Ship speed prediction based on machine learning for efficient shipping operation

Ameen M. Bassam[a,*], Alexander B. Phillips[c], Stephen R. Turnock[b], Philip A. Wilson[b]

[a] *Naval Architecture and Marine Engineering Department, Faculty of Engineering, Port Said University, Port Fouad, Egypt*
[b] *Fluid Structure Interactions Group, University of Southampton, Boldrewood Innovation Campus, SO16 7QF, UK*
[c] *National Oceanography Centre, Southampton, SO14 3ZH, UK*

Abstract

Optimizing ship operational performance has generated considerable research interest recently to reduce fuel consumption and its associated cost and emissions. One of the key factors to optimize ship design and operation is an accurate prediction of ship speed due to its significant influence on the ship operational efficiency. Traditional methods of ship speed estimation include theoretical calculations, numerical modeling, simulation, or experimental work which can be expensive, time-consuming, have limitations and uncertainties, or it can't be applied to ships under different operational conditions. Therefore, in this study, a data-driven machine learning approach is investigated for ship speed prediction through regression utilizing a high-quality publicly-accessible ship operational dataset of the *'M/S Smyril'* ferry. Employed regression algorithms include linear regression, regression trees with different sizes, regression trees ensembles, Gaussian process regression, and support vector machines using different covariance functions implemented in MATLAB and compared in terms of speed prediction accuracy. A comprehensive data preprocessing pipeline of operational features selection, extraction, engineering and scaling is also proposed. Moreover, cross validation, sensitivity analyses, correlation analyses, and numerical simulations are performed. It has been demonstrated that the proposed approach can provide accurate prediction of ship speed under real operational conditions and help in optimizing ship operational parameters.

Keywords: Ship speed prediction, Ship energy efficiency, Machine learning, Regression, MATLAB

1. Introduction

With more than 80% of the world trade handled by shipping, more stringent regulations are introduced by the International Maritime Organization (IMO) to improve ships operational efficiency and reduce its greenhouse gas (GHG) emissions. However, in spite of implementing stricter regulations, the total GHG emissions from ships as well as the shipping share percentage to the global emissions have increased by 9.6% and 4.7% respectively between 2012 and 2018 according to the latest IMO GHG study (Faber et al., 2020). Therefore, in order to control and reduce these emissions, IMO has adopted mandatory operational and technical measures which include the Energy Efficiency Design Index (EEDI) for new ship design and the Ship Energy Efficiency Management Plan (SEEMP) for all ships. The EEDI targets a minimum CO_2 emissions per cargo carried for newly built ships through implementing design-based solutions. Meanwhile, the SEEMP seeks to improve the operational energy efficiency of ships using operational strategies and practices of ship management (Rehmatulla et al., 2017; Bazari and Longva, 2011).

Among the various EEDI and SEEMP measures available, the speed based measures are increasing in popularity for improving ships energy efficiency and reducing GHG emissions (Capezza et al., 2019). This is mainly due to the fact that, a small speed adjustment can result in a significant improvement to the ship fuel

*Corresponding author
Email address: ab2e12@alumni.soton.ac.uk, ameen@eng.psu.edu.eg (Ameen M. Bassam)

consumption and energy efficiency (Smith et al., 2011). Moreover, applying speed based EEDI and SEEMP measures such as speed optimization, voyage execution, or speed reduction for new and existing ships does not require an upfront capital or investment costs and payback periods. However, it should be noted that altering ship speed can impact the voyage duration and associated costs which affects the ship productivity and total income (Capezza et al., 2019; Smith et al., 2011). Therefore, predicting ship speed in design stage and during operation is an essential element in evaluating the efficiency of EEDI and SEEMP measures.

Ship speed prediction is of significant importance in the decision making processes and has many implementations in the maritime industry. For example, for more accurate fuel consumption and emissions calculations, ship speed is the most principal operational parameter to be determined Bialystocki and Konovessis (2016). In addition, it has been shown that ship operating speed is a trend key driver of emissions and its growth rate (Faber et al., 2020). Moreover, in ship routing and voyage planning problems, an accurate ship speed prediction is essential to estimate the ship expected time of arrival (ETA) and satisfy the calling ports time windows constrains (Zis et al., 2020). Also, ship speed is a key factor in developing and operating ship trajectory planning and collision avoidance strategies for safer navigation especially in narrow channels or heavy traffic areas (Cockcroft and Lameijer, 2003).

The overall concept of the ship energy system can be explained as the fuel energy is converted into useful thrust by the propeller through the propulsive machinery to overcome the ship total resistance at a specific ship speed. Therefore, ship speed prediction and calculation depend on the characteristics of the ship hull, propulsion machinery, propeller, and the surrounding environment (Molland et al., 2011; Journée, 1976). This issue can be approached in different ways; experimentally using ship model tests or full scale ship speed trails (ITTC, 2014a,b), numerically by modeling the flow field around the ship hull using various computational fluid dynamic (CFD) techniques (Choi et al., 2009), from in-service propeller shaft measurements (Dalheim and Steen, 2021), or statistically using for example regression based methods to learn and estimate the relations between ship speed/power and other hull, operational, and environmental parameters. Measurements of these parameters can be obtained from model and full scale tests as in Holtrop work (Holtrop, 1984) or recorded during normal ship operation over a period of time using measuring instruments (Mao et al., 2016).

The recent advancement in sensor technologies, data acquisitions and storage systems enables the monitoring of ship operational performance to be more efficient and reliable owing to the higher data quality and integrity (Shenoi et al., 2015). The proper processing and analysis of this data can provide a deeper insight into the ship operational performance, extract valuable information from it, and uncover the correlation and patterns between the measured data. For these purposes, machine learning and statistical approaches have gained substantial momentum in shipping industry in the recent decades (Petersen et al., 2012a; Soner et al., 2019). This is because statistical and data-driven models can deal with high-dimensional and non-linear data such as the ship operational data without a priori knowledge of the ship underlying basic physics (Coraddu et al., 2017, 2015). Also, due to their nature, statistical and machine learning approaches have more prediction robustness and easier information extraction from sensor data compared to theoretical and parametric approaches (Coraddu et al., 2015; Soner et al., 2019).

The literature review in the area of ship operational performance monitoring through data analysis is dominated by modeling, predicting, and optimizing of the ship fuel consumption for economic as well as environmental reasons (Soner et al., 2018; Gkerekos et al., 2019; Uyanık et al., 2020; Parkes et al., 2018); however, there have been a few studies that investigated ship speed prediction based on available ship operational data. For monitoring and analyzing the operational performance of a ferry in terms of ship speed and fuel consumption, Gaussian Processes (GP) and neural network models were compared in (Petersen et al., 2012a). Based on the same dataset, the Ridge and LASSO regression models were also compared in (Soner et al., 2019). Moreover, tree based regression models were proposed in (Soner et al., 2018) for the same ship showing a comparable performance with the aforementioned models. Meanwhile, the operational performance of a containership was modeled in terms of ship speed and and engine power using GP model in (Yoo and Kim, 2019). In another study for speed prediction of a container ship, a preliminary investigation of the linear regression, autoregressive and the mixed effects models was conducted in (Mao et al., 2016) using a limited amount of operational data. Linear regression was also compared to the generalized additive and projection pursuit regression models for speed prediction in (Brandsæter and Vanem, 2018). Furthermore, speed prediction through regression was proposed for weather routing optimization study in (Krata and Szlapczynska, 2018), for modeling ship maneuverability in (Wang et al., 2015), and for navigation safety and collision avoidance of ice class ships in (Similä and Lensu, 2018).

From the above, it can be seen that ship speed prediction is of great concern for different purposes with different approaches being considered. Although ship speed can be predicted mathematically during design stage or be measured directly during operation using satellite based technologies such as global positioning or automatic identification systems, the purpose of this work is to predict ship speed based on measured real ship operational data as inputs to a machine learning model. This can be extended to provide deeper insights into the relation between ship speed and other ship operational parameters which is essential for operational optimization and decision support purposes. As a result, helping decisions makers and shipping companies to move towards more efficient operation environmentally and economically.

