



# Estimating the effect of biofouling on ship shaft power based on sensor measurements

Haakon Bakka, Hanne Rognebakke, Ingrid Glad, Ingrid Hobæk Haff & Erik Vanem

To cite this article: Haakon Bakka, Hanne Rognebakke, Ingrid Glad, Ingrid Hobæk Haff & Erik Vanem (2022): Estimating the effect of biofouling on ship shaft power based on sensor measurements, Ship Technology Research, DOI: [10.1080/09377255.2022.2159108](https://doi.org/10.1080/09377255.2022.2159108)

To link to this article: <https://doi.org/10.1080/09377255.2022.2159108>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 24 Dec 2022.



Submit your article to this journal [↗](#)



Article views: 166



View related articles [↗](#)



View Crossmark data [↗](#)

## Estimating the effect of biofouling on ship shaft power based on sensor measurements

Haakon Bakka <sup>a</sup>, Hanne Rognebakke<sup>b</sup>, Ingrid Glad<sup>a</sup>, Ingrid Hobæk Haff<sup>a</sup> and Erik Vanem<sup>a,c</sup>

<sup>a</sup>Department of Mathematics, University of Oslo, Oslo, Norway; <sup>b</sup>Norwegian Computing Centre, Oslo, Norway; <sup>c</sup>DNV, Oslo, Norway

### ABSTRACT

Marine biofouling on a ship's hull and propeller increases the resistance of the ship moving through water and reduces the propulsion efficiency of the ship. Estimating the effect of fouling is difficult, as the biomass is rarely measured. In this paper, we present a new data-driven model for the total shaft power use of a large containership, in order to estimate the unobserved effect of fouling. Due to the limitations of both physical models and machine learning models, we develop a Bayesian generalized additive model for our purpose. We discuss issues of representative training data for the model. Further, we subset and subsample the data to a representative sample. Models are compared by out-of-sample predictive quality, physical appropriateness, and through autocorrelation of residuals. The Bayesian generalized additive model combined with computational inference using integrated nested Laplace approximations gives a robust estimate of the biofouling effect over time. It also allows a decomposition of the total shaft power use into effects of speed, weather, and other conditions. This model can be used to understand the effectiveness and timing of different hull and propeller treatments.

### ARTICLE HISTORY

Received 7 January 2022  
Accepted 6 December 2022

### KEYWORDS

Biofouling; fouling; Bayesian; GAM; sensor; machine learning; statistical modelling

## 1. Introduction

Marine biofouling (fouling) on a ship's hull and propeller increases the roughness of the surfaces in contact with water, and thus the frictional resistance of the ship (Schultz et al. 2011). This significantly increases the resistance of the ship moving through water and seriously influences the propulsion efficiency and fuel use of the ship. The efficiency lost due to biofouling can be 5–15% of the total shaft power and fuel usage (Molland et al. 2014; Wang and Lutsey 2013; Hakim et al. 2017). Coraddu et al. (2019) state that this contributes to approximately 2.7% of the global anthropogenic carbon dioxide (CO<sub>2</sub>) emissions (Smith et al. 2015).

Fouling is removed by proper hull and propeller cleaning in order to maintain efficiency, but such cleaning is costly and time consuming. Moreover, cleaning may damage the anti-fouling system and hence accelerate further marine growth. Therefore, hull and propeller cleaning should only be performed when necessary and the optimal time to clean the hull is a trade-off between the cost of cleaning and the efficiency gain achieved. In order to make qualified decisions regarding cleaning, one needs models to estimate the effect of biofouling on the ship resistance and propulsion efficiency, and to separate the effect of fouling from other factors such as ship speed, loading condition and weather conditions. Such models can also be used to analyse the

effectiveness of different anti-fouling systems, including the decision of which paint to use.

There has recently been a regulatory push for emission reduction and more environmentally friendly shipping. There are global energy-efficiency requirements from the International Maritime Organization (IMO) as well as global and regional caps on air pollution from ships (MARPOL annex VI (IMO 2020)). In 2018, IMO adopted a strategy for the reduction of GHG emissions from ships, aiming at reducing total GHG emissions from international shipping by at least 50% by 2050 (IMO 2018). Hence efficiency is high on the agenda in the maritime industries, and optimized interventions to reduce the effect of biofouling could contribute to reduce emissions and meet IMO goals. A number of technical measures to reduce fuel consumption in shipping, both in design and operation, are discussed in Sharifi et al. (2017). Furthermore, the International Convention on the Control of Harmful Anti-Fouling Systems on Ships (IMO 2005), adopted in 2001 and entered into force in 2008, prohibits the use of some anti-fouling systems that have previously been effectively used to reduce biofouling. Hence, stricter requirements on anti-fouling systems contribute to the challenge of optimizing interventions to reduce biofouling.

Due to the importance of fuel efficiency and emission reduction in shipping, a number of numerical

models have been developed to assess ship performance. Such models include various components related to, e.g. calm water resistance, air resistance, added resistance in waves and wind, propulsion and propeller efficiency, see e.g. Tvette et al. (2020). The effect of biofouling would be an important part of such models, but, as pointed out in Tvette et al. (2020), this effect is still not fully understood and there is a lack of reliable models, see also Yeginbayeva and Atlar (2018). Hence, this paper proposes a data-driven model to account for the effect of biofouling, and this can be incorporated as an important sub-model to improve the accuracy of ship performance models.

For studying propulsion efficiency and fuel consumption, models for relevant physical relations are often proposed. A ship's speed-power curve is typically determined by sea trials and model tests prior to commissioning. Sea trials are ideally carried out under steady conditions, with deep waters, no wind, no current and no waves and with specified draught and trim conditions. Different methods can be used to correct the effect of waves and wind during such trials (Tsujiimoto et al. 2021). Power curves are important in the calculation of the attained energy-efficiency design index (EEDI) (Tu et al. 2018) and in the overall assessment of a ship performance and fuel consumption (Jalkanen et al. 2009; Tillig et al. 2018; Bialystocki and Konovessis 2016).

Data-driven models for fuel consumption have become more and more popular due to their ability to take into account many phenomena by using large amounts of data together with limited assumptions based on physical knowledge about the problem. They can also take into account ship-specific and environmental phenomena better than purely physical models. A review of various data-driven methods can be found in Gkerekos et al. (2019), who attempt to find efficient ways to implement the fuel consumption model. This includes Linear Regression (both with and without regularization), Decision Tree Regressors, Random Forest Regressors, Extra Trees Regressors, Support Vector Regressors, K-Nearest Neighbours, Artificial Neural Networks and ensemble method algorithms. One of the conclusions of Gkerekos et al. (2019) was that the quality of the model output correlates with the quality of its training input. This is an important finding that applies to all data-driven models. A critical assumption is that the conditions under which the model is trained are representative for the time period where the model is used to perform predictions. Various data-driven regression models were explored by Brandsæter and Vanem (2018) to account for the effect of environmental conditions on ship speed. Newer studies that use various data-driven methods can be found in Laurie et al. (2021). They use linear regression, decision trees (AdaBoost), K-nearest neighbours, artificial neural network and random forest to predict shaft power to identify

performance deterioration due to fouling. Kim et al. (2021) use linear regression and an artificial neural network to predict fuel consumption. The importance of data preprocessing and filtering when using data-driven models to predict ship propulsion is highlighted in Karagiannidis and Themelis (2021).

The aim of this paper is to analyse, based on sensor data collected from ships in operation, the effect of biofouling on the performance of the ship, and the additional propulsion power that is needed due to this effect. Flexible Bayesian models based on the sensor data will be established, avoiding the pitfalls of the machine learning methods mentioned above. As we will use some physical knowledge when we formulate the model, we first outline the necessary background related to the resistance of a ship in water. Thereafter, we comment on our experiences with the machine learning approaches mentioned above and advocate the use of more interpretable models.

