Ship speed prediction based on full scale sensor measurements of shaft thrust and environmental conditions

Andreas Brandsæter[a,b,1], Erik Vanem[a,b]

[a] *DNV GL, Veritasveien 1, N-1363 Høvik, Norway*
[b] *Department of Mathematics, University of Oslo, P.B.1053, Blindern, N-0316 Oslo, Norway*

Abstract

The primary goal of this study is to adapt and validate various regression models to predict a ship's speed through water based on relevant and available full scale sensor measurements from a ship, including measurements of external environmental forces. The wind is measured by on-board wind sensors, and the effect of the waves is measured by motion reference units (MRUs) installed on the ship, measuring motions in six degrees of freedom; three translational motions and rotations about these. Accurate speed estimates, which relate directly to the estimates of the propulsion efficiency, fuel efficiency and pollution, are vital to be able to optimize ship design and operation. We demonstrate how regression models such as linear regression, projection pursuit (PPT) and generalized additive models (GAM) can be easily implemented for this application. A simple regression model based on the well-established relationship between ship speed and shaft thrust represent a benchmark model towards which the other models are compared.

Keywords: , Ship speed estimation, Computational Methods/Numerical Analysis, Sensor data analytics, Ship Resistance and Propulsion, Performance measure quantification, Statistical modelling, Energy efficiency

1. Introduction

Accurate estimates of ship propulsion and fuel efficiency are important to be able to optimize ship design and operation. Deviations between the measured ship speed and the speed predictions based on propulsion power and other internal and external conditions can be indicative of an anomaly, such as e.g. hull, propeller or engine damage. Furthermore, the effect of modifications can be quantified. This can include modifications of the ship hull, such as for example hull cleaning or bow optimization, installation of new equipment such as kites, fixed sails or batteries for machinery optimization, propeller optimization such as contra rotating propeller and various efficiency improvement devices, and operational optimization measures such as weather routing and trim and draft optimization. The logistics planning can also be optimized with accurate of time of arrival estimation.

Due to the complexity of a modern ship and its exposure to external factors such as wind, waves and currents, estimating the ship efficiency accurately is not an easy task. Various methods are described in literature and used by the industry. The methods can be divided into four main groups as suggested by (Petersen et al., 2012):

1. Traditional and standard series methods which typically rely on a set of parameters describing the hull (Savitsky, 1964; Øyan, 2012; Holtrop and Mennen, 1982),
2. regression based methods based on a set of sensor measurements (Petersen et al., 2012; Bocchetti et al., 2015; Mao et al., 2016),
3. direct model tests in test tanks (Chuang and Steen, 2011), and
4. computational fluid dynamics (CFD) (Peri et al., 2001; Sadat-Hosseini et al., 2013; Ozdemir and Barlas, 2016).

The methods are often ordered on a scale between methods that are governed by physical laws and empirical or data driven methods (sometimes referred to as black box methods) that are based on statistical inference of historical data. Due to the fact that the data driven methods require little knowledge of the physical system (Coraddu et al., 2017) and there is no need to manually build a model of the data, these methods can be easily implemented in marine operations; making such technologies a lean alternative to complex tailor-made analytics (Brandsæter et al., 2016). At the same time, the data driven methods can be unsatisfactory in terms of physical explanation and it might require a significant amount of data to be sufficiently accurate (Vanem et al., 2017; Petersen et al., 2012).

The primary goal is to survey various regression models to estimate the ship speed through water based on relevant and available sensor measurements of the shaft thrust and external environmental forces from wind, waves and currents. The wind is measured by on-board wind sen-

Email address: andreas.brandsaeter@dnvgl.com (Andreas Brandsæter)

sors, and the effect of the waves is measured by motion reference units (MRUs) installed on the ship, measuring motions in six degrees of freedom; three translational motions and rotations about these, and the sea currents are incorporated in the measure of ship speed through water. The speed through water y relates directly to the propulsion efficiency which is commonly defined as $e_{prop} = \frac{y}{p}$, where p is the propulsion power. It is also linked directly to the fuel efficiency $e_{fuel} = \frac{y}{f}$, where f is the energy in the consumed fuel, as defined by (Petersen et al., 2012).

We work towards a better understanding of how the external conditions affect the ship's speed, propulsion and fuel efficiency, and aim to be able to quantify these effects. Several case studies using different methods are described in the recent literature, both with main focus on propulsion efficiency estimation(Petersen et al., 2012; Vanem et al., 2017; Chuang and Steen, 2011; Øyan, 2012; Holtrop and Mennen, 1982; Mao et al., 2016) and fuel efficiency and emission estimation (Trodden et al., 2015; Bialystocki and Konovessis, 2016; Coraddu et al., 2017; Rakke, 2016; Bocchetti et al., 2015).

2. Methodology

2.1. Regression models

We employ various regression models to describe and predict the data, i.e., linear regression models, generalized additive models (GAM) and projection pursuit regression (PPR) models. Other models including various regression trees and kernel density estimation were also explored to some extent. We were not able to tune these models to provide accurate predictions, hence they are omitted. In the following, a brief introduction to each of the models will be provided. Reference is made to textbooks such as (Hastie et al., 2009) for a more thorough introduction.

The response variable will be denoted Y and the explanatory variables will be denoted $\mathbf{X} = (X_1, X_2, \ldots, X_P)$. The basic problem is to construct a prediction rule for predicting Y conditioned on the explanatory variables based on a stochastic model on the form

$$Y = f(\mathbf{X}) + \varepsilon, \tag{1}$$

where ε represents stochastic white noise and is often modelled as a zero-mean Gaussian variable. Different models for the $f(\cdot)$ function give rise to different regression models. Assuming N observations, the observed values are y_j and $\mathbf{x}_j$ for $j = 1, \ldots, N$.

2.1.1. Linear regression models

In linear regression one assumes a linear model on the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_P X_P + \varepsilon. \tag{2}$$

The error made in such predictions are referred to as the residuals, and the residual sum of squares (RSS) will be a function of the model parameters. It is defined as

$$RSS(\hat{\beta}) = \sum_{j=1}^{N} (y_j - \hat{y}_j)^2 \tag{3}$$

and the fitted model parameters $\hat{\beta}_p, p = 0, \ldots, P$ are estimated from the data. In ordinary least squares, these estimators are found as

$$\hat{\beta}^{OLS} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{Y}. \tag{4}$$

In some cases one may obtain better predictions by setting some of the coefficients to zero. This will introduce bias, but may reduce mean square error by reducing variance. Deleting some coefficients gives a simpler model that may also be easier to interpret if P, the number of explanatory variables, is large, but interpretation could be affected by highly correlated covariates. Variable subset selection keeps the intercept in the model and selects a subset of the explanatory variables and ignores the rest. Compared to the standard linear regression, the accuracy when applying best subset linear regression was found to be similar, but somewhat poorer, and we omit the results here. Another possibility is to shrink some of the coefficients by for example using ridge regression, Lasso, least angle regression, principle component regression and partial least squares (Vanem et al., 2017).

2.1.2. Generalized additive models (GAM)

The generalized additive model has the form

$$Y = \alpha + \sum_{p=1}^{P} f_p(X_p) + \varepsilon \tag{5}$$

where $f_p(\cdot)$ are smooth functions of one covariate (Wood, 2006). Estimation of the smooth functions to fit the data as well as possible and to be as smooth as possible can be formulated as minimization of a penalized sum of squares, where a tuning parameter λ is introduced to control the degree of smoothing as follows:

$$\begin{aligned} PRSS(\alpha, f_p) = &\sum_{j=1}^{N} \left(y_j - \alpha - \sum_{p=1}^{P} f_p(x_{jp}) \right) \\ &+ \lambda \sum_{p=1}^{P} \int_t f_p''(t_p)^2 dt_p. \end{aligned} \tag{6}$$

If the tuning parameter $\lambda \to \infty$, the estimate for f will approach a straight line, while choosing $\lambda = 0$ results in an un-penalized regression spline estimate. We choose the tuning parameter λ using generalized cross validation (GCV), applying the mgcv package in R (Wood, 2017).

To impose additional smoothing, we inflate the model degrees of freedom in the GCV by choosing a constant multiplier γ (Wood, 2017). We report results using

$$\gamma = 0.5 * log(P) \tag{7}$$

where P is the number of explanatory variables. We refer to this model as GAM smooth.

A GAM model based on a subset of the variables is also fitted and the results are reported. The subset contains the variables believed to be most important, but it is not necessarily the optimal subset. The following variables are selected: Shaft thurst, Heave (M0), Heave (T02), Surge (T02), Sway (M0), Roll (T02), Yaw (T02), Wind-x, Wind-y and trim. This model is referred to as GAM selected.