Much of the current literature utilizes different ship types and datasets, with different data acquisition systems, processing techniques, and data scaling methods. Therefore, due to this inconsistency, it is inconclusive which model is more accurate in terms of ship speed prediction. To remedy this gap, the aim of this work is to train and validate various conventional regression models to examine and compare their prediction accuracy of ship speed using a high quality ship operational dataset. The studied machine learning regression models include Multiple Linear Regression (MLR), Regression trees with different sizes, Ensembles of trees using both bagging and boosting techniques, Gaussian Process Regression (GPR), and Support Vector Machine (SVM) models using different covariance functions and kernels. These algorithms are the most commonly used and they are chosen for their robustness, efficiency, power, and accuracy. Also, sensitivity analysis of different data preprocessing methods (Data scaling) as well as different number of data splits for cross validation are performed to assess its effect on the statistical performance of different regression models. Furthermore, a correlation analysis and computational experiments are conducted to study and examine the relation between ship speed and other ship operational parameters.

The paper is organized as follow; Section 2 introduces the examined ship and dataset preprocessing. Section 3 describes the used methodology, the studied regression models, and their validation and evaluation. Meanwhile, Section 4 shows the results and discussion. Finally, Section 5 presents the work conclusions, recommendation, and future work.

2. Ship & data description

This work utilizes the existing publicly available sensor data from the domestic ferry the *'M/S Smyril'* operating around the Faroe Islands. The ferry's specification is provided in Table 1. An automated on-board data acquisition system recorded for the ferry's two or three trips per day over a period of nearly two months from February 16th to April 21th 2010 completing, approximately 250 trips (Propulsion modelling, 2021; Petersen et al., 2012b). To improve the quality and representation of the collected dataset, the following data preprocessing has been undertaken as part of this study.

Table 1: Specifications of the M/S Smyril ferry

Parameter	Value
Length	123 m
Breadth	22.7 m
Draft	5.6 m
Passenger capacity	975
Car capacity	970 m / 200 cars
Service speed	21 kn
Main engines	4 * MAN B&W 7L32/40

2.1. Data preprocessing

2.1.1. Feature selection & extraction

One of the most commonly used data preprocessing technique is feature selection which is used to identify the important variables within the dataset and remove the unnecessary features. This, consequently, results in reducing the data dimensionality and the modeling computational cost, and improving the model performance. Therefore, a correlation analysis is performed to show the interrelation between the ship speed as the independent variable and other operational variables as presented in Table 2. Linear correlation is usually used to express the relationship between variables as in (Uyanık et al., 2020; Gkerekos et al., 2019; Brandsæter and Vanem, 2018). However, a nonlinear relationship may exist and not be captured. Therefore, distance correlation is also deployed to test the nonlinear correlation between various variables. The linear coefficient value ranges between 1 and -1, while distance correlation coefficient ranges from 0 to 1. For both coefficients, a value of 0 indicates no correlation between variables and a value close to 1 implies a strong relationship. Meanwhile, the linear correlation sign indicates the direction of the correlation trend.

Table 2: Correlation of ship speed to other operational variables

Variables	Linear correlation	Distance correlation
Port propeller pitch	0.9067	0.8832
Starboard propeller pitch	0.8862	0.8632
Port rudder angle	0.6689	0.7212
Headwind	0.1370	0.1904
Crosswind	-0.0003	0.0921
Starboard rudder angle	-0.2020	0.6099
Trim angle	-0.2392	0.3108
Draft	-0.4609	0.4011

According to Table 2, the port and starboard propeller pitch as well as the port and starboard rudder angle have significant effect on the ship operational speed. This is due to the fact that varying the propeller pitch varies the provided propeller thrust and ship speed while the altering the rudder angle affects the ship resistance and, accordingly, the ship speed. As can be noticed in Table 2, the dependence power between the ship speed and starboard rudder angle is detected by the distance correlation more than the linear correlation due to its nonlinearity nature. Moreover, the ship trim angle and draft have a high impact on the ship speed since it can affect the ship resistance and consequently the ship operational performance. Therefore, a trim optimization at different draft conditions can be conducted to further improve the ship operational efficiency. Likewise, for reducing the load on the engine bearings and shafting system, the impact of optimizing the propeller controllable pitch on the ship speed can be examined for decision support. On the contrary, headwind and crosswind variables show insignificant correlation, which is mostly due to the fact that the utilized dataset covers about two months of operation alternating within a narrow range.

Measurements of the selected operational variables, which are used to train the ship speed prediction models, were carried out as follows: the ship speed measured using a Doppler speed log, trim angle measured using an inclinometer, port and starboard water level measurements measured using two radars placed on the ship sides, port and starboard propeller pitch, port and starboard rudder angle, wind angle and direction as presented in Table 3. These variables have the most significant effect on ship speed and operational performance (Soner et al., 2018; Yoo and Kim, 2019).

The selected parameters were firstly extracted from the raw data and arranged into separate voyages. Then, due to the different sampling frequencies of the measurements as shown in Table 3, the extracted data was resampled at an average frequency of 1 Hz and averaged over 10 minutes windows or intervals as suggested in (Pedersen and Larsen, 2009; Leifsson et al., 2008). The resulted total data size is 2654 observations of each feature for the given dataset.

Table 3: Selected parameters and measurement frequency (Propulsion modelling, 2021)

Feature	Unit	Measurement frequency (Hz)
Ship speed	Kn	1
Trim angle	degree	3
Port water level measurement	m	3
Starboard water level measurement	m	3
Port propeller pitch	$Volt$	1
Starboard propeller pitch	$Volt$	1
Port rudder angle	$Volt$	1
Starboard rudder angle	$Volt$	1
Wind angle	degree	0.5
Wind speed	m/s	0.5

2.1.2. Feature engineering

In order to improve the performance of machine learning algorithms, new features can be engineered from the raw data to better represent the ship operational data. For example, the port and starboard water level measurements were transformed into draft amidships as a function of the radars heights, angles, and distance from the midship. Also, the inclinometer readings were corrected to have the real ship trim angle. Moreover, the wind speed V_{wind} and angle θ_{wind} measurements were transformed into two new features called the headwind V_{hw} and crosswind V_{cw} to eliminate the circular discontinuity issue of the wind direction when it passes between $0°$ and $360°$ using Equations 1.

$$\begin{aligned} V_{hw} &= V_{wind}.\cos(\theta_{wind}) \\ V_{cw} &= V_{wind}.\sin(\theta_{wind}) \end{aligned} \tag{1}$$

Figure 1 shows the processed data after features selection, extraction and engineering against ship speed. As expected and reported in the literature, there are nonlinear relationships between ship speed and other ship operational parameters. It can, however, be noted that ship speed is proportional to the propeller pitch.

As shown in Figure 1, the ship draft is spread between 5 and $6m$ while the trim angle corresponds to a trim between -1.5 to $1.25m$ (Petersen et al., 2012a; Soner et al., 2019). Regarding the ship speed, its average value varies mostly between 15 and $20Kn$ as shown in Figure 2. This is due to the fact that the ship slows only at dock while loading and unloading. Also, the wind speed tends to increase with the ship speed as shown in Figure 3 with an average headwind speed of $11m/s$.

2.1.3. Feature scaling

Since different ship operational parameters have different ranges and units, features scaling is an important preprocessing step. Hence, different features can be comparable to each other and contribute equally to the machine learning objective functions. Standardization and normalization are two common scaling methods and both are introduced into this study to test their impact on different regression models. Standardization scales the features data to unit variance and removes its mean according to Equation 2. Meanwhile, normalization scales the features data between 0 and 1 using min-max scaling as in Equation 3.

$$x_s = \frac{x - \mu}{\sigma} \tag{2}$$

$$x_m = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where x_s and x_m are the dimensionless standardized and normalized values of the actual variable x respectively. μ is the mean value of the variable x entire data, σ is its standard deviation, x_{min} and x_{max} are the minimum and maximum values of the variable x respectively. The processed data is then used to train and validate different predictions models as will be discussed in the following sections.