In Section 2, we present the data, while the Bayesian generalized additive model is introduced in Section 3, the results in Section 4, and finally we conclude the paper in Section 5.

### 1.1. Ship resistance

The resistance of a ship moving through water is typically described by the relationship between speed and power demand, and is expressed by a speed-power curve. Essentially, a ship's resistance is a sum of the resistance from several sources, such as water and air resistance. It is influenced by the ship's speed, displacement and hull form. In particular, the frictional resistance from water depends on the ship's wetted area, wave conditions, speed and fouling on the hull.

A simple model for the relationship between ship speed, displacement and propulsion power demand is expressed through the Admiralty coefficient (Brown and Aldridge 2019). The Admiralty coefficient,  $A$ , is a constant for a given ship and is defined as follows

$$A = \frac{\nabla^{\frac{2}{3}} v^3}{P}, \quad (1)$$

where  $\nabla$  denotes the displacement,  $v$  is the ship speed and  $P$  is the propulsion power need. It is generally acknowledged that this is an approximate model. Further, it assumes a cubic relationship between power demand and speed, given displacement. Hence, the Admiralty law translates to the following approximate relationship between power need and ship speed

$$P \approx \frac{\nabla^{\frac{2}{3}} v^3}{A} = C_A v^3, \quad (2)$$

where  $C_A$  is a constant (depending on displacement). Another simplified model for the relationship between

power need and speed is the propeller law (Brown and Aldridge 2019), which is based on the power required to drive a propeller with speed  $v_p$

$$P = cv_p^n, \quad (3)$$

where  $c$  and  $n$  are coefficients defining the relationship. It is seen that for  $n = 3$  and  $c = C_A$ , and by replacing the engine power and propeller speed with the ship power need and ship speed, the two laws are in general agreement. Resistance-based models take other factors into account and will typically be on the form (Brown and Aldridge 2019)

$$P = \frac{\rho C_T S v^3}{2\eta_T}, \quad (4)$$

where  $C_T$  is the total resistance coefficient,  $\rho$  is the sea water density,  $S$  is the hull-wetted surface area and  $\eta_T$  denotes the engine efficiency. A model for the instantaneous power used by a ship as a function of velocity and technical parameters of the ship is expressed as follows in Jalkanen et al. (2009),

$$P_{\text{transient}} = (CF + CR + CA + CAA) \left( \frac{1}{2} v^3 S \right) \frac{1}{\varepsilon_0}, \quad (5)$$

where  $CF$ ,  $CR$ ,  $CA$  and  $CAA$  are the frictional, residual, appendage and air resistances, respectively,  $S$  is the wet surface area and  $\varepsilon_0$  is the propulsive coefficient.

In summary, different models based on physical considerations suggest using a cubic relationship between propulsion power and speed, but these models are all acknowledged to be approximations, see e.g. Tu et al. (2018). The statistical modelling presented in this paper will use an adaptive spline for the speed-power relationship, to alleviate potential inaccuracies in the cubic model, especially for slow or fast speeds. However, for the effect of biofouling, we will assume a proportionality with speed to the power three.

### 1.2. Machine learning by training on clean hull data and estimating the fouling effect from residuals

One intuitive approach to the problem of estimating the effect of fouling is to use a two-step procedure as in Coraddu et al. (2019) in order to obtain predictions of fouling. First, we use a Random Forest (RF) (Breiman 2001) to build a model for shaft power based on the available covariates. This RF is trained on clean hull data and used as a reference model. Secondly, we use the estimated RF model to predict shaft power in the entire period and compare this prediction with observed shaft power. The difference between the predicted and observed shaft power is then assumed to represent the biofouling effect. The critical assumptions here are that the estimated model is fitted to data where marine fouling is essentially absent and

that the combination of covariate values in the training period is representative for the entire time period. In the training period with clean hull data, we need to observe the ship in many operational and environmental conditions, and these conditions must be representative for the entire time period. It is not clear how to define the training period in order to balance the need of a clean hull and a sufficient number of representative samples, and for some cases, there exists no good training dataset. In our case, the prediction model varied a lot depending on how we defined the training period. If we use the first year of data to train the model, we find that the ship speed was overall much lower in the training interval than in the succeeding period (see Figure 1 and some additional details in Appendix). Hence, the training period was not long enough to contain representative values for the ship speed. At the same time, the training period was already too large, as our main results show that the hull was far from clean at the end of the period.

The problems we encountered when fitting RF models show that when using black box machine learning models, we depend heavily on having representative training data. It is also difficult, if not impossible, to correct for this in retrospect. In fact, the results of the RF model were so unreliable and unstable that they were not even comparable to the results of the main model in this paper. We believe this was not a problem with RF, as both Gkerekos et al. (2019) and Laurie et al. (2021) had good results with RF, but a problem with the approach of training on clean hull data. The main model we propose will be based on a completely different approach; on fitting an interpretable model to the entire dataset.

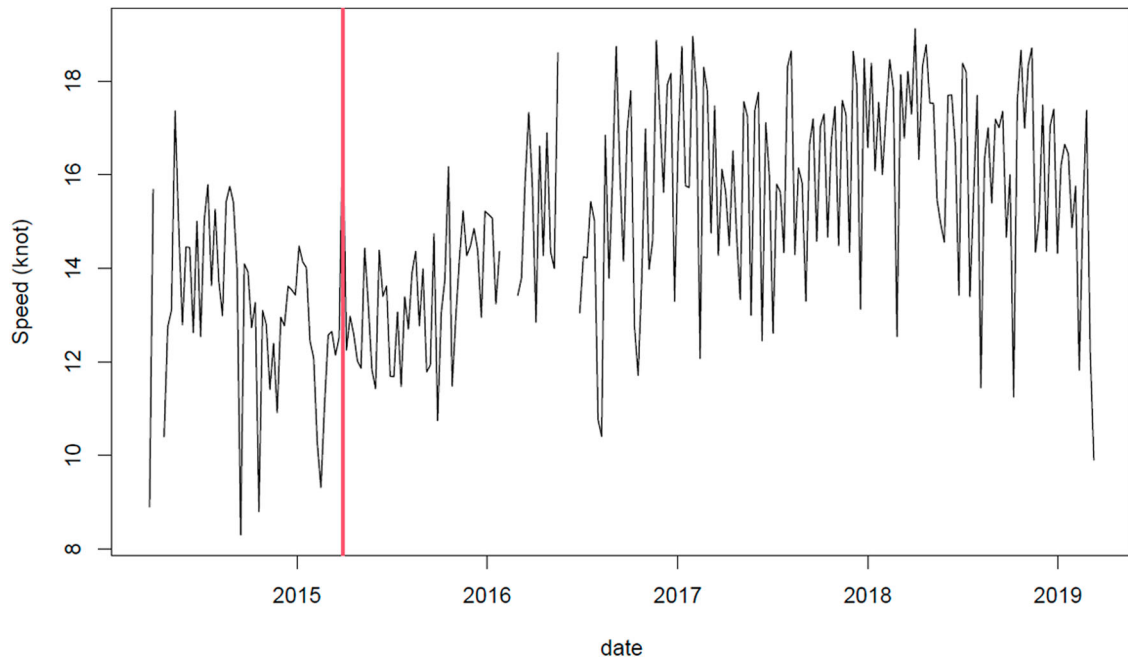
### 1.3. Interpretable, flexible additive models

Instead of dividing the dataset into training and prediction sets, we chose to fit an interpretable model to the entire dataset, to infer the unobserved effect of fouling. Since we do not have a measured covariate for the fouling level, this is not a typical regression problem. Instead, we assume a flexible parametric form for the effect of fouling, and use this as a time-varying model component in a generalized additive model (GAM) (Hastie and Tibshirani 1986).

