trim Configuration of Marine Vessels with Respect to Fuel Usage. Thesis (Masters). University of Iceland.

Pratt, J.W. 1959. Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures. Journal of the American Statistical Association, 54, pp. 655–667.

Rawson K.J., Tupper E.C. 2001. Basic Ship Theory 2. 5th ed. Oxford: Butterworth - Heinemann.

Schultz, M.P. 2007. Effects of coating roughness and biofouling on ship resistance and powering. Biofouling, 23(5), pp. 331–341.

Ship and Bunker. 2020. News and Intelligence for the Maritime Industry: World Bunker Prices [online] 13th April 2020. Available from: https://shipandbunker.com/prices#IFO380. [Accessed 13 August 2020].

Soner O., Akyuz E., Celik M. 2018. Use of tree-based methods in ship performance monitoring under operating conditions. Ocean Engineering, 166, pp. 302–310.

Song S., Demirel Y.K., Atlar M. 2020. Penalty of hull and propeller fouling on ship self-propulsion performance. Applied Ocean Research, 94.

Terazono E., Hume N. 2017. New Shipping Fuel Regulation Set to Hit Commodities. Financial Times. [online]. Available from: https://www.ft.com/content/d0ae63c4-452f-11e785199f94ee97d996. [Accessed 27 January 2019].

Uzun D., Demirel Y.K., Coraddu A., Turan O. 2019. Time-dependent biofouling growth model for predicting the effects of biofouling on ship resistance and powering. Ocean Engineering, 191.

Wilcoxon, F. 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin, 1, pp. 80–83.