2.1.3. Projection pursuit regression (PPR) models

A further generalization of the generalized additive model can be obtained where the smooth functions are allowed to be non-linear smooth functions of some linear combinations of the covariates $\mathbf{X}$. Projection pursuit regression models are such models on the form

$$\begin{aligned} f(\mathbf{x}) &= \beta_0 + \sum_{m=1}^{M} g_m(\boldsymbol{\omega}_m^T \mathbf{X}) \\ &= \beta_0 + \sum_{m=1}^{M} g_m \left(\sum_{j=1}^{p} \omega_{mj} X_j \right). \end{aligned} \tag{8}$$

The g_m's are smooth functions (called ridge functions, varying only in the direction defined by the vector $\boldsymbol{\omega}_m$) estimated from the data together with the direction vectors $\boldsymbol{\omega}_m$. The direction vectors are normalized, $\|\boldsymbol{\omega}_m\| = 1$ and one must choose parameter M and smoothness of each of the g_m's (tuning parameters). Note that the scalar variable $V_m = \boldsymbol{\omega}_m^T \mathbf{X}$ is the projection of $\boldsymbol{X}$ onto the unit vector $\boldsymbol{\omega}_m$. Projection pursuit regression models are able to handle interaction effects between variables and are in fact universal approximators in that they can approximate almost any function for large M. We report results from projection pursuit models using 3 and 10 terms.

2.2. Baseline regression method

The above mentioned methods use wind and waves measurements as well as shaft thrust as explanatory variables. We compare the accuracy of these methods with predictions produced using a baseline method which does not take weather effects into account. The baseline model is based on the following well known relationship between the ship's speed and power demand in calm sea conditions (Tupper, 2004)

$$A_c \approx \frac{\Delta^{2/3} Y^3}{P} \propto \frac{Y^2}{S}. \tag{9}$$

Here, A_c denotes the admiralty coefficient which is assumed to be constant for the ship, and Δ denotes the ships weight displacement which we also assume to be constant. The speed through water [knots] is denoted by Y and the installed power [kW] is denoted by P. The shaft thrust [kN], denoted by S, is assumed to be approximately proportional to power over speed.

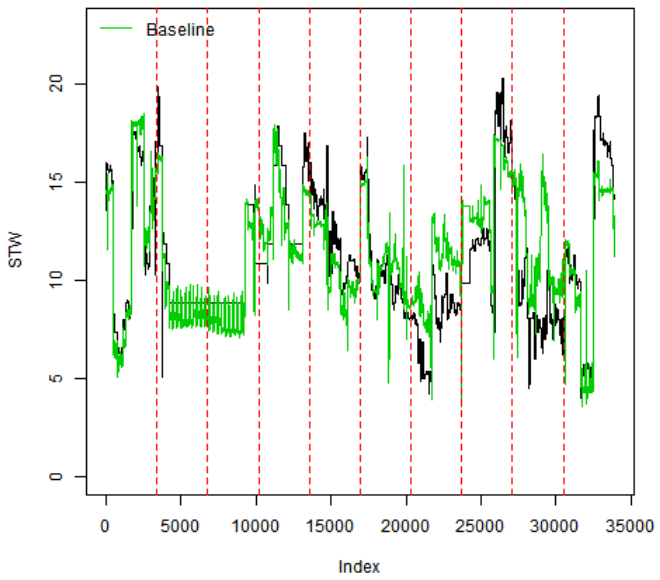


Figure 1: Prediction using the baseline method. The black line shows the observed speed through water, and the green line shows the speed through water predicted by the baseline method. The red dotted vertical lines indicates the boundry between the different folds. Here we use 10-fold cross validation.

Based on the this relationship we construct the following simple baseline method for the ship's speed through water:

$$Y_j = \beta \sqrt{S_j} + \epsilon_j \tag{10}$$

where S_j is the measured shaft thrust, ϵ_j is the prediction error, and β is a constant calculated from the training data

$$\beta = \frac{\sum_{i \in \mathcal{D}_t} Y_j^2}{\sum_{i \in \mathcal{D}_t} S_i} = \frac{\bar{Y^2}}{\bar{S}} \tag{11}$$

where $\mathcal{D}_t$ denotes the training dataset. This β minimizes the sum of squares:

$$\sum_{i \in \mathcal{D}_t} \left(Y_i - \beta \sqrt{(S_i)} \right)^2. \tag{12}$$

We refer to this method as the baseline method. Due to the difficulties comparing our results with previous work (Petersen et al., 2012), we use this method to benchmark the different models explored in this paper.

The speed through water (STW) predictions produced by the baseline method are displayed in figure 1. Here, we use 10-fold cross validation, and the 10 folds are distinguished in the figure by the vertical red dotted lines, see section 2.3 for further explanation.

2.3. Cross validation

It is well known that when we evaluate predications from a statistical model on the dataset used to train the model,

our accuracy estimates tend to be overoptimistic (Arlot and Celisse, 2010). To address this issue, a basic approach is to divide the dataset $\mathcal{D}$ into two exclusive parts $\mathcal{D}_t$ and $\mathcal{D}_k$; where one part $\mathcal{D}_t$ is used to train the model, and the other $\mathcal{D}_k$ is reserved for validation. To build robust and accurate models we ideally want to use all data available. The same applies to testing; we want to test our models in all situations, not only on a subset. Cross validation introduces various methods of repetitively splitting the data into training and validation datasets. Each of the splits provides a validation estimate, and by averaging over all the estimates we get a cross validation estimate. A range of different splitting techniques can be applied, providing different cross validation estimates. See for example (Arlot and Celisse, 2010; Kohavi, 1995) for a brief overview of the most common splitting techniques.

2.3.1. Standard K-fold cross validation

In this study we restrain to K-fold cross validation, which in its standard form splits the original dataset $\mathcal{D}$ into K subsets (folds) $\mathcal{D}_1, \ldots, \mathcal{D}_K$, as described in (Arlot and Celisse, 2010). Here, we choose the sets to be mutually exclusive with equal size. For each $k \in 1, 2, \ldots, K$ the models are trained on $\mathcal{D}_t = \mathcal{D} \setminus \mathcal{D}_k$, and tested on $\mathcal{D}_k$. Furthermore, we experiment with several K-s and report results with $K = 10$, 20, 50 and 3385. When $K = 3385$, each fold contains 10 points.

2.3.2. Modified K-fold cross validation

Cross validation is applicable to almost any algorithms and frameworks, involving regression, classification and many others (Arlot and Celisse, 2010). The only assumptions needed is that the data is identically distributed and that the training and validation sets are independent. The data itself do not need to be independent (Arlot and Celisse, 2010). In the case where the data is not independent, as is often the case with date retrieved from continuous sensor measurements, the independence between the training and validation datasets can be controlled by choosing $\mathcal{D}_t$ and $\mathcal{D}_k$ such that

$$min_{i \in \mathcal{D}_t, j \in \mathcal{D}_k} |i - j| > h$$

where h is a parameter $h > 0$.

Modified cross validation excludes from the training data the data in the folds which are adjacent to the validation set, that is for each $k \in 1, \ldots, K$ the models that are tested on $\mathcal{D}_k$ are trained on $\mathcal{D} \setminus \{\mathcal{D}_{k-1} \cup \mathcal{D}_k \cup \mathcal{D}_{k+1}\}$.

2.3.3. Repeated K-fold cross validation

To make sure that the results are not strongly dependent on how the folds are selected, we repeatedly run the K-fold cross validation with new selections. When we run K-fold cross validation with L repetitions, we divided the original dataset $\mathcal{D}$ into K subsets (folds) in L different ways, such that no folds selected are equal.

That is, we choose $\mathcal{D}_k^l$, the dataset of the k-th in the l-th repetition, such that $\mathcal{D}_k^l \neq \mathcal{D}_k^{\dot{l}}$ for all $\dot{l} \neq l$ and $\dot{k} \neq l$.

2.4. Model comparison and evaluation

2.4.1. Root mean squared error (RMSE)

After fitting the various models on the training data $\mathcal{D}_t$, the performance of the predictions are evaluated based on the root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(\hat{Y}_j - Y_j \right)^2} \tag{13}$$

where $\hat{Y}$ denotes the predicted speed through water based on shaft thrust and external forces, while Y is the direct speed measurement. N denotes the number of data points in the validation set. The model with the lowest RMSE will be preferred.

