Ocean Engineering 130 (2017) 351–370

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Review

Vessels fuel consumption forecast and trim optimisation: A data analytics perspective

Andrea Coraddu[a,*], Luca Oneto[b], Francesco Baldi[c], Davide Anguita[d]

[a] DAMEN Shipyard Singapore, R&D Department, 29 Tuas Crescent, 638720, Singapore
[b] DIBRIS – University of Genova, Via Opera Pia 13, I-16145 Genova, Italy
[c] Department of Shipping and Marine Technology, Chalmers University of Technology, SE-412 96, Gothenburg, Sweden
[d] DIBRIS – University of Genova, Via Opera Pia 13, I-16145 Genova, Italy

ARTICLE INFO

Keywords:
Naval propulsion plant
Fuel consumption
Trim optimisation
Ship efficiency
Sensors data collection
Numerical models
Data analytics
White Box Models
Black Box Models
Gray Box Model

ABSTRACT

In this paper the authors investigate the problems of predicting the fuel consumption and of providing the best value for the trim of a vessel in real operations based on data measured by the onboard automation systems. Three different approaches for the prediction of the fuel consumption are compared: White, Black and Gray Box Models. White Box Models (WBM) are based on the knowledge of the physical underling processes. Black Box Models (BBMs) build upon statistical inference procedures based on the historical data collection. Finally, the authors propose two different Gray Box Model (GBM) which are able to exploit both mechanistic knowledge of the underlying physical principles and available measurements. Based on these predictive models of the fuel consumption a new strategy for the optimisation of the trim of a vessel is proposed. Results on real world operational data show that the BBM is able to remarkably improve a state-of-the-art WBM, while the GBM is able to encapsulate the a-priory knowledge of the WBM into the BBM so to achieve the same performance of the latter but requiring less historical data. Moreover, results show that the GBM can be used as an effective tool for optimising the trim of a vessel in real operational conditions.

1. Introduction

Shipping is a relatively efficient mean of transport when compared to other transport modes. Despite its efficiency, however, shipping contributes significantly to air pollution, mainly in the form of sulphur oxides, nitrogen oxides, particulate matter, and carbon dioxide. For the latter, the contribution from shipping to global emissions is required to decrease significantly in the coming years. Since greenhouse gas emissions from the combustion of oil-based fuels are directly proportional to fuel consumption, improving ship energy efficiency is one of the possible solutions to this issue. Measures for the improvement of ship energy efficiency are normally divided into design and operational measures. While the former have been associated to larger saving potential, the latter can still provide a significant reduction in fuel consumption, while requiring a much more limited capital investment. However, the large amount of variables influencing ship energy efficiency makes it hard to assess ship performance in relation to a standard baseline. Operational measures include, among others, improvement in voyage execution, reduction of auxiliary power consumption, weather routing, optimised hull and propeller polishing schedule, slow steaming, and trim optimisation.

Among the above mentioned fuel saving measures, trim optimisation has been extensively discussed in the past. It is well known, from hydrodynamics principles, that the trim of the vessel can significantly influence its fuel consumption (Altosole et al., 2016). In most cases principles of trim optimisation are applied at design stage for an a priori prediction and optimisation of operational behaviour by means of full scale RANS simulations. The matrices of conditions is usually designed with a combinations of draft, trim and speed. However, weather conditions can influence the optimal value of the trim. Taking these aspects into account when selecting the appropriate trim can therefore lead to significant, cost-free savings in terms of fuel required for ship propulsion. Previous work in scientific literature related to trim optimisation have focused on three main alternative strategies: White-Box numerical Models (WBMs), and black-box numerical models (BBMs). WBMs describe the behaviour of the ship resistance, propeller characteristics and engine performances based on governing physical laws and taking into account their mutual interactions (Martelli et al., 2014; Coraddu et al., 2011). The higher the detail in the modelling of the physical equations which describe the different phenomena, the higher the expected accuracy of the results and the computational time required for the simulation. WBMs are generally rather tolerant to

* Corresponding author.
E-mail addresses: andrea.coraddu@damen.com (A. Coraddu), luca.oneto@unige.it (L. Oneto), francesco.baldi@chalmers.se (F. Baldi), davide.anguita@unige.it (D. Anguita).

http://dx.doi.org/10.1016/j.oceaneng.2016.11.058
Received 22 September 2015; Received in revised form 12 October 2016; Accepted 28 November 2016
Available online 13 December 2016
0029-8018/

extrapolation and do not require extensive amount of operational measurements; on the other hand, when employing models that are computationally fast enough to be used for online optimisation, the expected accuracy in the prediction of operational variables is relatively low. In addition, the construction of the model is a process that requires competence in the field, and availability of technical details which are often not easy to get access to. Examples of the use of WBMs for the optimisation of ship trim are (Lee et al., 2014), who employed advanced CFD methods, and Moustafa et al. (2015) who employed simpler empirical models (Holtrop-Mannen) for the estimation of possible gains from trim optimisation. Differently from WBMs, BBMs (also known as data driven models (Vapnik, 1998b)), make use of statistical inference procedures based on historical data collection. These methods do not require any a-priory knowledge of the physical system and allow exploiting even measurements whose role might be important for the calculation of the predicted variables but might not be captured by simple physical models. On the other hand, the model resulting from a black-box approach is not supported by any physical interpretation (Shawe-Taylor and Cristianini, 2004), and a significant amount of data (both in terms of number of different measured variables and of length of the time series) are required for building reliable models (Petersen et al., 2012b; Coraddu et al., 2014). As an example, in Petersen et al. (2012a) an application of BBMs is proposed (in particular of artificial neural network) to the prediction of the fuel consumption of a ferry and applied to the problem of trim optimisation. Gray-box models (GBMs) have been proposed as a way to combine the advantage of WBMs and BBMs (Coraddu et al., 2015). According to the GBMs principles, an existing WBM is improved using data-driven techniques, either in order to calculate uncertain parameters or by adding a black-box component to the model output (Leifsson et al., 2008). GBMs allow exploiting both the mechanistic knowledge of the underlying physical principles and available measurements. The proposed models are more accurate than WBMs with similar computational time requirements, and require a smaller amount of historical data when compared to a pure BBMs.

The aim of this paper is to propose the application of a gray-box modelling approach to the prediction of ship fuel consumption which can be used as a tool for online trim optimisation. In this framework the authors exploit Machine Learning techniques based on kernel methods and ensemble techniques (Shawe-Taylor and Cristianini, 2004; Breiman, 2001) so to improve an effective but simplified physical model (Coraddu et al., 2013) of the propulsion plant. The proposed model is tested on real data (Baldi et al., 2014) collected from a vessel during two years of on board sensors data acquisitions (e.g. sheep speed, axis rotational speed, torque, wind intensity and direction, temperatures, pressure, etc).

The paper is organised as follows: in Section 2 a general description of physical the problem is reported. In Section 3 a description of the mathematical framework is presented. The proposed WBMs, BBMs and GBMs for the prediction of the fuel consumption are reported in Section 4 and tested on data collected from a vessel during two years of on board sensors data acquisitions. In Section 5 the physical plausibility of the different models is discussed and in Section 6 is shown how to use these models in order to optimise the trim of a vessel. In Section 7 the conclusions of the paper are drown.

2. Problem description

2.1. Ship description

In this paper, the authors propose the utilisation of a predictive model of the fuel consumption for the online optimisation of the trim of a vessel. To show the potential of the proposed method, it is here tested on a Handymax chemical / product tanker. A conceptual representation of the ship propulsion plant is shown in Fig. 1, while relevant ship features are presented in Table 1. The ship systems consists of two main engines (MaK 8M32C four-stroke Diesel engines) rated 3840 kW each and designed for operation at 600 rpm. The two engines are connected to one common gearbox; the gearbox has two outputs: a controllable pitch propeller designed for operations at 105 rpm for ship propulsion; and a shaft generator (rated 3200 kW) used for fulfilling on board auxiliary power demand. Auxiliary power can also be generated by two auxiliary engines rated 682 kW each. Auxiliary heat demand is fulfilled by a combination of exhaust gas boilers and auxiliary oil-fired boilers.

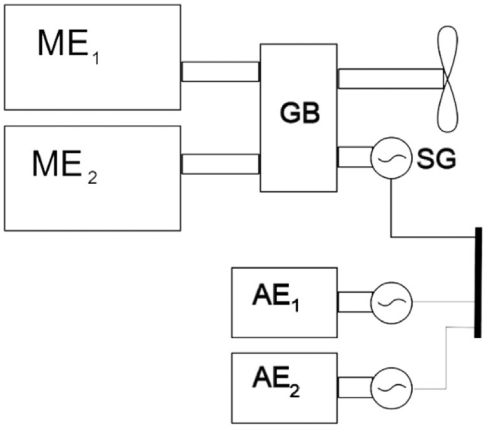

Fig. 1. Conceptual representation of the ship propulsion system.

Table 1
Main features of the case study ship.

Ship feature	Value	Unit
Deadweight	47,000	[t]
Installed power (Main Engines)	3840 (×2)	[kW]
Installed power (Auxiliary Engines)	682 (×2)	[kW]
Shaft generator design power	3200	[kW]
Exhaust boilers design steam gen.	1400	[kg/h]
Auxiliary boilers design steam gen.	28,000	[kg/h]

The ship is mainly used in the spot market (i.e. based on short-term planning of ship logistics, as opposed to long-term agreements with cargo owners on fixed schedules and routes) and therefore operates according to a variable schedule, both in terms of time spent at sea and of ports visited. The variety of different routes is shown in Fig. 2. Figs. 3 and 4represent the observed ship operations for the selected time period. It can be seen that although the ship spends a significant part of time in port, most of ship operations are related to open sea transport, either in laden or ballast mode (see Fig. 3). The focus of this work lies in the optimisation of ship trim; consequently, only data points related to sailing operations are considered in this study. Operations of manoeuvring, cargo loading, cargo unloading, and port stays were therefore excluded from the original dataset. These transport phases happen at a broad range of speeds, as shown in Fig. 4, which provides additional evidence of the need for an efficient tool for the optimisation of ship operations in different operational conditions. Because of their specific trading pattern, tankers are normally used in

Fig. 2. Description of the ships routes.

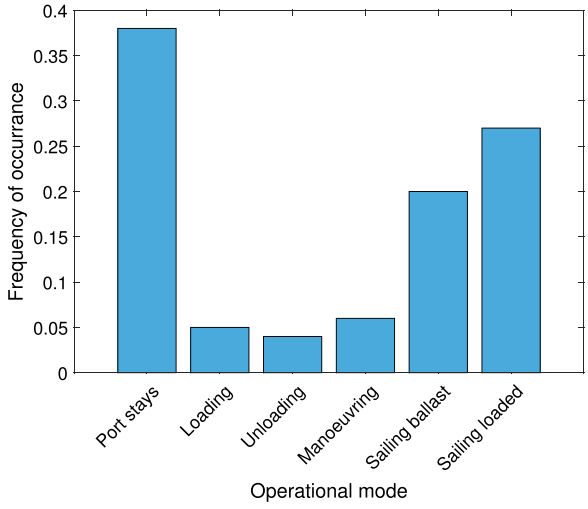

Fig. 3. Time spent in each operational mode for the selected vessel in the chosen period.

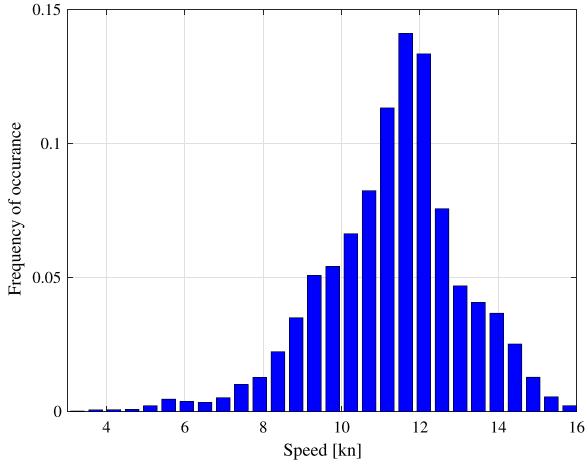

Fig. 4. Speed distribution during sailing time for the selected vessel in the chosen period.

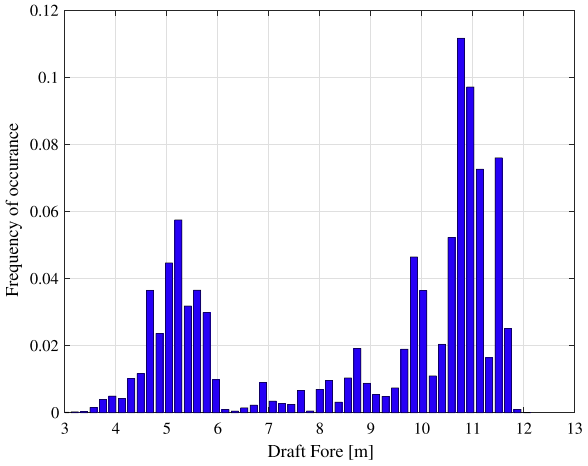

Fig. 5. Draft fore distribution during sailing time for the selected vessel in the chosen period.

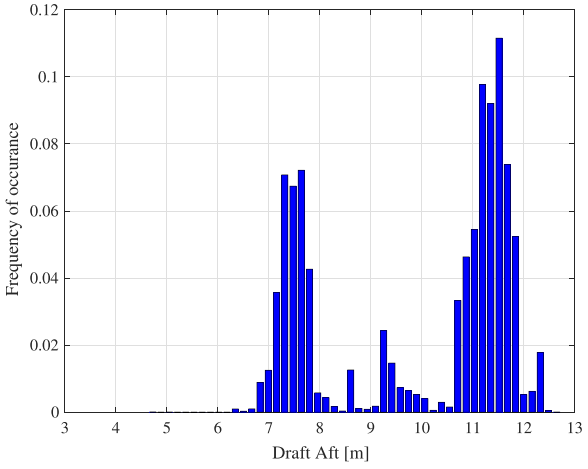

Fig. 6. Draft aft distribution during sailing time for the selected vessel in the chosen period.

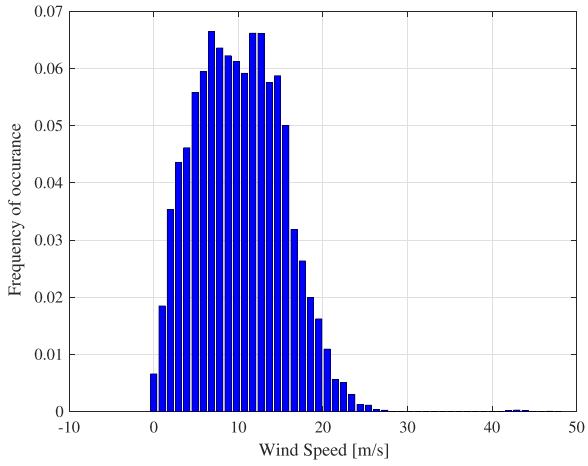

Fig. 7. Wind distribution during sailing time for the selected vessel in the chosen period.

two very distinct operational modes: laden (i.e. with full cargo holds, delivering liquid bulk cargo to the destination port), and ballast (with empty cargo holds, sailing to the port where the next cargo is available for loading). In reality, even when loaded, tankers vessels do not always sail with completely full holds tanks due to differences in order sizes. The ship's draught can consequently vary, depending on the operation, from 11 m when the ship is fully loaded to 6 m when cargo holds are completely empty. The distribution of ship fore and aft drafts during sailing time are presented in Figs. 5 and 6 respectively.