Figure 1

Click here to download high resolution image

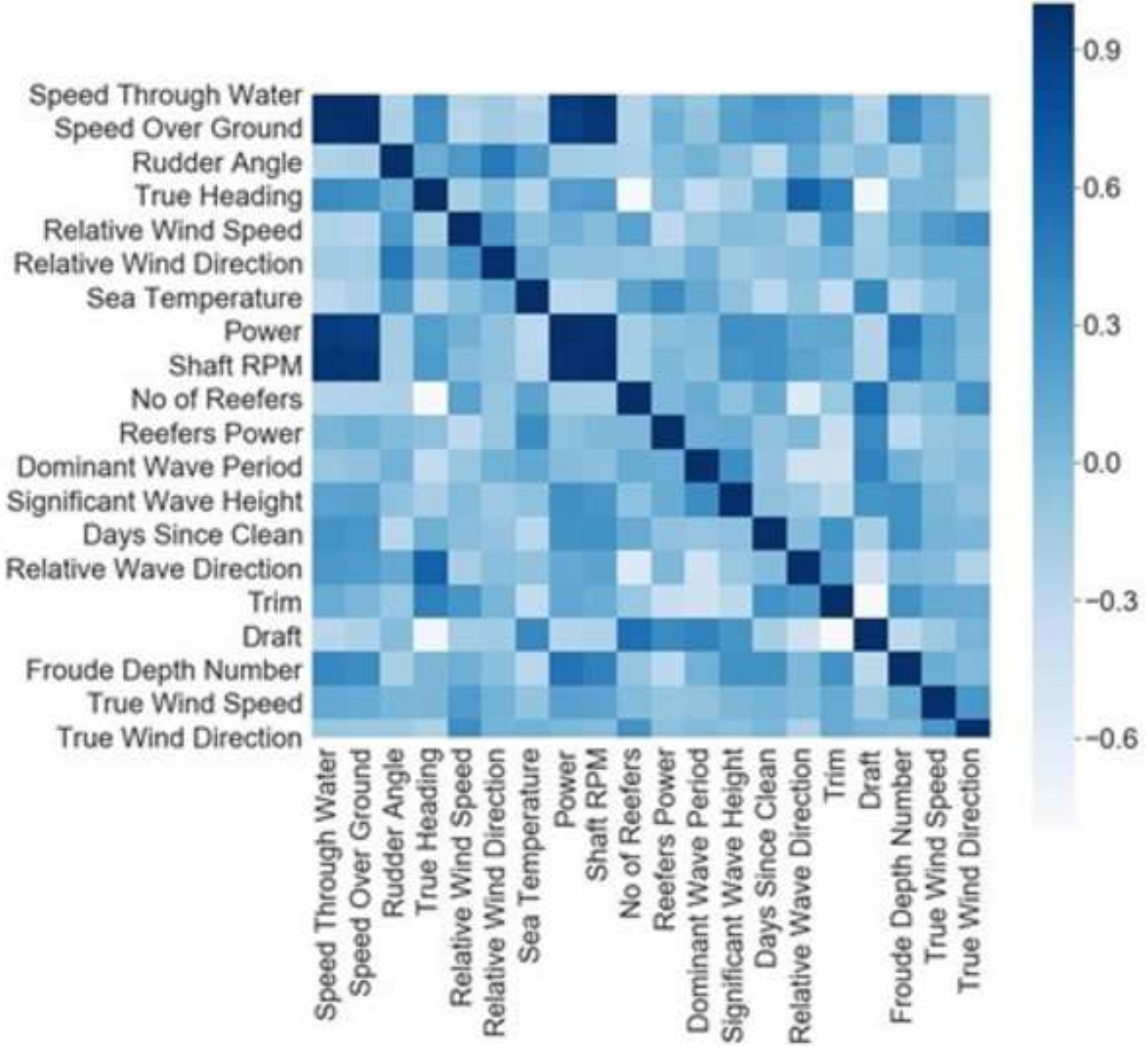

Figure 2