2.4.2. R squared

For initial model checking we use the coefficient of determination, R squared

$$R^2 = 1 - \frac{\sum_{i \in \mathcal{D}_t} (y_i - \hat{y}_i)^2}{\sum_{i \in \mathcal{D}_t} (y_i - \bar{y}_i)^2}. \tag{14}$$

This is a measure of how well the regression model explains the the total variation of the response variable. The coefficient varies from 0 to 1, where 1 indicates a perfect fit of the model to the data. The adjusted R^2 introduces a penalty for increasing (effective) number of parameters in order to avoid overly complex models.

2.5. Data description

This study is based on an extended version of the dataset used in (Vanem et al., 2017). For completeness, parts of the data description provided in (Vanem et al., 2017) is rendered in the following with minor modifications.

The dataset contains variables associated with the efficiency of the ship machinery system, such as the speed through water (knots), propulsion power [kW] and shaft thrust (N). The shaft thrust is assumed to be proportional to the propulsion power over speed over ground. Other variables included in the dataset are related to the ship's motions, wind speed relative to the ship and trim and draft. These variables represent external factors and are used to explain variation in the efficiency and ship speed.

The ship is installed with two motion reference units (MRUs) measuring the ship's motion in all six degrees of freedom (heave, surge, sway, roll, yaw and pitch). From the raw motion data recorded at 5 Hz various integrated parameters are stored every 30 seconds, calculated from the preceding 15 minutes time record of the motions. The integrated parameters reported by the system include the first five spectral moments of each motion signal (m_0, m_1, m_2, m_3 and m_4), the mean, standard deviation, skewness and kurtosis of the signal as well as the maximum and minimum values during the time window. Also the spectral peak period T_p and zero crossing period T_z were recorded.

Out of these parameters many are not relevant for the present analysis and are not considered here. Besides, some of the parameters carry redundant information and can be derived from other parameters, such as the standard deviation of the signal $\sigma = \sqrt{m_0}$ and $T_z = T_{02} = \sqrt{m_0/m_2}$. We therefore limit ourselves to consider m_0 and T_z for each of the degrees of freedom. The zeroth spectral moment $m_0 = \sigma^2$ is the total energy of the motion spectrum and indicates the magnitude of the ship motion. Note that for a wave record the significant wave height is usually defined as $H_s = 4\sqrt{m_0}$. Likewise, T_z indicates the typical period of the different motions. Since the periodic ship motions are primarily an effect of the waves, both m_0 and T_z can be considered as proxies for the real wave conditions in the sense that m_0 will be proportional to the real significant wave height and T_z will be similar to the typical period of the wave field. Moreover, the ship response in the different degrees of freedom will to a certain extent reflect the wave direction relative to the ship.

In addition to the ship motions, representing the effect of the waves, the wind speed relative to the ship is recorded; the wind component perpendicular to the ship (Wind-y) and the wind parallel to the ship (Wind-x). Wind-x is defined so that a positive value means wind blowing in the same direction as the ship speed. Two other parameters that are important for the hydrodynamic resistance of a ship is the draft and trim, which have also been recorded and included in the present analyses. The draft is defined as the vertical distance between the waterline and the bottom of the hull and is naturally related to the cargo level of the ship, while the trim is the difference between the forward and aft drafts.

For the analysis presented in this paper the original 30 second data were down-sampled to 5 minutes resolution by calculating the average values within each 5 minute window. This makes the dataset smaller and easier to handle, and reduces the time dependency.

The data have been collected from an ocean-going ship over approximately 10 months in normal operation. Due to data quality issues, a large fraction of the data were removed in initial cleaning and outlier removal. For example when the difference between the measured speed through water and speed over ground is significantly larger than reasonable current, at least one of the measured speed sensors must report wrong values. Although the measurement of the speed through water is known not to be most reliable (van den Boom and Hasselaar, 2014), it is difficult to know which reading is wrong, hence we remove the data point. After the initial filtering and outlier removal, the dataset used in the analysis contains about 33000 data points, which correspond to about 115 days of data.

Initially, 18 selected variables are included in the analysis. Trace plots of the data are shown in figure 2. The upper plot shows the speed and thrust data series, the next shows all the ship motion data and the two lower plots show the wind and the trim and draft data, respectively.

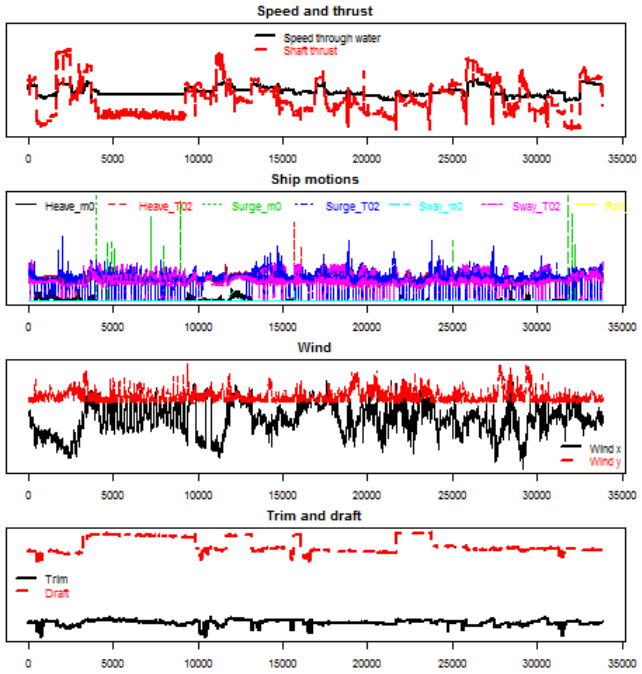


Figure 2: Trace plot of the data

A correlation plot showing the linear correlation between the various variables in the data set is shown in figure 3. It is seen that the a highly correlated structure is present in the dataset.

3. Analysis and results

If we fit the regression models introduced above using only shaft thrust as explanatory variable, the prediction accuracy, in terms of RMSE based on predictions using 10-fold cross validation, are similar but still worse than the accuracy achieved with the baseline method. Out of the analysed methods, the linear model achieves the highest average RMSE of 2.055, some 0.8 % higher than the baseline method. When we instead use the square root of the shaft thrust as explanatory variable when fitting a linear model, the obtained RMSE is even slightly higher. This suggests that we need more elaborate methods to be able to increase our accuracy.

The total RMSE of the different regression models where all the explanatory variables are utilized, also variables which describe external conditions, are shown in the boxplot in figure 4. The RMSE values per repetition $l \in 1, \ldots, 7$ are shown, using 10-fold cross validation. For each regression method the box displays the inner quartile range (the range between the 25th percentile and the 75th percentile), and the median is marked with a horizontal line. The mean value is indicated with a cross, while the actual values are marked with circles. The whiskers in this plot displays the maximal and minimal values.

We observe that the spread in the results are not very large, which indicate that the results are not heavily de-

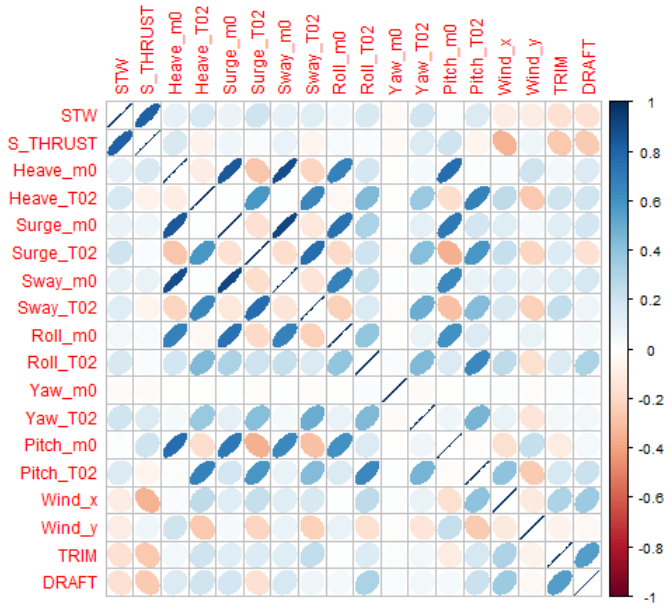


Figure 3: Correlation plot of the linear correlations in the data from the sensors signals used in the models. Dark blue color indicates strong correlations, white indicates no correlations, and dark red indicates strong negative correlation.