In addition to ship speed and draught, weather conditions are also known to have an influence on the optimal trim to be used when sailing, and can vary during ship operations. Fig. 7 represents wind speed which, in turn, is strongly correlated to the air resistance component of the total ship resistance.

Table 2
Measured values available from the continuous monitoring system.

Variable name	Unit
Time stamp	[YYYY-MM-DD, hh: mm:ss]
Latitude	[°, ′, ″]
Longitude	[°, ′, ″]
Fuel consumption (Main engines)	[kg/15 mins]
Auxiliary engines power output	[kW]
Shaft generator power	[kW]
Propeller shaft power	[kW]
Propeller speed	[kW]
Ship draft (forward)	[m]
Ship draft (aft)	[m]
Draft Port	[m]
Draft Starboard	[m]
Relative wind speed	[m/s]
Relative wind direction	[°]
GPS heading	[°]
Speed over ground (GPS)	[kn]
Speed through water (LOG)	[kn]
Sea depth	[m]
Sea Water Temperature	[°C]
CPP Setpoint	[%]
CPP Feedback	[%]
Fuel Density	[kg/m^3]
Fuel Temperature	[°C]
Ambient Pressure	[mbar]
Humidity	[%]
Dew Point Temperature	[°C]
Shaft Torque	[kN m]
Rudder Angle	[°]
Acceleration X Direction	[m/s^2]
Acceleration Y Direction	[m/s^2]
Acceleration Z Direction	[m/s^2]
GyroX	[rpm]
GyroY	[rpm]
GyroZ	[rpm]
Roll	[rad]
Pitch	[rad]
Yaw	[rad]

2.2. Data logging system

The ship under study is provided with a data logging system installed by an energy management provider which is used by the company both for on board monitoring and for land-based performance control. Table 2 summarises the available measurements from the continuous monitoring system.

The original data frequency measured by the monitoring system is 15 s. In order to provide easier data handling, the raw data are sent to the provider server, where they are processed into 15 min averages. The data processing is performed by the provider company and could not be influenced or modified by the authors.

Measured values come from on board sensors, whose accuracy and reliability cannot be ensured in the process. In particular, issues related to the measurement of speed through water (LOG speed) are well known and such measurements are often partly unreliable as a consequence of the fact that the flow through the measurement device can be easily disturbed by its interaction with the hull or by other environmental conditions. On the other hand measurements of speed over ground (GPS speed), although more reliable, do not include the influence of currents, which can be as strong as $2 \div 3$ knots depending on time and location and therefore influence ship power demand for propulsion. Fuel consumption is measured using a mass flow meter, which is known to be more accurate of the more common volume flow meters as it eliminates uncertainty on fuel density. It should be noted, however, that measurements of fuel specific energy content (LHV) were not available; variation of heavy fuel oil LHV is known to be in the order of ± 2 MJ/kg, which corresponds to a variation of $\pm 5\%$. Propeller speed, torque measurement and fuel mass flow accuracy were provided by the shipyard at respectively $\pm 0.1\%$, $\pm 1\%$ and $\pm 3\%$.

3. From inference to data analytics

Inference is the act or process of deriving logical conclusions from premises known or assumed to be true (MacKay, 2003). There are two main families of inference processes: deterministic, and statistical inference. The former studies the laws of valid inference, while the latter allows to draw conclusions in the presence of uncertainty, and therefore represents a generalisation of the former. Several different types of inference are commonly used when dealing with the conceptual representation of reality as shown in Fig. 8:

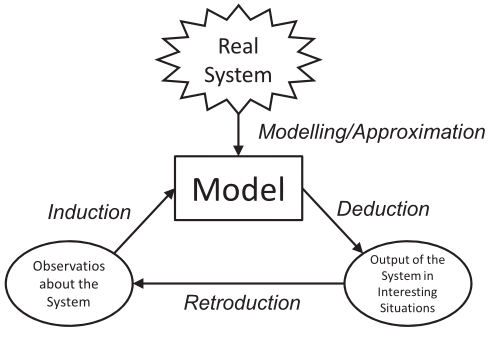

Fig. 8. Type of inference exploited in this paper.

- Modelling/approximation refers to the process of building a model of a real system based on the knowledge of the underlying laws of physics that are known to govern the behaviour of the system. Depending on the expected use and needs of the model, as well a on the available information, different levels of approximation can be used. Modelling/approximation of a real system only based on mechanistic knowledge can be categorised as deterministic inference (Howison, 2005; Lewis, 1988; MacKay, 2003);
- When the model is built by statistically elaborating observations of system inputs and outputs, the process belongs to the category of statistical induction. As the model is inferred based on measurements affected by different types of noise, this process is intrinsically under the effect of uncertainty and therefore belongs to the category of statistical inference (Vapnik, 1998a);
- The process of using an existing model to make predictions about the output of the system given a certain input is called deductive inference. This process can be both deterministic or probabilistic depending on how the model is formulated (Vapnik, 1998a);
- The process of actively modifying model inputs in order to obtain a desired output is normally referred to as retroduction (or abduction) (Josephson and Josephson, 1996).

The subject of this paper can hence be seen as the application of a general category of problems to a specific case. The physical laws governing ship propulsion are known and widely used in the dedicated literature with the purpose of modelling the ship behaviour. Moreover a series of historical data about the ship's propulsion system are available, and based on this it is possible to build a statistical model of the process (Coraddu et al., 2015, 2014, 2013) which again can be exploited to predict the behaviour of the system. In particular data analytics tools allow performing different levels of statistical modelling (Evans and Lindner, 2012):

- descriptive analytics tools allow understanding what happened to the system (e.g. what was the temperature of the cylinders of the engine in the last days). Descriptive analytics answers to the question 'What happened?'
- diagnostic analytics tools allow understanding why something happened to the system (e.g. the fuel consumption it too high and this is due to a the decay of the hull). Diagnostic analytics answers to the question 'Why did it happen?'
- predictive analytics tools allow making predictions about the system (e.g. when a new propeller is installed to reduce fuel consumption). Predictive analytics answers to the question 'What will happen?'
- prescriptive analytics tools allow understanding why the system behave in a particular way and how to force the system to be in a particular state (e.g. what is the best possible way to steer the ship in order to save fuel). Prescriptive analytics answers to the question 'How can we make it happen?'

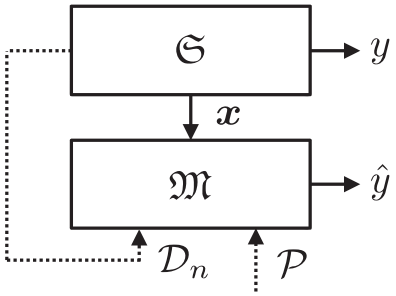

Fig. 9. Our regression problem.

Descriptive analytics is something very simple to implement, for example in Section 2 authors showed some compressed information coming from the historical data collection which can be interpreted as a descriptive analytic process. These tools are the least interesting ones since there is no additional knowledge extracted from the data (Jiawei and Kamber,, Evans and Lindner, 2012). Diagnostic analytics is a step forward where the authors try to understand what happened in the past, searching correlation in the data in order to get additional information from the data itself. Examples of these approach in the context of naval transportation system can be found in Coraddu et al. (2014); Widodo and Yang (2007); Jardine et al. (2006); Palmé et al. (2011). Finally, predictive and prescriptive analytics are the most complex approaches where a model of the system is built and studied in order to understand the accuracy and the properties of the model and make the system behave in a particular way. This is the most important analysis in practical applications since, even if diagnostic analytics allows improving the understanding of past and present conditions of the system, it is more important to predict the future and take action in order to prevent the occurrence of some event (Coraddu et al., 2014; Fumeo et al., 2015) (substitute a component before it fails) or to make some event happen (Petersen et al., 2012a) (reduce the fuel consumption of a ship).

For these reasons in the next sections a more rigorous framework is depicted together with the description of the approaches adopted for building predictive models. An assessment of their accuracy and properties is performed and a complete description about how to use these models to force the system in producing an output is provided.

3.1. Supervised learning

Using the conventional regression framework (Vapnik, 1998b; Shawe-Taylor and Cristianini, 2004) a set of data $\mathcal{D}_n = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$, with $\boldsymbol{x}_i \in \mathcal{X} \subseteq \mathbb{R}^d$ and $y_i \in \mathcal{Y} \subseteq \mathbb{R}$, are available from the automation system. Each tuple $(\boldsymbol{x}_i, y_i)$ is called sample and each element of the vector $\boldsymbol{x} \in \mathcal{X}$ is called feature.

When inferring a model starting from a real system, the goal is to provide an approximation $\mathfrak{M}: \mathcal{X} \rightarrow \mathcal{Y}$ of the unknown true model $\mathfrak{S}: \mathcal{X} \rightarrow \mathcal{Y}$. $\mathfrak{S}$ and $\mathfrak{M}$ are graphically represented in Fig. 9. It should be noted that the unknown model $\mathfrak{S}$ can be also seen, from a probabilistic point of view, as a conditional probability $\mathbb{P}(y|\boldsymbol{x})$ or, in other words, as the probability of the output y given the fact that we observed $\boldsymbol{x}$ as an input to $\mathfrak{S}$.

As previously described, in this paper three alternative modelling strategies are compared: white-, black-, and gray-box models:

- White Box Model (WBM): in this case the model $\mathfrak{M}_{\text{WBM}}$ is built based on a priori, mechanistic knowledge of $\mathfrak{S}$ (numerical description of the body hull, propulsion plant configuration, design information of the ship). The implementation of a WBM in this specific case is described in Section 4.1.
- Black Box Model (BBM): in this case the model $\mathfrak{M}_{\text{BBM}}$ is built based on a series of historical observation of $\mathfrak{S}$ (or in other words $\mathcal{D}_n$). In this paper, this is done by exploiting state of the art Machine Learning techniques as described in Section 4.2.
- Gray Box Model (GBM): in this case the WBM and BBM are combined in order to build a model $\mathfrak{M}_{\text{GBM}}$ that takes into account both a priori information and historical data $\mathcal{D}_n$ so to improve the performances of both the WBM and BBM models. The implementation of the GBM principle to the specific case of this work is described in Section 4.3.

3.2. Estimation of model accuracy

The accuracy of the model $\mathfrak{M}$ as a representation of the unknown system $\mathfrak{S}$ can be evaluated using different measures of accuracy (Elattar et al., 2010; Ghelardoni et al., 2013). In particular, given a

series data $\mathcal{T}_m = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_m, y_m)\}$,[1] the model will predict a series of outputs $\{\hat{y}_1, \ldots, \hat{y}_m\}$ given the inputs $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$. Based on these outputs it is possible to compute these performance indicators:

- mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{1}$$

- mean absolute percentage error (MAPE)

$$\mathrm{MAPE} = 100\frac{1}{m}\sum_{i=1}^{m}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

- mean square error (MSE)

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{3}$$

- normalised mean square error (NMSE)

$$\mathrm{NMSE} = \frac{1}{m\Delta}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2\Delta = \frac{1}{m}\sum_{i=1}^{m}(y_i - \bar{y})^2\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i \tag{4}$$

- relative error percentage (REP)

$$\mathrm{REP} = 100\sqrt{\frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m} y_i^2}} \tag{5}$$

- Pearson Product-Moment Correlation Coefficient (PPMCC) which allows to compute the correlation between the output of the system and the output of the model

$$\mathrm{PPMCC} = \frac{\sum_{i=1}^{m}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{m}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{m}(\hat{y}_i - \bar{\hat{y}})^2}}\bar{\hat{y}} = \frac{1}{m}\sum_{i=1}^{m}\hat{y}_i \tag{6}$$

Note that all these measures of accuracy are useful for giving an exhaustive description of the quality of the forecast (Ghelardoni et al., 2013).

3.3. Prescriptive analytics

Once the model $\mathfrak{M}$ of the system $\mathfrak{S}$ is available, it is possible to control its inputs in order to produce a desired output. In this particular application, the goal is to find the minimum for the fuel consumption by acting on the ship's trim while keeping all other model inputs unchanged.

This approach, however, requires additional care and understanding of the underlying physics of the system:

- With reference to the previous work from the authors (Coraddu et al., 2015), not all variables available as measurements can be used for building a predictive model. In this case, in particular, the power and torque at the propeller shaft had to be excluded from the input list (see Table 3). Changing the trim would consequently change ship resistance and, therefore, the power required for its propulsion. Therefore modifying the trim while keeping the propeller power constant would represent a conceptual error.
- Not all possible trim values are physically allowed, and in turn power values, therefore boundary values, based on a priori knowledge of the system, should be provided.
- Although GBMs are more reliable in the extrapolation phase, their accuracy is expected to be reduced if they are extrapolated too far for outside the boundaries of the original range $\mathcal{D}_n$. Extrapolation is therefore allowed (the use of GBMs proposed in this paper is also based on their improved performance for extrapolation compared to BBMs) but this operation should be performed with care.

[1] The set $\mathcal{T}_m$ must be a different set respect to $\mathcal{D}_n$ which has been used to built the model $\mathfrak{M}$ in the case of BBMs and GBMs (Ghelardoni et al., 2013; Anguita et al., 2012)

Table 3
Variable of Table 2 exploited to built the $\mathfrak{M}$.

id	Name	Type
1	Latitude	Input
2	Longitude	Input
3	Volume	Input
4	State	Input
5	Auxiliary consumed	Input
6	Auxiliary electrical power output	Input
8	Shaft rpm	Input
9	Ship draft (forward)	Input
10	Ship draft (aft)	Input
11	Relative wind speed	Input
12	Relative wind direction	Input
13	GPS heading	Input
14	GPS speed	Input
15	Log speed	Input
16	Shaft generator power	Input
17	Sea depth	Input
18	Draft Port	Input
19	Draft Starboard	Input
20	Sea Water Temperature	Input
21	CPP Setpoint	Input
22	CPP Feedback	Input
23	Fuel Density	Input
24	Fuel Temperature	Input
25	Ambient Pressure	Input
26	Humidity	Input
27	Dew Point Temperature	Input
29	Rudder Angle	Input
30	Acceleration *X* Direction	Input
32	Acceleration *Y* Direction	Input
32	Acceleration *Z* Direction	Input
33	GyroX	Input
34	GyroY	Input
35	GyroZ	Input
36	Roll	Input
37	Pitch	Input
38	Yaw	Input
39	True direction	Input
40	True speed	Input
41	Beaufort	Input
	Shaft Power	Output
	Shaft Torque	Output
	Main Engine consumption	Output
	Shaft power predicted by the WBM	Input GBMs
	Shaft Torque predicted by the WBM	Input GBMs
	Main Engine consumption predicted by the WBM	Input GBMs

Based on these considerations, in this paper a method for trim optimisation is proposed. WBM, BBM and GBM are presented and compared based on the accuracy metrics proposed in Section 3.2. Based on this comparison, one model is selected for further analysis, checked for physical plausibility (Section 5) and used for application to the problem of trim optimisation (Section 6).