Click here to download high resolution image

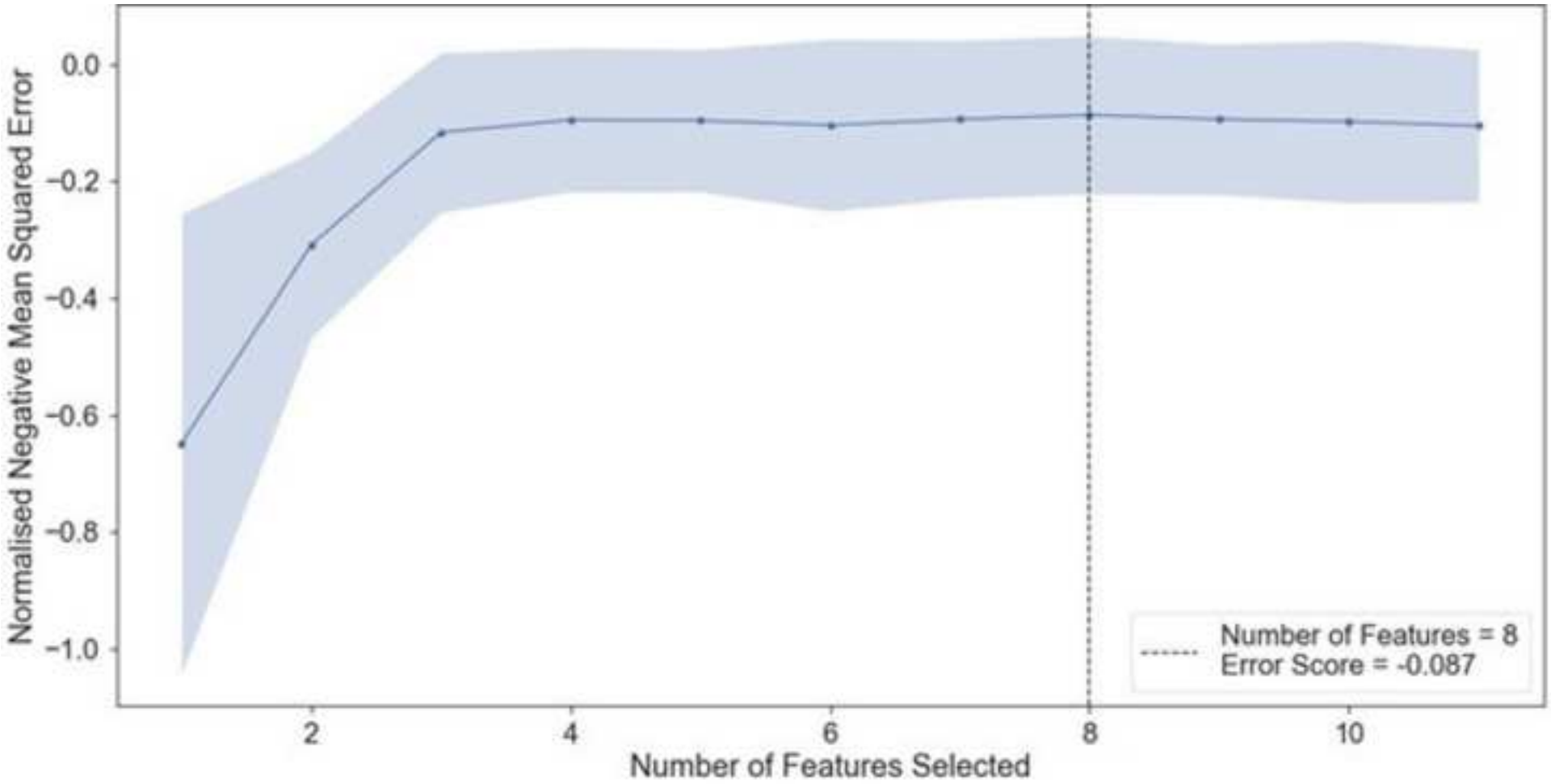

Figure 3

Click here to download high resolution image