Our proposed model is inherently interpretable, which is a great advantage. As Rudin (2019) states,

but trying to explain black box models, rather than creating models that are interpretable in the first place, is likely to perpetuate bad practices and can potentially cause great harm to society. The way forward is to design models that are inherently interpretable.

In the problem of estimating fouling, this is especially relevant. In this application, there is no



**Figure 1.** Weekly mean of ship speed for the ship data.

clear prediction target where we can choose whichever model is the best at attaining good predictions. Rather, we aim for a model that gives a good approximation of the true data generating mechanism, and then we use the fouling estimate from this model. In order to investigate whether our fouling estimate is good, we must be able to interpret and criticize the model. The main assumption in a GAM is that the model components  $f_j$  add together to give the predictor

$$\eta = f_1 + f_2 + f_3 + \dots + f_j,$$

which is then used to fit the observations  $y$  through a likelihood, or loss function,  $L(y|\eta)$ . This additive structure enables interpretable model building,

**Table 1.** The response and the covariates.

Code name	Text name	Details
Shaft_power_MW	Shaft Power	The response
Speed_knot	Ship Speed	Speed through water
Accel1_knp15m	Linear Acceleration	Units: Knots per 15 minutes
Accel2_knp15m	Turning Acceleration	
Draft_mean_m	Draft Mean	Average of draft forwards and aft
Draft_trim_m	Draft Trim	Half difference between forward and aft
Wave_dir	Wave Direction	
Wind_speed_knot	Wind Speed	Relative to ship
Tz_s	Wave Period	
Hs_m	Wave Height	
Clean.times	Cleaning Events	Hull or propeller cleaning or painting
		Not a covariate, but a list of timestamps
Cleaning_period	Cleaning Period	E.g. 1: Before the first Cleaning Event
Time_since_clean_days	Time_since_clean_days	

where relevant covariates and physical understanding is used to build the model components. The main purpose of using GAM for modelling fouling is to adjust for all covariates unrelated to fouling, and simultaneously model the residual effect of fouling.

## 2. Data

For this study, a 2500 TEU containership of about 26,000 gross tonnage with a length of approximately 200 m and breadth approximately 30 m has been analysed. We utilize data from several sources. The primary data source for ship performance is various sensors onboard the ship in operation. These data are supplemented by combining Automatic Identification System (AIS) data for the location of the vessel with weather data from the ERA 5 reanalysis (Hersbach et al. 2020) to obtain the prevailing weather and sea state conditions at all times. The covariates we used from the ERA 5 reanalysis were Wave Direction, Wind Speed, Wave Period and Wave Height. In addition, six dates at which cleaning was performed were obtained from the ship owner, where the last date was the end of our dataset. Unfortunately, we were not able to obtain precise information about the type of each cleaning, e.g. whether it was hull or propeller cleaning.

Based on these data sources, a large number of covariates were available. From these, we selected a set of covariates based on expert judgement and preliminary analysis. The covariates sea water temperature, air temperature, dew point temperature, and fraction of sea ice, were available from the ERA 5 reanalysis but excluded based on lack of significance in preliminary models. See Table 1 for the response and all the

included covariates, together with a brief explanation. Note that the last part of the code name gives the measurement unit. The six cleaning events were converted to the covariate `Cleaning_period`, an integer representing the six time periods before the cleaning events. We note that wind direction was not included directly, but used to compute Wind Speed relative to ship direction.

### 2.1. Data preprocessing and cleaning

We have approximately 5 years of data. The sampling frequency varied between the different variables, from 1 minute up to 15 minutes. In addition, the sampling frequency for the performance monitoring data varied throughout the period, but it was typically 15 minutes. Hence, we aggregated all variables into 15-min intervals using the mean value in the interval. We used the circular mean value for the variables that were measured in degrees.

In this paper, we study fouling by modeling shaft power as a function of ship speed, weather and other conditions. The fouling effect is a slowly varying component that is best explained in steady-state conditions, like when the ship is in transit. The behaviour of shaft power is more complex when we have low ship speed or high accelerations, which is typically when the ship is manoeuvring close to shore. Hence, to reduce noise in the data, we disregard all time points where the ship speed was lower than 5 knots, and similarly, we disregard all time points where the linear acceleration was greater than 1 knot per 15 minutes and the turning acceleration greater than 2.5 knots per 15 minutes.

## 3. Methods

### 3.1. Model definitions

We denote the models we propose as BGAM, for Bayesian generalized additive models. The main model we present, BGAM-main, is

$$y = \eta + \epsilon \quad (6)$$

$$\begin{aligned} \eta = X\beta + f_{\text{rw2}}(\text{Speed}) + f_2(\text{Wave\_dir}) \\ + \text{Speed}^3 h(t_*), \end{aligned} \quad (7)$$

where  $y$  is the observed Shaft Power in MW,  $\epsilon$  is the residuals,  $\eta$  is the predictor, and the other terms are referred to as model components. We use linear effects ( $X\beta$ ) for the variables Linear Acceleration, Turning Acceleration, Draft Trim, Draft Mean, Wind Speed, Wave Height and Wave Period. The  $\beta$ 's have approximately flat priors.

The effect  $f_{\text{rw2}}(\text{Speed})$  is for the Ship Speed, and is a second order spline, called a random walk order 2 spline (see e.g. Gómez-Rubio (2020) Chapter 9). We

chose this flexible spline, instead of e.g. a  $\text{Speed}^3$  curve, to account for any model misspecification at low or high speeds. The scaling parameter  $\sigma_{\text{rw2}}$  of this spline is given an exponential prior, based on the penalizing complexity (PC) framework (Simpson et al. 2017). The effect  $f_2$  is for Wave Direction and is a two-parameter trigonometric function,  $\alpha_1 \sin(x) + \alpha_2 \cos(x)$ . We could not use a linear effect for Wave Direction because the minimal angle and the maximum angle should have approximately the same effect, as they represent the same direction. The two parameters have approximately flat priors. The residuals are modelled as iid. Gaussian variables,

$$\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2). \quad (8)$$

The biofouling effect  $h(t_*)$  is a piecewise linear function in  $t_*$ , which is *time at rest since the last cleaning event*,

$$h(t_*) = a_k + b_k t_*. \quad (9)$$

The  $a_k$  and  $b_k$  are given approximately uniform priors. The fouling effect is multiplied with  $\text{Speed}^3$  following the theory in Section 1.1. The piecewise linear structure is due to our expectation of a jump in the fouling effect after a cleaning event, and then a gradual increase until the next cleaning event.