4. White, black and gray box models

4.1. White Box Models

A numerical model, the so called White Box Model (WBM), has been developed to evaluate the ship consumption, for different ship speed V, forward T_f, and aft T_a drafts in calm water scenario. The model is based on the knowledge of the ship's hull geometry, mass distribu-

Table 4
Main input quantities for ship resistance prediction.

Input	Symbolf	Unit
Length on waterline	L_{WL}	[m]
Breath on waterline	B_{WL}	[m]
Forward draught	T	[m]
Aft draught	T	[m]
Volume	∇	[m^3]
Wetted surface	S	[m^2]
Longitudinal position of center of buoyancy	L_{CB}	[m]
Breadth-Draught Ratio	B/T	
Mid-ship block coefficient	C_m	
Longitudinal prismatic coefficient	C_p	
Bow shape coefficient		
Section shape coefficient		

tion, propeller characteristics and main Diesel engine consumption map. The selected control variables (i.e. the system input which is under the user's control) taken into account are: the main engine revolution N, the forward and aft drafts. The control of these variables allow the ship to sail at the desired speed. The total ship's fuel consumption is used as model output.

The core of the procedure is the engine-propeller matching code utilised to evaluate the total ship fuel consumption and already tested as an effective tool in a previous work (Coraddu et al., 2011; Altosole et al. 2014,).

The prediction of ship resistance in calm water can be performed according to different approaches, normally divided in parametric approaches (Guldhammer and Harvald, 1974; Holtrop, 1984; Basin and Todd, 1964), and approaches based on computational fluid dynamics (CFD), such as the Reynolds averaged Navier-Stokes (RANS) or boundary element methods (BEM) (Hochkirch and Mallol,, Gaggero et al., 2010). In this study only parametric methods were considered because of their lower computational requirements. In particular, the Guldhammer Harvald method (Guldhammer and Harvald, 1974) was employed for the prediction of calm water resistance and, in particular, of the coefficient of total hull resistance in calm water C_T in Eq. (7). The inputs related to ship geometry used in the Guldhammer Harvald method are summarised in Table 4.

$$R_{\text{tot}} = \frac{1}{2} C_T \rho S V^2 \tag{7}$$

where ρ is the sea water density.

The equilibrium displacement is calculated, together with the necessary input variables required by he Guldhammer Harvald method (Guldhammer and Harvald, 1974) for each forward and aft drafts. The developed method reported in Coraddu et al. (2011) allows the authors to evaluate all the necessary inputs, reported in Table 4, taking into account the trim. With the proposed approach it is possible to compare the WBM model with the BBMs and GBMs ones. The propulsion coefficients have been corrected in magnitude as reported in Lützen and Kristensen (2012).

Propeller thrust and torque were computed offline for different pitch settings by means of a viscous method presented in Gaggero et al. (2010) and based on the knowledge of the geometrical features of the propeller. The calculated values were implemented in the matching code through the non dimensional thrust K_T and torque K_Q coefficients.

As reported in Fig. 1 a shaft generator is used for fulfilling on board auxiliary power demand. In order to optimise this feature the ship propulsion system has been set-up for working at fixed rpm using the pitch as control variable. Uncertainties about the available measures of propeller feedback and set point magnitude, reported in Table 2, forced the authors to impose the shaft rate of revolutions fixed and equal to 105 rpm. Once the displacement, shaft rate of revolutions and vessel speed are fixed, the advance coefficient J is defined together with the non dimensional thrust coefficient according to the following equations:

$$J = \frac{V(1-w)}{nD} \tag{8}$$

$$K_T = \frac{T}{\rho n^2 D^4} \tag{9}$$

where w is the wake factor, n is the propeller rate of revolution, D is the propeller diameter and T is the required thrust of the propeller. The engine-propeller matching code used in this work allows calculating the pitch ratio that provides the required thrust at the fixed shaft speed. Finally the delivered power P_d can be evaluated by means of the following quantities:

$$K_Q = \frac{Q}{\rho n^2 D^5} \tag{10}$$

$$\eta_0 = \frac{J k_T}{2\pi K_Q} \tag{11}$$

For a generic tuple of speed V_i, forward and aft drafts T_{fi}, T_{ai} values, the WBM model evaluates the propeller pitch ratio, which ensures the propulsion equilibrium between delivered and required thrust, and finally the associated fuel consumption. Starting from propeller torque, the engine brake power P_b is computed by the global efficiency of the drivetrain and it is then possible to evaluate the corresponding specific fuel consumption as reported in Coraddu et al. (2014).

A validation of the WBM model was performed based on the available measurements of delivered power at different displacement derived from model tests in calm water. The measured (P_{dh}) and predicted (P_{dn}) delivered power, together with the absolute percentage error of the model, are reported in Table 5. The results obtained with the WBM model clearly testify that the method used to derive a general representation of the relationship between vessel speed, displacement and delivered power in calm water scenarios can be innacurate. The Guldhammer and Harvard method was used for the determination of resistance in full awareness of its limitations in terms of accuracy and, to some extent, because of this inaccuracy. Nevertheless, this method is used in preference to RANS modelling for three reasons. Firstly, using a method for resistance prediction that is known to be limited for modern-day hulls actually serves to highlight the value of the Machine

Table 5
White Box Model validation.

Speed	$P_{dh}[KW]$	$P_{dn}[KW]$	error [%]
Δ = 25000 t			
10	1452.0	1548.9	6.7
12	2589.0	2653.8	2.5
14	4478.0	4465.9	0.3
16	7761.9	7430.3	4.3
Δ = 30000 t			
10	1528.0	1637.6	7.2
12	2763.0	2805.3	1.5
14	4768.0	4818.1	1.1
16	18,139.7	8139.1	0.0
Δ = 40000 t			
10	1768.0	1804.9	2.1
12	3100.0	3091.1	0.3
14	5196.0	5403.4	4.0
16	8760.2	9359.5	6.8
Δ = 50800 t			
10	1994.0	1815.5	9.0
12	3432.0	3106.3	9.5
14	5662.0	5458.5	3.6
16	9494.8	9546.7	0.5
Δ = 57100 t			
10	2162.0	1908.6	11.7
12	3690.0	3265.1	11.5
14	6089.0	5786.9	5.0
16	10,323.4	10,283.7	0.4

Learning technique, whereby it is able to optimise the model's outcome with the continual acquisition of real-world data. This is an important point that the paper is intended to highlight. Secondly, it could be applied rapidly for an enormous number of in-service operational data samples (1 measurement every 15 s for 2 years can produce 4.1 million operational states). To use RANS modelling for the determination of the resistance estimate in this framework would clearly be impractical and would massively limit the data that could be evaluated. Given the primary intent of the method is to utilize Machine Learning techniques with massive in-service data sets, RANS modelling would defeat the purpose. And finally, RANS is not used in this work as the presented framework is intended as an in-service operation optimisation methodology, whereas RANS modelling is best suited to a priori prediction and optimisation of operational behaviour. Indeed, the currently proposed framework can be complementary to an a priori RANS parametric study. While the RANS study can optimise the design before it enters service, the current framework can suggest any necessary changes required thereafter to optimise the vessel's operation as it takes in account (varying) real-world operating conditions that CFD cannot.

4.2. Black Box Models

Machine Learning (ML) approaches play a central role in extracting information from raw data collected from ship data logging systems. The learning process for ML approaches usually consists of two phases: (i) during the training phase, a set of data is used to induce a model that best fits them, according to some criteria; (ii) the trained model is used for prediction and control of the real system (feed-forward phase).

As the authors are targeting a regression problem (Vapnik, 1998b), the purpose is to find the best approximating function $h(\boldsymbol{x})$, where h: $\mathbb{R}^d \rightarrow \mathbb{R}$. During the training phase, the quality of the regressor $h(\boldsymbol{x})$ is measured according to a loss function $\ell(h(\boldsymbol{x}), y)$ (Lee et al., 1998), which calculates the discrepancy between the true and the estimated output (y and $\hat{y}$). The empirical error then computes the average discrepancy, reported by a model over $\mathcal{D}_n$:

$$\widehat{L}_n(h) = \frac{1}{n}\sum_{i=1}^{n} \ell(h(\boldsymbol{x}_i), y_i). \tag{12}$$

A simple criterium for selecting the final model during the training phase consists in choosing the approximating function that minimises the empirical error $\widehat{L}_n(h)$: this approach is known as Empirical Risk Minimisation (ERM) (Vapnik, 1998b). However, ERM is usually avoided in ML as it leads to severely overfitting the model on the training dataset (Anguita et al., 2011). A more effective approach consists in the minimisation of a cost function where the tradeoff between accuracy on the training data and a measure of the complexity of the selected approximating function is implemented (Tikhonov and Arsenin, 1979):

$$h^*: \quad \min_{h} \quad \widehat{L}_n(h) + \lambda\, C(h). \tag{13}$$

where $C(\cdot)$ is a complexity measure which depends on the selected ML approach and λ is a hyperparameter that must be set a priori and regulates the trade-off between the overfitting tendency, related to the minimisation of the empirical error, and the underfitting tendency, related to the minimisation of $C(\cdot)$. The optimal value for λ is problem-dependent, and tuning this hyperparameter is a non-trivial task (Anguita et al., 2011) and will be faced later in this section.

The approaches exploited in this paper are: the Regularised Least Squares (RLS) (Györfi, 2002), the Lasso Regression (LAR) (Tibshirani, 1996), and the Random Forrest (RF) (Breiman, 2001). These approaches are described in Appendix A.

4.3. Gray Box Models

GBMs are a combination of a WBMs and BBMs. This requires to modify the BBMs as defined in the previous section in a way to include the mechanistic knowledge of the system. Two approaches are tested and compared in this paper:

- a Naive approach (N-GBM) where the output of the WBM is used as a new feature that the BBM can use for training the model.
- an Advanced approach (A-GBM) where the regularisation process is changed in order to include some a-priori information (Decherchi et al., 2010; Anguita et al., 2011, 2011).

In the N-GBM case, the WBM can be seen as a function of the input $\boldsymbol{x}$. The WBM, that we call here $h_{\mathrm{WBM}}(\boldsymbol{x})$, allows the creation of a new dataset:

$$\mathcal{D}_n^{\mathrm{WBM},\mathcal{X}} = \left\{ \left(\begin{bmatrix} \boldsymbol{x}_1 \\ h_{\mathrm{WBM}}(\boldsymbol{x}_1) \end{bmatrix}, y_1 \right), \ldots, \left(\begin{bmatrix} \boldsymbol{x}_n \\ h_{\mathrm{WBM}}(\boldsymbol{x}_n) \end{bmatrix}, y_n \right) \right\}$$

Based on this new dataset a BBM can be generated $h_{\mathrm{BBM}}([\boldsymbol{x}^T | h_{\mathrm{WBM}}(\boldsymbol{x})]^T)$. According to this approach, every run of the GBM requires an initial run of the WBM in order to compute its output $h_{\mathrm{WBM}}(\boldsymbol{x})$, which allows evaluating $h_{\mathrm{BBM}}([\boldsymbol{x}^T | h_{\mathrm{WBM}}(\boldsymbol{x})]^T)$. This is the simplest approach for including new information into the learning process. Note that with this approach any of the previously cited BBMs (e.g. RLS, LAR or RM) can be used for building the corresponding N-GBM.

In the A-GBM case the WBM part of the model is assumed to be included in the $\boldsymbol{w}$ vector:

$$h_{\mathrm{WBM}}(\boldsymbol{x}) = \boldsymbol{w}_{\mathrm{WBM}}^T \boldsymbol{\phi}(\boldsymbol{x}), \tag{14}$$

According to Anguita et al. (2011, 2011), the regularisation process of Eq. (A.4) is modified to:

$$\boldsymbol{w}^*: \ \min_{\boldsymbol{w}} \frac{1}{n}\sum_{i=1}^{n} [\boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) - y_i]^2 + \lambda \|\boldsymbol{w} - \boldsymbol{w}_{\mathrm{WBM}}\|_2^2. \tag{15}$$

It is possible to prove that by exploiting the kernel trick the solution to this problem can be rewritten as:

$$h^*(\boldsymbol{x}) = h_{\mathrm{WBM}}(\boldsymbol{x}) + \sum_{i=1}^{n} \alpha_i^* K(\boldsymbol{x}_i, \boldsymbol{x}). \tag{16}$$

where

$$\boldsymbol{\alpha}^*: \quad \min_{\boldsymbol{\alpha}} \frac{1}{n}\sum_{i=1}^{n} \left[\sum_{j=1}^{n} \alpha_j K(\boldsymbol{x}_j, \boldsymbol{x}_i) + h_{\mathrm{WBM}}(\boldsymbol{x}_i) - y_i \right]^2 + \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j K(\boldsymbol{x}_j, \boldsymbol{x}_i), \tag{17}$$

The solution to this problem can be computed by solving the following linear system:

$$(K + n\lambda I)\boldsymbol{\alpha}^* = \boldsymbol{y} - \boldsymbol{h}_{\mathrm{WBM}}, \tag{18}$$

where $\boldsymbol{h}_{\mathrm{WBM}} = [h_{\mathrm{WBM}}(\boldsymbol{x}_1), \ldots, h_{\mathrm{WBM}}(\boldsymbol{x}_n)]^T$. Note that the solution does not depend on the form of $h_{\mathrm{WBM}}(\boldsymbol{x})$ so that any WBM can be used as $h_{\mathrm{WBM}}(\boldsymbol{x})$.

Another possible way of achieving the same solution of the problem of Eq. (18) is to create a new dataset:

$$\mathcal{D}_n^{\mathrm{WBM},\mathcal{Y}} = \{(\boldsymbol{x}_1, y_1 - h_{\mathrm{WBM}}(\boldsymbol{x}_1)), \ldots, (\boldsymbol{x}_n, y_n - h_{\mathrm{WBM}}(\boldsymbol{x}_n))\}$$

where the target is no longer the true label y but the true label minus the hint given by the a priori information included in $h_{\mathrm{WBM}}(\boldsymbol{x})$. This means finding a BBM that minimises the error of the WBM prediction.

It should be noted that the A-GBM is more theoretically justified in

Table 6
Indexes of performance of the WBM in predicting the Shaft Power, Shaft Torque, and Fuel Consumption.

