Marine dual fuel engines monitoring in the wild through weakly supervised data analytics

Andrea Coraddu [a],[*], Luca Oneto [b], Davide Ilardi [b], Sokratis Stoumpos [a], Gerasimos Theotokatos [c]

[a] *University of Strathclyde, 100 Montrose Street, Glasgow G4 0LZ, UK*
[b] *DIBRIS - University of Genova, Via Opera Pia 11, 16145, Genova, Italy*
[c] *Maritime Safety Research Centre, Department of Naval Architecture, Ocean & Marine Engineering, University of Strathclyde, 100 Montrose Street, Glasgow G4 0LZ, UK*

ARTICLE INFO

ABSTRACT

Background: Maritime transportation accounts for around 80% of the world freight movements, remarkably contributing to the global environmental footprint. Dual fuel engines, running on both gaseous and liquid fuels, represent a viable way toward the reduction of emissions at the cost of additional complexity in monitoring activities.

Motivation: Data-driven methods represent the frontier in research and in maritime industrial applications, and they usually require a large amount of labelled data, i.e., sensor measurements plus the associated engine status usually annotated by human operators, which are costly and seldomly available in the wild. Unlabelled samples, instead, are commonly, cheaply, and readily available.

Hypothesis: The enabling technology for data-driven methods is the availability of a network of sensors and an automation system able to capture and store the associated stream of data.

Methods: In this paper, we design and propose multiple alternatives toward the weakly supervised marine dual fuel engines data-driven monitoring. To this aim, we will rely on a Digital Twin of the dual fuel engine or on novelty detection algorithms and we will compare them against state-of-the-art fully supervised approaches.

Results: Results on data generated from a real-data validated simulator of a marine dual fuel engine demonstrate that the proposed weakly supervised monitoring approaches lead to a negligible loss in accuracy compared to costly and often unfeasible fully supervised ones supporting the validity of the proposal for its application in the wild.

Conclusion: The main outcome is a guideline for selecting the best data-driven dual fuel engine monitoring method according to the available data.

1. Introduction

Maritime transportation accounts for around 80% of global freight movements (Mangan, 2017). With very few exceptions, vessels are powered by internal combustion engines, burning conventional fossil fuels, producing a large amounts of undesired greenhouse and non-greenhouse emissions (Reitz et al., 2020). In fact, carbon dioxide, carbon monoxide, sulphur oxides, nitrogen oxides, methane, and particulate matter (including black carbon) negatively affect the climate, the environment, and the public health (Manisalidis et al., 2020).

The use of alternative fuels including natural gas, methanol, and biofuels has been proposed as a viable way toward the improvement of the environmental sustainability of the maritime transportation (Hansson et al., 2019; Zuo et al., 2020a; Jiaqiang et al., 2019). In particular, the use of Liquefied Natural Gas (LNG) as a fuel proved to be the most viable solution due to the lower LNG fuel price levels compared to other fossil fuels (Bae and Kim, 2017), the rapid development of the global LNG infrastructure (Thomson et al., 2015), as well as the clean nature of lean combustion, which leads to the reduction of the nitrogen oxides

---

due to the low carbon to hydrogen ratio whilst almost eliminating most of the particulate matter and sulphur oxide emissions (Hansson et al., 2019; Louis, 2001; Zuo et al., 2020b; Qian et al., 2020).

The economic and environmental benefits of using LNG led the marine engine manufacturers to the development of Dual Fuel (DF) versions of both two-stroke and four-stroke Diesel Engines (DEs) (Pavlenko et al., 2020; Zhong et al., 2020) as well as retrofitting kits for converting existing diesel engines (DE) to DF engines[1,2] Moreover, a pure natural gas fuelled vessel requires at least 48% more storage capacity (Boretti, 2019). Nevertheless, an obvious drawback of DF engines is the additional technological complexity of the engine fuel system as well as the monitoring, control and safety systems (Boretti, 2019; Wang et al., 2016) to keep their performance always at a desired level. In fact, the degradation of marine engines performance decreases their operational efficiency, leading to a higher fuel consumption and consequently to an increase in greenhouse emissions. For this reason, the implementation of efficient and effective monitoring strategies is of paramount importance to ensure availability, reliability, cost, and environmental sustainability (Gratsos et al., 2009; Lloyd and Cackette, 2001; Xu et al., 2002).

Marine engine manufacturers already provide turnkey monitoring solutions for their DEs. For example, MaK DICARE (CAT, 2019) remote engine monitoring system provides condition-specific maintenance suggestions comparing in real time the engine condition to the desired state and suggesting maintenance actions. Another example is MAN Computer Controlled Surveillance (MAN, 2019), a diagnostic tool for monitoring and storing DEs performance data, and trends dedicated to assisting users in evaluating the machinery status and performance. These systems are based on the knowledge of permissible operating engine parameters and actions are triggered when the monitored parameters exceed their boundaries. The final decision on the actions to undertake is usually left to the operators experience and knowledge, and this needs to be addressed for enabling highly automated or autonomous systems operation. In fact, operators need to be carefully trained, their decisions are biased by their experience, too many degrees of freedom are left to their judgement, their ability to exploit the automation data is limited, and the monitoring process may be stressful and time consuming.