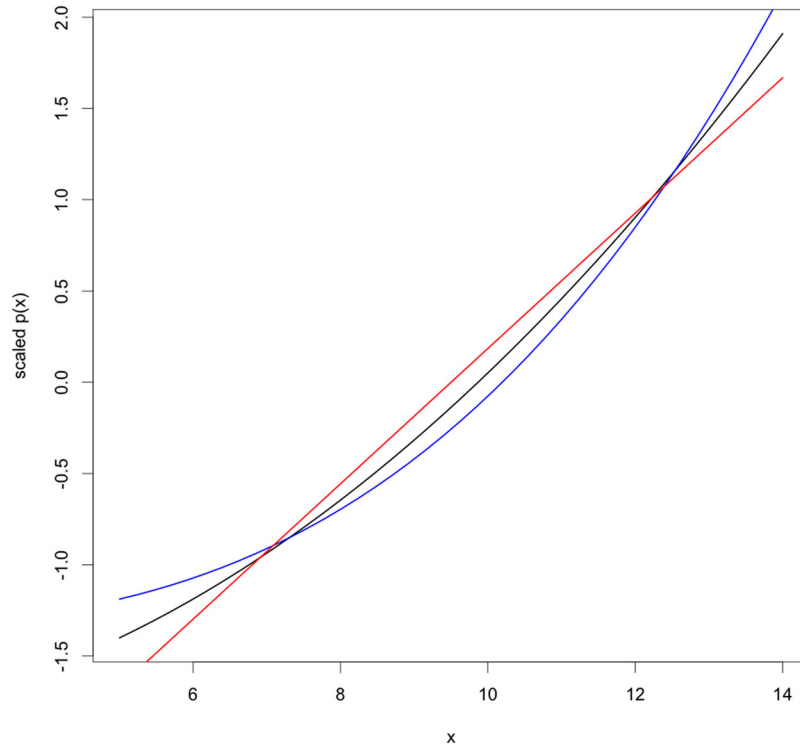
In addition to the main model, we present a few secondary models for comparison. The model BGAM-T is the same as BGAM-main, except that the time  $t_*$  used for the fouling effect, is clock time since the last cleaning (not just when the ship was standing still). The models BGAM-V1 and BGAM-V2 are the same as BGAM-main, except for the exponent of the Speed in  $\text{Speed}^3 h(t_*)$ , which is changed to  $\text{Speed}^1$  and  $\text{Speed}^2$ , respectively. From the physical principles in Section 1.1, we find it natural to use  $\text{Speed}^3$ , but we wanted to investigate alternatives. However, as we see in Figure 2, the effect of using a different exponent should not be large for this dataset.

### 3.2. Bayesian inference

We follow a Bayesian approach to inference for the models we propose. We denote the prior model  $\pi(y, \eta)$ , giving a joint distribution for the predictor  $\eta$  and the response  $y$ . This prior model is built as  $\pi(y|\eta)\pi(\eta)$ , where the first term is referred to as the (observation) likelihood, and the second as the prior. The prior of  $\eta$  is defined through the prior for all the components of  $\eta$  as described above. The posterior is

$$\pi(\eta|y) = \frac{\pi(y, \eta)}{\pi(y)}.$$

After training the model on a set of data, the model can predict new observations by integrating over the



**Figure 2.** Polynomials  $x$ ,  $x^2$ ,  $x^3$  evaluated at ship speeds in our dataset and rescaled to be similar. The red line is  $x$ , the black line is  $x^2$ , and the blue line is  $x^3$ .

parameters inferred by the old data

$$\pi(y_{\text{new}}|y_{\text{old}}) = \int_{\eta} \pi(y_{\text{new}}|\eta)\pi(\eta|y_{\text{old}})d\eta.$$

To interpret the model, we compute the posterior estimate and uncertainty for the individual model components that make up  $\eta$ .

We present one main model and several other candidate models. It is impossible to find the ‘one true model’, as no model is a perfect representation of reality. What we want is a good representation of reality that is useful for answering questions and provides reasonable uncertainty estimates. We approach this in three ways. The first is to build the model structure in line with a physical understanding of the process. The second is to study the quality of out-of-sample predictions, in Section 4.1. The third is to study whether the statistical assumptions are appropriate, e.g. whether the residuals are independent.

For computational inference, we use the R package INLA (Rue et al. 2009, 2017), short for Integrated Nested Laplace Approximations. The INLA algorithm is based on quadratic approximations of the likelihood function and sparse matrix representations of Gaussian Markov random fields (GMRFs) (Rue and Held 2005). These models can also be extended to spatial models within the same framework (Bakka et al. 2018; Krainski et al. 2018).

### 3.3. Subsampling to correct bias and variance

When we use a Gaussian likelihood,

$$\begin{aligned} \pi(y|\eta) &= \pi(\epsilon) = \prod_i^n \mathcal{N}(0, \sigma_\epsilon^2)(\epsilon_i), \\ &= -\frac{n}{2} \log(2\pi) + \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_i^n \epsilon_i^2\right) \end{aligned}$$

or, equivalently, a quadratic loss function, where the loss  $L$  is the negative log-likelihood

$$L(y) = \frac{1}{2\sigma_\epsilon^2} \sum_i^n (y_i - \eta_i)^2 = \frac{1}{2\sigma_\epsilon^2} \sum_i^n \epsilon_i^2,$$

we assume that the  $\epsilon_i$ s are independent of each other. Initial model exploration (with BGAM and RF) showed that  $\epsilon$  has a strong autocorrelation in time. Informally, this autocorrelation hints that we are reusing the data, as adjacent observations contain overlapping information about the system we are modelling. As an extreme example, consider measuring the ship variables for only a single day, but every millisecond, producing a large dataset. Such a dataset would not be informative of fouling as the data would be mostly redundant. Ignoring this problem could cause the models to have a local bias, where different observation frequencies, e.g. due to missing data, imply different levels of informativeness on different days. Worse, however, is that all uncertainty estimates would be rendered useless. Inference on repeated data gives results which are overconfident about their own certainty. Fitting models to the full dataset

gave very small uncertainty estimates which were clearly wildly inaccurate.

To reduce or eliminate this correlation, we subsampled the data. First, denote by subset 1 the full, processed data set. Then subset 2 is this data subsampled to every hour, subset 3 to every 4 hours, and subset 4 to every 24 hours (daily). Later we present results for the different levels of subsampling. Even with subset 4, there were missing data.

### 3.4. Size decomposition of additive models

Consider the model in Equation (7) as a sum of vectors, we can then take the length, or norm, of each of the vectors, and consider whether they sum up

$$\begin{aligned} \|y\|^2 = & \|X\beta\|^2 + \|f_{\text{rw2}}(\text{Speed})\|^2 \\ & + \|f_2(\text{Wave.dir})\|^2 + \|\text{Speed}^3 h(t_*)\|^2 \\ & + \|\epsilon\|^2. \end{aligned} \quad (10)$$

If we take  $\| - \|$  to be the Euclidean norm, after centering the vectors, i.e.  $\|y\| = \sqrt{\sum_i (y_i - \bar{y})^2}$ , and assume that all the vectors are parallel, i.e. correlation 1, then the equation holds. This  $\| - \|$  is the empirical standard deviation (StDev), and leads us to decompose the standard deviation of  $y$  in terms of standard deviations of the model components. Because the correlation is not 1, the sum of the standard deviations will not equal the standard deviation of  $y$ . However, presenting the numbers gives us an understanding of the relative sizes. Alternatively, if we take  $\| - \|$  to be the Euclidean norm squared, i.e.  $\|y\|^2 = \sum_i (y_i - \bar{y})^2$ , and assume that all the vectors are orthogonal, i.e. uncorrelated, then this equation is the Pythagorean theorem. This  $\| - \|$  is the empirical variance (Var), and leads us to decompose the variance of  $y$  in terms of variances of the model components. Because the variables are not independent, the sum of the variances will not equal the variance of  $y$ . However, presenting the numbers again gives us an understanding of the relative sizes.

**Table 2.** Out-of-sample model comparison, mean across training-test folds. CVC is defined in Section 3.5. The '90%' refers to the coverage of 90% prediction intervals.

Subset	Model	CVC RMSE	CVC MAE	CVC 90%	Cor <sub>e</sub>
3	BGAM-main	1.16 (1.3)	0.86 (0.98)	0.830 (0.80)	0.52
1	BGAM-main	1.12 (1.4)	0.83 (0.90)	0.820 (0.79)	0.77
2	BGAM-main	1.13 (1.4)	0.84 (0.90)	0.826 (0.80)	0.68
4	BGAM-main	1.24 (1.4)	0.96 (1.1)	0.847 (0.79)	0.22
3	BGAM-V1	1.13 (1.3)	0.86 (0.99)	0.842 (0.81)	0.53
3	BGAM-V2	1.16 (1.3)	0.87 (0.99)	0.832 (0.79)	0.52
3	BGAM-T	1.17 (1.4)	0.87 (1.1)	0.818 (0.73)	0.52
**	BGAM-main	1.09 (1.3)	0.80 (0.94)	0.839 (0.77)	0.47

In parenthesis, we show the highest value across the training-test folds for RMSE and MAE, and the lowest for coverage. The \*\* subset puts stricter limits on covariate values for both training and test, and then subsets to one observation every 4 h, hence the numbers cannot be fairly compared to the above.