Shaft Power					
MAE [KW]	MAPE [%]	MSE [KW^2]	NMSE	REP [%]	PPMCC
7.69e+02	17.85	1.00e+06	1.13	23.59	0.65
Shaft Torque					
MAE [Nm]	MAPE [%]	MSE [N^2m^2]	NMSE	REP [%]	PPMCC
6.54e+01	18.13	6.92e+03	0.94	22.01	0.22
Fuel Consumption					
MAE $\left[\frac{g}{KWh}\right]$	MAPE [%]	MSE $\left[\frac{g^2}{KW^2h^2}\right]$	NMSE	REP [%]	PPMCC
5.14e−02	20.95	3.94e−03	1.98	25.40	0.63

the regularisation context while the N-GBM is more intuitive since all the available knowledge is given as input to the BBM learning process. From a probabilistic point of view, the A-GMB changes the $\mathbb{P}(y|\boldsymbol{x})$ while the N-GBM modifies the whole joint probability $\mathbb{P}(y, \boldsymbol{x})$, hence deeply influencing the nature of the problem.

The (λ, γ) for RLS, the λ for LAR and d' for RF of the N-GBM and A-GBM are tuned with the BOO as described for the BBM, since both N-GBM and A-GBM basically require to build a BBM over a modified training set.

Note that the A-GBM approach may seems quite similar to the one of Transfer Learning (TL) where a model learned for solving a problem in one domain is used as hint in another domain (Pan and Yang, 2010), but there are two main differences between TL and A-GBM. In TL, contrarily to A-GBM, the two models are learned from the data and in TL the two models are build to solve two different problems in two different domains so to transfer the knowledge from one domain to another.

4.4. Model validation and computational requirements

The WBM was validated using the data described in Section 2 versus propeller shaft power, shaft torque, and total fuel consumption. The results of the validation are presented in Table 6. The results show that the WBM does not show sufficient accuracy when compared with operational measurements. The inability of the model to take into account the influence of the sea state (i.e. wind and waves) on the required propulsion power is considered to be the largest source of error for this model.

The BBMs built according to the RLS, LAR and RF methods were validated versus the same dataset as for the WBM validation procedure. However, in the case of the BBMs the $\mathcal{D}_n$ was divided in two sets $\mathcal{L}_{n_l}$ and $\mathcal{T}_{n_t}$ respectively for learning and test. The two sets were defined so that $\mathcal{D}_n = \mathcal{L}_{n_l} \cup \mathcal{T}_{n_t}$ and $\mathcal{L}_{n_l} \cap \mathcal{T}_{n_t} = \oslash$ in order to maintain the

Table 7
Indexes of performance of the different BBM for different n_l for the prediction of Shaft Power, Shaft Torque and Fuel Consumption.

Shaft Power	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF
n_l	MAE [KW]			MAPE [%]			MSE [KW^2]			NMSE			REP [%]			PPMCC		
10	**7.79e+02**	8.62e+02	7.83e+02	**17.52**	19.34	17.66	1.14e+06	1.28e+06	**1.14e+06**	1.29	1.42	**1.28**	25.18	27.96	**24.89**	0.23	**0.23**	0.23
20	5.20e+02	5.96e+02	**5.05e+02**	12.27	14.14	**12.10**	4.89e+05	5.52e+05	**4.69e+05**	0.55	0.63	**0.54**	16.47	18.17	**16.01**	0.71	0.71	**0.72**
50	4.08e+02	4.68e+02	**3.59e+02**	10.30	11.70	**8.90**	2.87e+05	3.22e+05	**2.51e+05**	0.32	0.37	**0.28**	12.62	14.30	**10.91**	0.84	0.84	**0.84**
100	2.92e+02	3.34e+02	**2.34e+02**	7.35	8.33	**6.03**	1.57e+05	1.79e+05	**1.26e+05**	0.18	0.20	**0.14**	9.33	10.59	**7.81**	0.92	**0.93**	0.92
200	2.05e+02	2.34e+02	**1.56e+02**	5.21	5.82	**3.98**	9.50e+04	1.07e+05	**7.25e+04**	0.11	0.12	**0.08**	7.26	8.38	**5.64**	0.95	0.95	**0.96**
500	1.45e+02	1.65e+02	**1.05e+02**	3.62	3.99	**2.53**	5.45e+04	6.12e+04	**3.94e+04**	0.06	0.07	**0.04**	5.50	6.06	**3.95**	0.97	**0.98**	0.97
1000	1.05e+02	1.20e+02	**7.67e+01**	2.62	3.02	**1.90**	3.50e+04	3.99e+04	**2.47e+04**	0.04	0.04	**0.03**	4.41	5.07	**3.18**	0.98	0.98	**0.99**
Shaft Torque																		
n_l	MAE [Nm]			MAPE [%]			MSE [N^2m^2]			NMSE			REP [%]			PPMCC		
10	**6.60e+01**	7.48e+01	6.69e+01	19.13	21.96	**18.95**	7.02e+03	8.05e+03	**6.94e+03**	0.96	1.09	**0.95**	**22.22**	24.95	22.66	0.21	0.21	**0.21**
20	6.20e+01	6.99e+01	**6.18e+01**	17.90	20.57	**17.51**	6.00e+03	6.76e+03	**5.93e+03**	0.82	0.93	**0.80**	20.53	22.67	**19.77**	0.42	0.42	**0.42**
50	3.74e+01	4.21e+01	**3.20e+01**	10.92	12.33	**9.70**	2.65e+03	3.01e+03	**2.31e+03**	0.36	0.41	**0.32**	13.64	15.47	**11.85**	0.83	0.83	**0.83**
100	2.67e+01	2.95e+01	**2.22e+01**	7.40	8.16	**6.07**	1.39e+03	1.56e+03	**1.15e+03**	0.19	0.22	**0.16**	9.89	11.07	**8.18**	0.92	0.93	**0.93**
200	1.95e+01	2.21e+01	**1.51e+01**	5.36	6.15	**4.21**	8.05e+02	8.87e+02	**6.18e+02**	0.11	0.12	**0.08**	7.52	8.45	**5.70**	0.95	0.95	**0.95**
500	1.33e+01	1.51e+01	**9.61e+00**	3.75	4.14	**2.74**	4.46e+02	5.14e+02	**3.27e+02**	0.06	0.07	**0.04**	5.60	6.23	**4.06**	0.97	**0.97**	0.97
1000	1.01e+01	1.12e+01	**7.23e+00**	2.83	3.26	**2.05**	3.08e+02	3.42e+02	**2.16e+02**	0.04	0.05	**0.03**	4.65	5.24	**3.34**	0.98	**0.99**	0.98
Fuel Consumption																		
n_l	MAE $\left[\frac{g}{KWh}\right]$			MAPE [%]			MSE $\left[\frac{g^2}{KW^2h^2}\right]$			NMSE			REP [%]			PPMCC		
10	**3.54e−02**	3.92e−02	3.59e−02	14.12	15.98	**14.11**	**2.25e-03**	2.50e−03	2.28e−03	**1.13**	1.28	1.16	**19.21**	21.33	19.62	0.23	0.23	**0.23**
20	3.26e−02	3.68e−02	**3.14e−02**	12.88	14.37	**12.75**	1.96e−03	2.19e−03	**1.92e−03**	0.98	1.09	**0.94**	17.91	20.18	**17.84**	0.38	**0.38**	0.38
50	2.40e−02	2.74e−02	**2.07e−02**	9.39	10.58	**8.00**	1.26e−03	1.41e−03	**1.09e−03**	0.63	0.72	**0.55**	14.38	16.39	**12.80**	0.65	**0.66**	0.65
100	1.53e−02	1.74e−02	**1.25e−02**	6.25	6.90	**5.19**	5.22e−04	5.77e−04	**4.24e−04**	0.26	0.30	**0.22**	9.25	10.54	**7.57**	0.87	**0.88**	0.88
200	1.04e−02	1.16e−02	**8.03e−03**	4.35	5.00	**3.27**	2.71e−04	3.09e−04	**2.07e−04**	0.14	0.15	**0.10**	6.67	7.41	**5.09**	0.93	0.94	**0.94**
500	7.68e−03	8.81e−03	**5.44e−03**	3.28	3.67	**2.34**	1.83e−04	2.04e−04	**1.31e−04**	0.09	0.10	**0.06**	5.47	6.20	**3.83**	0.95	**0.96**	0.95
1000	6.30e−03	7.24e−03	**4.62e−03**	2.70	3.01	**1.95**	1.52e−04	1.67e−04	**1.10e−04**	0.08	0.09	**0.06**	4.99	5.64	**3.56**	0.96	0.96	**0.96**

Table 8
Indexes of performance of the different N-GBM for different n_l for the prediction of Shaft Power, Shaft Torque and Fuel Consumption.

n_l	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF
Shaft power																		
n_l	MAE [KW]			MAPE [%]			MSE [KW^2]			NMSE			REP [%]			PPMCC		
10	3.91e+02	4.75e+02	**3.12e+02**	9.10	10.67	**7.22**	5.93e+05	7.08e+05	**4.76e+05**	0.66	0.79	**0.51**	12.89	15.50	**10.28**	0.23	0.23	**0.23**
20	2.71e+02	3.24e+02	**2.10e+02**	6.35	7.49	**5.12**	2.50e+05	3.02e+05	**1.96e+05**	0.28	0.33	**0.22**	8.48	9.97	**6.72**	0.72	**0.72**	0.72
50	2.06e+02	2.49e+02	**1.66e+02**	5.28	6.25	**4.15**	1.46e+05	1.73e+05	**1.19e+05**	0.16	0.20	**0.13**	6.53	7.63	**5.23**	**0.85**	0.85	0.85
100	1.22e+02	1.79e+02	**1.21e+02**	**2.94**	4.44	3.05	6.47e+04	9.81e+04	**6.45e+04**	0.07	0.11	**0.07**	**3.87**	5.72	3.89	0.93	0.93	**0.93**
200	8.34e+01	1.03e+02	**6.39e+01**	2.13	2.72	**1.57**	3.82e+04	4.91e+04	**2.91e+04**	0.04	0.05	**0.03**	2.96	3.73	**2.25**	0.96	**0.96**	0.96
500	5.79e+01	7.32e+01	**4.55e+01**	1.45	1.82	**1.14**	2.23e+04	2.82e+04	**1.71e+04**	0.03	0.03	**0.02**	2.26	2.83	**1.69**	**0.98**	0.98	0.98
1000	4.39e+01	5.47e+01	**3.18e+01**	1.10	1.32	**0.79**	1.46e+04	1.78e+04	**1.06e+04**	0.02	0.02	**0.01**	1.79	2.22	**1.35**	0.99	**0.99**	0.99
Shaft Torque																		
n_l	MAE [Nm]			MAPE [%]			MSE [N^{2}m^2]			NMSE			REP [%]			PPMCC		
10	3.34e+01	3.99e+01	**2.74e+01**	9.72	11.93	**7.72**	3.60e+03	4.31e+03	**2.89e+03**	0.49	0.59	**0.40**	11.34	13.47	**8.93**	0.21	0.21	**0.21**
20	3.15e+01	3.84e+01	**2.49e+01**	9.13	11.19	**7.18**	3.09e+03	3.72e+03	**2.47e+03**	0.43	0.50	**0.34**	10.29	12.43	**8.59**	0.43	**0.43**	0.43
50	1.89e+01	2.34e+01	**1.51e+01**	5.59	6.80	**4.55**	1.37e+03	1.60e+03	**1.10e+03**	0.18	0.22	**0.15**	6.92	8.40	**5.52**	0.84	0.84	**0.84**
100	**1.10e+01**	1.64e+01	1.10e+01	**3.00**	4.49	3.04	5.82e+02	8.58e+02	**5.59e+02**	0.08	0.12	**0.08**	4.15	5.98	**4.14**	**0.93**	0.93	0.93
200	7.89e+00	1.00e+01	**5.93e+00**	2.18	2.76	**1.62**	3.32e+02	4.22e+02	**2.49e+02**	0.05	0.06	**0.03**	3.08	3.92	**2.33**	0.96	**0.96**	0.96
500	5.54e+00	6.71e+00	**4.00e+00**	1.53	1.90	**1.15**	1.87e+02	2.25e+02	**1.39e+02**	0.03	0.03	**0.02**	2.31	2.90	**1.69**	0.98	0.98	**0.98**
1000	4.06e+00	5.23e+00	**3.11e+00**	1.13	1.42	**0.85**	1.29e+02	1.60e+02	**9.45e+01**	0.02	0.02	**0.01**	1.89	2.44	**1.46**	0.99	0.99	**0.99**
Fuel Consumption																		
n_l	MAE $\left[\frac{g}{KWh}\right]$			MAPE [%]			MSE $\left[\frac{g^2}{KW^2h^2}\right]$			NMSE			REP [%]			PPMCC		
10	1.77e−02	2.16e−02	**1.45e−02**	7.20	8.88	**5.86**	1.15e−03	1.37e−03	**9.38e−04**	0.57	0.69	**0.46**	9.64	11.79	**7.70**	0.24	**0.24**	0.24
20	1.70e−02	1.97e−02	**1.35e−02**	6.66	8.05	**5.28**	9.88e−04	1.22e−03	**7.93e−04**	0.52	0.61	**0.41**	9.40	10.81	**7.25**	**0.38**	0.38	0.38
50	1.25e−02	1.46e−02	**9.63e−03**	4.90	5.77	**3.79**	6.47e−04	7.63e−04	**5.11e−04**	0.32	0.39	**0.26**	7.40	8.99	**5.93**	0.66	0.66	**0.66**
100	6.20e−03	9.52e−03	**6.16e−03**	2.57	3.80	**2.52**	2.16e−04	3.28e−04	**2.09e−04**	0.11	0.16	**0.11**	**3.78**	5.68	3.86	**0.88**	0.88	0.88
200	4.18e−03	5.20e−03	**3.25e−03**	1.80	2.28	**1.31**	1.11e−04	1.42e−04	**8.40e−05**	0.06	0.07	**0.04**	2.69	3.45	**2.03**	0.94	0.94	**0.94**
500	3.15e−03	3.94e−03	**2.35e−03**	1.36	1.67	**1.00**	7.53e−05	9.17e−05	**5.64e−05**	0.04	0.05	**0.03**	2.28	2.84	**1.67**	0.96	0.97	**0.97**
1000	2.61e−03	3.29e−03	**1.97e−03**	1.12	1.42	**0.83**	6.21e−05	7.74e−05	**4.71e−05**	0.03	0.04	**0.02**	2.03	2.54	**1.55**	0.97	**0.97**	0.97

independence of the two sets.