State-of-the-art methods try to overcome the limitations of exploiting the human in the loop for monitoring activities by exploiting instead numerical methods (Kowalski et al., 2017; Cipollini et al., 2018a,b). For this purpose, a gold-standard solution is to compare the engine behaviour in dynamic conditions with the normal (expected) behaviour provided by an accurate Digital Twin Grimmelius et al. (2007). This solution enables the identification of unexpected behaviour and to establish trends in temporal performance variation. Numerical methods play a central role in developing an accurate Digital Twin of the engine for the prediction of key performance parameters. In particular, engine modelling has been performed by employing commercial or custom-made tools based on first principles and thermo-physical processes fundamentals. A number of engine models of varying complexity are reported in the pertinent literature (Xiang et al., 2019; Baldi et al., 2015; Reitz and Rutland, 1995). Detailed modelling approaches (of the zero-dimensional to three-dimensional type) result in computationally demanding simulations and consequently are unsuitable for real-time engine monitoring applications (Stoumpos et al., 2018, 2020).

To develop models suitable for effective monitoring in real operational conditions, two main alternatives exist. The first one is to exploit approximate but computationally efficient first-principle models of the mean value type (Geertsma et al., 2017, 2018) or the combined mean

value/0D type (Baldi et al., 2015), whilst the second one is to exploit the historical data acquired by the modern automation systems to build accurate data-driven models. The first approach results in faster but often inaccurate predictions, which limits their effectiveness for engine monitoring in real time operation (Geertsma et al., 2018). The second approach represents the frontier in both research and industrial application and is highly dependent on the availability of an adequate amount of historical data (Talaat et al., 2018; Cipollini et al., 2018b). Their enabling technology is then the availability of a network of sensors and an automation system able to capture and store the associated stream of data, which are nowadays readily available. This data, which is often called unlabelled as there is usually no associated annotation about the status of the engines, are commonly, cheaply, and readily available (Munim et al., 2020). In fact, these annotations are costly and rarely available in the wild since they require the supervision of the engine performed by an human operator (Nixon et al., 2018). In some cases, this labelling activity requires to reduce the operation or to eventually stop the vessel or to maintain the engine. Consequently, unlabelled data is available in large quantities for a large period of time with a very high frequency and just a very small amount of these data are actually labelled during planned maintenance (every few years) or during exceptional disruptions (few time in the life of a vessel).

In this respect, this study aims at designing and proposing multiple alternatives for the reduction of the use of labelled data toward a weekly supervised monitoring for marine DF engines targeting to reduce as much as possible, the necessity of labelled data at least to a realistic level which is realistic to retrieve in the wild. Furthermore, this study proves that the preceding proposal does not compromise the modelling accuracy below a level that prevents their use in real operations.

The approach employed in this study includes the following three steps: (a) the Fully Supervised Performance estimation; (b) the Fully Supervised Health Status Estimation, and; (c) the Weakly Supervised Health Status Estimation. The Fully Supervised Performance Estimation step includes the design of a Digital Twin, exploiting state-of-the-art supervised data-driven methods for enabling the prediction of the engine performance and emissions parameters based on the control variables (e.g. engine load and engine speed), in healthy engine conditions. This step actually does not employ labelled data; instead it employs the acquired data from engine operation under healthy conditions. The Fully Supervised Health Status Estimation step focuses on developing models capable of classifying the status of the engines as healthy or faulty and it is accomplished by employing two approaches. The first one employs the Digital Twin developed in the first step to estimate the deviation (drift) of the parameters of the actual engine operation (based on the acquired data) from the respective Digital Twin predicted parameters. The second one exploits state-of-the-art supervised data-driven methods to classify the status of the investigated engine based on the control and performance parameters. This step requires labelled data with the engine under healthy and faulty conditions. The Weakly Supervised Health Status Estimation step focuses on reducing the amount of labelled data required to build the models developed in the second step by employing two approaches. The first one focuses on the estimation of the engine performance parameters variation from the respective parameters calculated by the Digital Twin by employing a limited amount of labelled data for tuning the drift detection model. The second one, instead, will exploit state-of-the-art unsupervised data-driven methods to detect abnormal conditions (anomalies) of the investigated engine by employing as input the considered control and performance parameters. The weakly supervised health status estimation step employs the models trained just with data acquired under the engine healthy conditions from the engine monitoring system. These models are subsequently fine tuned with a very small amount of labelled data. Fig. 1 depicts our proposal with a simple graphical representation.

Based on the preceding methodology, this study contributes to the better understanding of the effects of using, in multiple methods,

---

[1] https://www.man-es.com/discover/pioneering-retrofit-to-reduce-diesel-emissions.