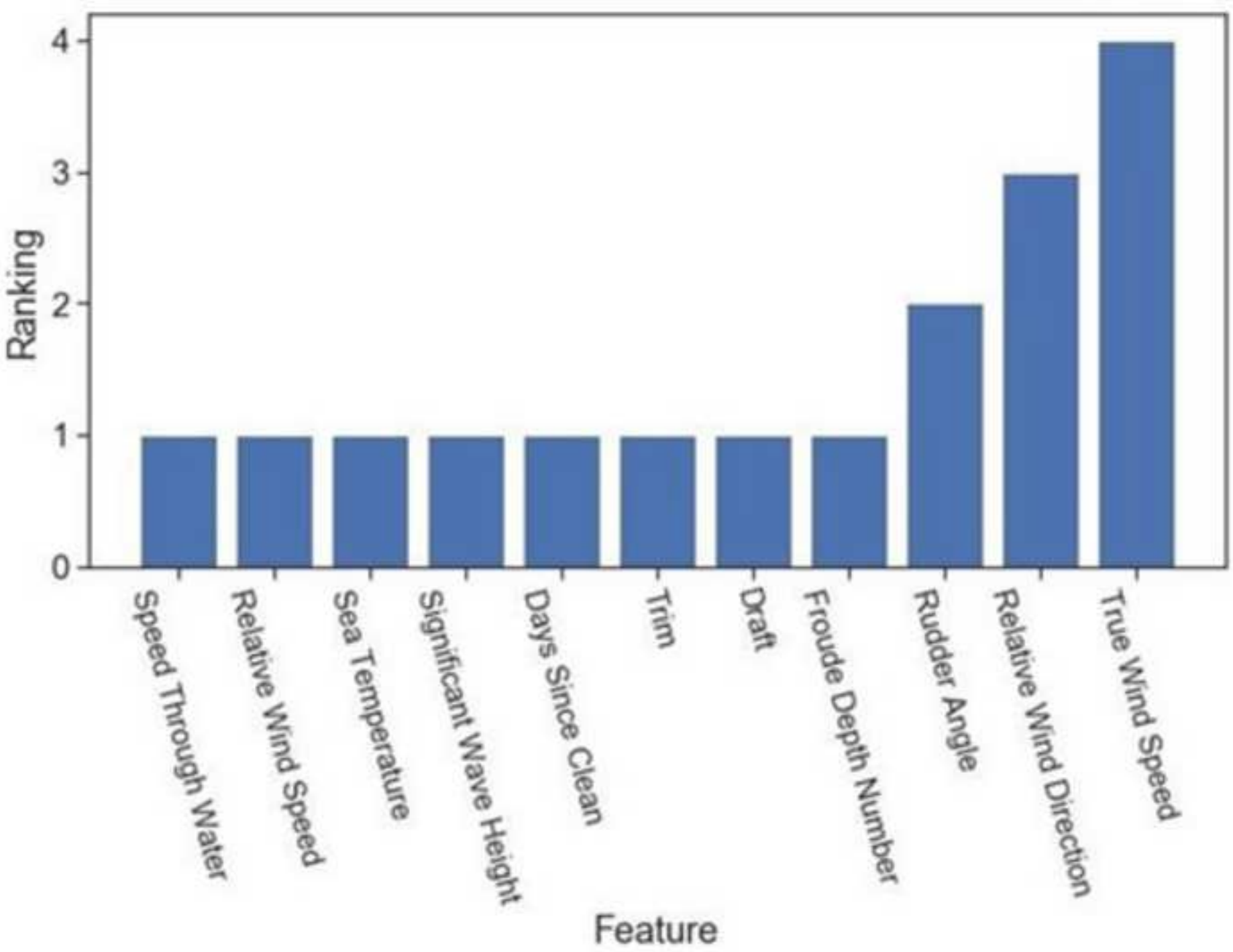

Figure 4

Click here to download high resolution image