The process of splitting the full dataset in a learning set and test set is repeated 30 times so to obtain statistical relevant results. The results are reported for different sizes of $\mathcal{L}_{n_l}$ with $n_l \in \{10, 20, 50, 100, 200, 500, 1000, 2000, 5000\}$ sampled randomly from the full sized dataset of 100,000 observations. The optimisation procedure is repeated for different values of both hyperparameters (γ and λ), where their values are taken based on a 60 points equally spaced in logarithmic scale in the range $[10^{-6}, 10^{3}]$ and the best set of hyperparameters is selected according to the BOO (Section 4.2). The same has been done for d' in RF.

Authors report in Table 7 the results of the accuracy over the Shaft Power, Shaft Torque and Fuel Consumption. Note that the different BBMs remarkably outperform the WBM since they are able to take into account all the available information measured by the on board sensors. Among the different BBMs, the RF-based model shows the most promising results compared to the RLS and LAR.

Also in the case of the GBM, analogously to the procedure adopted for the BBM, the original dataset $\mathcal{D}_n$ is divided in two sets $\mathcal{L}_{n_l}$ and $\mathcal{T}_{n_t}$ and λ, γ and d' are chosen according to the BOO procedure. The results of the accuracy over Shaft Power, Shaft Torque, and Fuel Consumption is reported in Table 8 for the N-GBM and in Table 9 for the A-GBM. Note that both the N-GBM and A-GBM models improve the BBM just by few percentage points. In particular the improvement decreases as n increases.

From the results it is possible to note that the WBM, as expected, has the lowest performance in terms of prediction accuracy. On the other hand, the GBMs outperform the BBMs by a smaller percentage. The MAPE of the BBM, N-GBM and A-GBM for different values of n_l are reported in Fig. 10, while Tables 10–12 shows the percentage of improvement for the different indexes of performance of the BBM, N-GBM and A-GBM over the WBM. Note that, for the RF, the N-GBM outperform both the A-GBM and BBM. From the results of Fig. 10 and Tables 10–12 it is possible to note how the WBM, even if not so accurate, can help the GBM in obtaining higher accuracy, with respect to the BBM, by using almost half of the data given a required accuracy. This is a critical issue in real word applications where the collection of labeled data can be expensive or at least requires a long period of in-service operational time of the vessel (Coraddu et al., 2015).

As final remark, it is useful to briefly discuss the computational requirements of the three methods. The proposed WBM is computational inexpensive respect to the BBM. Im particular, the time needed to built the model and make prediction with the WBM is negligible respect to the time needed by the BBM. The computational requirements of the GBM, instead, are, by construction, the sum of the requirements of the WBM and BBM. This means that the GBM, although more accurate of the BBM, requires a computational effort which is approximately the same of the BBM.

5. Feature selection

Once a model is built and has been confirmed to be a sufficiently

Table 9
Indexes of performance of the different A-GBM for different n_l for the prediction of Shaft Power, Shaft Torque and Fuel Consumption.

n_l	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF	RLS	LAR	RF
Shaft power																		
n_l	MAE [KW]			MAPE [%]			MSE [KW^2]			NMSE			REP [%]			PPMCC		
10	**3.64e+02**	4.00e+02	4.06e+02	**8.16**	8.92	8.88	**5.37e+05**	5.89e+05	5.91e+05	**0.59**	0.67	0.66	**11.65**	13.10	13.02	**0.23**	0.23	0.23
20	**2.38e+02**	2.64e+02	2.68e+02	**5.76**	6.20	6.25	**2.30e+05**	2.49e+05	2.53e+05	**0.26**	0.28	0.28	**7.55**	8.56	8.45	0.72	0.72	**0.72**
50	**1.85e+02**	2.12e+02	2.14e+02	**4.78**	5.15	5.36	**1.33e+05**	1.51e+05	1.48e+05	**0.15**	0.17	0.16	**5.88**	6.32	6.60	**0.85**	0.85	0.85
100	**1.06e+02**	1.49e+02	1.19e+02	**2.61**	3.83	2.94	**5.51e+04**	7.88e+04	6.43e+04	**0.06**	0.09	0.07	**3.37**	4.84	3.88	0.93	0.93	**0.93**
200	**7.29e+01**	9.66e+01	8.50e+01	**1.91**	2.36	2.10	**3.42e+04**	4.39e+04	3.82e+04	**0.04**	0.05	0.04	**2.61**	3.34	2.97	0.96	**0.96**	0.96
500	**5.26e+01**	6.82e+01	5.91e+01	**1.30**	1.69	1.50	**1.97e+04**	2.50e+04	2.20e+04	**0.02**	0.03	0.02	**2.01**	2.57	2.25	**0.98**	0.98	0.98
1000	**3.72e+01**	4.86e+01	4.30e+01	**0.92**	1.23	1.10	**1.25e+04**	1.59e+04	1.43e+04	**0.01**	0.02	0.02	**1.61**	2.07	1.76	**0.99**	0.99	0.99
Shaft Torque																		
n_l	MAE [Nm]			MAPE [%]			MSE [N^{2}m^2]			NMSE			REP [%]			PPMCC		
10	**3.02e+01**	3.31e+01	3.43e+01	**8.80**	9.86	9.71	**3.25e+03**	3.57e+03	3.56e+03	**0.44**	0.49	0.49	**10.19**	11.22	11.62	0.21	**0.21**	0.21
20	**2.89e+01**	3.13e+01	3.24e+01	**8.06**	9.13	9.15	**2.77e+03**	3.05e+03	3.07e+03	**0.39**	0.43	0.42	**9.56**	10.54	10.60	0.43	**0.43**	0.43
50	**1.72e+01**	1.93e+01	1.90e+01	**4.97**	5.69	5.52	**1.19e+03**	1.37e+03	1.37e+03	**0.17**	0.19	0.19	**6.23**	6.88	6.95	0.84	0.84	**0.84**
100	**9.47e+00**	1.40e+01	1.11e+01	**2.64**	3.76	3.06	**5.09e+02**	7.16e+02	5.60e+02	**0.07**	0.10	0.08	**3.63**	5.06	4.06	0.93	0.93	**0.94**
200	**7.04e+00**	9.05e+00	7.95e+00	**1.90**	2.45	2.22	**2.83e+02**	3.78e+02	3.35e+02	**0.04**	0.05	0.04	**2.70**	3.49	3.12	**0.96**	0.96	0.96
500	**4.78e+00**	6.12e+00	5.34e+00	**1.34**	1.70	1.55	**1.64e+02**	2.11e+02	1.80e+02	**0.02**	0.03	0.03	**2.00**	2.54	2.31	0.98	0.98	**0.98**
1000	**3.61e+00**	4.56e+00	4.17e+00	**1.01**	1.28	1.15	**1.10e+02**	1.39e+02	1.26e+02	**0.01**	0.02	0.02	**1.66**	2.14	1.92	0.99	0.99	**0.99**
Fuel Consumption																		
n_l	MAE $\left[\frac{g}{KWh}\right]$			MAPE [%]			MSE $\left[\frac{g^2}{KW^2h^2}\right]$			NMSE			REP [%]			PPMCC		
10	**1.65e−02**	1.78e−02	1.80e−02	**6.67**	7.37	7.08	**1.05e−03**	1.14e−03	1.15e−03	**0.52**	0.58	0.58	**9.00**	10.03	9.97	0.24	**0.24**	0.24
20	**1.50e−02**	1.71e−02	1.64e−02	**5.88**	6.62	6.46	**9.13e−04**	1.01e−03	1.02e−03	**0.45**	0.50	0.50	**8.29**	9.29	9.22	**0.38**	0.38	0.38
50	**1.12e−02**	1.25e−02	1.25e−02	**4.35**	4.83	4.79	**5.78e−04**	6.33e−04	6.59e−04	**0.29**	0.33	0.33	**6.78**	7.30	7.51	0.66	**0.66**	0.66
100	**5.43e−03**	7.82e−03	6.41e−03	**2.21**	3.15	2.62	**1.84e−04**	2.72e−04	2.14e−04	**0.10**	0.13	0.11	**3.27**	4.70	3.87	**0.88**	0.88	0.88
200	**3.78e−03**	4.87e−03	4.30e−03	**1.58**	2.05	1.79	**9.81e−05**	1.27e−04	1.09e−04	**0.05**	0.06	0.06	**2.37**	3.13	2.73	**0.95**	0.95	0.94
500	**2.78e−03**	3.55e−03	3.10e−03	**1.17**	1.52	1.37	**6.51e−05**	8.55e−05	7.42e−05	**0.03**	0.04	0.04	**2.00**	2.58	2.23	0.97	0.96	**0.97**
1000	**2.26e−03**	2.91e−03	2.55e−03	**0.95**	1.28	1.13	**5.48e−05**	6.92e−05	6.20e−05	**0.03**	0.04	0.03	**1.77**	2.32	2.06	**0.97**	0.97	0.97

accurate representation of the real system of interest, it can be interesting to investigate how the model $\mathfrak{M}$ is affected by the different features that have been used in the model identification phase.

In data analytics this procedure is called feature selection or feature ranking (Guyon and Elisseeff, 2003; Friedman et al., 2001; Chang and Lin, 2008; Yoon et al., 2005; Hong, 1997). This process allows detecting if the importance of those features, that are known to be relevant from a theoretical perspective, is appropriately described by $\mathfrak{M}$. The failure of the statistical model to properly account for the relevant features might indicate poor quality in the measurements. Feature selection therefore represents an important step of model verification, since the proposed model $\mathfrak{M}$ should generate results consistently with the available knowledge of the physical system under exam. This is particularly important in the case of BBM (and, to a more limited extent, for GBM), since they do not make use of any mechanistic knowledge of the system and might therefore lead to non-physical results (e.g. mass or energy unbalances). Feature selection also allows checking the statistical robustness of the employed methods.

In this paper, three different methods for feature ranking are applied:

- Brute Force Method (BFM), which searches for the optimal solution. This is the most accurate method but also the most computationally expensive (see Section 5.1) (Gieseke et al., 2014; Friedman et al., 2001).
- Regularisation Based Method (RBM) which works by building the BBM which automatically discarding the features that do not significantly contribute to the model output (for example by building an ad-hoc regularisers (Friedman et al., 2001; Zou et al., 2007; Zou and Hastie, 2005; Simon et al., 2013; De Mol et al., 2009; Ng, 2004; Meinshausen and Bühlmann, 2010)). In this paper, the Lasso Regularisation technique was used (see Section 5.2).
- Random Forest based method (RFM) uses a combination of Decision Tree methods together with the permutation test (Good, 2013) in order to perform the selection and the ranking of the features (Friedman et al., 2001; Sugumaran et al., 2007; Kohavi and John, 1997).

5.1. Brute force method

According to the Brute Force method (BFM) for feature selection, the k most important features of the model can be identified as follows:

- a first version of the model $\mathfrak{M}$ including all the available features is built. The full model is tested against a test set $\widehat{L}_{\text{Test}}$;
- for a given k, a set of new models is built for all possible configurations including feature k. For every possible configuration, which are $\binom{d}{k}$, a new model $\mathfrak{M}^j$ is built where $j \in \{1, \ldots, \binom{d}{k}\}$ together with its error on the test set $\widehat{L}^j_{\text{Test}}$;
- the smaller is the difference between $\widehat{L}^j_{\text{Test}}$ and $\widehat{L}_{\text{Test}}$, the greater is the importance of that set of features.

Given its high computational demands, this approach is not feasible

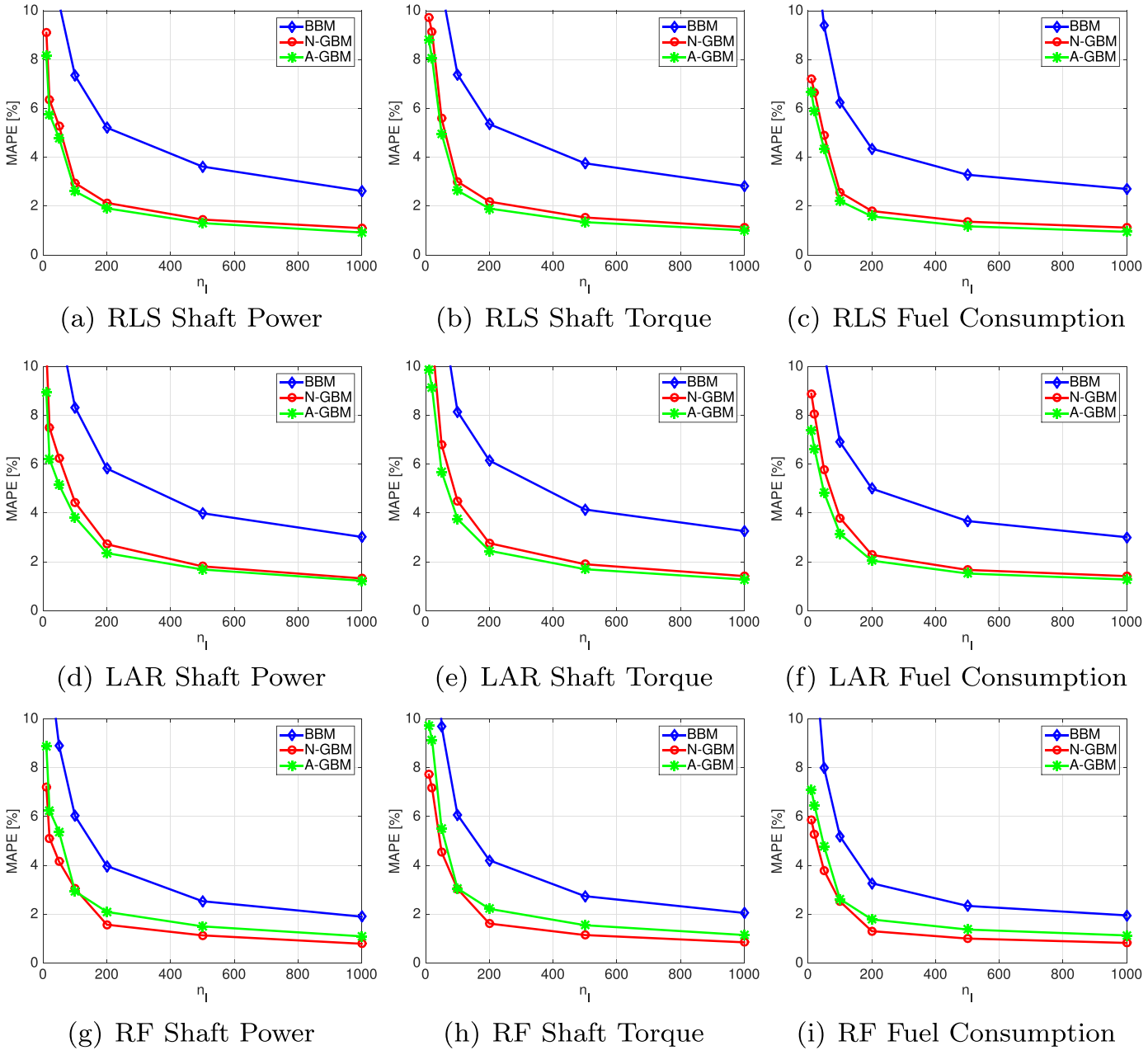

Fig. 10. Shaft Power, Shaft Torque, and Fuel Consumption MAPE of the BBM, N-GBM and A-GBM for different n_l.

for $d > 15 \div 20$. A solution for reducing the required computational time is to adopt a greedy procedure:

- a first version of the model $\mathfrak{M}$ including all the available features is built. The full model is tested against a test set $\widehat{L}_{\text{Test}}$;
- given a feature j_1, the model $\mathfrak{M}$ is built which only includes that feature. The error against the test set ($\widehat{L}_{\text{Test}}^{j_1}$) can now be calculated;
- the same procedure is performed for each feature $j_1 \in \{1, \ldots, d\}$;
- the smaller is the difference between $\widehat{L}_{\text{Test}}^{j_1}$ and $\widehat{L}_{\text{Test}}$ the grater is the importance of the features j_1

$$j_1^* = \arg \min_{j_1 \in \{1,\ldots,d\}} \widehat{L}_{\text{Test}} - \widehat{L}_{\text{Test}}^{j_1} \tag{19}$$

- this procedure is repeated by adding to j_1^* all the other features one at the time for finding the second most important feature $j_2^* \in \{1, \ldots, d\} \backslash j_1^*$. This operation is repeated until the required size (k) of the ranking is achieved.