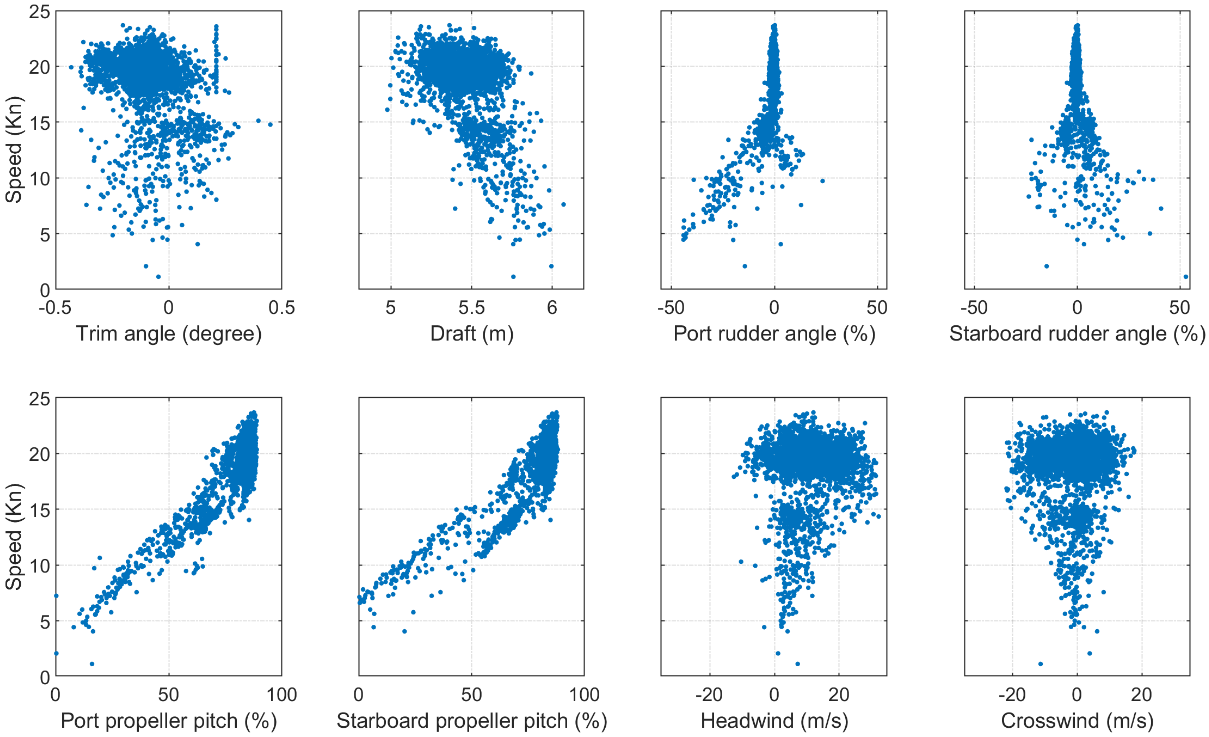

Figure 1: Scatter plots of ship speed versus processed ship operational parameter

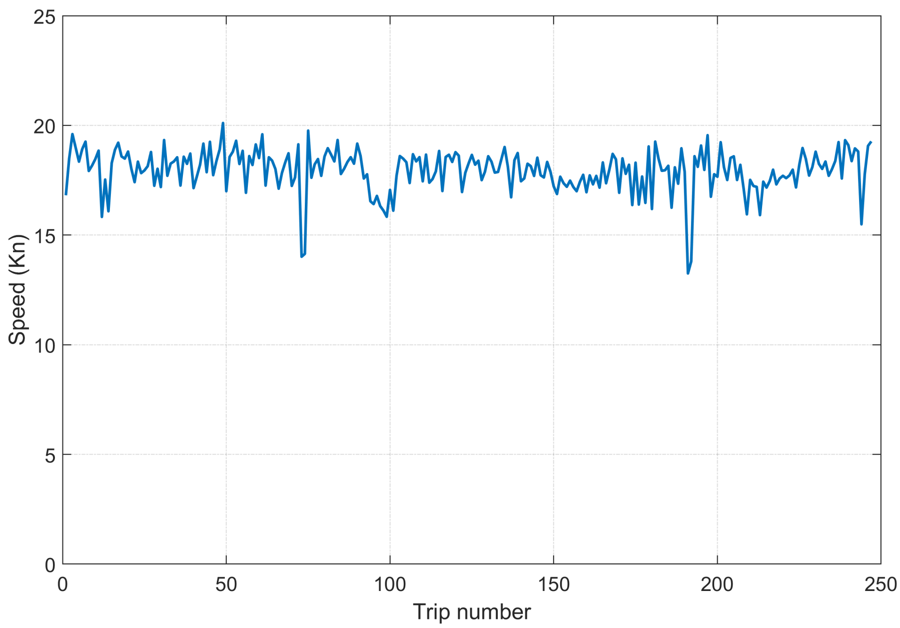

Figure 2: Average ship speed for each trip

3. Prediction models

In reality, ship speed is influenced by many factors including ship operational conditions (e.g. draft, trim) as well as environmental conditions which makes it difficult to be modeled using conventional approaches.

Therefore, machine learning regression techniques are applied in this research to predict ship speed as a function of measured ship operational parameters and surrounding environment conditions. These parameters are used to train regression models to construct a mapping function from input variables to infer the output ship speed variable, and then make prediction for new data. In the following sections, a diverse set of common regression models are described which vary in their level of complexity and accuracy. These models are then employed and compared for ship speed prediction.

3.1. Multiple linear regression (MLR) model

MLR is an extension of linear regression that assumes a linear relationship between the response variable (y_i) and the predictor variables (x_{i1} to x_{ip}) as shown in Equation 11.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon \tag{4}$$

where (β_0) is the constant term in the model, (β_1 to β_p) are the corresponding coefficients of the predictor variables (x_{ip}), and (ϵ) is the error term of the model. Due to its advantages of simplicity and ease of interpretation, MLR is one of the most popular parametric models and it is normally used as a reference to compare other models performance (Gkerekos et al., 2019).

3.2. Regression trees

Tree-based regression model is one of the advanced and accurate non-parametric statistical model which is suitable for non-linear parameters such as ship operational datasets (Soner et al., 2018). Building a regression tree for prediction involves two main steps; dividing the predictor variables into distinct non-overlapping regions (R_1 to R_j). Then, predictions are made from the mean response values of the training observations for every observation in different regions $R_{(1..j)}$ (James et al., 2013). Meanwhile, the main goal is to find the regions that minimizes the Residual Sum of Square (RSS) shown in Equation 5.

$$\text{RSS} = \sum_{j=1}^{J} \sum_{i \epsilon R_j} (y_i - \hat{y}_{R_j})^2 \tag{5}$$

where $\hat{y}_{R_j}$ is the training observations mean response within the jth region and $j = 1, 2, .., p$ and p represents the number of regions or leaves of the tree. The number of these regions or leaves defines the regression tree size and its level of accuracy, flexibility, and robustness. Whilst a fine regression tree can produce more accurate results, the overfitting risk increases. In contrast, a coarse regression tree has lower training accuracy but it is more robust with lower variance (James et al., 2013). Consequently, choosing the regression tree size is essential to have balance between the model complexity, speed, accuracy, and overfitting risk.

3.3. Ensembles of trees

Despite the advantages of regression tree models, they can suffer from high variance, bias and overfitting. Therefore, multiple regression trees can be combined to build an ensemble of trees to improve the predictive performance of the model (James et al., 2013). Two of the most popular ensemble techniques to aggregate many regression trees are bagging and boosting. Bagging or bootstrap aggregating uses multiple separate training sets from the original training dataset randomly with replacement to train different regression trees. The predictions of different trees ($\hat{f}^{*1}(x)$ to $\hat{f}^{*B}(x)$) are then calculated and averaged as follows which reduces the variance compared to a single regression tree (James et al., 2013).

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{*b}(x) \tag{6}$$

where $\hat{f}_{bag}(x)$ is the average prediction of all the regression trees and B is the number of the separate training sets and trees. On the other hand, boosting technique grows the number of trees B sequentially where each tree utilizes a modified version of the whole dataset using information from the previously grown tree (James et al., 2013). The learning process improves the prediction performance of each tree from $\hat{f}^1(x)$ to $\hat{f}^B(x)$ by updating the observations' weights of the training dataset without bootstrap sampling and the boosted model output $\hat{f}_{boost}(x)$ is given as follows:

$$\hat{f}_{boost}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x) \tag{7}$$

where λ is the shrinkage parameter which controls the rate of the boosting learning process. The shrinkage parameter and the B number of trees for both bagging and boosting methods are determined by cross-validation as explained later.