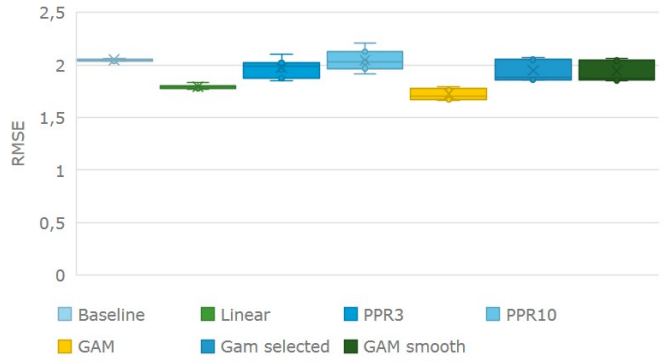


Figure 4: Boxplot of the RMSE per repetition $l \in 1, \ldots, 7$ for the different regression models, using 10-fold cross validation. For each regression method the box displays the inner quartile range, and the median is marked with a horizontal line. The mean value is indicated with a cross, while the actual values are marked with circles. The whiskers in this plot displays the maximal and minimal values.

pendent on how we chose to select the folds. The RMSE of both the linear model and the full GAM model achieves RMSE values significantly below the baseline model, with mean values 12 and 16 % below the baseline model respectively. The smoothed GAM and the GAM on selected explanatory variables achieves average values about 5% below the baseline method. The same applies to the PPR with 3 terms, while the average RMSE of the PPR with 10 terms are about the same as the baseline method.

Figure 5 shows the density of the residuals of the predictions by the baseline, linear, PPR with 3 terms and GAM models. Here we use 10-fold cross validation without repetition. The figure indicate that the predictions are not biased, except for the baseline method which tend to predict too high values.

Figure 6 shows a scatter plot with shaft thrust on the horizontal axis and speed through water on the vertical axis. The observed values are marked in black, while the predicted values based on the baseline, linear, GAM and PPR with 3 terms models are marked in green, red, blue and grey respectively. We note that the predictions based on the linear, PPR and GAM models have a good spread, while the baseline model does not take notice of the variations which we believe are induced by the external conditions.

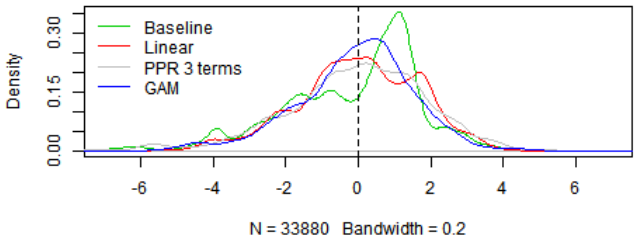


Figure 5: Densityplot of the residuals for the different models, using 10-fold cross validation without repetition. The bandwidth is set to 0.2.

3.1. Fold specific results

A boxplot of the fold specific RMSE values of the different regression models is displayed in figure 7. Here we have used 10-fold cross validation without repetition. The

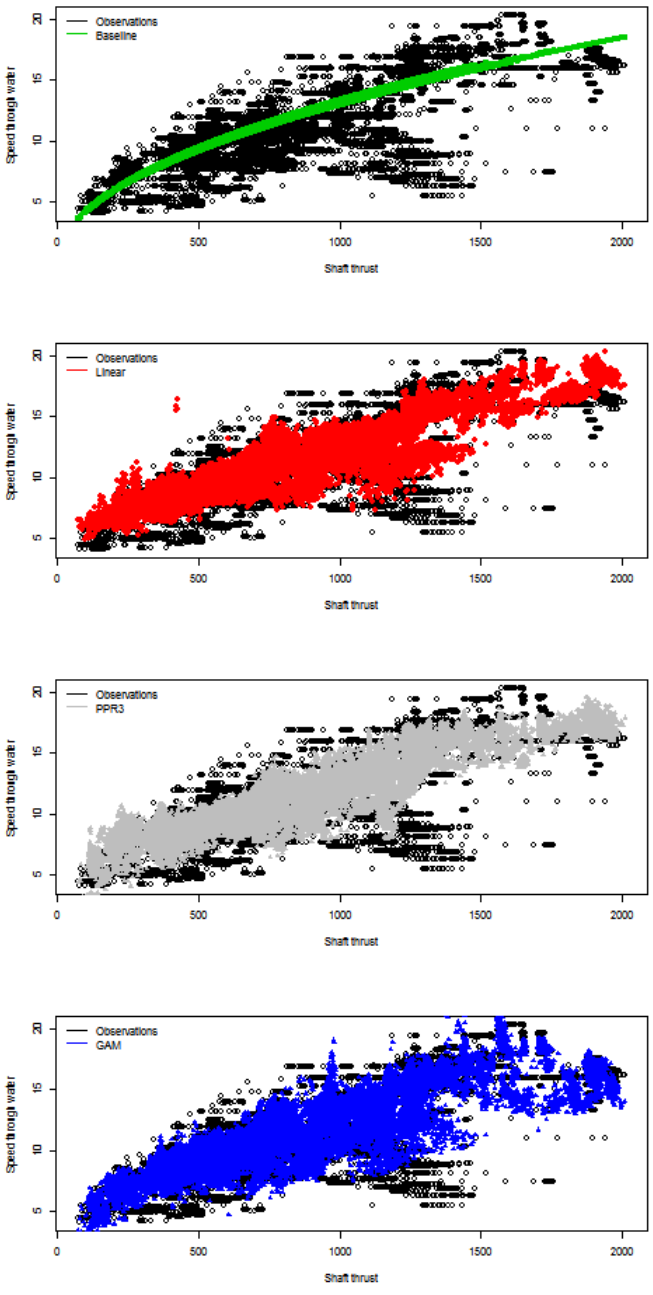


Figure 6: Scatter plot of the shaft thrust vs the speed through water and the values predicted using the different models, using 10-fold cross validation

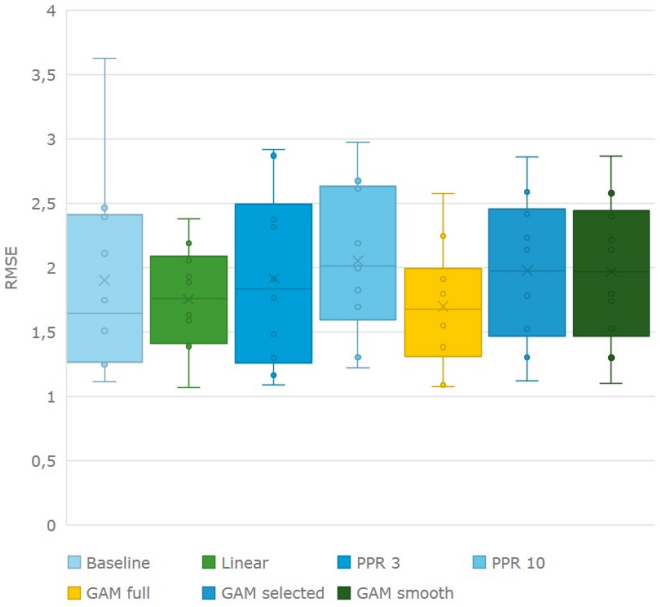


Figure 7: Boxplot of the RMSE per fold $k \in 1, \ldots, 10$, without repetition, for each regression model.

GAM model achieves the lowest 75th percentile, at 1.9, while the baseline method has a slightly lower median and 25th percentile compared to the GAM model. The linear model achieves the lowest maximal RMSE value at 2.4, which indicates robust performance. The results indicate that the baseline method performs well in many cases, but as we observe from the high maximal value, which is some 50 % higher than the maximum value reported for the linear model, it is not robust in all situations.

The R^2 values, which are calculated for each fold, are displayed in figure 8. The coefficient varies from 0 to 1, where 1 indicates a perfect fit of the model to the data. The PPR10 achieves the highest value, followed by the full GAM model, and the other PPR and GAM's. The lowest R^2 calculated for PPR10 and the GAM model is 0.89, which indicate that the total variation of the response variable is quite well explained by the regression model.

3.2. Comparing folds with calm and harsh weather

To demonstrate the importance of representative training data, and and the different regression models' strengths and weaknesses in this respect, we select two periods which represent calm and harsh weather and investigates these in somewhat greater depth.

Data from three selected weather sensors; heave, pitch and wind perpendicular to the ship are displayed in figure 9. Figure 10 shows a scatter plot of the wind component and the pitch motion. The figures indicate that calm weather are well represented by fold 2, while the weather conditions seam to be more severe in fold 9. Fold 2 and 9 are marked in magenta and cyan respectively.