[2] https://www.wartsila.com/media/news/22-10-2019-wartsila-retrofit-will-reduce-environmental-impact-of-ferry-operating-in-ecologically-sensitive-waters-2555695.
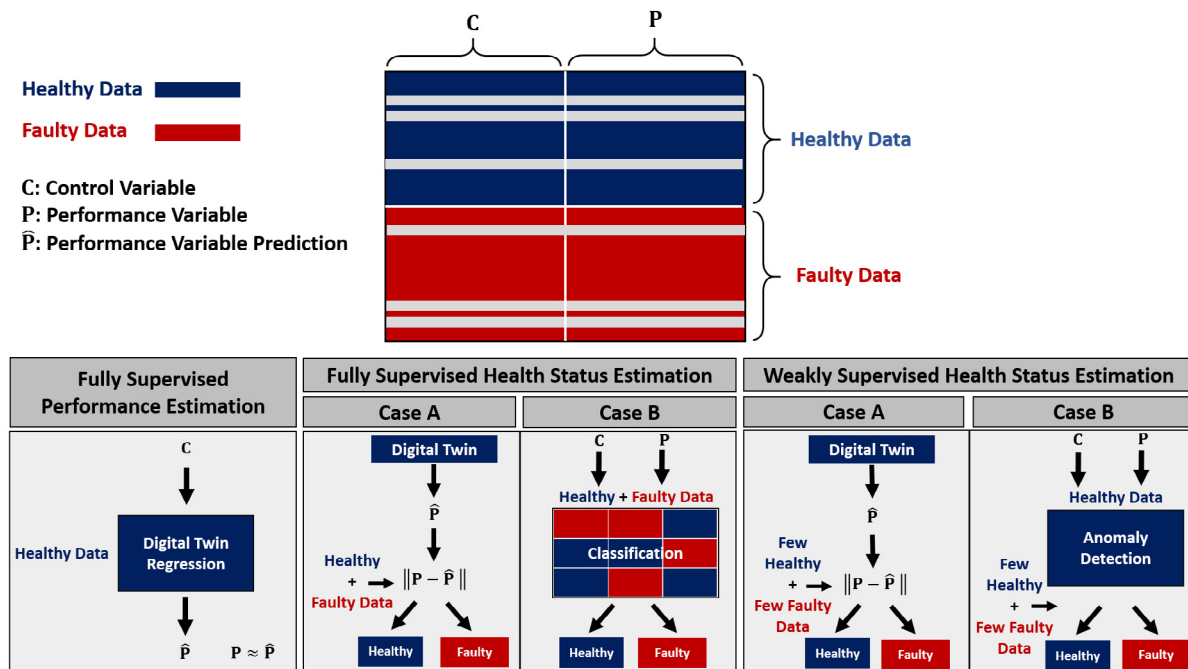
Fig. 1. Our proposal in a Nutshell.

labelled and unlabelled data as well as to quantify the methods accuracy deterioration in cases when the available labelled data is limited. Furthermore, it is demonstrated whether it is possible to monitor the investigated engine status by employing a weakly supervised method with a realistic amount of data.

It must be noted that a large amount of labelled data acquired from marine DE and DF engines pertaining to faulty conditions are not currently available in the literature. To overcome this limitation, this study employs data generated from a validated simulator of a marine DF engine capable of simulating both healthy and faulty conditions (Stoumpos et al., 2020). This study demonstrate that the proposed weakly supervised monitoring approaches lead to a negligible deterioration of the prediction accuracy compared with the costly and often unfeasible fully supervised ones, supporting the validity of the proposal for its application in the wild.

The rest of the paper is organised as follows. Section 2 provides an overview of the state-of-the-art on engine modelling with a particular focus on maritime applications. Section 3 introduces the engine simulator employed to generate the required data. Section 4 describes how the dataset used in this study has been generated by using the simulator described in Section 3. Section 5 presents our proposal. Section 6 tests and demonstrates its validity employing the data described in Section 4. Section 7 summarises the main findings of this study.

2. Related work

The marine DEs and DF engines design, development, optimisation, and monitoring procedures are nowadays increasingly based on mathematical modelling, numerical simulations, and data-driven models, rather than on experiments and prototyping (Lebedevas et al., 2020). He and Rutland (2004), for example, developed a general DEs simulation tool with a small computer resource footprint for engine design based on Artificial Neural Networks.

Karlsson et al. (2010) and Stewart and Borrelli (2008) point out the necessity of sophisticated control systems based on numerical and data-driven models to respect the tightened legislation on emissions and to provide a fast response to the operators to guarantee safe operation and low fuel consumption. Hsieh and Wang (2009), Herceg et al. (2006) and Zhao et al. (2013) developed a new Model Predictive Control

method showing its effectiveness at a cost of an increased computational complexity. Brzozowska et al. (2005), instead, identified six independent engine control variables to approximate the measurement results of the toxicity of exhaust gas compounds.

A considerable number of the past and current research studies focuses on DEs behaviour (Wang et al., 2020; Wakode and Kanase-Patil, 2017; Syed et al., 2017) and very few studies have been published dealing with marine DF engines (e.g. Stoumpos et al. (2020)), on which this study is primarily focused on. Some DF engines monitoring approaches have been developed for applications of the automotive industry (Zhai and Yu, 2009; Shayler et al., 2000; Song et al., 2018; Ahmed et al., 2015).

Whilst a number of studies have attempted to develop models for monitoring the DEs performance and emissions (Shin et al., 2020; Cai et al., 2017; Li et al., 2012), the majority of these works are conducted in a simplified scenario, since they are based upon data collected from experiments produced under controlled laboratory conditions when usually the DE is not subjected to anomaly behaviour or faulty conditions (Özener et al., 2013; Liu et al., 2018; Turkson et al., 2016). Moreover, very few of these studies covered the interaction of the engine performance monitoring with the maintenance and failures (Muchiri et al., 2014). In fact, apart from monitoring purposes, it is becoming increasingly important to develop fault detection systems capable of detecting and preventing the system's failures to reduce maintenance costs and prevent accidents (Xi et al., 2018).