3.4. Gaussian process regression (GPR) models

Implementing a Gaussian process (GP) for regression purposes has been proposed considerably due to its power, efficiency, and accuracy. Also, GP-based regression models can describe the uncertainty and non-linearity between the dataset parameters through a nonparametric approach (Rasmussen and Williams, 2006). Therefore, GPR models are proposed for ship operational data analysis and ship speed prediction (Yoo and Kim, 2019; Petersen et al., 2012a).

A GP is a collection of random variables where any finite collection of which are described by a joint Gaussian probability distribution. Whereas in GPR, the function of variables $f(x)$ is assumed to be distributed as a GP which is defined by its mean function $m(x)$ and covariance function $k(x, x')$ as follows.

$$f(x) \sim GP(m(x), k(x, x')) \tag{8}$$

One of the most popular covariance function and commonly used is the squared exponential (SE) or the radial basis function (RBF) (Rasmussen and Williams, 2006; Yuan and Nian, 2018). This covariance function $k_{SE}(r)$ or $k_{SE}(x, x')$ is very smooth due to its infinitely differentiable nature and it can be written as follows.

$$k_{SE}(r) = \exp(-\frac{r^2}{2l^2}) = \exp(-\frac{|x - x'|^2}{2l^2}) \tag{9}$$

where x and x' are the training and testing points pairs respectively and l defines the characteristic length-scale for the input values. It should, however, be noted that the assumed SE smoothness may not be realistic to model some physical systems. Therefore, the Matérn class of Gaussian process is recommended because it includes a parameter (ν) that can control the learned function smoothness (Rasmussen and Williams, 2006; Stein, 1999). One type of the Matérn class functions is the exponential covariance function obtained when $\nu = 1/2$ which is a continuous but not differentiable function as the SE function and it can be defined as follows:

$$k_{\nu=1/2}(r) = \exp(-\frac{r}{l}) \tag{10}$$

Both functions, the exponential and squared exponential, are common and widely used and they are implemented in this study to be compared in terms of their accuracy of ship speed prediction.

3.5. Support vector machine (SVM) models

Due to its robustness, accuracy, power, and generalization ability, SVM is one of the most attractive supervised learning model proposed for many fields which can be used for classification and regression (Uyanık et al., 2020; Awad and Khanna, 2015). In regression problems, SVMs are built as regressors which try to fit a hyperplane or a function that predicts a continuous target value within a tolerance margin or a decision boundary based on the training samples. The objectives of adjusting this margin is to minimize the prediction error and balance it with the model complexity and robustness.

SVM is also a kernel based technique which extends its functionality by using different kernel functions depending on the data type. Therefore, SVM can be a parametric model using a linear kernel or a non-parametric model using an RBF kernel. Popular kernel functions include: linear kernel in Equation 11, polynomial kernel in Equation 12, and Gaussian kernel in Equation 14.

$$k(x, x') = x^T x' \tag{11}$$

$$k(x, x') = (1 + x^T x')^d \tag{12}$$

$$k(x, x') = \exp(-\gamma ||x - x'||^2) \tag{13}$$

where d is the polynomial degree of kernel and γ is the Gaussian kernel scale hyperparameters which can be adjusted to enhance the SVM model performance. Therefore, different polynomial degrees and kernel scales are investigated in this study to find the optimal model configuration.

3.6. Prediction performance evaluation

In order to examine the predictive accuracy of the employed models and measure its effectiveness, prediction performance indices can be used. However, training the prediction models and testing its performance using the same dataset can give overoptimistic results (Arlot et al., 2010). Therefore, a validation scheme is implemented to split the dataset into a training dataset to train the prediction models, and a test dataset to validate its performance after training. Then, a number of performance measures can be used to compare all the models performance as explained in the following subsections.

3.6.1. Cross validation

Validation reduces the risk of overfitting and ensures the generalization capabilities of the trained prediction models and the robustness of its hyperparameters' values. However, partitioning the available data into training and test datasets reduces the available data for training as well as for validating the models. To address this issue, cross validation is applied in the form of K-folding which is widely used and preferred in the literature (Uyanık et al., 2020; Gkerekos et al., 2019).

In K-fold cross validation method, the dataset is split into k subsets or folds where K-1 subsets are combined and used to train the prediction models and the remaining subset is used for validation. This process is repeated for K times where each time, one of the K folds is used as a validation subset. Finally, the average validation error of all the K runs are obtained. By using this technique, most of the data is used for training as well as for validation which reduces the bias and variance and gives a good estimation of the predictive accuracy of the studied models. Nevertheless, different suitable values of K are reported in the literature including 4 in (Uyanık et al., 2020; Leifsson et al., 2008), 5 in (Yan et al., 2020; Hu et al., 2019), 10 in (Soner et al., 2018, 2019), 20, 30, and 50 in (Brandsæter and Vanem, 2018; Coraddu et al., 2017). Therefore, a sensitivity analysis of different K values is made in this study.

3.6.2. Coefficient of determination R^2

The coefficient of determination R^2 explains the variation of the measured response variable y_i as a function of the response variable prediction made by the trained model $\hat{y}_i$ and the average value of the response variable $\bar{y}_i$ as follows

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y}_i)} \tag{14}$$

where n is the number of samples. R^2 value varies normally from 0 to 1 where a higher value means a better fit of the trained model to the data.

3.6.3. Mean squared error (MSE)

This index measures the mean of the square of all errors between the predicted and the measured values of the response variable as shown in the following equation.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{15}$$

By squaring the errors, MSE is always positive and it gives more weight to high errors. The lower the MSE, the better the prediction model performance.

3.6.4. Root mean square error (RMSE)

RMSE is the most commonly used and easily interpreted statistic, as it has the same unit of the studied variable that better reflects the prediction models performance. It is calculated by taking the square root of the MSE as shown in the following equation.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{16}$$

3.6.5. Mean absolute error (MAE)

This criterion is similar to the RMSE. However, it is more robust and less sensitive to data outliers compared to the MSE. MAE corresponds to the average of all the absolute errors and can be calculated as follows.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{17}$$

3.6.6. Methodology implementation & parameters

The discussed prediction models as well as the examined ferry dataset are modeled mathematically and implemented in MATLAB environment in order to compare the performance of these models in terms of its accuracy of predicting the ship speed. First, the dataset collected from the ferry *'M/S Smyril'* is loaded into the MATLAB workspace, the selected parameters are extracted, resampled, averaged, and engineered as explained in 2. Regarding the data scaling, a quantitative analysis is performed to study the sensitivity of various prediction models to different scaling techniques. Therefore, the examined regression models are trained with the raw, standardized, and normalized versions of the ferry dataset to show its impact on the performance and results of different models. These datasets are then split into training and testing datasets to validate the prediction models according to the K-fold cross validation method where K values of 5, 10, 30, and 50 are used to study the effect of this parameter on the the performance of different models. Moreover, to get more consistent results, models training is repeated 10 times as suggested in (Gkerekos et al., 2019) and results are averaged. Finally, the performance metrics are computed to evaluate and compare different models in terms of prediction accuracy of ship speed.

For the tree-based regression model, three different regression tree sizes are studied to investigate the trade-off between the tree model accuracy and complexity. A minimum leaf size, that indicates the number of variable observations in a tree leaf, of 4, 12, and 36 are used which corresponds to fine, medium, and coarse regression trees respectively as suggested in (MATLAB, 2021). For the ensembles of trees model, the two most popular techniques of aggregating regression trees, which are bagging and boosting, are investigated with a shrinkage parameter or learning rate of 0.1 and a minimum leaf size of 8 as recommended in (MATLAB, 2021). Regarding the SVM regression model hyperparameters, the polynomial kernel degree d is set to 2 and 3 for the quadratic and the cubic SVM models respectively which are compared to the linear SVM model. For the SVM Gaussian kernel scale, γ is set to $\sqrt{P}/4$, $\sqrt{P}$, and $4\sqrt{P}$ which correspond to fine, medium, and coarse Gaussian SVM models respectively where P is the number of the trained model predictors (MATLAB, 2021). The aforementioned models are integrated in the 'Regression Learner App' of

MATLAB's Statistics and machine learning toolbox. In this study, MATLAB R2019b is used on a desktop computer (Intel Core i7, 3.4 GHz, Memory 16 GB).