### 3.5. Cross validation for out of sample performance

We use out of sample predictions to study the performance of the models. We name the scheme CVC, for cross validation of cleaning intervals. Here, for each cleaning interval, we fit the model to all the data except for the last 1/3 of observations in that cleaning interval. Because we have 6 cleaning intervals, we get 6 training-test combination sets.

To summarize the predictive quality we use the mean squared error (MSE)

$$\sqrt{\frac{1}{n_k} \sum_i^{n_k} (y_{i,\text{predicted}} - y_{i,\text{true}})^2},$$

and mean absolute error MAE

$$\frac{1}{n_k} \sum_i^{n_k} |y_{i,\text{predicted}} - y_{i,\text{true}}|,$$

where we for each of the  $k$  test sets sum over  $i \in [1, 2, \dots, n_k]$ ,  $n_k$  being the number of observations in that test set. To summarize the calibration of prediction uncertainty we present the coverage of 90% prediction intervals.

## 4. Results and discussion

In this section, we compare the different candidate models, we select the main model, and then we detail the results of the main model.

### 4.1. Model comparison and selection

We give an overview of all the models' out-of-sample predictive quality in Table 2, with the assessment described in Section 3.5. The first row of the table contains the model we will give detailed results on. We also report the measure of the single worst fold in parenthesis. Because we are mainly interested in robust estimation of the fouling effect, the worst performance across the folds is as interesting as the mean performance. If this were a prediction problem, and we knew that our predictive scheme and our predictive RMSE summary matched exactly the goal of the model, we would have picked the model with the lowest RMSE. However, in this application, we want to estimate the hidden fouling effect, thus we want a model with good overall behaviour.

In Table 2, the four first lines are with the same model, but with different subsets (see Section 3.3). The less subsetting, the higher the mean RMSE and MAE for prediction accuracy. This is expected due to using more data for prediction. However, the worst-case RMSE is the best for BGAM-main on subset 3. For the smaller subsets, both the 90% coverage is better (larger), and the autocorrelation of the residuals



is better (smaller). We chose subset 3 as a good trade-off between predictive accuracy, reasonable coverage and low residual autocorrelation. However, we find subset 4 to also be a reasonable choice, due to the residual autocorrelation being substantially lower.

Lines 5 and 6 in Table 2 represent the models with different exponents for the Speed multiplier in the fouling component. We observe that using a smaller exponent (1 or 2) gives better predictive performance and coverage. However, the improvement is minimal, so due to the physical arguments in Section 1.1, we did not select one of these models.

BGAM-T is clearly worse than the main model, which is the reason why we used  $t^*$  (time at rest) in this paper. Our experimentation showed that  $t^*$  was superior to using clock time across a wide range of models.

The last row of Table 2 uses a different approach to subsetting, where more of the data was removed for being ‘extreme’ before the data was subsampled. Here, Ship Speed > 7 knots, Linear Acceleration < 1, Turning Acceleration < 2.5, Scaled Wind Speed < 1.9 and Scaled Wave Height < 1.9. A scaled variable has the mean removed and is divided by the standard deviation. After this, the observations were subsetted to every 4 hours (similarly to subset 3). We note that the performance is improved across all criteria. However, the results from this row are not directly comparable with the other rows, due to the test set being substantially different; informally, this is an easier prediction task.

#### 4.2. Results from the selected Bayesian generalized additive model

We present the results from the main model (BGAM-main) fitted to subset 3. We present the posterior estimates of the various model components and the variance decomposition in Table 3. The variance decomposition is used to understand how much each model component contributes to the model, and to compare effect sizes, as introduced in Section 3.4. The row named Sum is the sum of the above variances (similar for Stdev). If the model components were independent, the sum of variances should equal the variance of the predictor, which they do not, hence they are not independent. If the model components were completely dependent, the sum of Stdev should equal the Stdev of the predictor, which they do not, hence they are not completely dependent. The predictor variance and the residual variance do sum to the total Shaft Power variance, meaning that the predictor and the residuals are (approximately) independent.

Care must be taken when interpreting the size of the model components according to the variance decomposition. As an example, take the residuals.

With a variance contribution of 4.1%, one could intuitively say that the model explains 95.9% of the data variability and that 4.1% is not explained by the model. However, by looking at the Stdev decomposition we see that the variability of the residuals is roughly 1/5 of the variability of the Shaft Power, which does not agree with only 4.1% residual variation. When comparing sizes, it is important to have the right measurement unit, which in this case is MW, and not  $MW^2$  as it would be for the variance.

From the variance decomposition we conclude that the Speed covariate is by far the most important predictor of Shaft Power, which comes as no surprise. The only other component more important than Fouling is the Wind Speed. Fouling is more important than any other covariate, even more important than the Draft. Using the Stdev decomposition, we see that the Fouling effect represents roughly 15% of the total variation of the raw data, measured in MW.

For the linear effects, we see that they are all statistically significant at the level of 95% credible intervals. From the posterior estimates of the linear effects, we see that increased acceleration, Trim, Wind Speed and Wave Height all lead to increased shaft power use, which is sensible. A larger Draft Trim leads to slightly less shaft power use. This is a questionable relationship, and cannot be true for extreme values of the Draft Trim variable. The Draft Trim effect could be related to other operational conditions that are strongly correlated with Draft Trim. However, the effect is small, approximately 1% of the data variability. A larger wave period results in slightly less shaft power use, but again the effect is very small.

Figure 3 shows the estimated effect of ship speed. This effect is adjusted for weather, draft, fouling, and other effects. In reality, this should be a strictly increasing function, with an increasing derivative, which is close to true for this graph. However, there are some discrepancies between the theory and this graph, especially near the endpoints. This can be due to unmeasured effects or biases at low and high speeds, and the estimate is not trustworthy at these extreme values. In the statistical model, an unmeasured effect at low speed can be accounted for by this model component. This is in contrast to models with strong physical constraints, where such unmeasured effects are carried into the estimation of fouling instead. From the figure, we see that the uncertainty is large at the endpoints. This implies that there is not enough data in these regions to support a good estimate. However, it also implies that the estimate in these regions will not greatly influence the model.

Figure 4 shows the estimated fouling effect. This effect is more complicated to interpret than other effects as it depends on both speed and fouling. To make interpretation easier, we infer the fouling effect at 12 knots and show this in Figure 5. This figure

**Table 3.** Table of variance (and Stdev) decomposition, in percentage of Shaft Power variance (or Stdev), for BGAM-main on subset 3, together with standard deviations measured in MW, as detailed in Section 3.4.