Figure 11 shows a scatter plot with shaft thrust on the horizontal axis and speed through water on the vertical

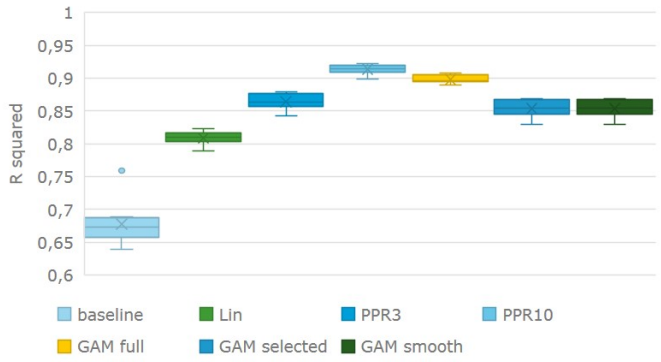


Figure 8: Boxplot of the adjusted R^2 per fold $k \in 1, \ldots, 10$, without repetition, for each regression model. The whiskers extend upwards to the largest element that is less than 1.5 times the inner quartile range higher than the upper quantile, and downwards to the lowest element. Elements outside this range are considered outliers, and are marked with a dot, as seen for one of the values using the baseline model.

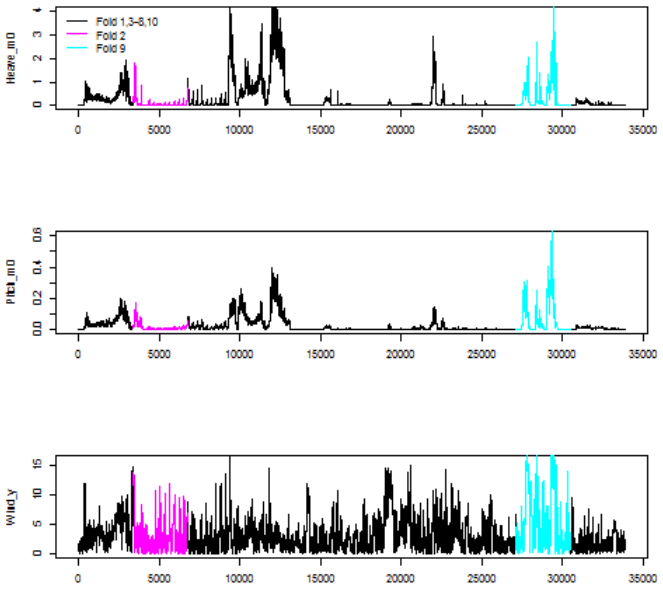


Figure 9: Traceplot of data from selected weather sensors; heave, pitch and wind perpendicular to the ship. The dataset is divided into 10 mutually exclusive folds with equal size. Fold 2 and 9 of are higlighted in magenta and cyan respectively

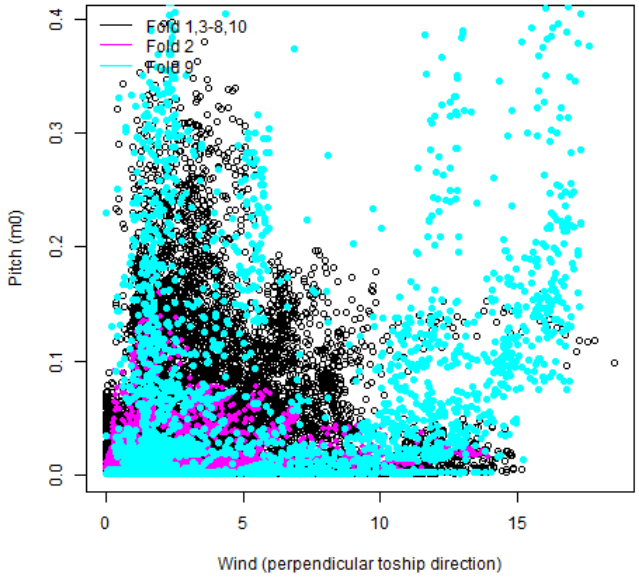


Figure 10: Scatter plot of the wind component perpendicular to the ship and the pitch motion. The dataset is divided into 10 mutually exclusive folds with equal size. Fold 2 and 9 of are higlighted in magenta and cyan respectively

axis. Again, fold 2 and 9 are highlighted. The observed values from the other folds are marked in black, while the predicted values based on the baseline models are marked in green. We observe that the observed values of fold 2 lie close to the baseline predictions. This indicates that the baseline method performs well in many cases, including calm weahter (fold 2), but we also observe that the baseline method is unable to achieve accurate predictions in harsh weather (fold 9). This is supported by the fold specific RMSE calculation for the baseline method, as reported in figure 12. It shows that the RMSE of fold 2 is low, while the reported RMSE of fold 9, where the weather conditions are more severe, is high.

The observed speed through water and corresponding predictions produced by the baseline, linear and GAM models on fold 2 and 9 are shown in figure 13.

From table 1, we see that in terms of RMSE, the baseline method outperforms both the GAM and the linear model in calm weather (fold 2), while in harsh weather (fold 9), the RMSE of both the linear and the GAM method is more than 40 % lower than the baseline method.

3.3. Modified K-fold cross validation

In the modified version of K-fold cross validation the data in the folds which are adjacent to the validation set are excluded from the training set. That is for each $k \in 1, 2, \ldots, K$ the models that are tested on $\mathcal{D}_k$ are trained on $\mathcal{D} \setminus \{\mathcal{D}_{k-1} \cup \mathcal{D}_k \cup \mathcal{D}_{k+1}\}$. For example, the training set used for predictions of the data points in fold 11 in

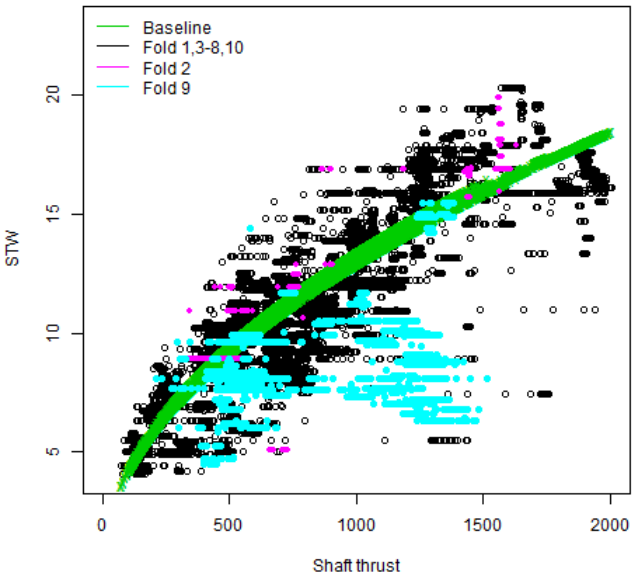


Figure 11: Scatter plot of the shaft thrust and speed through water. The green makers are the predictions based on the baseline method. The magenta and cyan markers show the observed data in fold 2 and 9 respectively. The data from the remaining folds are marked in black.

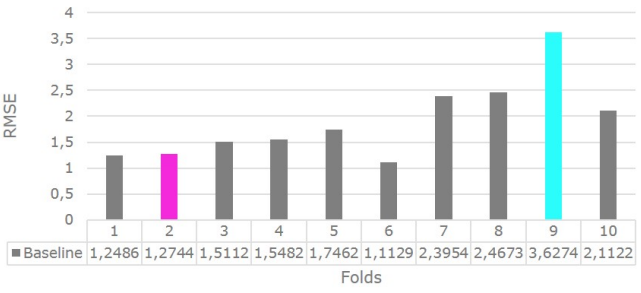


Figure 12: RMSE of the baseline prediction, for 10 different folds. The RMSE value of fold 2, representing calm weahter conditions, and fold 9, representing harsh weather conditions, are highlighted in magenta and cyan respectively.

Table 1: RMSE for fold 2 and 9, using 10-fold cross validation, without repetition

Regression method	Fold 2 RMSE	%	Fold 9 RMSE	%	Total RMSE	%
Baseline	1.27	0	3.80	0	2.06	0
Linear	1.39	9	2.06	-43	1.80	-12
GAM full	1.39	9	2.25	-38	1.76	-14

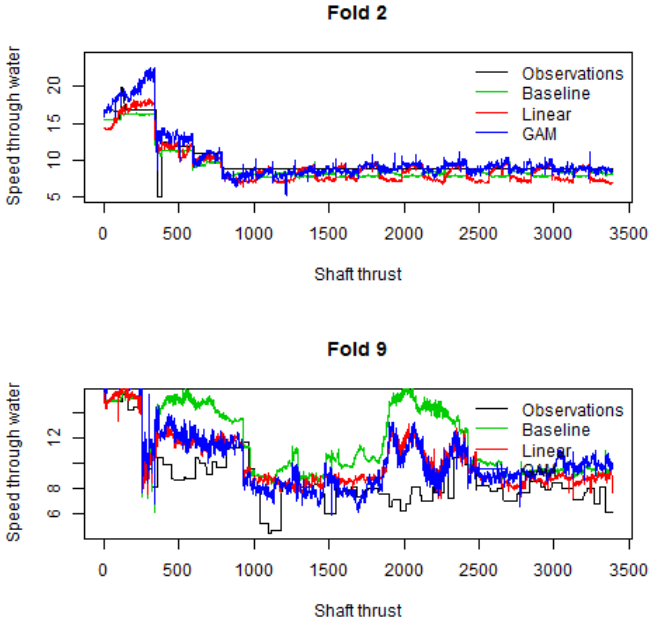


Figure 13: Prediction with different regression models on fold 2 (calm weather) and 9 (harsh weather), using 10-fold cross validation.

modified 30-fold cross validation consists of data from fold 1-9 and 13-30.