Greedy methods are more time efficient compared to brute force methods, but do not ensure the full correctness of the result.

In this paper, several different models were proposed and are here tested for feature ranking. These models are: BBM, N-GBM and A-GBM with RLS, LAR and RF for a total of nine possibilities. It should be noted that for the N-GBM there is another feature which is the WBM (see Table 3).

5.2. Regularisation based method

The brute force method is a quite powerful approach but it requires

Table 10
Percentage of improvement for the different indexes of performance of the BBM, N-GBM and A-GBM over the WBM built with different RLS for the prediction of Shaft Power, Shaft Torque and Fuel Consumption.

	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM
Shaft Power																		
n_l	MAE [KW]			MAPE [%]			MSE [KW^2]			NMSE			REP [%]			PPMCC		
10	−1	49	**53**	2	49	**54**	−14	41	**46**	−14	41	**48**	−7	45	**51**	−65	−64	**−64**
20	32	65	**69**	31	64	**68**	51	75	**77**	51	75	**77**	30	64	**68**	10	**11**	11
50	47	73	**76**	42	70	**73**	71	85	**87**	71	86	**87**	46	72	**75**	29	**31**	30
100	62	84	**86**	59	84	**85**	84	94	**94**	84	93	**94**	60	84	**86**	41	**43**	43
200	73	89	**91**	71	88	**89**	90	96	**97**	91	96	**97**	69	87	**89**	46	48	**48**
500	81	92	**93**	80	92	**93**	95	98	**98**	95	98	**98**	77	90	**91**	49	**51**	51
1000	86	94	**95**	85	94	**95**	96	99	**99**	97	99	**99**	81	92	**93**	51	**53**	53
Shaft Torque																		
n_l	MAE [Nm]			MAPE [%]			MSE [N^{2}m^2]			NMSE			REP [%]			PPMCC		
10	−1	49	**54**	−6	46	**51**	−1	48	**53**	−3	48	**53**	−1	48	**54**	−6	−5	**−5**
20	5	52	**56**	1	50	**56**	13	55	**60**	12	54	**59**	7	53	**57**	91	94	**94**
50	43	71	**74**	40	69	**73**	62	80	**83**	61	80	**82**	38	69	**72**	275	280	**280**
100	59	83	**86**	59	83	**85**	80	92	**93**	80	92	**93**	55	81	**84**	319	**325**	324
200	70	88	**89**	70	88	**90**	88	95	**96**	88	95	**96**	66	86	**88**	331	337	**338**
500	80	92	**93**	79	92	**93**	94	97	**98**	93	97	**98**	75	89	**91**	341	**347**	347
1000	85	94	**94**	84	94	**94**	96	98	**98**	95	98	**98**	79	91	**92**	345	350	**350**
Fuel Consumption																		
n_l	MAE $\left[\frac{g}{KWh}\right]$			MAPE [%]			MSE $\left[\frac{g^2}{KW^2h^2}\right]$			NMSE			REP [%]			PPMCC		
10	31	66	**68**	33	66	**68**	43	71	**73**	43	71	**74**	24	62	**65**	−63	**−63**	−63
20	36	67	**71**	39	68	**72**	50	75	**77**	50	74	**77**	30	63	**67**	−40	**−39**	−39
50	53	76	**78**	55	77	**79**	68	84	**85**	68	84	**85**	43	71	**73**	4	5	**5**
100	70	88	**89**	70	88	**89**	87	95	**95**	87	94	**95**	64	85	**87**	38	**40**	40
200	80	92	**93**	79	91	**92**	93	97	**98**	93	97	**97**	74	89	**91**	48	50	**50**
500	85	94	**95**	84	93	**94**	95	98	**98**	95	98	**98**	78	91	**92**	51	53	**53**
1000	88	95	**96**	87	95	**95**	96	98	**99**	96	98	**99**	80	92	**93**	53	54	**55**

a significant computational effort. The Lasso Regression can be used for ranking the importance of the features with lower computational demand.

However, the results of the Lasso Regression method are strongly influenced by the training dataset and by the choice of the hyperparameters used in the learning phase. Meinshausen and Bühlmann (2010); Zou et al. (2007); Zou and Hastie (2005). For this reason given the best value λ^* of the hyperparameter selected with the BOO procedure another bootstrap procedure is applied in order to improve the reliability of the feature selection method: n samples are extracted from $\mathcal{D}_n$, the model is built with LAR and λ^* and the features are selected. The bootstrap is repeated several times and features are ranked based on how many times each feature is selected as important by the LAR method (Meinshausen and Bühlmann, 2010).

In this work, the LAR method for feature selection was used in three different kind of models (BBM, N-GBM and A-GBM). Similarly to the case of Brute Force Methods, in the N-GBM case the WBM represents an additional feature (see Table 3).

5.3. Random forest based method

In addition to its use for regression models, the Random Forest (RF) method can also be used to perform a very stable Feature Selection procedure. The procedure can be described as follows: in every tree grown in the forest the error on the out-of-bag must be kept. Then a random permutation of the values of variable j must be performed in the samples of the out-of-bag and the error on the out-of-bag must be kept again. Subtract the error on the untouched out-of-bag data with the error over the permuted out-of-bag samples. The average of this value over all the trees in the forest is the raw importance score for variable j. This approach is inspired by the permutation test (Good, 2013) which is quite used in literature, is computationally inexpensive in the case of Random Forest, and has shown to be quite effective in real wold applications (Deng et al., 2011; White and Liu, 1994). The results for the RF feature selection method are reported for all the models (BBM, N-BBM and A-GBM).

5.4. Results

In Table 13 all the results of the feature selection method are reported. Authors decided to provide just the seven most informative features not to compromise the readability of the tables. From the tables it is possible to draw the following considerations:

- all methods identify the same variables as the most relevant for the model, thus confirming the validity of the modelling procedure. This also allows to trust the reliability of the information contained historical data.
- the BF methods are the most stable, closely followed by the RF methods.
- the WBM is always among the seven most important features for GBMs. This suggests that the N-GBM is able to take into account the

Table 11
Percentage of improvement for the different indexes of performance of the BBM, N-GBM and A-GBM over the WBM built with different approaches LAR for the prediction of Shaft Power, Shaft Torque and Fuel Consumption.

	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM
LAR Shaft Power																		
n_l	MAE [KW]			MAPE [%]			MSE [KW^2]			NMSE			REP [%]			PPMCC		
10	−12	38	**48**	−8	40	**50**	−28	29	**41**	−26	30	**41**	−19	34	**44**	−65	−64	**−64**
20	22	58	**66**	21	58	**65**	45	70	**75**	44	71	**76**	23	58	**64**	10	**11**	11
50	39	68	**72**	34	65	**71**	68	83	**85**	67	82	**85**	39	68	**73**	29	**31**	30
100	57	77	**81**	53	75	**79**	82	90	**92**	82	90	**92**	55	76	**79**	42	**43**	43
200	70	87	**87**	67	85	**87**	89	95	**96**	89	95	**96**	64	84	**86**	47	48	**48**
500	79	90	**91**	78	90	**91**	94	97	**98**	94	97	**98**	74	88	**89**	50	**51**	51
1000	84	93	**94**	83	93	**93**	96	98	**98**	96	98	**98**	78	91	**91**	52	**53**	53
Shaft Torque																		
n_l	MAE [Nm]			MAPE [%]			MSE [N^{2}m^2]			NMSE			REP [%]			PPMCC		
10	−14	39	**49**	−21	34	**46**	−16	38	**48**	−16	37	**47**	−13	39	**49**	−5	−5	**−5**
20	−7	41	**52**	−13	38	**50**	2	46	**56**	1	47	**54**	−3	44	**52**	92	94	**94**
50	36	64	**70**	32	62	**69**	57	77	**80**	57	77	**80**	30	62	**69**	276	**280**	280
100	55	75	**79**	55	75	**79**	77	88	**90**	77	87	**90**	50	73	**77**	321	**324**	324
200	66	85	**86**	66	85	**86**	87	94	**95**	87	94	**95**	62	82	**84**	333	**338**	337
500	77	90	**91**	77	89	**91**	93	97	**97**	93	97	**97**	72	87	**88**	343	346	**347**
1000	83	92	**93**	82	92	**93**	95	98	**98**	95	98	**98**	76	89	**90**	348	**351**	350
Fuel Consumption																		
n_l	MAE $\left[\frac{g}{KWh}\right]$			MAPE [%]			MSE $\left[\frac{g^2}{KW^2h^2}\right]$			NMSE			REP [%]			PPMCC		
10	24	58	**65**	24	58	**65**	37	65	**71**	35	65	**71**	16	54	**60**	−63	**−63**	−63
20	28	62	**67**	31	62	**68**	44	69	**74**	45	69	**75**	21	57	**63**	−40	**−39**	−40
50	47	72	**76**	50	72	**77**	64	81	**84**	64	80	**83**	35	65	**71**	4	5	**5**
100	66	81	**85**	67	82	**85**	85	92	**93**	85	92	**93**	59	78	**82**	39	**40**	40
200	77	90	**91**	76	89	**90**	92	96	**97**	92	96	**97**	71	86	**88**	49	50	**50**
500	83	92	**93**	82	92	**93**	95	98	**98**	95	98	**98**	76	89	**90**	52	**53**	53
1000	86	94	**94**	86	93	**94**	96	98	**98**	96	98	**98**	78	90	**91**	53	**55**	54

information generated by the WBM and use it appropriately, confirming the results of the previous section which underlined the improved performance of GBMs compared to BBMs.

From a physical point of view the results of the feature selection identify the propeller pitch (both setpoint and feedback) and the ship speed (both GPS and LOG) as the most important variables for the prediction, which is what to be expected from this type of ship propulsion system. Propeller speed is not among the most important features, as it is normally kept constant during ship operations and therefore has very limited impact from a modelling perspective. The ship draft (fore and aft) are normally selected as important variables (5th-6th), which also reflects physical expectations from the system as the draft influences both ship resistance and, to a minor extent, propeller performance. As expected the shaft generator power, for this propulsion plants configuration, plays an important role for the prediction. In addition to this, some variables that could be expected to contribute significantly to the overall performance are missing. In particular, wind speed and direction are generally used for estimating the impact of the sea state, but are not included among the five most relevant features by any feature selection method. This suggests that either the sea state has a less significant impact on the ship's fuel consumption compared to what originally expected, or that wind speed and direction are not appropriate predictors for modelling this type of effects, contrarily to what often assumed in relevant literature. One possible additional explanation to the absence of wind speed and direction from the important variables is that the influence of the sea state is already accounted for by the propeller pitch ratio, which is expected to vary as a consequence of both ship speed and ship added resistance. As matter of fact in order to keep constant speed profile, the on board automation system should be designed to change the pitch settings and the fuel consumption rate to take into account time domain variation of boundary conditions such as wind and sea state conditions. Under this assumption the relevant information about added resistance and wind intensity are already included in the propeller pitch ratio. The same conclusions were found by an independent method in Soares and Figari (2009).

6. Trim optimisation

Of all the models proposed in the previous part of this paper, the N-GBM based on RF features the best accuracy properties and best physical plausibility and is therefore used for the trim optimisation problem. For each sample of the dataset the trim is defined as the difference between the draft forward (x_9) and the draft aft (x_{10}), as reported in Table 2. In Fig. 11 the trim distribution during sailing time is depicted. It is possible to note that vessel is normal navigating trimmed by the stern. Nevertheless, the range of feasible trim allowed is in the range observed from the available dataset, hence both positive and negative value. In order to meet the requirements expressed in

Table 12
Percentage of improvement for the different indexes of performance of the BBM, N-GBM and A-GBM over the WBM built with different approaches RF for the prediction of Shaft Power, Shaft Torque and Fuel Consumption.