4. Results & analysis

As shown in Table 4, the GPR model with the the Matérn class or exponential kernel yields the best results with an R^2 of 0.91 and RMSE of 0.91 kn utilizing the raw dataset. This indicates that controlling the smoothness behavior of the stochastic processes realization as in the Matérn class can be beneficial for modeling realistic physical systems. The GPR model with the squared exponential kernel and the bagged trees ensemble model provide accurate results with an R^2 of 0.88. On the other hand, the SVM model using a fine Gaussian kernel performs the lowest estimation with an R^2 of 0.51 due to its small-scale kernel function. It follows that a rapid variations in the SVM response function which causes the model to overfit and doesn't perform accurately in the low ship speed region as illustrated in Figure 3. By increasing the Gaussian kernel scale value γ, a less complicated SVM model can be obtained with better prediction performance as demonstrated by the medium Gaussian SVM model. However, the prediction errors start to increase again for large γ value which results in a rigid SVM response function with higher probability of underfitting as shown in Table 4 and Figure 3 for the coarse Gaussian SVM model.

Table 4: Performance measures for different machine learning approaches using 5-fold cross validation

	Raw data				Standardized data				Normalized data			
	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
Linear	1.17	0.86	1.37	0.92	0.38	0.86	0.142	0.295	0.052	0.86	0.0027	0.0406
Regression trees:												
Fine tree	1.19	0.86	1.41	0.87	0.39	0.85	0.154	0.291	0.053	0.85	0.0028	0.0396
Medium tree	1.14	0.87	1.29	0.84	0.37	0.86	0.138	0.278	0.051	0.86	0.0026	0.0383
Coarse tree	1.16	0.86	1.35	0.87	0.38	0.86	0.142	0.285	0.052	0.86	0.0027	0.0395
Trees Ensemble:												
Boosted trees	1.31	0.82	1.72	1.04	0.34	0.89	0.115	0.259	0.057	0.83	0.0033	0.0448
Bagged trees	1.06	0.88	1.12	0.78	0.34	0.88	0.117	0.252	0.047	0.86	0.0022	0.0348
GPR:												
SE	1.06	0.88	1.11	0.80	0.34	0.89	0.114	0.257	0.046	0.89	0.0022	0.0353
Exponential	0.91	0.91	0.84	0.68	0.29	0.91	0.086	0.220	0.041	0.91	0.0017	0.0305
SVM:												
Linear	1.18	0.86	1.38	0.91	0.38	0.86	0.143	0.294	0.052	0.86	0.0027	0.0405
Quadratic	1.14	0.87	1.30	0.85	0.37	0.86	0.136	0.276	0.051	0.86	0.0026	0.0379
Cubic	1.54	0.76	2.37	0.86	0.40	0.84	0.159	0.274	0.061	0.80	0.0037	0.0380
Fine Gaussian	2.17	0.51	4.71	1.25	0.70	0.51	0.492	0.406	0.097	0.51	0.0093	0.0560
Medium Gaussian	1.11	0.87	1.23	0.80	0.36	0.87	0.126	0.256	0.049	0.87	0.0024	0.0354
Coarse Gaussian	1.13	0.87	1.28	0.87	0.36	0.87	0.132	0.280	0.050	0.87	0.0025	0.0386

In the same way, the medium regression tree model performs better than the fine and coarse tree models as shown in Table 4. This is due to the fact that there is a trade-off between the regression tree size and the model performance. Therefore, a coarse regression tree with fewer large leaves results in spreading the speed predictions over fewer large regions compared to the medium and fine tree models as shown in Figure 3. Meanwhile, a very leafy fine tree can result in overfitting and lower its generalization capability. It should be also mentioned that the studied prediction models also perform similarly to Figure 3 when utilizing the standardized and normalized versions of the dataset.

Regarding the ensemble of trees models, the bagging technique provides better results than the boosting technique using the raw and normalized datasets as detailed in Table 4. Meanwhile, by standardizing the dataset before training, the boosted trees model performance gets better and provides comparable performance to the bagged trees model. This can be explained by the fact that data standardization improves the data consistency which enhances the sequential learning process of the boosted trees model.

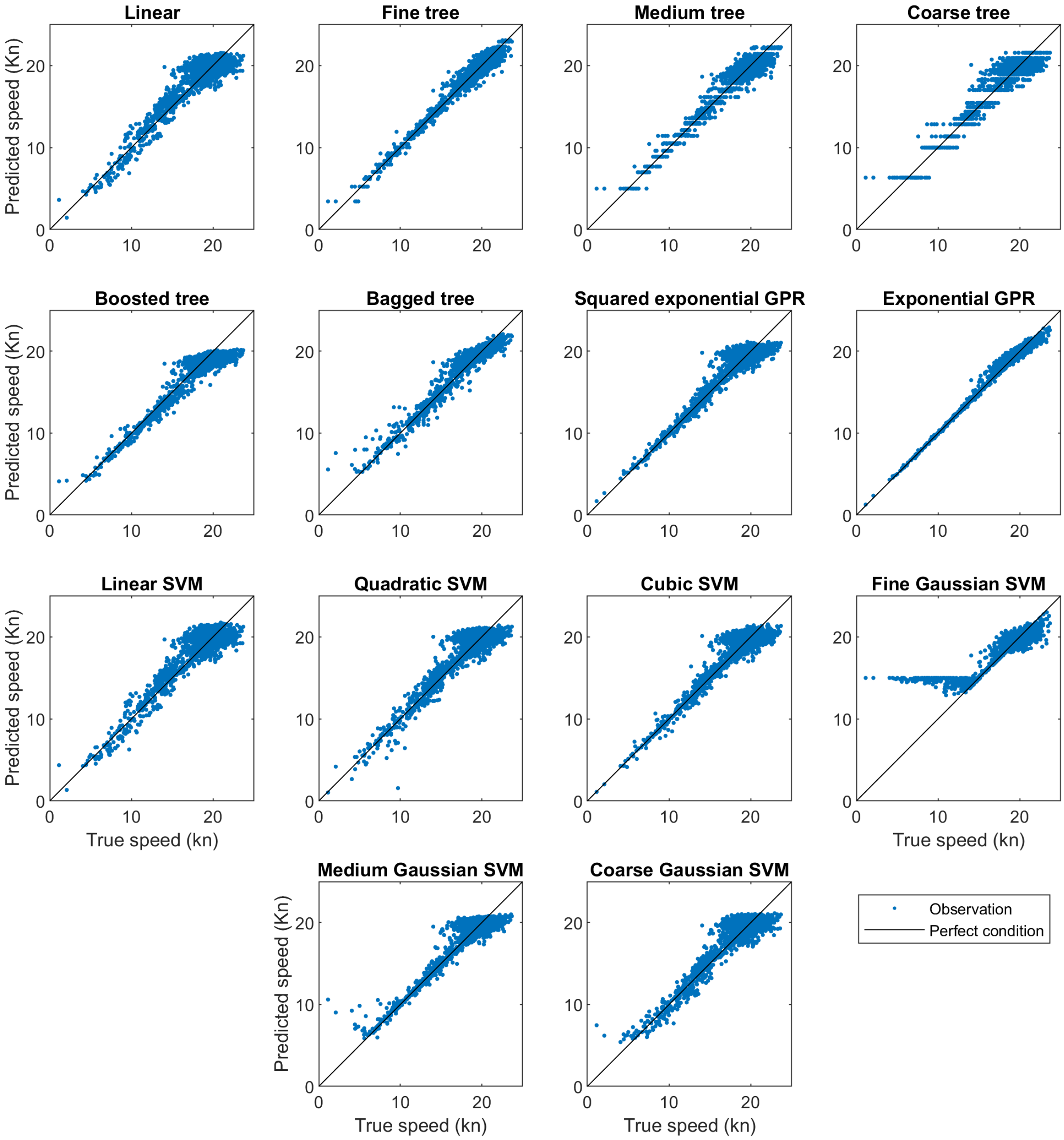

Figure 3: Predicted ship speed versus actual ship speed measurements for different machine learning approaches using raw data and 5-fold cross validation

Therefore, this indicates that the boosted trees model is more sensitive to the dataset variables range and feature scaling than the bagged trees model.