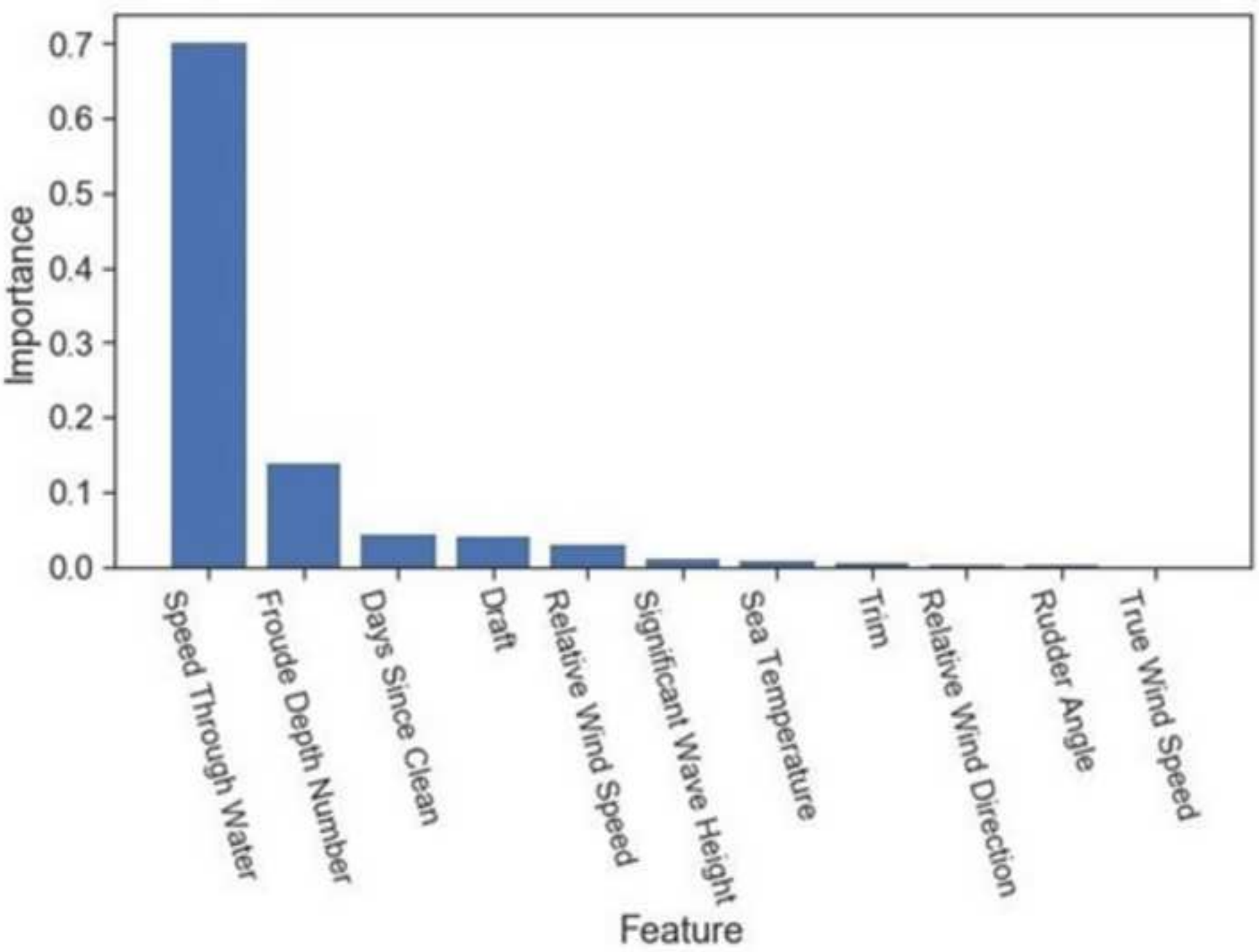

Figure 5

Click here to download high resolution image

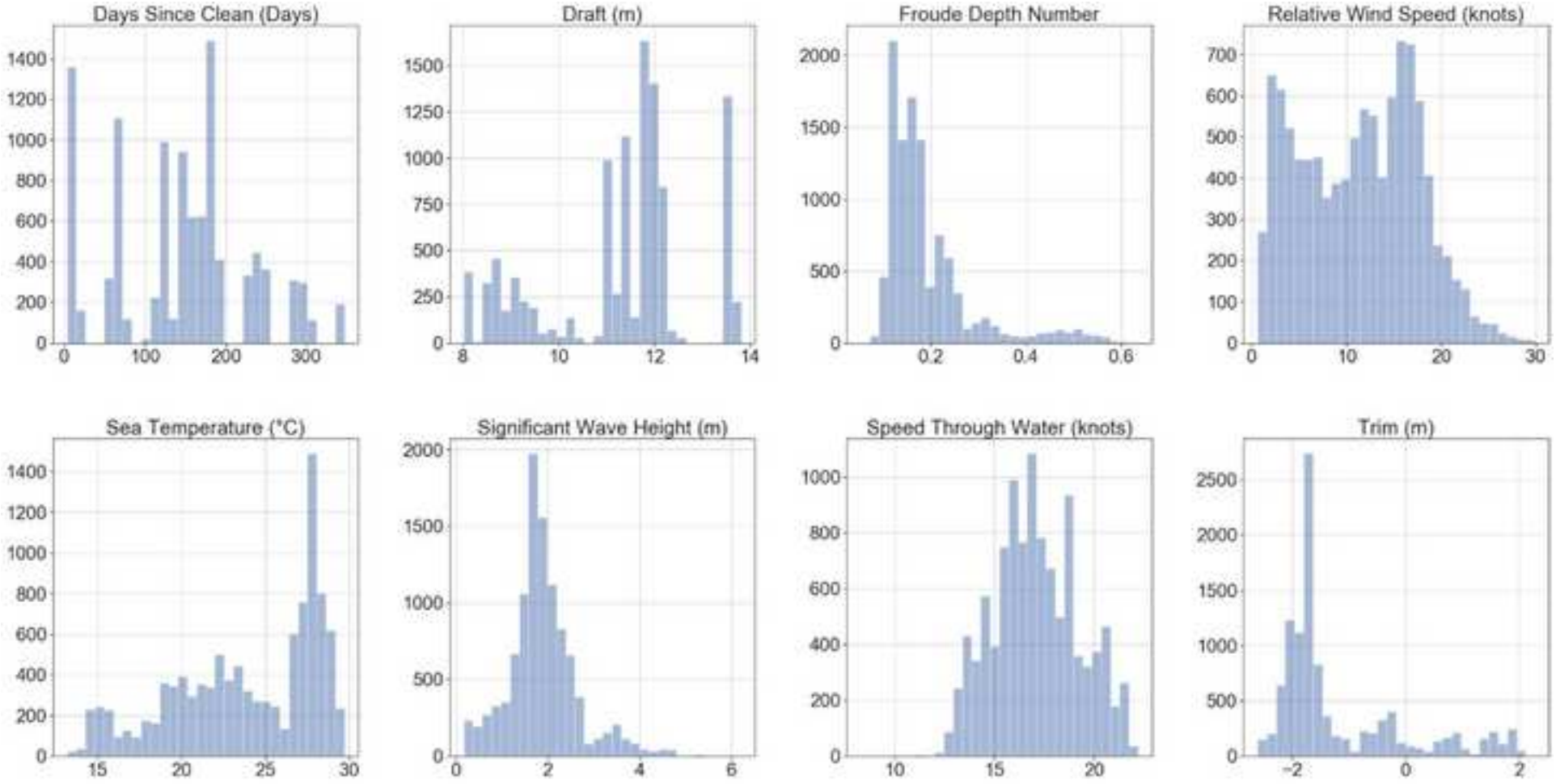