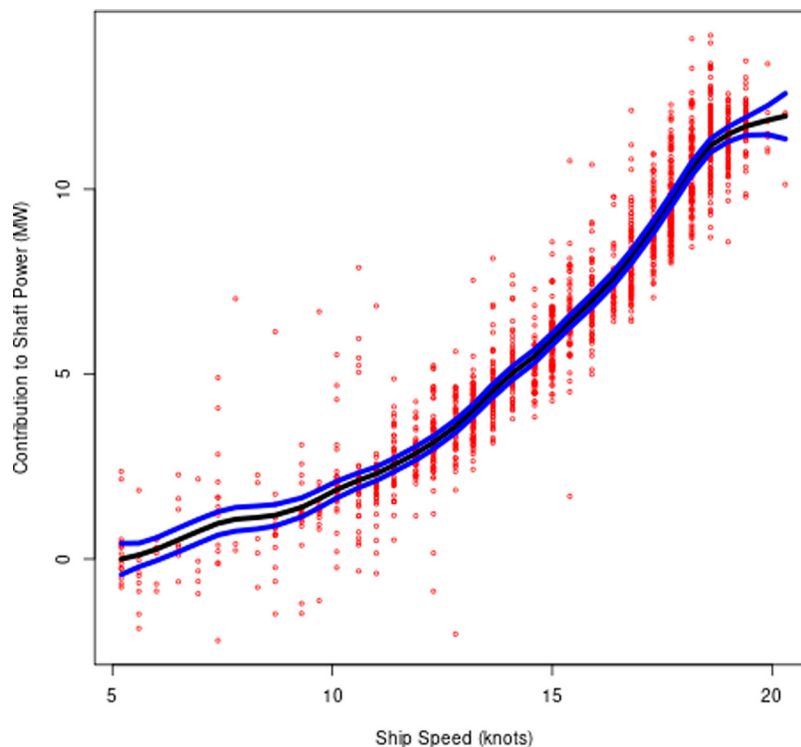
Model component	Var	Stdev	Description
Speed_knot	56.9%	75.4% (3.35 MW)	Spline, increasing super-linear
Fouling	2.2%	14.9% (0.66 MW)	PWS
Accel1_knp15m	<0.1%	1.2% (0.05 MW)	Linear, $\beta = 0.235$ (0.144, 0.327)
Accel2_knp15m	<0.1%	0.8% (0.03 MW)	Linear, $\beta = 0.059$ (0.020, 0.099)
Draft_trim_m	<0.1%	1.1% (0.05 MW)	Linear, $\beta = -0.090$ (-0.150, -0.030)
Draft_mean_m	0.8%	8.8% (0.39 MW)	Linear, $\beta = 0.282$ (0.257, 0.306)
Wind_speed_knot	2.2%	14.8% (0.66 MW)	Linear, $\beta = 0.054$ (0.052, 0.056)
Tz_s	<0.1%	1.6% (0.07 MW)	Linear, $\beta = -0.045$ (-0.066, -0.023)
Hs_m	0.5%	7.3% (0.32 MW)	Linear, $\beta = 0.382$ (0.341, 0.424)
Wave_dir	<0.1%	1.2% (0.05 MW)	Sinusoid, max at 184 degrees
Sum	62.7%	127.0% (5.70 MW)	Sum of above percentages
Predictor	95.9%	97.9% (4.35 MW)	Model structure
Residuals	4.1%	20.2% (0.90 MW)	Model noise
Shaft_power_MW	100%	100% (4.44 MW)	Observations

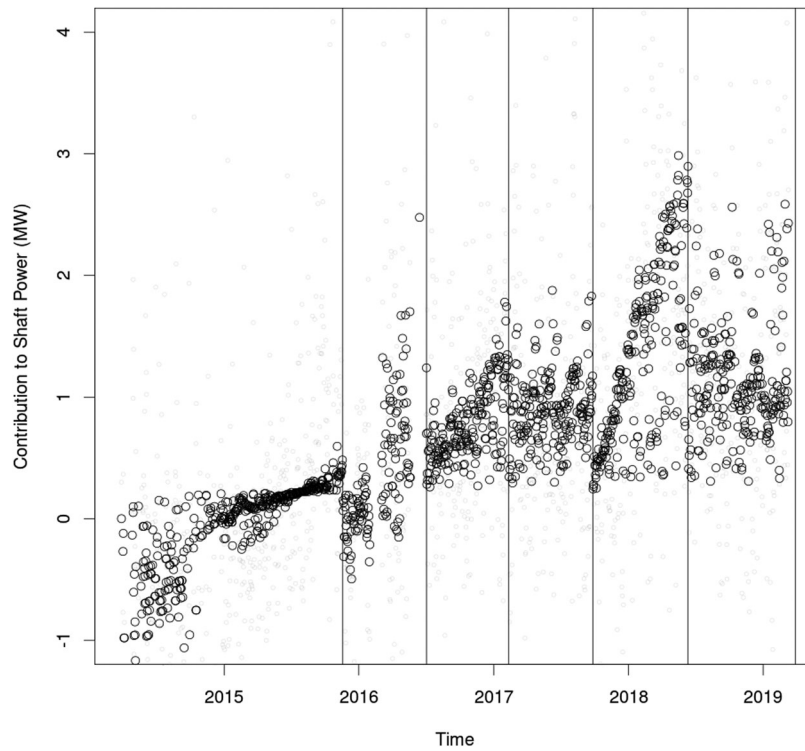
The last column gives parameter estimates and 95% credible intervals for linear effects, otherwise a description.

shows what the fouling effect would have been, according to our model, had the ship been going at 12 knots. The estimates are linear in  $t_*$ , which is the time the ship was at rest. The slopes of the fouling trends are different in the different time intervals. This could be due to operational conditions causing either faster biological growth or a larger impact of the fouling under some operational conditions. The general trend is that the biofouling effect worsens over time, is reduced by cleaning, but never resets to the low effect of a new ship. The most important source of efficiency loss is the gradual loss, i.e. that the ship never returns to its initial biofouling effect. This gradual worsening might be partly due to other factors than biofouling, if there are other age-related inefficiencies, e.g. small damages to the hull, or inefficiencies in the propeller, these will be included in our

global biofouling effect. If a cleaning event completely restores the hull to the conditions of a new ship, the biofouling effect in our plots still may not revert to zero after the cleaning, due to these other inefficiencies. The red points in Figure 5 shows the residuals added to the estimated effect. These illustrate the variability from which the estimates and uncertainty are computed. Note that the blue lines are not prediction intervals, and so the red points are not supposed to be comprised by the blue lines.

Figure 6 illustrates how the fouling effect varies with different ship speeds. Each piecewise function represents one speed scenario, the lowest speed gives the smallest fouling effect, and the highest speed gives the largest fouling effect. This model implies that the fouling effect becomes extremely problematic at very high speeds.

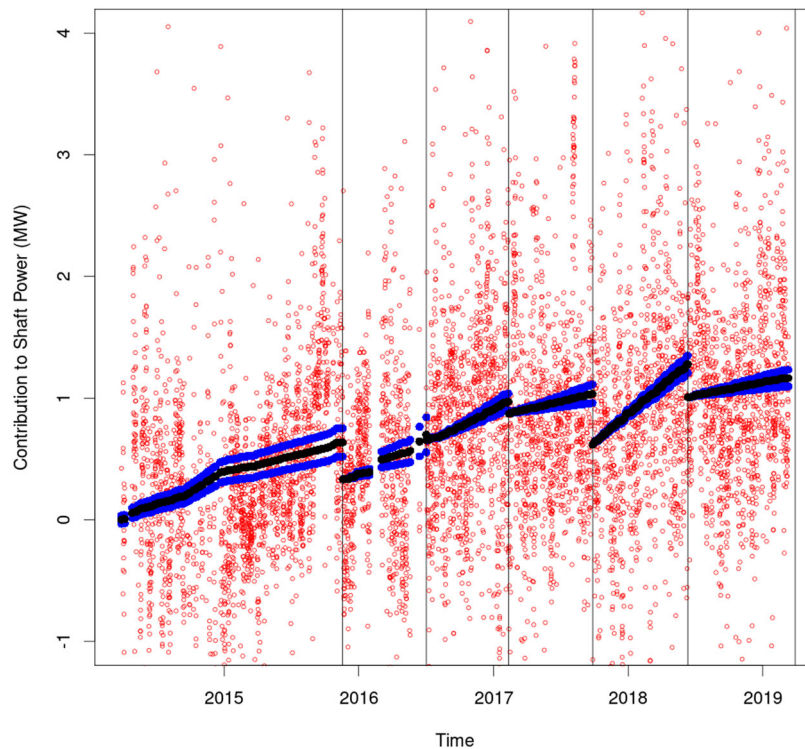
**Figure 3.** Estimated Shaft Power as a function of ship speed, with residuals added in red, from the model BGAM-main on subset 4.