Data-driven approaches represent an effective choice in this context, since they have proven to increase the reliability and decrease the probability of producing false alarms (Ahmed et al., 2015). Ahmed et al. (2015), for instance, proposed a combustion-related fault detection method based on the calculation of the instantaneous angular speed of a diesel engine. Li et al. (2012) developed a fault diagnosis technique for marine DE utilising the information fusion of vibrations and wear particle analyses (which are typically studied separately) by employing a new Independent Component Analysis to identify the engine's vibration source signals collected from multi channel sensors. This study also employs a neural network for the integration of the features extracted to detect faults in a supervised learning manner. Zabihi-Hesari et al. (2019) discussed the importance of vibration monitoring for detecting machine anomalies presenting a novel fault detection method based on
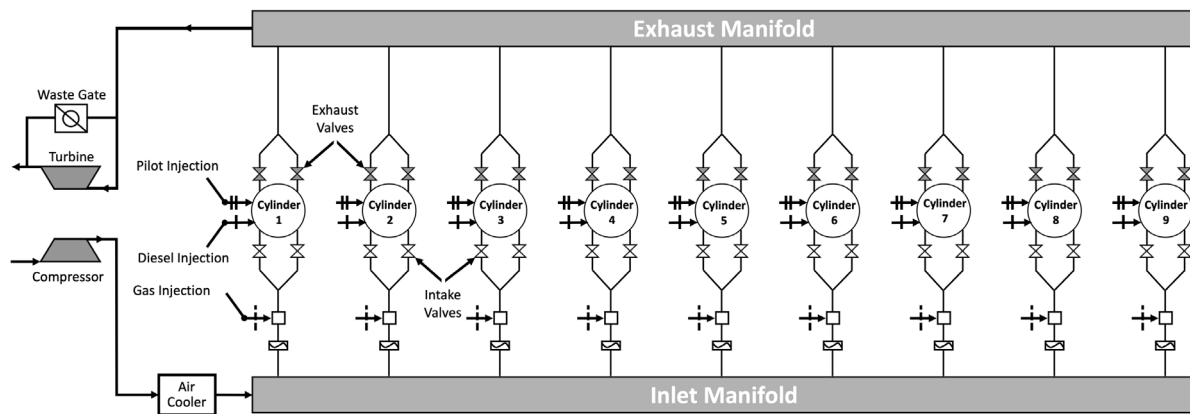
Fig. 2. Wärtsilä 9L50DF Engine Layout.

Table 1
Wärtsilä 9L50DF engine characteristics.

Engine characteristics	Unit	Value
Model	–	9L50DF
MCR Power	[kW]	8775
MCR Speed	[rpm]	514
No. of cylinders		9
BMEP at MCR	[bar]	20
Bore / Stroke	[mm]	500 / 580
Brake-specific fuel consumption (BSFC) at MCR (Diesel mode)	[g/kWh]	190
Brake-specific energy consumption (BSEC) at MCR (Gas mode)	[kJ/kWh]	7300
IMO compliance at: diesel mode / gas mode	–	TIER II /TIER III
Turbocharger units	–	1

the power spectral density of vibration signal to differentiate between faulty and normal DE status. Subsequently, an artificial neural network was used to identify the fault location and the sensor placement. Cai et al. (2017) proposed another fault detection approach for overcoming the state-of-the-art DE fault prediction methods limitations. This study employed a data-driven approach based on support vector machine along with rule mining algorithms to acquire information about the general status of the engine, without focusing on a specific signal nor a specific fault, ensuring a complete fault diagnosis, independent from experts' judgement. Likewise, Song et al. (2018) proposed a novel fault prediction method for city buses powered by LNG by collecting different data types, such as bus daily schedules, maintenance occurrences, and relevant weather conditions and subsequently applying a random forest algorithm. Ahmed et al. (2015) trained an artificial neural network employing a new estimation strategy known as the smooth variable structure filter to detect engine's faults. This approach demonstrated stability and generalisation accuracy exhibiting improved performance compared with the first order back propagation algorithm and similar performance compared with the extended Kalman filter.

3. Physical model description

3.1. The considered DF engine

This study investigates the Wärtsilä 9L50DF engine, which is a marine four-stroke, turbocharged and intercooled DF engine (Wärtsilä, 2012) that is employed for ship propulsion or electrical generation, in the latter case as part of a generator set (Wärtsilä, 2012). The engine is capable of operating in two distinct modes: (i) the gas mode running on natural gas and liquid pilot fuel, usually Light Fuel Oil, that is injected in the engine cylinders for initiating the combustion of the premixed natural gas–air mixture; and (ii) the diesel mode, in which either Heavy or Light Fuel Oil is used as the main fuel. Both liquid fuels (pilot and main) are injected within the engine cylinder (depending on the operating mode) by a combined fuel injector located in the cylinder head. The natural gas is injected in the inlet port of each cylinder during

the intake valve opening period, and subsequently is mixed with the air; the mixture of air and gas is trapped in the engine cylinder. The natural gas injection starts after the closing of the cylinder exhaust valve for preventing natural gas leakages to the exhaust manifold. This engine operates employing the lean-burn combustion concept, which is associated with low nitrogen oxides emissions and thermal loading due to the reduced in-cylinder peak temperature level. The engine complies with the IMO Tier III nitrogen oxides limits when operating in the gas mode, whereas the engine diesel mode operation is compliant with the Tier II limits.