Figure 6

Click here to download high resolution image

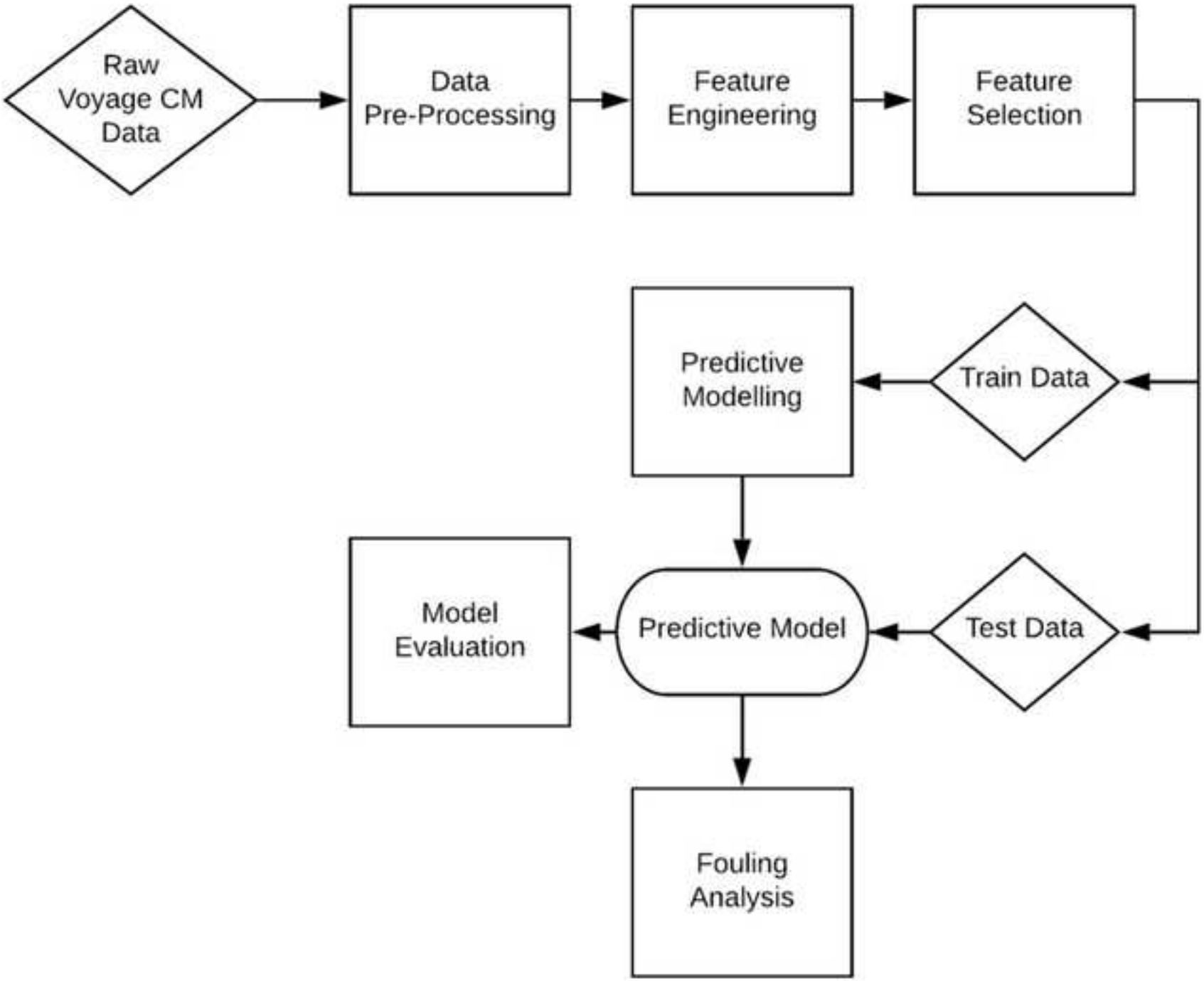

Figure 7