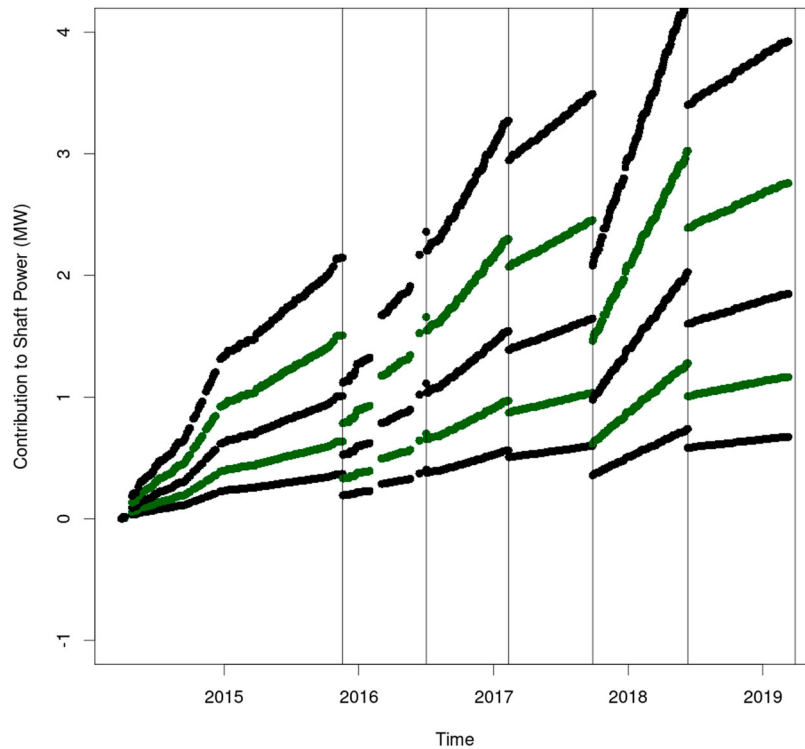
**Figure 4.** Estimated fouling effect in black, from the model BGAM-main on the subset 4 data. Uncertainty is not shown. The estimated fouling effect includes the term  $\text{Speed}^3$ ; hence, it varies quickly in time because the ship speed varies quickly in time. The vertical lines show the six cleaning events.

We considered presenting the model applied to subset 4, instead of subset 3. Therefore, we show the inferred fouling effect in that model in Figure 7. The general trends of the fouling effect are very similar to those of subset 3, but the uncertainty is significantly

larger. Additionally, our results for the fouling effect from the other models in Table 2 were qualitatively similar to the effects presented in Figures 5 and 7, hence the estimation of the fouling effect is very robust.



**Figure 5.** Estimated fouling in black at a constant 12 knots speed, from the model BGAM-main on the subset 3 data. The back lines are linear in  $t_*$ , but not linear in time. Uncertainty is shown as blue lines. The red points show the residuals  $\epsilon_i$  added to the estimates (black lines). The vertical lines show the six cleaning events.

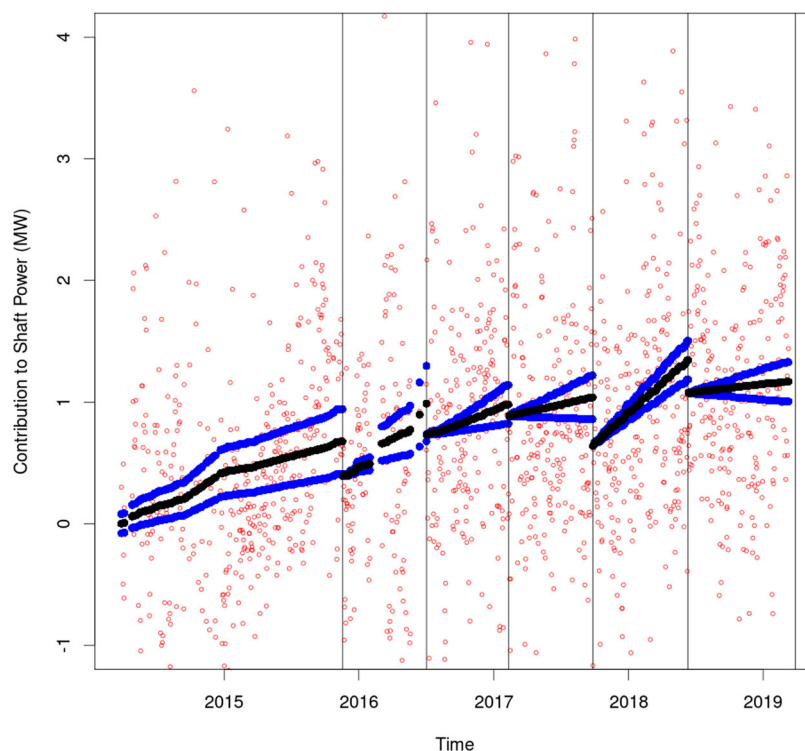


**Figure 6.** Estimated fouling at constant speeds, assuming the same conditions as in the dataset. The five curves are with speed 10 knots (lowest black), 12 knots (lowest green), 14 knots (middle black), 16 knots (highest green), and 18 knots (highest black). The results are from the model BGAM-main on the subset 3 data.

## 5. Conclusion

In this paper, we presented a Bayesian Generalised Additive model (BGAM) for modelling shaftpower as a sum of model components and for obtaining a robust estimate of the fouling effect.

The BGAM presented here is a good tool for decomposing shaft power into model components in order to describe how shaft power increases or decreases as a result of the speed, operational conditions and weather conditions. We have shown how



**Figure 7.** Estimated fouling in black at a constant 12 knots speed, from the model BGAM-main on the subset 4 data. Uncertainty is shown as blue lines, and residuals as red points.

to deal with too few or too many samples on any given day by fitting the model jointly to a subsampled dataset with missing values.

The BGAM is a good tool for estimating the fouling effect. We get robust and statistically significant estimates of the fouling effect, together with reasonable uncertainty bands. There is an obvious positive short-term effect of every cleaning, and the cleaning is more impactful when the fouling is larger (see Figure 5). After each cleaning, we get a visualization of how fast the fouling effect returns, and we observe a gradual worsening throughout the lifetime of the ship.

The BGAM can be a useful tool for decision makers as an input to cost/benefit analysis of cleaning. For any ship where similar data are collected, a figure similar to our Figure 5 can be produced and used to argue for less or more cleanings. Such an analysis can also be used to retrospectively study the effect of different types of cleaning on the overall fouling trend. The practice of reducing speed when there is a lot of fouling can be quantified for improved decision making on the choice of speed. More efficient strategies for dealing with fouling will give both economic and environmental benefits. These models can also increase the understanding of the environmental impacts of fouling through increased energy demand.

There is more work that can be done to improve the BGAM. We have presented a model of limited complexity on data which we believe are attainable for a large range of ships. In situations with less noisy or larger datasets, one should consider more covariates, more complex interactions between covariates, and also, other functional forms for  $h(t)$ . For example, sea and air temperature may be included. Additionally, all the covariates measured here have some level of measurement uncertainty. If one can quantify it, it is possible to extend the model to incorporate measurement uncertainty, through the use of multiple likelihoods. We can gain a deeper understanding of the fouling and biological growth by measuring the amount/type of biological material before and after each cleaning. Then we could combine the results from BGAMs on several ships to understand how fast the biomass grows on the hull and propeller, and again combine this with AIS and weather data to understand under which conditions the biomass grows faster and slower.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This research is funded by the Norwegian Research Council research-based innovation centre BigInsight, project no. 237718. It has also been carried out in cooperation with

the VERDE project, funded by the Norwegian Research Council under project no. 282293.