The engine achieves stable combustion conditions (avoiding the instabilities of knocking and misfiring) by adjusting the airfuel ratio via an electronically controlled exhaust waste gate, which bypasses a part of the exhaust gas along the turbocharger turbine. In addition, the gas admission valves as well as the diesel fuel injectors are electronically controlled to adjust the engine power output and speed in both operating modes.

In this study, the investigated engine was considered as a part of a generator set operating at a constant speed of 514 r/min. The engine detailed description is reported in the manufacturer product guide (Wärtsilä, 2012). The main engine characteristics are illustrated in Table 1. The engine layout and components interconnections are presented in Fig. 2.

3.2. Engine modelling description

The investigated engine and its control system have been modelled in the GT-ISE software, which is a state-of-the-art tool for engine modelling and analysis extensively used from academia and industry (Gamma-Technologies, 2019). The complete engine model was developed utilising the following assemblies of the GT-ISE software: (i) engine thermodynamic model; (ii) user input; (iii) engine control system; (iv) alarms and monitoring system.

The investigated engine model was developed in previous authors' studies (Stoumpos et al., 2018, 2020) and is capable of simulating both the engine steady state and transient conditions in both gas and

Table 2
Envelope of engine input parameters employed in the DoE design.

Parameter	Unit	Diesel mode	Gas mode
Load	[%]	25, 50, 75, 100	25, 50, 75, 100
Fuel — diesel:1/gas:2	[–]	1	2
Ambient temperature	[K]	298.1, 308.1, 318.1	298.1, 308.1, 318.1
Air cooler cooling water temperature	[K]	298.1, 308.1, 318.1	298.1, 308.1, 318.1
GVU gas pressure	[bar]	5.92	5.92, 5.96, 6.00
Diesel fuel LHV	[MJ/kg]	42.0, 43.3, 44.6, 46.0	42.6
Gas LHV	[MJ/kg]	50.0	45.0, 47.5, 50.0

Table 3
Control variables (C) and performance variables (P).

Type	Name	Unit
C_1	Ambient temperature	[K]
C_2	Air cooler temperature	[K]
C_3	GVU gas pressure	[bar]
C_4	Heating value of Diesel	[MJ/kg]
C_5	Heating value of gas	[MJ/kg]
C_6	Engine load	[kW]
C_7	Engine speed (constant)	[rpm]
P_1	Fuel rack position (Diesel mode)	
P_2	Main gas pressure	[bar]
P_3	Max cylinder pressure	[bar]
P_4	Charged air pressure (Inlet section)	[bar]
P_5	Exhaust gas temperature Turbocharger — inlet	[K]
P_6	Turbocharger speed	[rpm]
P_7	Waste gate opening	[mm^2]
P_8	Nitrogen oxide	[g/kWh]
P_9	Carbon dioxide	[g/kWh]
P_{10}	BSFC	[g/kWh]

diesel operating modes. As it is reported in these studies, the model was extensively validated based on experimental data available in the literature for both steady-state conditions and transient conditions with modes switching. In addition, this model was employed in Theotokatos et al. (2020) to investigate the influence of the exhaust gas waste gate valve behaviour on the engine operation. Therefore, only a succinct summary of the modelling principles is provided herein.

The engine cylinder processes were modelled by using the zero-dimensional approach that employs the energy and mass conservation equations along with the ideal gas equation for the calculation of the cylinder working media temperature, mass, pressure, and mixture composition. An one-zone approach was used for modelling the gas exchange and compression processes, whereas a two-zone approach (an unburned mixture zone and a burned gas zone generated after the combustion start) was employed for modelling the combustion and expansion processes (Merker et al., 2005).

The cylinder gas-to-wall heat transfer coefficient was calculated by employing the Woschni equation (Woschni, 1967). The Chen-Flynn friction model (Rakopoulos and Giakoumis, 2007) was employed for calculating the engine friction mean effective pressure. The extended Zeldovich mechanism, which is described in Lavoie et al. (1970), Hanson and Salimian (1984), Heywood (1988) for the estimation of the nitrogen oxide emissions in both operating modes.

The combustion model employs a single-Wiebe function to calculate the heat release rate at the diesel mode, whereas the ignition delay is calculated by considering the Sitkey equation (Merker et al., 2005; Sitkei, 2013). For the gas operating mode, the combustion is modelled by employing a triple-Wiebe function (Xu et al., 2017) assuming that each function represents the following consecutive combustion phases: (a) the premixed combustion of a portion of the pilot fuel; (b) the diffusive combustion of the remaining pilot fuel and the rapid burning of the gaseous fuel, and; (c) the cylinder residuals tail combustion (Karim, 2015). The ignition delay is approximated by using the data reported in Christen and Brand (2013) and Sixel et al. (2016).