We observe that the predictions for the corresponding folds in the standard 10-fold and the modified 30-fold cross validation are similar. This is supported by the fold specific RMSE calculations, displayed in table 2. We note, however, that the RMSE for the predictions made by the GAM model on fold 4-6 in the 30-fold cross validation differ significantly from the RMSE calculation using standard 10-fold cross validation for the corresponding fold.

4. Discussion

4.1. Changing the numbers of folds in the cross validation

In figure 14 the RMSE of the different regression models are displayed with standard $K-$fold cross validation for K =10, 20, 50 and 3385. In addition results using 30-fold modified cross validation is presented.

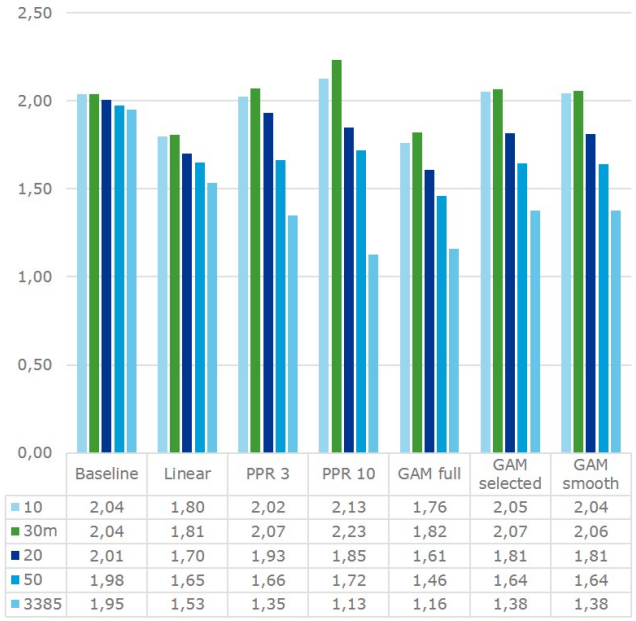
Clearly, the calculated RMSE values decrease with increasing number of folds in the cross validations. This is not surprising, due to the implicit increase in available training data. But it illustrates, however, that insufficiency in training data can lead to inaccurate predictions (Vanem et al., 2017; Petersen et al., 2012).

4.2. Ensemble methods

By combining the before-mentioned methods we are able to achieve slightly increased accuracy. For example, on average the GAM model achieves the best predictions, but we have observed that in cases where the pitch motion is higher than usual, as is the case in fold 9, the linear model

Table 2: RMSE for fold 4-6 and 25-27, in modified 30-fold cross validation

Regression method	Fold 4-6 RMSE	%	Fold 25-27 RMSE	%	Total RMSE	%
Baseline	1.26	0	3.57	0	2.04	0
Linear	1.40	11	2.00	-44	1.81	-11
GAM full	1.58	25	2.16	-40	1.82	-11



	Baseline	Linear	PPR 3	PPR 10	GAM full	GAM selected	GAM smooth
10	2,04	1,80	2,02	2,13	1,76	2,05	2,04
30m	2,04	1,81	2,07	2,23	1,82	2,07	2,06
20	2,01	1,70	1,93	1,85	1,61	1,81	1,81
50	1,98	1,65	1,66	1,72	1,46	1,64	1,64
3385	1,95	1,53	1,35	1,13	1,16	1,38	1,38

Figure 14: RMSE for the different regression mehods, with $K = 5$, 10, 20, 50 and 3385 folds. The dark blue bars, marked with 30 m, is the results of the modified 30 fold corss validation.

performs better. The ensemble model selects the prediction based on the GAM model when the pitch motion is below the mean value, and reversely, when the pitch motion is above the mean value, the prediction based on the linear model is selected.

Furthermore, if we are in an anomaly detection setting, we might want to make use of the information we have on the speed through water. We might then create an ensemble method which uses one model for speed through water above a given threshold, and another model for speeds below. As an example, we observe that the baseline model achieves better prediction accuracy at high ship speeds compared to the results of the GAM model. Hence we can use the baseline model at high speeds (for example above 11 knots) and use the GAM for lower speeds (below 11 knots).

Table 3 reports the RMSE values for the two ensemble methods described above. As reference, the results of the baseline model, the full GAM model and the linear model are also reported. The percentage relative difference between the results based on the baseline method and the other methods are reported in right hand column.

This method seems a bit ad-hoc and might be prone for over fitting. Hence, we merely mention it here, and have not investigated it to great depth. When prediction is not the aim, and we do have information about the speed, a simple ensemble method could prove to be useful. Also, comparing the predictions produced by the different methods could be used as an indicator of the prediction accuracy.

4.3. Insufficient training data

As illustrated above, the amount and quality of the training data available are critical for the methods explored here. For example, when a ship is entering a type of operation that is not well represented in the training data this will often cause inaccurate predictions. Training data can possibly be "borrowed" from sister ships or other ships with similar design. When the ships are not identical by design, the training data can possibly be reused after some modifications detailed by for example simulation software such as (Dimopoulos et al., 2014; Tillig et al., 2016). Notwithstanding, it should be noted that this of course can both be work intensive and might introduce biases and inaccuracies.

4.4. Operational mode selection

Typical operational modes of a ship include transit (in different speeds and loading conditions), harbour, stand by (with or without anchor) and dynamic positioning. During the different modes, the behaviour of the ship changes substantially, and it might therefore be advantageous to select among different methods based on the current operational mode. The training data should be divided and used to fit different models. This might result in reduced computational efforts and increased accuracy (Al-Dahidi et al., 2014; Baraldi et al., 2012; Brandsæter et al., 2016).

5. Conclusions

Accurate ship speed estimates, and accompanied propulsion and fuel efficiency estimates, are vital to be able to optimize ship design and operation. This paper points at some of the strengths and weaknesses associated with the use of standard statistical methods for speed predictions. The results are compared and benchmarked with respect to a simple model based on the well established relationship between ship speed and shaft thrust.

Table 3: RMSE values, including ensemble method, using standard 10-fold cross validation

Method	RMSE	%
Baseline	2.04	0
Linear	1.80	11.9
GAM	1.76	13.7
Ensemble: GAM-Linear (based on pitch)	1.74	-14.7
Ensemble: GAM-Baseline (based on STW)	1.70	-16.7

In many cases, especially in calm weather conditions, the baseline method performs well in terms of prediction accuracy. Furthermore, the various regression models explored were not able to outperform the simple baseline method when shaft thrust was the only explanatory variable, but when environmental conditions were included, the accuracy of the predictions were significantly increased.

By the models investigated in this work both the generalized additive model (GAM) and the linear models prove most useful, with increased accuracy of 16 and 12 % compared to the baseline model respectively, using 10-fold cross validation.

When the cross validation was performed on higher number of folds, the relative difference increased significantly, essentially for the GAM model, which when validated on the 50-fold and the 3385-fold cross validation, achieved an accuracy increase of 26 and 41 % respectively, compared to the baseline method.

The lower RMSE achieved by the GAM model in the case where the $K-$fold cross validation were performed on higher K-s might indicate that the GAM would perform even better with a more extensive and more representative dataset available. Also the projection pursuit models were satisfactory with respect to accuracy when evaluated on a high number of folds, but they were not able to produce accurate predictions when the number of folds in the K-fold cross validations were in the lower range.

It is not surprising that the accuracy increases when the number of folds increase, since this makes more data available when training the model. Nevertheless, the large accuracy increase points at the importance of large, relevant, high quality datasets, which is difficult to obtain.

6. Acknowledgements

The research is partly funded by the Research Council of Norway, project number 237718 and 251396. We would like to thank Ingrid Kristine Glad (University of Oslo) and Magne Aldrin (Norwegian Computing Center) for good discussions.

Appendix A. Additional figures

Appendix A.1. Increasing the number of folds

Figure A.15 shows a boxplot of the fold specific RMSE using 50-fold cross validation without repetition.