### ORCID

Haakon Bakka  <http://orcid.org/0000-0001-8272-865X>

### Data availability statement

Data have been provided by Klaveness, and cannot be accessed without their explicit permission, contact the authors for assistance.

### References

- Bakka H, Rue H, Fuglstad G-A, Riebler A, Bolin D, Illian J, Krainski E, Simpson D, Lindgren F. 2018. Spatial modeling with R-INLA: a review. *WIREs Comput Stat.* 10(6): e1443.
- Bialystocki N, Konovessis D. 2016. On the estimation of ship's fuel consumption and speed curve: a statistical approach. *J Ocean Eng Sci.* 1(2):157–166.
- Brandsæter A, Vanem E. 2018. Ship speed prediction based on full scale sensor measurements of shaft thrust and environmental conditions. *Ocean Eng.* 162:316–330.
- Breiman L. 2001. Random forests. *Mach Learn.* 45(1):5–32.
- Brown IN, Aldridge MF. 2019. Power models and average ship parameter effects on marine emissions inventories. *J Air Waste Manag Assoc.* 69(6):752–763.
- Coraddu A, Oneto L, Baldi F, Cipollini F, Atlar M, Savio S. 2019. Data-driven ship digital twin for estimating the speed loss caused by the marine fouling. *Ocean Eng.* 186:106063.
- Gkerekos C, Lazakis I, Theotokatos G. 2019. Machine learning models for predicting ship main engine fuel oil consumption: a comparative study. *Ocean Eng.* 188:106282.
- Gómez-Rubio V (2020). *Bayesian inference with INLA.* Boca Raton (FL): CRC Press.
- Hakim M, Utama I, Nugroho B, Yusim A, Baithal M, Suastika I. 2017. Review of correlation between marine fouling and fuel consumption on a ship. In: *Proceeding of SENTA: 2nd Conference on Marine Technolog, Surabaya, Indonesiay.* p. 122–129.
- Hastie T, Tibshirani R. 1986. Generalized additive models. *Stat Sci.* 1(3):297–310.
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, et al. 2020. The ERA5 global reanalysis. *Q J R Meteorol Soc.* 146(730):1999–2049.
- IMO. 2005. International convention on the control of harmful anti-fouling systems on ships. London: International Maritime Organization.
- IMO. 2018. Adoption of the initial imo strategy on reduction of GHG emissions from ships and existing imo activity related to reducing GHG emissions in the shipping sector. [accessed 2022 Dec 21]. [https://unfccc.int/sites/default/files/resource/250\\_IMO%20submission\\_Talanoa%20Dialogue\\_April%202018.pdf](https://unfccc.int/sites/default/files/resource/250_IMO%20submission_Talanoa%20Dialogue_April%202018.pdf).
- IMO. 2020. International convention for the prevention of pollution from ships (MARPOL). London: International Maritime Organization.
- Jalkanen J-P, Brink A, Kalli J, Pettersson H, Kukkonen J, Stipa T. 2009. A modelling system for the exhaust emissions of marine traffic and its application in the baltic sea area. *Atmos Chem Phys.* 9(23):9209–9223.

- Karagiannidis P, Themelis N. 2021. Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss. *Ocean Eng.* 222:108616.
- Kim Y-R, Jung M, Park J-B. 2021. Development of a fuel consumption prediction model based on machine learning using ship in-service data. *J Mar Sci Eng.* 9(2):137.
- Krainski E, Gómez-Rubio V, Bakka H, Lenzi A, Castro-Camilo D, Simpson D, Lindgren F, Rue H (2018). Advanced spatial modeling with stochastic partial differential equations using R and INLA. Boca Raton (FL): Chapman and Hall/CRC.
- Laurie A, Anderlini E, Dietz J, Thomas G. 2021. Machine learning for shaft power prediction and analysis of fouling related performance deterioration. *Ocean Eng.* 234:108886.
- Molland A, Turnock S, Hudson D, Utama I. 2014. Reducing ship emissions: a review of potential practical improvements in the propulsive efficiency of future ships. *Trans R Inst Naval Arch Part A Int J Mar Eng.* 156:175–188.
- Rudin C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 1(5):206–215.
- Rue H, Held L (2005). Gaussian Markov random fields: theory and applications, volume 104 of monographs on statistics and applied probability. London: Chapman & Hall.
- Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J R Stat Soc Ser B (Stat Methodol).* 71(2):319–392.
- Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. 2017. Bayesian computing with INLA: a review. *Annu Rev Stat Appl.* 4:395–421.
- Schultz M, Bendick J, Holm E, Hertel W. 2011. Economic impact of biofouling on a naval surface ship. *Biofouling.* 27(1):87–98.
- Sharifi Y, Ghassemi H, Zanganeh H. 2017. Various innovative technologic devices in shipping energy saving and diminish fuel consumption. *Int J Phys.* 5(1):21–29.
- Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. 2017. Penalising model component complexity: a principled, practical approach to constructing priors. *Stat Sci.* 32(1):1–28.
- Smith TW, Jalkanen J, Anderson B, Corbett J, Faber J, Hanayama S, O'keeffe E, Parker S, Johansson L, Aldous L. 2015. Third IMO GHG study 2014. London: International Maritime Organization (IMO). [accessed 2022 Dec 21]. <https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Third%20Greenhouse%20Gas%20Study/GHG3%20Executive%20Summary%20and%20Report.pdf>.
- Tillig F, Ringsberg JW, Mao W, Ramne B. 2018. Analysis of uncertainties in the prediction of ships' fuel consumption—from early design to operation conditions. *Ships Offsh Struct.* 13(sup1):13–24.
- Tsujimoto M, Yasukawa H, Yamamoto K, Lee TI. 2021. Validation of added resistance in waves by tank tests and sea trial data. *Ship Technol Res.* 1–12. <https://doi.org/10.1080/09377255.2021.1925047>.
- Tu H, Yang Y, Zhang L, Xie D, Lyu X, Song L, Guan Y-m., Sun J. 2018. A modified admiralty coefficient for estimating power curves in EEDI calculations. *Ocean Eng.* 150:309–317.
- Tvete HA, Guo B, Liang Q, Brinks H. 2020. A modelling system for power consumption of marine traffic. In: Proceedings of 39th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2020); August 3–7; Virtual (Online). American Society of Mechanical Engineers (ASME).
- Wang H, Lutsey N. 2013. Long-term potential for increased shipping efficiency through the adoption of industry-leading practices. Washington (DC): The International Council on Clean Transportation. [accessed 2022 Dec 21]. [https://theicct.org/sites/default/files/publications/ICCT\\_ShipEfficiency\\_20130723.pdf](https://theicct.org/sites/default/files/publications/ICCT_ShipEfficiency_20130723.pdf).
- Yeginbayeva IA, Atlar M. 2018. An experimental investigation into the surface and hydrodynamic characteristics of marine coatings with mimicked hull roughness ranges. *Biofouling.* 34(9):1001–1019.

## Appendix. Random forest model

We tried different lengths of training periods for the model, from the first half year of data to the first whole year of data. We must assume that the fouling is small in this period. It is possible to include data from periods after the ship has been cleaned. However, from the results with the BGAM model, we see that after cleaning, the ship never returns close to its initial state. We have tested a number of combinations of variables to include in the model. The results for the different variable combinations were quite similar, but changing the training period slightly altered the results. The reason for this is probably that the conditions experienced by the ship have changed after the training period. Figure 1 shows weekly mean of Ship Speed, which is the most important variable in the model. We see that the speed is overall much lower in the training period than in the succeeding period. Hence, the final results for the estimated fouling effect are quite unstable and not reliable. This illustrates the importance of having representative training data when using machine learning methods on complicated datasets.