The cumulative burned fuel is calculated according to Gamma-Technologies (2019)

$$x_b = \sum_{i=1}^{3} FF_i \left(1 - exp \left(-\alpha \left(\frac{\theta - \theta_{SC,i}}{\Delta \theta_i} \right)_i^{m_i+1} \right) \right), \tag{1}$$

where the subscript i denotes the Wiebe function; FF denotes the weight of each Wiebe function; a is the Wiebe function parameter (considered 6.9); θ denotes the crank angle, θ_{SC} denotes the start of combustion; $\Delta \theta$ is the combustion duration and m denotes the Wiebe function shape factor.

The heat release rate is calculated according to Eq. (2), which employs the fuel burning rate (time derivative of the cumulative burned fuel from Eq. (1)) and the total energy from all the injected fuels

$$\dot{Q}_b = \dot{x}_b \sum_{j=1}^{3} m_{f,j} \cdot LHV_j, \tag{2}$$

where $\dot{x}_b$ denotes the fuel burning rate; j denotes the fuel (main diesel, pilot diesel, or natural gas); m_f is the burned fuel amount; LHV denotes the fuel lower heating value. It must be noted that only one Wiebe function ($i = 1$) with weight equal to 1 was used for modelling the combustion in the diesel operating mode.

The mass flow rates through the engine intake and exhaust valves are calculated by using the quasi-steady adiabatic flow equation considering the respective profiles (equivalent area versus crank angle) and pressure ratios (Heywood, 1988). The engine crank shaft rotational speed is modelled by employing the angular momentum conservation equation.

The engine inlet and exhaust manifolds were modelled by using a one-dimensional approach that employs the mass, momentum, and energy conversation equations for the calculation of the pressure, velocity, temperature, and composition of the working media (air or exhaust gas) along the manifolds length (Gamma-Technologies, 2019).

The model calibration process included the tuning of the Wiebe functions parameters at the considered operating points to achieve adequate accuracy in comparison with the engine parameters experimentally obtained in the engine trials.

3.3. Engine abnormalities modelling

The model described in Section 3.2 has been updated to take into account the performance degradation of the DE air cooler due to fouling and corrosion mainly in the cooling water side.

As reported in Joshi et al. (2009), the degradation/fouling of the charge-air cooler in turbocharged DEs has a significant impact on engine performance resulting in lower air cooler effectiveness and higher pressure drop, which, in turn, reduce the pressure and increase the temperature at the engine inlet manifold. As a result, the engine power output reduces having an adverse effect on the engine fuel consumption and emissions.

In agreement with previous studies, (Müller-Steinhagen, 1999), the effect of fouling and corrosion in the cooling water and air sides is simulated by considering a reduced value by 30% for the overall heat transfer coefficient compared with the respective values for healthy conditions. This resulted in temperature levels of the inlet manifolds close to the manufacturer upper limit, beyond which an alarm is

activated. Typically, fouling in the air side of the air cooler also causes decrease of the overall heat transfer coefficient and additionally increase of the friction factor (resulting in increase of the air cooler pressure drop). However, the latter effect was not considered in this study.

4. Dataset generation

For the reason described in the introduction and for confidentiality issues, datasets corresponding to the investigated marine DF engine under faulty conditions were not available to the authors, therefore, this study employed the physical model of high fidelity that was developed and validated in previous authors' studies (Stoumpos et al., 2018), which was briefly described in Section 3.2, to generate it. For this purpose, multiple simulation runs, corresponding to different scenario, were performed collecting engine control and performance parameters to generate this dataset.

Two of these datasets correspond to the engine operation at healthy conditions in both the diesel mode ($DB_{Healthy}^{DM}$) and the gas mode ($DB_{Healthy}^{GM}$). These datasets are referred as the "healthy status" datasets henceforth. Moreover, a third dataset (DB_{Faulty}^{DM}) is created and contains the control and the performance variables corresponding to the degraded conditions described in Section 3.3 and the engine diesel mode. This dataset is referred as the "faulty status" dataset henceforth.

The simulation runs were performed for the engine steady state conditions in the diesel and gas modes at 25%, 50%, 75%, and 100% loads considering the following engine operating parameters varying within the ranges provided in Table 2: (a) ambient temperature; (b) diesel fuel LHV; (c) gas fuel LHV (for the gas mode); (d) Gas Valve Unit (GVU) pressure (for the gas mode); and (e) air cooler coolant temperature. In this respect, a large number of simulation runs for all combinations of possible engine operating conditions were performed. For the engine model simulation runs design, the integrated GT-ISE software Design of Experiments (DoE) tool was used. The parametric runs were designed based on the full factorial DoE method (Antony, 2014), imposing for each considered parameter the minimum and maximum values (factor).

The Control and performance variables collected from the physical model are listed in Table 3. Note that these variables are the ones that can be actually measured by current automation systems.

5. Methods

In this section, the proposed models are presented, which address the problem described in Section 1 exploiting the datasets described in Section 4.