In order to study the sensitivity of other regression models to the dataset scale and appropriately compare their performance, the RMSE value index is calculated for different models using the raw, standardized, and normalized versions of the dataset. The value index number is calculated as a percentage of the RMSE of different models compared to the lowest RMSE as a base value as shown in Figure 4.

Since the exponential GPR model achieves the lowest RMSE, it is used as a baseline with a value index

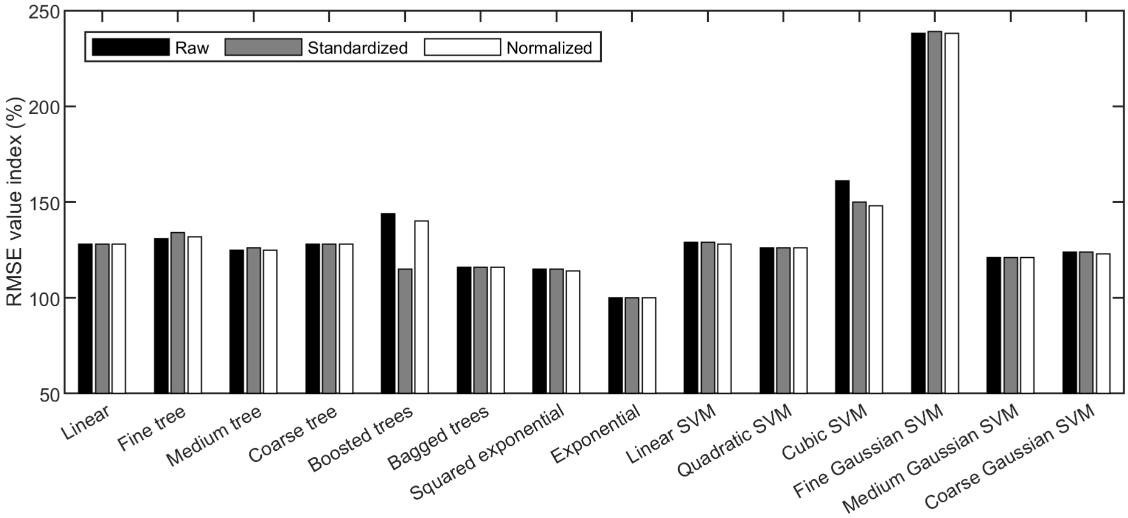

Figure 4: RMSE index value percentage of different machine learning approaches change from the base value using 5-fold cross validation

of 100%. As can be found in Figure 4, algorithms such as the boosted trees and the SVM model with cubic kernel are more sensitive to the range of their input values than other models. Accordingly, standardizing the dataset before training the boosted tree instead of using the raw dataset can reduce the RMSE value index by 20%. Meanwhile, normalizing the dataset before training the cubic kernel SVM can reduce the RMSE value index by 8% compared with using the raw dataset as well. Regarding other models, a slight accuracy improvement can be achieved by suitably processing the dataset before training the regression models.

The statistical performance of regression models can also be affected by the size of the training and test dataset which is decided by the k-fold cross validation. Therefore, different values of K is used to split the dataset into K folds to train and validate different regression models to study the impact of this parameter on their accuracy in terms of RMSE as shown in Figure 5.

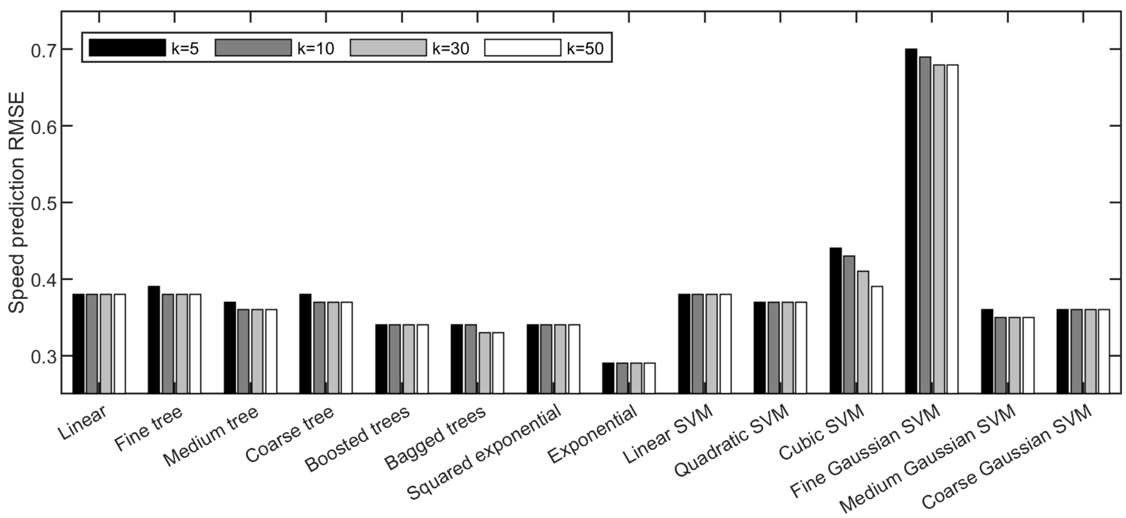

Figure 5: RMSE of different machine learning approaches using the standardized data with K values of 5, 10, 30, and 50

As indicated in Figure 5, increasing the number of folds K reduces the calculated RMSE of different machine learning approaches. This is because, as explained earlier, using higher values of K increases the size of the training dataset which improves the models statistical performance. However, this improvement is more significant for the cubic and fine Gaussian SVM models which their RMSE are reduced by 11% and 3% respectively by increasing K from 5 to 50. For other approaches, increasing the number of folds K results in a RMSE reduction of less than 3%. Nonetheless, longer training time and higher computational cost are required by hard interpretability models such as GPR, quadratic and cubic SVM models as a result of increasing the number of data splits as detailed in Table 5. Therefore, a K value of 10 can be considered in further studies to manage the trade-off between the models predictive quality and computational complexity.

Table 5: Required training time for different machine learning approaches with K values of 5, 10, 30, and 50

	Training time (minute)			
	K=5	K=10	K=30	K=50
Linear	<1	<1	<1	<1
Regression trees				
Fine tree	<1	<1	<1	<1
Medium tree	<1	<1	<1	<1
Coarse tree	<1	<1	<1	<1
Trees Ensemble				
Boosted trees	<1	<1	<1	<1
Bagged trees	<1	<1	<1	<1
GPR				
SE	1.3	2.4	6.4	9.7
Exponential	1.8	3.3	8.7	12.9
SVM				
Linear	<1	<1	<1	<1
Quadratic	<1	<1	1.2	1.9
Cubic	2.8	5.2	16.6	27
Fine Gaussian	<1	<1	<1	<1
Medium Gaussian	<1	<1	<1	<1
Coarse Gaussian	<1	<1	<1	<1

4.1. Computational experiments

In order to demonstrate the functionality of the proposed methodology in the optimization and decision-making processes, different computational experiments are conducted to estimate the impacts of changing significant operational parameters on the ship speed. Operational parameters such as the propellers pitch and ship drafts are used as inputs to the trained exponential GPR model, due to their significant effect on the ship speed, and the results are compared with the model predicted speed using the real ship operational data and the real measured ship speed.

In the first case, two propellers pitch values of 95% and 70% are used instead of the real propellers pitch values while using other real operational data of trim, draft, rudders angles, and environmental conditions. As shown in Figure 6, increasing the propellers pitch value results in higher ship speed. However, the propeller and engine rotational speeds should be taken into consideration while selecting the propeller pitch for higher operational efficiency of the ship propellers and engines.