Click here to download high resolution image

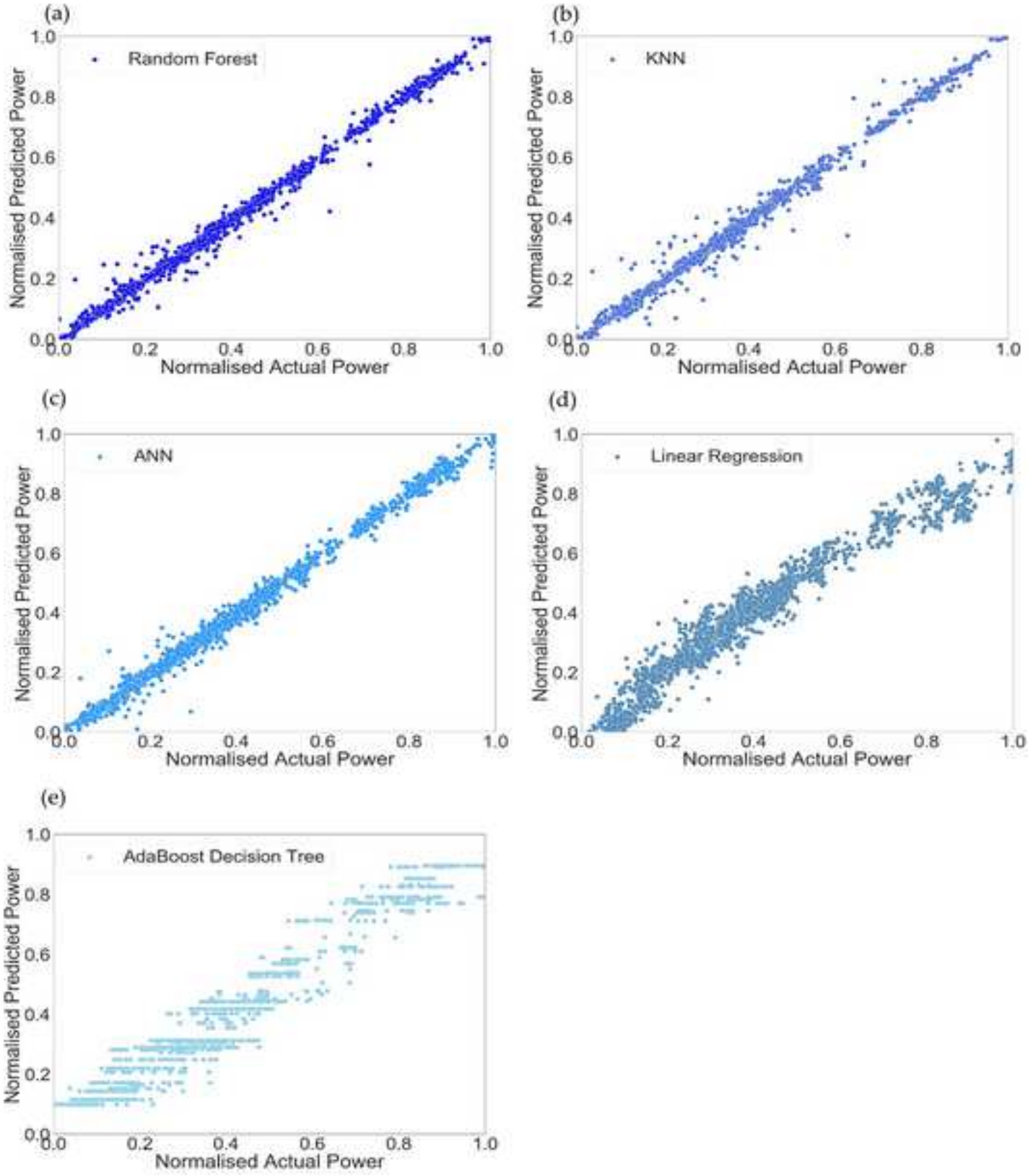

Figure 8

Click here to download