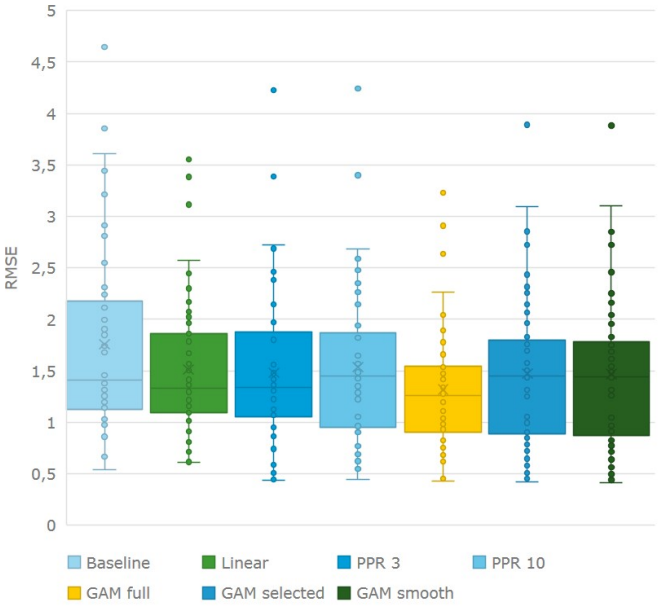


Figure A.15: Boxplot of the RMSE per fold $k \in 1, \ldots, 50$, without repetition, for each regression model.

Appendix A.2. Other performance measures

The mean error $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)$ is reported in the boxplot in figure A.16. Again, we use 10-fold cross validation with 7 repetitions. The figure indicate that the predictions are not biased, except for the baseline method which tend to predict too high values.

The relative mean absolute error (MAE) is reported in the boxplot in figure A.17. Here we report the results per $l \in 1, \ldots, 7$ for the different regression models, using 10-fold cross validation. The GAM model achieves the lowest value, for each of the different cross validations.

The relative mean absolute percentage error (MAPE) is reported in the boxplot in figure A.18. Again, we report the results per repetition $l \in 1, \ldots, 7$ for the different regression models, using 10-fold cross validation.

Appendix A.3. Intermediate results

In figure A.19 a diagnostics plot of the linear model is shown. Here, all data are used for training, leaving no data for model validation. In this situation the obtained adjusted R squared value is 0.807.

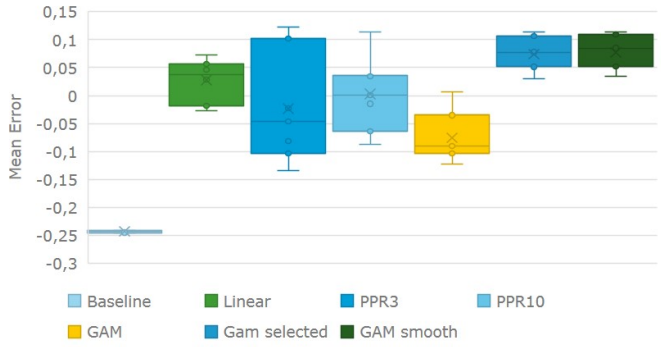


Figure A.16: Boxplot of the total mean error per repetition $l \in 1, \ldots, 7$ for the different regression models, using 10-fold cross validation.

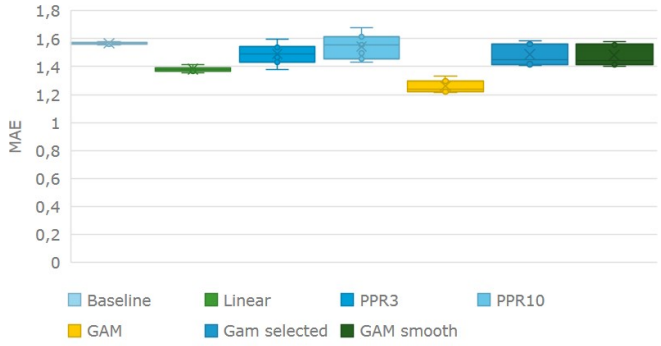


Figure A.17: Boxplot of the mean absolute error (MAE) per repetition $l \in 1, \ldots, 7$ for the different regression models, using 10-fold cross validation.

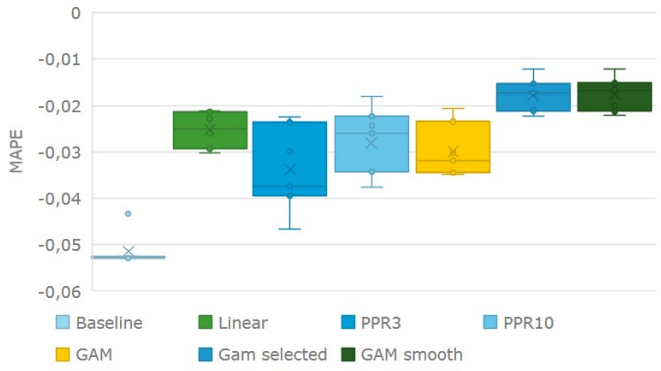


Figure A.18: Boxplot of the mean absolute percentage error (MAPE) per repetition $l \in 1, \ldots, 7$ for the different regression models, using 10-fold cross validation.

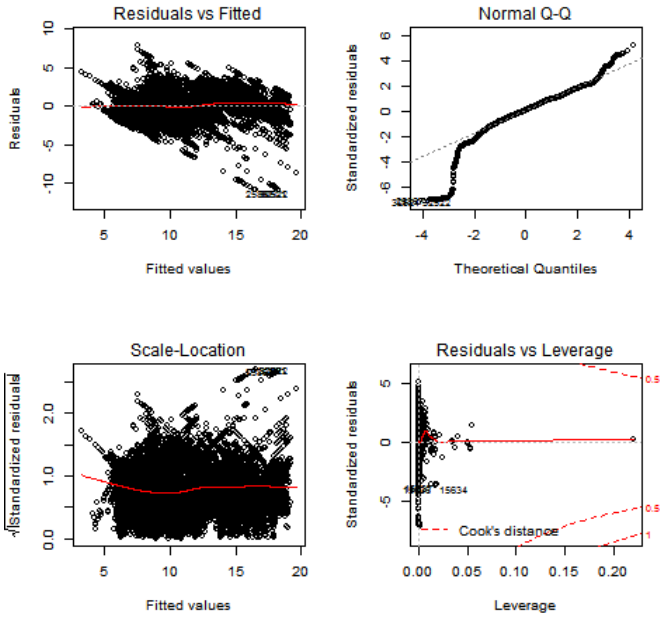


Figure A.19: Diagnostics plot of the linear model.

A diagnostics plot for the GAM model is shown in figure A.20. Again, all data are used for model fitting. When this is done for the GAM model, the adjusted R squared value is 0.894. The estimated functions for the GAM model using a selection of the available explanatory variables are displayed in figure A.21.

Figure A.22 displays the estimated functions for the PPR model with 10 terms. The adjusted R squared value is 0.891.

References

Al-Dahidi, S., Baraldi, P., Di Maio, F., Zio, E., 2014. Quantification of signal reconstruction uncertainty in fault detection systems, Second European Conference of the Prognostics and Health Management Society.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statist. Surv. 4, 40–79. URL: http://dx.doi.org/10.1214/09-SS054, doi:10.1214/09-SS054.

Baraldi, P., Di Maio, F., Pappaglione, L., Zio, E., Seraoui, R., 2012. Condition monitoring of electrical power plant components during operational transients. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, SAGE 226, 568–583.

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. URL: //www.sciencedirect.com/science/article/pii/S2468013315300127, doi:http://dx.doi.org/10.1016/j.joes.2016.02.001.

Bocchetti, D., Lepore, A., Palumbo, B., Vitiello, L., 2015. A statistical approach to ship fuel consumption monitoring. Journal of Ship Research 59, 162–171. doi:http://dx.doi.org/10.5957/JOSR.59.3.150012.

van den Boom, H., Hasselaar, T., 2014. Ship speed-power performance assessment, SNAME Annual Meeting, Houston, TX.