In the second case, two different values of ship draft of $5m$ and $6.5m$ are simulated while using other real ship operational data of trim, propellers pitch, rudders angle, and environmental conditions. As shown in Figure 7, increasing the ship draft reduces the ship speed as a result of increasing the ship resistance, and correspondingly higher ship speed are obtained at relatively small ship draft of $5m$. On the other hand, lighter ship drafts may increase ship resistance at inappropriate trim angles. Therefore, optimization of ship draft/trim combination should be made for more efficient ship operation.

Another case to further test the generalization capability of the developed ship speed prediction model is conducted by adapting a new unseen test case where multiple operational parameters are changed. It

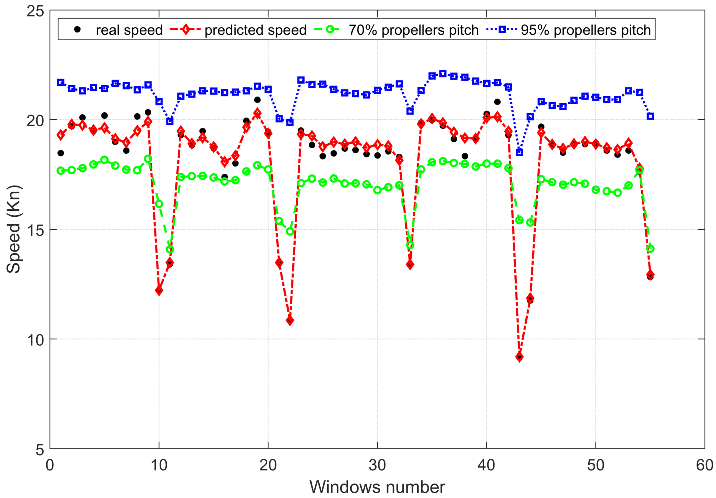

Figure 6: Real vs predicted ship speed at different values of propellers pitch using the exponential GPR model with 5-fold cross validation

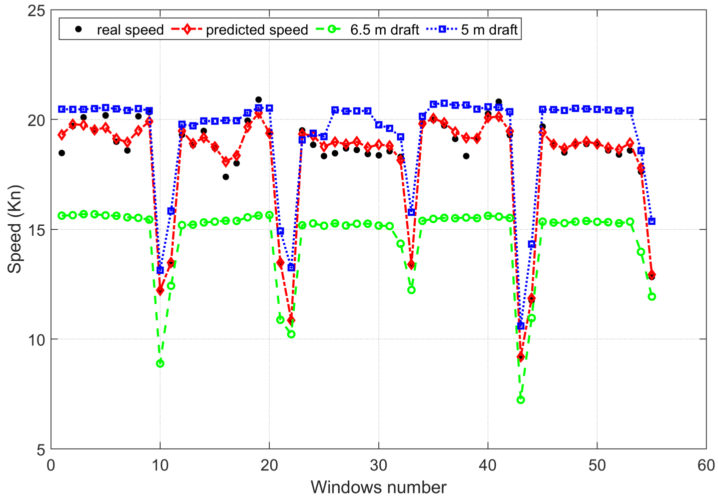

Figure 7: Real vs predicted ship speed at different values of ship draft using the exponential GPR model with 5-fold cross validation

is assumed that the ship is loaded to a draft of $6m$ sailing in a headwind of 20 m/s with an average trim angle of 0.05 degree aft. As a result, a reduced propellers pitch of 75% is selected for more efficient shipping operation by properly loading the main engines without having to increase its speed or heavily running the propellers. Then, it is assumed that the wind speed is changed during sailing to 5 m/s headwind which allows the propellers pitch to be increased to 90% gradually to avoid any operational delays. The predicted ship speed for this case is presented in Figure 8 using the exponential GPR model with the propellers pitch change given in the top plot. However, no real data is available to validate this scenario. Therefore, the exponential GPR model performance is validated with respect to the prediction of one accurate model of each regression algorithm type as illustrated in Figure 8 which also shows the response of different models to the input signals of propeller pitch, wind speed, and rudder angle.

According to the obtained results, the developed methodology can help ship operators and decision makers with evaluating the effect of changing the operational parameters on the ship speed. Consequently, it can help in creating more advanced models of voyage tracking and monitoring or optimization of different

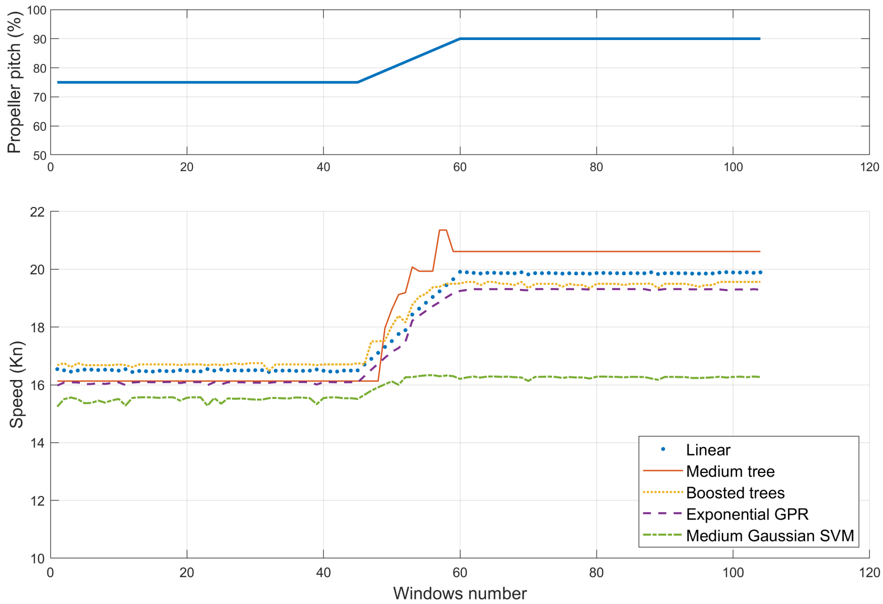

Figure 8: Predicted ship speed for the test case using the exponential GPR model with 5-fold cross validation compared to other regression models

ship operational parameters.

5. Conclusions

In recent years there has been a growing interest in monitoring and optimizing ship operation for better sustainability and profitability which requires accurate speed prediction. Among different approaches of ship speed prediction, machine learning and statistical methods have gained substantial momentum in shipping industry driven by the advances in computer power and the increasing operational data availability. This allows data-driven models based on machine learning to raise its responsiveness, analytical and prediction capabilities with more accuracy by extracting hidden information from the collected datasets.

A performance comparison of the most commonly used machine learning regression algorithms in terms of their prediction accuracy of ship speed in real operational conditions utilizing a high quality operational dataset of a ferry has been presented in this paper. For this, a framework for data preprocessing is provided which includes the selection and extraction of the operational features having significant influence on the ship speed. Moreover, new features have been engineered as well for better statistical performance of the studied models. Features transformation and scaling have been also made before training the machine learning regression algorithms. Then, cross validation has been made to avoid overfitting and assess the models generalization capability to new data. This paper also provides useful insights into the effect of different data scaling techniques on the prediction accuracy of the regression models. Also, a sensitivity analysis of different folds number and data splits for the cross validation has been made. Furthermore, the effect of changing different operational parameters on the ship speed is investigated through a correlation analysis using different techniques. The main findings can be summarized as follows:

- The studied regression models can accurately predict the ship speed with good accuracy except for the SVM with fine Gaussian kernel which had only R^2 of 0.51.
- The GPR method with the Matérn kernel function outperformed all other models in predicting ship speed with an R^2 of 0.91 but with more required training time.

- Multiple linear regression which is a considerably simpler algorithm has provided comparable accurate results.
- Regression trees and trees ensemble models have yielded accurate ship speed prediction with lower computational time. It should, however, be noted that the ensemble boosted trees was sensitive to the data scaling technique which affected its prediction accuracy.
- The performed sensitivity analysis showed that the SVM algorithms can be sensitive to the data scaling technique as well as the cross validation number of folds depending on the used covariance function. An accuracy increase of 11% and 3% has been achieved in the RMSE of the cubic and fine Gaussian SVM models respectively by increasing the cross validation fold number from 5 to 50.
- Other model performance hasn't improved noticeably by changing the fold number. Therefore, a 10-fold cross validation can be recommended for computationally efficient model performance in terms of prediction accuracy and complexity.
- Computational experiments have been conducted using the proposed methodology to manage the ship operational parameters and evaluate its effect on the ship speed where the simulation results were rational.