	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM	BBM	N-GMB	A-GBM
Shaft Power																		
n_l	MAE [KW]			MAPE [%]			MSE [KW^2]			NMSE			REP [%]			PPMCC		
10	−2	**59**	47	1	**60**	50	−14	**52**	41	−14	**54**	42	−5	**56**	45	−65	**−64**	−64
20	34	**73**	65	32	**71**	65	53	**80**	75	52	**80**	75	32	**72**	64	10	11	**11**
50	53	**78**	72	50	**77**	70	75	**88**	85	75	**88**	85	54	**78**	72	30	30	**30**
100	70	84	**85**	66	83	**84**	87	94	**94**	87	94	**94**	67	84	**84**	42	**43**	43
200	80	**92**	89	78	**91**	88	93	**97**	96	93	**97**	96	76	**90**	87	47	48	**48**
500	86	**94**	92	86	**94**	92	96	**98**	98	96	**98**	98	83	**93**	90	49	51	**51**
1000	90	**96**	94	89	**96**	94	98	**99**	99	97	**99**	99	87	**94**	93	52	**53**	53
Shaft Torque																		
n_l	MAE [Nm]			MAPE [%]			MSE [N^2m^2]			NMSE			REP [%]			PPMCC		
10	−2	**58**	48	−5	**57**	46	−0	**58**	49	−1	**57**	48	−3	**59**	47	−5	**−5**	−5
20	5	**62**	51	3	**60**	50	14	**64**	56	15	**64**	55	10	**61**	52	92	**94**	94
50	51	**77**	71	46	**75**	70	67	**84**	80	66	**84**	80	46	**75**	68	276	280	**280**
100	66	**83**	83	67	**83**	83	83	**92**	92	83	**92**	92	63	81	**82**	322	324	**325**
200	77	**91**	88	77	**91**	88	91	**96**	95	91	**96**	95	74	**89**	86	334	337	**337**
500	85	**94**	92	85	**94**	91	95	**98**	97	95	**98**	97	82	**92**	90	342	**348**	347
1000	89	**95**	94	89	**95**	94	97	**99**	98	97	**99**	98	85	**93**	91	346	**352**	351
Fuel Consumption																		
n_l	MAE $\left[\frac{g}{KWh}\right]$			MAPE [%]			MSE $\left[\frac{g^2}{KW^2h^2}\right]$			NMSE			REP [%]			PPMCC		
10	30	**72**	65	33	**72**	66	42	**76**	71	41	**77**	71	23	**70**	61	−63	**−63**	−63
20	39	**74**	68	39	**75**	69	51	**80**	74	53	**79**	75	30	**71**	64	−40	**−40**	−40
50	60	**81**	76	62	**82**	77	72	**87**	83	72	**87**	83	50	**77**	70	4	**5**	5
100	76	**88**	88	75	**88**	87	89	**95**	95	89	**95**	95	70	**85**	85	39	40	**40**
200	84	**94**	92	84	**94**	91	95	**98**	97	95	**98**	97	80	**92**	89	49	50	**50**
500	89	**95**	94	89	**95**	93	97	**99**	98	97	**99**	98	85	**93**	91	51	53	**54**
1000	91	**96**	95	91	**96**	95	97	**99**	98	97	**99**	98	86	**94**	92	53	54	**54**

Section 3.3 the following is considered:

- Variables that are influenced by the trim, such as propeller power and torque, were excluded from the model (see Table 3).
- For each pair of ship speed and displacement, the trim is only allowed to vary in the range observed from the available dataset, extended by δ%. This allows, for every pair, to limit the extrapolation and therefore to ensure additional reliability of the optimisation results. Basically we solved and optimisation problem by searching for the best value of trim, in the range defined by the trim, respect to the fuel consumption. Since we had only one variable to optimise we performed a grid search.

In Table 14 the Fuel Consumption percentage reduction with the trim Optimisation technique is reported for different values of δ. As expected, the optimisation procedure always leads to a reduction in fuel consumption. The improvement that can be achieved via trim optimisation increases when δ is increased, although this tendency seems to stabilise for $\delta > 5\%$.

According to the results of this model, improvements exceeding 2% in fuel consumption can be achieved by applying the model for trim optimisation to the selected vessel. It should be noted that trim optimisation can be performed at near to zero cost on board, since it does not require the installation of any additional equipment.

Future work in this area will include testing trim optimisation system here proposed on a real vessel, in order to check the validity of the model and the performance of the optimisation tool. In particular, we are planning to implement onboard our method on some vessels in order to check the quality of the trim suggestions in real operational conditions on a large time horizon (one or two years).

7. Conclusion

In this paper, the authors have shown how data driven, or black-box, models can outperform state-of-the-art numerical, or white-box, models which exploits the physical knowledge of the system in the task of predicting the fuel consumption of a naval propulsion plant. Based on these models new approaches for modelling the system have been developed, namely the gray-box models, which are able to exploit the advantages of two philosophy: gray-box models are able to obtain the same performances of the black-box but requiring less historical data thanks to the knowledge embedded in the white box models. The proposed methodologies have been tested on real world historical data collected from a real vessel during two years of on board sensors data

Table 13
Feature ranking with the BF, LAR and RF with just the seven most informative features.

	BFM									RBM			RFM		
	RLS			LAR			RF								
Ranking	BBM	N-GBM	A-GBM	BBM	N-GBM	A-GBM	BBM	N-GBM	A-GBM	BBM	N-GBM	A-GBM	BBM	N-GBM	A-GBM
	Shaft rpm														
1^o	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
2^o	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
3^o	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
4^o	16	16	16	16	16	16	16	16	16	21	21	9	16	16	16
5^o	21	WBM	21	21	WBM	21	21	WBM	21	9	WBM	21	21	WBM	21
6^o	9	9	9	9	9	9	9	9	9	16	9	10	9	9	9
7^o	10	10	10	10	10	10	10	10	10	10	10	16	10	10	10
	Shaft Power														
1^o	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
2^o	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
3^o	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
4^o	16	16	16	16	16	16	16	16	16	19	9	16	16	16	16
5^o	9	9	9	9	9	9	9	9	9	9	16	10	9	9	9
6^o	19	WBM	19	19	WBM	19	19	WBM	19	16	10	19	19	WBM	19
7^o	10	10	10	10	10	10	10	10	10	10	WBM	9	10	10	10
	Fuel Consumption														
1^o	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
2^o	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
3^o	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
4^o	21	21	21	21	21	21	21	21	21	21	9	16	21	21	21
5^o	9	9	9	9	9	9	9	9	9	9	16	9	9	9	9
6^o	16	WBM	16	16	WBM	16	16	WBM	16	16	WBM	21	16	WBM	16
7^o	10	10	10	10	10	10	10	10	10	27	27	27	10	10	10

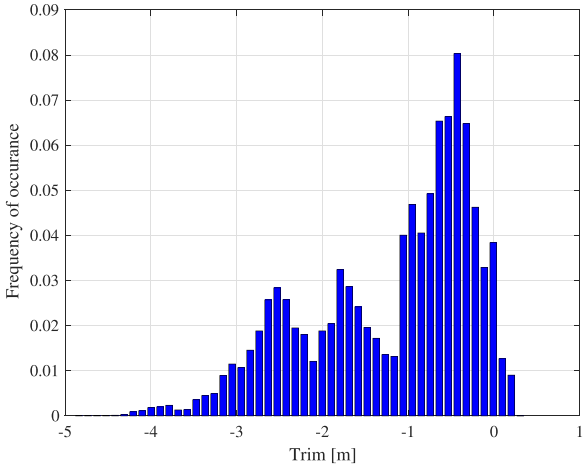

Fig. 11. Trim distribution during sailing time for the selected vessel in the chosen period.

Table 14
Fuel Consumption percentage reduction with the trim Optimisation technique.

δ	% reduction
0%	0.52 ± 0.12
1%	1.45 ± 0.32
2%	1.72 ± 0.51
5%	2.22 ± 0.67
10%	2.30 ± 0.64

acquisitions, and the physical plausibility of the models has been checked through a feature ranking process. Feature ranking allows improving the understanding of black- and gray-box models as for these model physical principles are only partly accounted for. Thanks to the high accuracy and physical plausibility of the developed models, the authors have been able to propose a trim optimisation technique which exploits the predictive power of the proposed models for the online selection of the best configuration of the trim for reducing the fuel consumption. Results have shown to be very promising and they should be further verified by implementing our proposal on the onboard system of a vessel.

Appendix A. Learning algorithms and model selection

In RLS, approximation functions are defined as

$$h(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}), \tag{A.1}$$

where a non-linear mapping $\boldsymbol{\phi}$: $\mathbb{R}^d \rightarrow \mathbb{R}^D$, $D \gg d$, is applied so that non-linearity is pursued while still coping with linear models.

For RLS, Problem (13) is configured as follows. The complexity of the approximation function is measured as

$$C(h) = \|\boldsymbol{w}\|_2^2 \tag{A.2}$$

i.e. the Euclidean norm of the set of weights describing the regressor, which is a quite standard complexity measure in ML (Tikhonov and Arsenin, 1979). Regarding the loss function, the Mean Squared Error (MSE) loss is adopted:

$$\widehat{L}_n(h) = \frac{1}{n}\sum_{i=1}^{n} \ell(h(\boldsymbol{x}_i), y_i) = \frac{1}{n}\sum_{i=1}^{n} [h(\boldsymbol{x}_i) - y_i]^2. \tag{A.3}$$

Consequently, Problem (13) can be reformulated as:

$$\boldsymbol{w}^*: \quad \min_{\boldsymbol{w}} \quad \frac{1}{n}\sum_{i=1}^{n} [\boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) - y_i]^2 + \lambda \|\boldsymbol{w}\|_2^2. \tag{A.4}$$

By exploiting the Representer Theorem (Schölkopf et al., 2001), the solution h^* of the RLS Problem (A.4) can be expressed as a linear combination of the samples projected in the space defined by $\boldsymbol{\phi}$:

$$h^*(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}(\boldsymbol{x}_i)^T \boldsymbol{\phi}(\boldsymbol{x}). \tag{A.5}$$

It is worth underlining that, according to the kernel trick (Scholkopf, 2001; Bolón-Canedo et al., 2015), it is possible to reformulate $h^*(\boldsymbol{x})$ without an explicit knowledge of $\boldsymbol{\phi}$ by using a proper kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x}_i)^T \boldsymbol{\phi}(\boldsymbol{x})$:

$$h^*(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}). \tag{A.6}$$

Of the several kernel functions which can be found in literature (Scholkopf, 2001), the Gaussian kernel is often used as it enables learning every possible function (Keerthi and Lin, 2003; Oneto et al., 2015):

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}, \tag{A.7}$$

where γ is an hyperparameter which regulates the non-linearity of the solution (Oneto et al., 2015) and must be set a priori, analogously to λ. Small values of γ lead the optimisation to converge to simpler functions $h(x)$ (note that for $\gamma = 0$ the model becomes constant, but with a sufficiently small value of γ the model becomes a linear regressor (Oneto et al., 2015; Keerthi and Lin, 2003)), while high values of γ allow higher complexity of $h(x)$.

Finally, the RLS Problem (A.4) can be reformulated by exploiting kernels:

$$\boldsymbol{\alpha}^*: \quad \min_{\boldsymbol{\alpha}} \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n} \alpha_j K(\boldsymbol{x}_j, \boldsymbol{x}_i) - y_i\right]^2 + \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j K(\boldsymbol{x}_j, \boldsymbol{x}_i). \tag{A.8}$$

Given $\boldsymbol{y} = [y_1, \ldots, y_n]^T$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$, the matrix K such that $K_{i,j} = K_{ji} = K(\boldsymbol{x}_j, \boldsymbol{x}_i)$, and the Identity matrix $I \in \mathbb{R}^{n \times n}$, a matrix-based formulation of Problem (A.8) can be obtained:

$$\boldsymbol{\alpha}^*: \quad \min_{\boldsymbol{\alpha}} \quad \frac{1}{n}\|K\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2 + \lambda \boldsymbol{\alpha}^T K \boldsymbol{\alpha} \tag{A.9}$$

By setting the derivative with respect to $\boldsymbol{\alpha}$ equal to zero, $\boldsymbol{\alpha}$ can be found by solving the following linear system:

$$(K + n\lambda I)\boldsymbol{\alpha}^* = \boldsymbol{y}. \tag{A.10}$$

Effective solvers have been developed throughout the years, allowing to efficiently solve the problem of Eq. (A.10) even when very large sets of training data are available (Young, 2003).

In LAR, instead, approximation functions are defined as

$$h(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b, \tag{A.11}$$

which are linear functions in the original space $\mathbb{R}^d$.

For LAR, Problem (13) is configured as follows. The complexity of the approximation function is measured as

$$C(h) = \| \boldsymbol{w} \|_1 \tag{A.12}$$

i.e. the Manhattan norm of the set of weights describing the regressor (Tibshirani, 1996).

Regarding the loss function, the Mean Squared Error (MSE) loss is again adopted. Consequently, Problem (13) can be reformulated as:

$$\boldsymbol{w}^*: \quad \min_{\boldsymbol{w}} \quad \frac{1}{n}\sum_{i=1}^{n} [\boldsymbol{w}^T \boldsymbol{x} - y_i]^2 + \lambda \| \boldsymbol{w} \|_1. \tag{A.13}$$

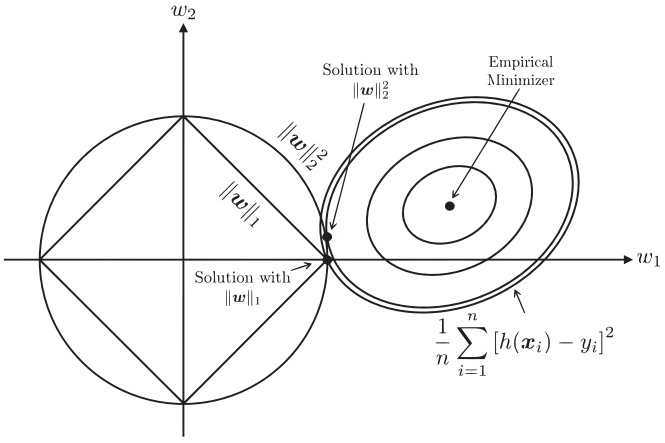

Fig. A12. Manhattan Norm VS Euclidean Norm.

As depicted in Fig. A12 the Manhattan norm is quite different from the Euclidean one since it allows increasing the sparsity of the solution. In other words the solution will tend to fall on the edge of the square, forcing some weights of $\boldsymbol{w}$ to be zero. Hence, the Manhattan norm allows both regularising the function and discarding features that are not sufficiently relevant to the model. This property is particularly useful in the feature selection process (Tibshirani, 1996).

Two main approaches can be used to compute the solutions of Problem (A.13): the LARS algorithms (Efron et al., 2004) and the pathwise coordinate descent (Friedman et al., 2010). In this paper, the LARS algorithm is exploited because of its straight-forward implementation (Efron et al., 2004).

The performance of RLS (or LAR) models depends on the quality of the hyperparameters tuning procedure. As highlighted while presenting this approach, the parameters $\boldsymbol{\alpha}^*$, $\hat{\boldsymbol{\alpha}}^*$, and $\check{\boldsymbol{\alpha}}^*$ (or $\boldsymbol{w}$) result from an optimisation procedure which requires the a priori setting of the tuples of hyperparameters (λ, γ) (or λ). The phase in which the problem of selecting the best value of the hyperparameter is addressed is called model selection phase (Anguita et al., 2011). The most effective model selection approaches consist in performing an exhaustive hyperparameters grid search: the optimisation problem for RLS (or LAR) is solved several times for different values of γ and λ, and the best pair of hyperparameters is chosen according to some criteria.

For the optimal choice of the hyperparameters γ and λ, in this paper the authors exploit the Bootstrap technique (BOO) (Anguita et al., 2012, 2000; Efron and Tibshirani, 1993). This technique represents an improvement of the well–known k–Fold Cross Validation (KCV) (Anguita et al., 2009, 2012; Kohavi, 1995) where the original dataset is split into k independent subsets (namely, the folds), each one consisting of n/k samples: $(k-1)$ parts are used, in turn, as a training set, and the remaining fold is exploited as a validation set. The procedure is iterated k times.

The standard Bootstrap (Anguita et al., 2012) method is a pure resampling technique: at each j-th step, a training set $\mathcal{D}_{\mathrm{TR}}^j$, with the same cardinality of the original one, is built by sampling the patterns in $\mathcal{D}_n$ with replacement. The remaining data $\mathcal{D}_{\mathrm{VL}}^j$, which consists, on average, of approximately 36.8% of the original dataset, are used as validation set. The procedure is then repeated several times $N_B \in [1, \binom{2n-1}{n}]$ in order to obtain statistically sound results (Efron and Tibshirani, 1993).