The problem described in Section 1 can be associated to a conventional ML framework (Shalev-Shwartz and Ben-David, 2014), in which one has to consider an input space $\mathcal{X} \subseteq \mathbb{R}^d$ and an output space $\mathcal{Y}$ and the goal estimating the unknown rule $\mu : \mathcal{X} \rightarrow \mathcal{Y}$ which associates an element $y \in \mathcal{Y}$ to an element $\boldsymbol{x} \in \mathcal{X}$. ML techniques estimates μ through a learning algorithm $\mathscr{A}_{\mathcal{H}} : \mathcal{D}_n \times \mathcal{F} \rightarrow f$, characterised by its set of hyperparameters $\mathcal{H}$, which maps a series of examples of the input/output relation contained in a datasets of n samples $\mathcal{D}_n : \left\{ (\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n) \right\}$ into a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ chosen in a set of possible ones $\mathcal{F}$ (see Section 5.1).

When $\mathcal{D}_n$ contains both elements in the input space and the associated elements in the output space the problems is named supervised (Shalev-Shwartz and Ben-David, 2014). Regression and classification are two of the most popular examples of supervised ML problems. In regression $\mathcal{Y} \subseteq \mathbb{R}$ and the elements in the $\mathcal{Y}$ have an associated notion of distance, while in classification $\mathcal{Y} \in \{C_1, \ldots, C_c\}$ and the elements in $\mathcal{Y}$ have no associated notion of distance. Binary classification is a particular example of classification problem where $\mathcal{Y} \in \{\pm 1\}$. When $\mathcal{D}_n$ contains just elements in the input space, which means that it is not explicitly known the associated element of the output space (or

we know it for just few elements), it has to be assumed that similar inputs are associated with outputs where the concept of similarity is something that needs to be defined based on μ. In this last case, the ML problems are called unsupervised (or weekly supervised) (Shalev-Shwartz and Ben-David, 2014). Anomaly (novelty, outlier) detection is a common example of unsupervised (or weekly supervised) learning problem where the unknown $y \in \mathcal{Y}$ can assume only two possible values: -1 for "non-anomaly" and $+1$ for "anomaly".

The error that f commits in approximating μ is measured with reference to a metric (see Section 5.2) $M(\cdot, \cdot)$. This error need to be optimised and estimated during the Model Selection and Error Estimation phases (Oneto, 2020) (see Section 5.3) since, obviously, the error that f commits over $\mathcal{D}_n$, is optimistically biased since $\mathcal{D}_n$ has been used, together with $\mathcal{F}$, for building f itself. For this reason, another set of fresh data, composed of t samples and called test set $\mathcal{T}_t = \{(\boldsymbol{x}_1', y_1'), \ldots, (\boldsymbol{x}_t', y_t')\}$, needs to be exploited. Note that, $\boldsymbol{x}_i' \in \mathcal{X}$ and $y_i' \in \mathcal{Y}$ with $i \in \{1, \ldots, m\}$, and the association of y_i^t to $\boldsymbol{x}_i^t$ is again made based on μ. Moreover, both for supervised and unsupervised (or weekly supervised) problems $\mathcal{T}_t$ must contain both $\boldsymbol{x}_i' \in \mathcal{X}$ and $y_i' \in \mathcal{Y}$ with $i \in \{1, \ldots, m\}$ to estimate the error of f.

5.1. Machine learning algorithms

In this study, we will exploit different state-of-the-art algorithms for both supervised and weakly supervised problems.

For the supervised case we will exploit: Random Forests (RF) (Breiman, 2001), Kernel Methods (KM) (Shawe-Taylor and Cristianini, 2004), and Neural Networks (NN) (Goodfellow et al., 2016). RF, KM, and NN are three state of the art methods coming from the three main families of supervised data driven models, able to deal both with regression and classification problems, that have shown to perform quite well in many different applications (Fernández-Delgado et al., 2014; Wainberg et al., 2016). RF (Breiman, 2001) technique relies on the simple fact that combining the output of several classifiers results in a much better performance than using any one of them alone. RF combine many decision trees in order to obtain effective predictors, which have limited hyperparameter sensitivity and high numerical robustness. RF have two main hyperparameters: the numbers of trees h_1 and the numbers of predictor to sample in each node creation h_2. KMs are a family of ML techniques which exploits the Kernel trick for distances in order to extend linear techniques to the solution of non-linear problems (Scholkopf, 2001). In the case of classification, KMs select f as the function which minimises the trade-off between the sum of the accuracy over the data, namely the empirical error, and the complexity of the solution, namely the regularisation term. The hyperparameters of the KM $\mathcal{H}^{KM}$ are: the kernel, which is usually fixed and in this paper author chose the Gaussian Kernel for the reasons described in Keerthi and Lin (2003), its hyperparameter h_1 and the regularisation hyperparameter h_2. NN (Goodfellow et al., 2016) combine together many simple models of a human brain neuron, called perceptrons, in order to build a complex network. The neurons are organised in stacked layers connected together by weights that are learned based on the available data via backpropagation. Since the data employed in this study has no particular structure, a simple shallow network is a natural choice (Goodfellow et al., 2016). In particular, a NN can be interpreted as a KM where the kernel is not chosen but learned from data. Then the hyperparameters of an NN $\mathcal{H}^{NN}$ are the number of neuron in the hidden layer h_1 and the regularisation hyperparameter on the last layer h_2.