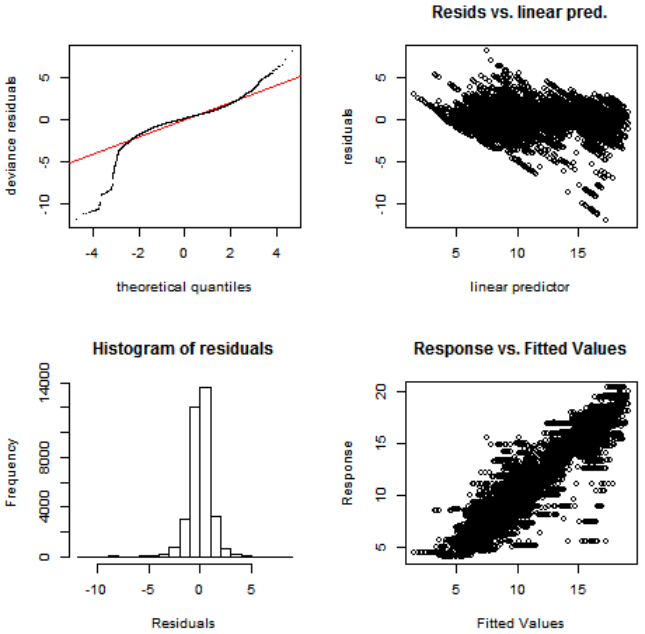


Figure A.20: Diagnostics plot of the GAM model.

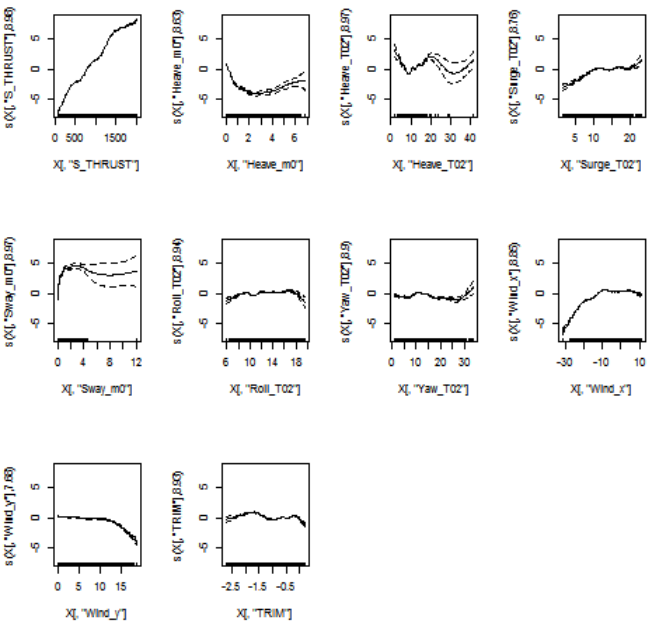


Figure A.21: Estimated functions for the GAM model using a selection of the available explanatory variables.

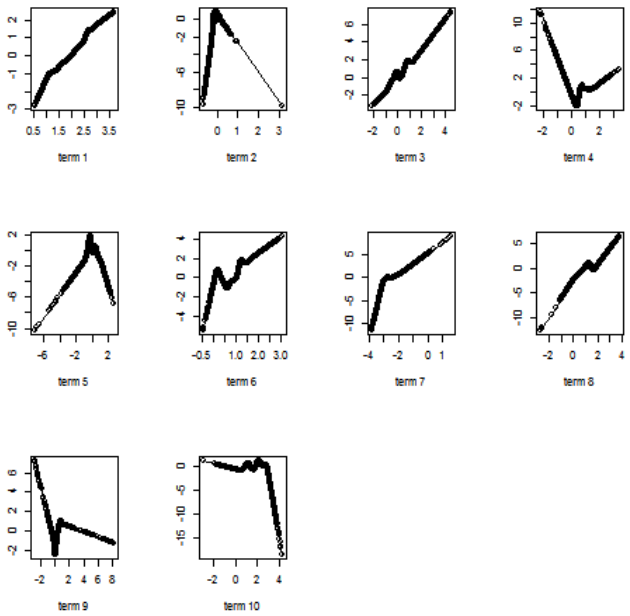


Figure A.22: Estimated functions for the PPR model with 10 terms.

Brandsæter, A., Manno, G., Vanem, E., Glad, I.K., 2016. An application of sensor-based anomaly detection in the maritime industry, in: 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 1–8. doi:10.1109/ICPHM.2016.7811910.

Chuang, Z., Steen, S., 2011. Prediction of speed loss of a ship in waves, in: Proceedings of Second International Symposium on Marine Propulsors smp'11 :, Institute for Fluid Dynamics and Ship Theory (FDS) - Hamburg University of Technology (TUHH), German Society for Maritime Technology (STG).

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. URL: //www.sciencedirect.com/science/article/pii/S0029801816305571, doi:http://dx.doi.org/10.1016/j.oceaneng.2016.11.058.

Dimopoulos, G.G., Georgopoulou, C.A., Stefanatos, I.C., Zymaris, A.S., Kakalis, N.M., 2014. A general-purpose process modelling framework for marine energy systems. Energy Conversion and Management 86, 325–339.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference and prediction. 2 ed., Springer. URL: http://www-stat.stanford.edu/~tibs/ElemStatLearn/.

Holtrop, J., Mennen, G., 1982. An approximate power prediction method, International Shipbuilding Progress.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. pp. 1137–1143. URL: http://dl.acm.org/citation.cfm?id=1643031.1643047.

Mao, W., Rychlik, I., Wallin, J., Storhaug, G., 2016. Statistical models for the speed prediction of a container ship. Ocean Engineering 126, 152 – 162. URL: http://www.sciencedirect.com/science/article/pii/S0029801816303699, doi:http://dx.doi.org/10.1016/j.oceaneng.2016.08.033.

Øyan, E., 2012. Speed and powering prediction for ships based on model testing.

Ozdemir, Y.H., Barlas, B., 2016. Numerical study of ship motions and added resistance in regular incident waves of

{KVLCC2} model. International Journal of Naval Architecture and Ocean Engineering , –URL: http://www.sciencedirect.com/science/article/pii/S2092678216305076, doi:http://dx.doi.org/10.1016/j.ijnaoe.2016.09.001.

Peri, D., Rossetti, M., Campana, E.F., 2001. Design optimization of ship hulls via cfd techniques. Journal of Ship Research 45, 140–149. URL: http://www.ingentaconnect.com/content/sname/jsr/2001/00000045/00000002/art00006.

Petersen, J.P., Jacobsen, D.J., Winther, O., 2012. Statistical modelling for ship propulsion efficiency. Journal of Marine Science and Technology 17, 30–39. URL: http://dx.doi.org/10.1007/s00773-011-0151-0, doi:10.1007/s00773-011-0151-0.

Rakke, S.G., 2016. Ship emissions calculation from ais.

Sadat-Hosseini, H., Wu, P.C., Carrica, P.M., Kim, H., Toda, Y., Stern, F., 2013. {CFD} verification and validation of added resistance and motions of {KVLCC2} with fixed and free surge in short and long head waves. Ocean Engineering 59, 240 – 273. URL: http://www.sciencedirect.com/science/article/pii/S0029801812004258, doi:http://dx.doi.org/10.1016/j.oceaneng.2012.12.016.

Savitsky, D., 1964. Hydrodynamic design og planning hulls, Marine Technology.

Tillig, F., Ringsberg, J., Mao, W., Ramne, B., 2016. A generic energy systems model for efficient ship design and operation. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment URL: http://dx.doi.org/10.1177/1475090216680672, arXiv:http://dx.doi.org/10.1177/1475090216680672.

Trodden, D., Murphy, A., Pazouki, K., Sargeant, J., 2015. Fuel usage data analysis for efficient shipping operations. Ocean Engineering 110, Part B, 75 – 84. URL: //www.sciencedirect.com/science/article/pii/S0029801815005004, doi:http://dx.doi.org/10.1016/j.oceaneng.2015.09.028. energy Efficient Ship Design and Operations.

Tupper, E. (Ed.), 2004. Introduction to Naval Architecture. Fourth edition ed., Butterworth-Heinemann, Oxford. URL: //www.sciencedirect.com/science/article/pii/B9780750665544500008, doi:http://dx.doi.org/10.1016/B978-075066554-4/50000-8.

Vanem, E., Brandsæter, A., Gramstad, O., 2017. Regression models for the effect of environmental conditions on the efficiency of ship machinery systems, in: Bedford, M.R..T. (Ed.), Risk, Reliability and Safety : Innovating Theory and Practice : Proceedings of ESREL 2016 (Glasgow, Scotland, 25-29 September 2016), Lesley Walls.

Wood, S., 2017. Mixed gam computation vehicle with gcv/aic/reml smoothness estimation, february 8, 2017. https://cran.r-project.org/web/packages/mgcv/mgcv.pdf. Accessed: 2017-02-22.

Wood, S.N., 2006. Generalized additive models: an introduction with R. Texts in Statistical Science, Chapman Hall/CRC. URL: http://opus.bath.ac.uk/7011/.