high resolution image

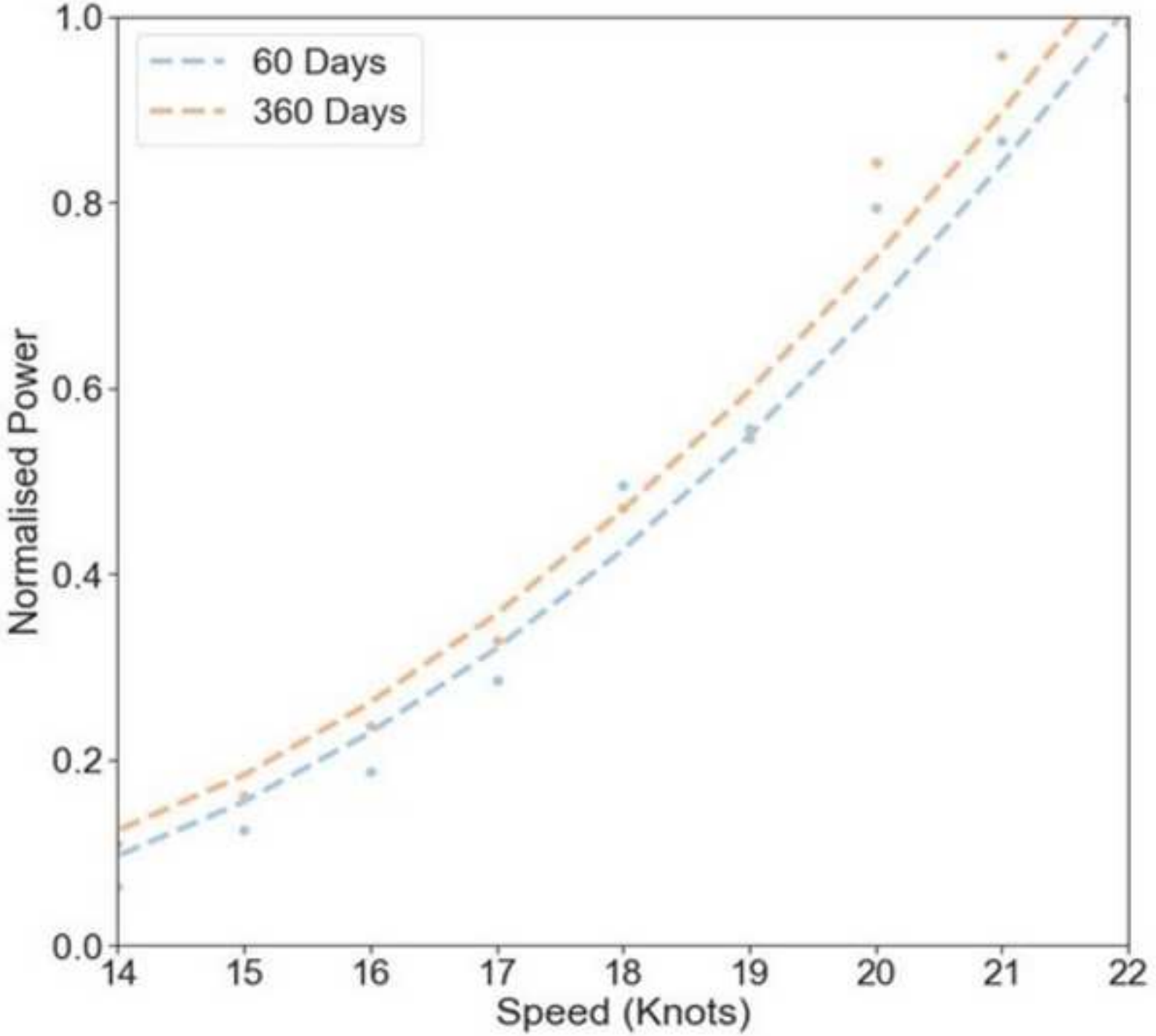

Figure 9

Click here to download high resolution image