By accurately predicting ship speed, the outcomes of this paper can help ship management companies in creating further advanced models for the purposes of route optimization, ship tracking, voyage planning, etc. Also, the proposed methodology can be applied without difficulty to any ship type at different operational conditions or to predict or optimize other important operational parameters such as ship trim or propeller controllable pitch which can be part of the future work. Performance comparison with artificial neural networks should be also made in future studies considering the suitable network architecture, number of layers, neurons, etc.

6. Acknowledgement

The authors would like to thank Jóan Petur Petersen for the assistance and the provision of data utilized in this study.

References

Arlot, S., Celisse, A., et al., 2010. A survey of cross-validation procedures for model selection. Statistics surveys 4, 40–79.

Awad, M., Khanna, R., 2015. Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer nature.

Bazari, Z., Longva, T., 2011. Assessment of IMO mandated energy efficiency measures for international shipping. International Maritime Organization .

Bialystocki, N., Konovessis, D., 2016. On the estimation of ship's fuel consumption and speed curve: a statistical approach. Journal of Ocean Engineering and Science 1, 157–166.

Brandsæter, A., Vanem, E., 2018. Ship speed prediction based on full scale sensor measurements of shaft thrust and environmental conditions. Ocean Engineering 162, 316–330.

Capezza, C., Coleman, S., Lepore, A., Palumbo, B., Vitiello, L., 2019. Ship fuel consumption monitoring and fault detection via partial least squares and control charts of navigation data. Transportation Research Part D: Transport and Environment 67, 375–387.

Choi, J.E., Kim, J.H., Lee, H.G., Choi, B.J., Lee, D.H., 2009. Computational predictions of ship-speed performance. Journal of marine science and technology 14, 322–333.

Cockcroft, A.N., Lameijer, J.N.F., 2003. Guide to the collision avoidance rules. Elsevier.

Coraddu, A., Oneto, L., Baldi, F., Anguita, D., 2015. Ship efficiency forecast based on sensors data collection: Improving numerical models through data analytics, in: OCEANS 2015-Genova, IEEE. pp. 1–10.

Coraddu, A., Oneto, L., Baldi, F., Anguita, D., 2017. Vessels fuel consumption forecast and trim optimisation: a data analytics perspective. Ocean Engineering 130, 351–370.

Dalheim, Ø.Ø., Steen, S., 2021. Uncertainty in the real-time estimation of ship speed through water. Ocean Engineering 235, 109423.

Faber, J., Hanayama, S., Zhang, S., Pereda, P., Comer, B., Hauerhof, E., van der Loeff, W.S., Smith, T., Zhang, Y., Kosaka, H., Adachi, M., Bonello, J.M., Galbraith, C., Gong, Z., Hirata, K., Hummels, D., Kleijn, A., Lee, D.S., Liu, Y., Lucchesi, A., Mao, X., Muraoka, E., Osipova, L., Qian, H., Rutherford, D., de la Fuente, S.S., Yuan, H., Perico, C.V., Wu, L., Sun, D., Yoo, D., Xing, H., 2020. Fourth IMO GHG Study. Technical Report. CE Delft. Delft, Netherlands.

Gkerekos, C., Lazakis, I., Theotokatos, G., 2019. Machine learning models for predicting ship main engine fuel oil consumption: A comparative study. Ocean Engineering 188, 106282.

Holtrop, J., 1984. A statistical re-analysis of resistance and propulsion data. International shipbuilding progress 31, 272–276.

Hu, Z., Jin, Y., Hu, Q., Sen, S., Zhou, T., Osman, M.T., 2019. Prediction of fuel consumption for enroute ship based on machine learning. IEEE Access 7, 119497–119505.

ITTC, 2014a. Recommended procedures and guidelines – 1978 ITTC performance prediction method (7.5 – 02 – 03 – 01.4). Technical Report.

ITTC, 2014b. Recommended procedures and guidelines – Analysis of speed/power trail data (7.5 – 04 – 01 – 01.2). Technical Report.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning. volume 112. Springer.

Journée, J.M., 1976. Prediction of speed and behaviour of a ship in a sea-way. International Shipbuilding Progress 23, 285–301.

Krata, P., Szlapczynska, J., 2018. Ship weather routing optimization with dynamic constraints based on reliable synchronous roll prediction. Ocean Engineering 150, 124–137.

Leifsson, L.Þ., Sævarsdóttir, H., Sigurðsson, S.Þ., Vésteinsson, A., 2008. Grey-box modeling of an ocean vessel for operational optimization. Simulation Modelling Practice and Theory 16, 923–932.

Mao, W., Rychlik, I., Wallin, J., Storhaug, G., 2016. Statistical models for the speed prediction of a container ship. Ocean engineering 126, 152–162.

MATLAB, 2021. https://www.mathworks.com/help/stats/choose-regression-model-options.html. Accessed: 2021-07-01.

Molland, A., Turnock, S.R., Hudson, D., 2011. Ship resistance and propulsion: practical estimation of ship propulsive power. Cambridge University Press.

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

Pedersen, B.P., Larsen, J., 2009. Prediction of full-scale propulsion power using artificial neural networks, in: Proceedings of the 8th international conference on computer and IT applications in the maritime industries (COMPIT09), Budapest, Hungary May, pp. 10–12.

Petersen, J.P., Jacobsen, D.J., Winther, O., 2012a. Statistical modelling for ship propulsion efficiency. Journal of marine science and technology 17, 30–39.

Petersen, J.P., Winther, O., Jacobsen, D.J., 2012b. A machine-learning approach to predict main energy consumption under realistic operational conditions. Ship Technology Research 59, 64–72.

Propulsion modelling, 2021. http://cogsys.imm.dtu.dk/propulsionmodelling/data.html. Accessed: 2021-03-01.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press.

Rehmatulla, N., Calleya, J., Smith, T., 2017. The implementation of technical energy efficiency and CO2 emission reduction measures in shipping. Ocean Engineering 139, 184–197.

Shenoi, R., Bowker, J., Dzielendziak, A.S., Lidtke, A.K., Zhu, G., Cheng, F., Argyos, D., Fang, I., Gonzalez, J., Johnson, S., et al., 2015. Global marine technology trends 2030 .

Similä, M., Lensu, M., 2018. Estimating the speed of ice-going ships by integrating sar imagery and ship data from an automatic identification system. Remote Sensing 10, 1132.

Smith, T., Parker, S., Rehmatulla, N., 2011. On the speed of ships, in: International Conference on Technologies, Operations, Logistics and Modelling for Low Carbon Shipping, LCS2011. University of Strathclyde, Glasgow, UK, pp. 22–24.

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

Soner, O., Akyuz, E., Celik, M., 2019. Statistical modelling of ship operational performance monitoring problem. Journal of Marine Science and Technology 24, 543–552.

Stein, M.L., 1999. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.

Uyanık, T., Karatuğ, Ç., Arslanoğlu, Y., 2020. Machine learning approach to ship fuel consumption: A case of container vessel. Transportation Research Part D: Transport and Environment 84, 102389.

Wang, X.G., Zou, Z.J., Hou, X.R., Feng, X., 2015. System identification modelling of ship manoeuvring motion based on support vector regression. Journal of Hydrodynamics, Ser. B 27, 502–512.

Yan, R., Wang, S., Du, Y., 2020. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. Transportation Research Part E: Logistics and Transportation Review 138, 101930.

Yoo, B., Kim, J., 2019. Probabilistic modeling of ship powering performance using full-scale operational data. Applied Ocean Research 82, 1–9.

Yuan, J., Nian, V., 2018. Ship energy consumption prediction with gaussian process metamodel. Energy Procedia 152, 655–660.

Zis, T.P., Psaraftis, H.N., Ding, L., 2020. Ship weather routing: A taxonomy and survey. Ocean Engineering 213, 107697.