According to the Bootstrap technique, at each j-th step the available dataset $\mathcal{D}_n$ is split in two sets:

- A training Set: $\mathcal{D}_{\mathrm{TR}}^j$
- A validation Set: $\mathcal{D}_{\mathrm{VL}}^j$

In order to select the best pair of hyperparameters (λ^*, γ^*) (or λ^*) among all the available ones $\mathcal{G} = \{(\lambda_1, \gamma_1), (\lambda_2, \gamma_2), \cdots\}$ (or $\mathcal{G} = \{\lambda_1, \lambda_2, \cdots\}$) for the algorithm for RLS (or LAR) the following optimisation procedure is required:

- for each $\mathcal{D}_{\mathrm{TR}}^j$ and for each tuple (λ_i, γ_i) (or λ_i) with $i \in \{1, 2, \cdots\}$ the optimisation problem of Eq. (A.10) (or Eq. (A.13)) is solved and the solution $h_j^i(\boldsymbol{x})$ is found
- using the validation set $\mathcal{D}_{\mathrm{VL}}^j$ for searching the (λ^*, γ^*) (or λ^*) $\in \mathcal{G}$

$$\begin{matrix}(\lambda^*, \gamma^*) \\ \text{or } \lambda^*\end{matrix} = \arg \min_{\substack{\{(\lambda_1,\gamma_1),(\lambda_2,\gamma_2),\ldots,(\lambda_i,\gamma_i),\cdots\} \\ \text{or } \{\lambda_1,\lambda_2,\ldots,\lambda_i,\cdots\}}} \frac{1}{N_B}\sum_{j=1}^{N_B}\frac{1}{|\mathcal{D}_{\mathrm{VL}}^j|}\sum_{(\boldsymbol{x},y)\in\mathcal{D}_{\mathrm{VL}}^j}[h(\boldsymbol{x})-y]^2. \tag{A.14}$$

Once the best tuple is found, the final model is trained on the whole set $\mathcal{D}_n$ by running the learning procedure with the best values of the hyperparameters (Anguita et al., 2009).

Another learning algorithm tested for building the BBM is the Random Forest (RF) (Breiman, 2001; Fernández-Delgado et al., 2014). Random Forests grows many regression trees (Quinlan, 1987). To classify a new object from an input vector each of the trees of the forest is applied to the vector. Each tree gives an output and the forest chooses the mode of the votes (over all the trees in the forest). Each single tree is grown by following this procedure: (I) n samples are sampled (with replacement) from the original $\mathcal{D}_n$, (II) $d' \ll d$ features are chosen randomly out of the d and the best split on these d' is used to split the node, (III) each tree is grown to the largest possible extent, without any pruning. In the original paper (Breiman, 2001) it was shown that the forest error rate depends on two elements: the correlation between any couples of trees in the forest (increasing the correlation increases the forest error rate) and the strength of each individual tree in the forest (reducing the error rate of each tree decreases the forest error rate). Reducing d' reduces both the correlation and the strength. Increasing it increases both. Somewhere in between is an optimal range

of d' - usually quite wide so this is not usually considered as an hyperparameter. Note that, since we used a bootstrap procedure by sampling n samples with replacement from the original $\mathcal{D}_n$, we can use the error on the remaining part of the data (this is called out-of-bag error) to chose the best d'.

References

Altosole, M., Campora, U., Martelli, M., Figari, M., 2014. Performance decay analysis of a marine gas turbine propulsion system. J. Ship Res. 58 (3).

Altosole, M., Figari, M., Ferrari, A., Bruzzone, D., Vernengo, G., 2016. Experimental and numerical investigation of draught and trim effects on the energy efficiency of a displacement mono-hull. In: Proceedings of the International Ocean and Polar Engineering Conference.

Anguita, D., Boni, A., Ridella, S., 2000. Evaluating the generalization ability of support vector machines through the bootstrap. Neural Process. Lett. 11 (1), 51–58.

Anguita, D., Ghio, A., Oneto, L., Ridella, S., 2012. In-sample and out-of-sample model selection and error estimation for support vector machines. IEEE Trans. Neural Netw. Learn. Syst. 23 (9), 1390–1406.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S., 2012. The k in k-fold cross validation. In: Proceedings of the European Symposium on Artificial Neural Networks.

Anguita, D., Ghio, A., Oneto, L., Ridella, S., 2011. In-sample model selection for support vector machines. In: Proceedings of the International Joint Conference on Neural Networks.

Anguita, D., Ghio, A., Oneto, L., Ridella, S., 2011. Selecting the hypothesis space for improving the generalization ability of support vector machines. In: Proceedings of the International Joint Conference on Neural Networks.

Anguita, D., Ghio, A., Ridella, S., Sterpi, D., 2009. K-fold cross validation for error rate estimate in support vector machines. In: Proceedings of the International Conference on Data Mining.

Anguita, D., Ridella, G.A., S., Sterpi, D., 2009. K-fold cross validation for error rate estimate in support vector machines. In: Proceedings of the International Conference on Data Mining.

Baldi, F., Johnson, H., Gabrielii, C., Andersson, K., 2014. Energy and exergy analysis of ship energy systems-the case study of a chemical tanker. In: Proceedings of the 27th ECOS, International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems.

Basin., D.W.T.H., Todd. F.H., 1964. Series 60 Methodical Experiments With Models of Single-screw Merchant Ships. Washington.

Bolón-Canedo V., Donini, M., Aiolli, F., 2015. Feature and kernel learning. In: Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Chang, Y.W., Lin, C.J., 2008. Feature ranking using linear svm. Causa. Predict. Chall. Chall. Mach. Learn. 2, 47.

Coraddu, A., Figari, M., Ghio, A., Oneto, L., Savio, S., 2014. A sustainability analytics matlab tool to predict ship energy consumption. In: International Conference on Computer Applications and Information Technology in the Maritime Industries.

Coraddu, A., Figari, M., Savio, S., Villa, D., Orlandi, A., 2013. Integration of seakeeping and powering computational techniques with meteo-marine forecasting data for in-service ship energy assessment. In: Developments in Maritime Transportation and Exploitation of Sea Resources.

Coraddu, A., Gaggero, S., Figari, M., Villa, D., 2011. A new approach in engine-propeller matching. In: Sustainable Maritime Transportation and Exploitation of Sea Resources, vol. 1, CRC Press – Taylor and Francis Group, pp. 631–637.

Coraddu, A., Gualeni, P., Villa, D., 2011. Investigation about wave profile effects on ship sability, In: IMAM 2011 International Maritime Association of the Mediterranean – Sustainable Maritime Transportation and Exploration of the Sea Resources. vol. 1, pp. 143–149.

Coraddu, A., Oneto, L., Baldi, F., Anguita, D., 2015. A ship efficiency forecast based on sensors data collection: improving numerical models through data analytics. In: OCEANS.

Coraddu, A., Oneto, L., Ghio, A., Savio, S., Anguita, D., Figari, M., 2014. Machine learning approaches for improving condition-based maintenance of naval propulsion plants. In: Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment. ⟨http://dx.doi.org/10.1177/1475090214540874⟩.

De Mol, C., De Vito, E., Rosasco, L., 2009. Elastic-net regularization in learning theory. J. Complex. 25 (2), 201–230.

Decherchi, S., Ridella, S., Zunino, R., Gastaldo, P., Anguita, D., 2010. Using unsupervised analysis to constrain generalization bounds for support vector classifiers. IEEE Trans. Neural Netw. 21 (3), 424–438.

Deng, H., Runger, G., Tuv, E., 2011. Bias of importance measures for multi-valued attributes and solutions. Artificial Neural Networks and Machine Learning–ICANN 2011. pp. 293–300.

Efron, B., Tibshirani, R., 1993. An Introduction to The Bootstrap. CRC Press, London.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32 (2), 407–499.

Elattar, E.E., Goulermas, J., Wu, Q.H., 2010. Electric load forecasting based on locally weighted support vector regression. IEEE Trans. Syst. Man, Cybern. Part C: Appl. Rev. 40 (4), 438–447.

Evans, J.R., Lindner, C.H., 2012. Business analytics: the next frontier for decision sciences. Decis. Line 43 (2), 4–6.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15 (1), 3133–3181.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer Series in Statistics Springer, Berlin.

Fumeo, E., Oneto, L., Anguita, D., 2015. Condition based maintenance in railway transportation systems based on big data streaming analysis. In: INNS Conference on Big Data.

Gaggero, S., Villa, D., Brizzolara, S., 2010. Rans and panel method for unsteady flow propeller analysis. J. Hydrodyn. Ser. B 22 (5), 564–569.

Ghelardoni, L., Ghio, A., Anguita, D., 2013. Energy load forecasting using empirical mode decomposition and support vector regression. IEEE Trans. Smart Grid 4 (1), 549–556.

Gieseke, F., Polsterer, K.L., Oancea, C.E., Igel, C., 2014. Speedy greedy feature selection: Better redshift estimation via massive parallelism. In: Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.

Good, P., 2013. Permutation Tests: a Practical Guide to Resampling Methods For Testing Hypotheses. Springer Science & Business Media, Berlin.

Guldhammer, H., Harvald, S.A., 1974. Ship Resistance: Effect of Form and Principal Dimensions. Akademisk Forlag, Copenhagen.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Györfi, L., 2002. A Distribution-Free Theory of Nonparametric Regression. Springer, Berlin.

Hochkirch, K., Mallol, B., 2013. On the importance of fullscale cfd simulations for ships. In: Proceedings of the International Conference on Computer Applications and Information Technology in the Maritime Industries.

Holtrop, J., 1984. A statistical re-analysis of resistance and propulsion data. Int. Shipbuild. Prog. 31 (363), 272–276.

Hong, S.J., 1997. Use of contextual information for feature ranking and discretization. IEEE Trans. Knowl. Data Eng. 9 (5), 718–730.

Howison, S., 2005. Practical Applied Mathematics: Modelling, Analysis, Approximation. Cambridge University Press, Cambridge.

Jardine, A.K., Lin, D., Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. 20 (7), 1483–1510.

Jiawei, H., Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann.

Josephson, J.R., Josephson, S.G., 1996. Abductive Inference: Computation, Philosophy, Technology. Cambridge University Press, Cambridge.

Keerthi, S.S., Lin, C.-J., 2003. Asymptotic behaviors of support vector machines with gaussian kernel. Neural Comput. 15 (7), 1667–1689.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intell. 97 (1), 273–324.

Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the International Joint Conference on Artificial Intelligence.

Lee, W.S., Bartlett, P.L., Williamson, R.C., 1998. The importance of convexity in learning with squared loss. IEEE Trans. Inf. Theory 44 (5), 1974–1980.

Lee, J., Yoo, S., Choi, S., Kim, H., Hong, C., Seo, J., 2014. Development and application of trim optimization and parametric study using an evaluation system (solution) based on the rans for improvement of eeoi. In: Proceedings of the International Conference on Ocean, Offshore and Arctic Engineering.

Leifsson, L., Saevarsdottir, H., Sigurdsson, S., Vesteinsson, A., 2008. Grey-box modeling of an ocean vessel for operational optimization. Simul. Model. Pract. Theory 16, 923–932.

Lewis, E.V., 1988. Principles of naval architecture. Soc. Nav. Archit..

Lützen, M., Kristensen, H., 2012. A model for prediction of propulsion power and emissions – tankers and bulk carriers. In: Proceedings of the World Maritime Technology Conference.

MacKay, D.J.C., 2003. Information Theory, Inference and Learning Algorithms. Cambridge university Press, Cambridge.

Martelli, M., Viviani, M., Altosole, M., Figari, M., Vignolo, S., 2014. Numerical modelling of propulsion, control and ship motions in 6 degrees of freedom. In: Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment.

Meinshausen, N., Bühlmann, P., 2010. Stability selection. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 72 (4), 417–473.

Moustafa, M.M., Yehia, W., Hussein, A.W., 2015. Energy efficient operation of bulk carriers by trim optimization. In: Proceedings of theInternational Conference on Ships and Shipping Research.

Ng, A.Y., 2004. Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In: Proceedings of the International Conference on Machine Learning.

Oneto, L., Ghio, A., Ridella, S., Anguita, D., 2015. Support vector machines and strictly positive definite kernel: the regularization hyperparameter is more important than the kernel hyperparameters. In: Proceedings of the International Joint Conference on Neural Networks.

Palmé, T., Breuhaus, P., Assadi, M., Klein, A., Kim, M., 2011. New alstom monitoring tools leveraging artificial neural network technologies. In: Turbo Expo: Turbine Technical Conference and Exposition.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Petersen, J.P., Winther, O., Jacobsen, D.J., 2012a. A machine-learning approach to predict main energy consumption under realistic operational conditions. Ship Technol. Res. 59 (1), 64–72.
Petersen, J.P., Jacobsen, D.J., Winther, O., 2012b. Statistical modelling for ship propulsion efficiency. J. Mar. Sci. Technol. 17 (1), 30–39.
Quinlan, J.R., 1987. Simplifying decision trees. Int. J. Man-Mach. Stud. 27 (3), 221–234.
Scholkopf, B., 2001. The kernel trick for distances. In: Neural Information Processing Systems.
Schölkopf, B., Herbrich, R., Smola, A.J., 2001. A generalized representer theorem. In: Computational Learning Theory.
Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge.
Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2013. A sparse-group lasso. J. Comput. Graph. Stat. 22 (2), 231–245.
Soares, C.G., Figari, M., 2009. Fuel consumption and exhaust emissions reduction by dynamic propeller pitch control. In: Analysis and Design of Marine Structures, CRC Press.
Sugumaran, V., Muralidharan, V., Ramachandran, K., 2007. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mech. Syst. Signal Process. 21 (2), 930–942.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.), 267–288.
Tikhonov, A., Arsenin, V.Y., 1979. Methods for Solving Ill-Posed Problems. Nauka, Moscow.
Vapnik, V.N., 1998a. Statistical Learning Theory. Wiley-Interscience, New York.
Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.
White, A.P., Liu, W.Z., 1994. Technical note: bias in information-based measures in decision tree induction. Mach. Learn. 15 (3), 321–329.
Widodo, A., Yang, B.-S., 2007. Support vector machine in machine condition monitoring and fault diagnosis. Mech. Syst. Signal Process. 21 (6), 2560–2574.
Yoon, H., Yang, K., Shahabi, C., 2005. Feature subset selection and feature ranking for multivariate time series. IEEE Trans. Knowl. Data Eng. 17 (9), 1186–1198.
Young, D.M., 2003. Iterative Solution of Large Linear Systems, DoverPublications.com.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 67 (2), 301–320.
Zou, H., Hastie, T., Tibshirani, R., 2007. On the degrees of freedom of the lasso. Ann. Stat. 35 (5), 2173–2192.