For the weekly supervised problems, authors will exploit the most known and effective techniques for solving these problems according to Swersky et al. (2016). In particular (Swersky et al., 2016) shows that two anomaly detection methods based on KM and K-Nearest Neighbourhood (KNN) respectively, are the top choices in this context. In particular One-Class Support Vector Machines (OCSVM) (Shawe-Taylor and Cristianini, 2004) is a boundary-based anomaly detection method,

inspired by KM, which enclose the inlier class in a minimum volume hypersphere and like KM can also be extended to non-linearly transformed spaces using the Kernel trick. The hyperparameters OCSVM $\mathcal{H}^{\text{OCSVM}}$ are the same as the ones of KM. The Global KNN (GKNN), inspired by the KNN, has been originally introduced as an unsupervised distance-based outlier detection method (Ramaswamy et al., 2000). The hyperparameter GKNN $\mathcal{H}^{\text{GKNN}}$ is the number of neighbours to be considered h_1.

It is worth noting that there is no really particular strength or weakness which distinguishes the ML methods. What changes is the functional form of the different models and, as studied by the no-free-lunch theorems (Wolpert, 1996), the only way to verify which is the best method for a particular application is to test it empirically on real data.

5.2. Metrics

In this work, many state-of-the-art techniques will be tested and their performances will be compared. In order to perform this analysis, authors have to define different metrics of performance $M(\cdot, \cdot)$, able to well characterise the quality of the different models.

In regression the most natural choice metric is surely the Mean Absolute Percentage Error (MAPE) which can be computed as follows

$$M(f, \mathcal{T}_t) = \frac{100}{t} \sum_{(\boldsymbol{x}, y) \in \mathcal{T}_t} |f(\boldsymbol{x}) - y|, \tag{3}$$

In the binary classification, the most natural metric is the one which counts the number of misclassified samples, namely the percentage of misclassifications

$$M(f, \mathcal{T}_t) = \frac{100}{t} \sum_{(\boldsymbol{x}, y) \in \mathcal{T}_t} [f(\boldsymbol{x}) \neq y]. \tag{4}$$

where the Iverson bracket notation is exploited. Moreover, this measure will be also used for the anomaly detection problems since, also in this case, a binary output is considered (non-anomaly or anomaly). Based on this metric it is possible to define other important indexes of performance (Powers, 2011) like the Confusion Matrix, which measures four different quantities

• the percentage of true negative

$$\text{TN} = \frac{100}{t} \sum_{(\boldsymbol{x}, y) \in \mathcal{T}_t} [f(\boldsymbol{x}) = y \wedge y = -1]; \tag{5}$$

• the percentage of true positive

$$\text{TP} = \frac{100}{t} \sum_{(\boldsymbol{x}, y) \in \mathcal{T}_t} [f(\boldsymbol{x}) = y \wedge y = +1]; \tag{6}$$

• the percentage of false negative

$$\text{FN} = \frac{100}{t} \sum_{(\boldsymbol{x}, y) \in \mathcal{T}_t} [f(\boldsymbol{x}) \neq y \wedge y = -1]; \tag{7}$$

• the percentage of false positive

$$\text{FP} = \frac{100}{t} \sum_{(\boldsymbol{x}, y) \in \mathcal{T}_t} [f(\boldsymbol{x}) \neq y \wedge y = +1]. \tag{8}$$

5.3. Model selection and error estimation

Model Selection (MS) and Error Estimation (EE) techniques address the problem of tuning and estimating the performance of a learning algorithm.

Since the hyperparameters $\mathcal{H}$ influence the performance of $\mathcal{A}_{\mathcal{H}}$, a proper MS procedure needs to be adopted (Oneto, 2020). In this work the Bootstrap procedure will be exploited. Bootstrap rely on a simple idea: the original dataset $\mathcal{D}_n$ is resampled r times with replacement to build a training set of size equal to the original one $\mathcal{L}_l^i$ while the remaining samples are kept in the validation set $\mathcal{V}_v^i$, with $i \in \{1, \dots, r\}$.

In order to perform the MS phase, i.e. to select the best combination of the hyperparameters $\mathcal{H}$ in the set of possible ones $\mathfrak{H} = \{\mathcal{H}_1, \mathcal{H}_2, \dots\}$ using the algorithm $\mathcal{A}_{\mathcal{H}}$, the hyperparameters which minimise the performance of the model measured with the preferred metric $M(\cdot, \cdot)$, trained on the training set, have been searched on the validation set, in formula:

$$\mathcal{H}^* : \quad \min_{\mathcal{H} \in \mathfrak{H}} \frac{1}{r} \sum_{i=1}^{r} M(\mathcal{A}_{\mathcal{H}, \mathcal{L}_l^i}, \mathcal{V}_v^i), \tag{9}$$

where $\mathcal{A}_{\mathcal{H}, \mathcal{L}_l^i}$ is a model built with the algorithm $\mathcal{A}$ with its set of hyperparameters $\mathcal{H}$ and with the data $\mathcal{L}_l^i$. Since the data in $\mathcal{L}_l^i$ are independent from the ones in $\mathcal{V}_v^i$, the idea is that $\mathcal{H}^*$ should be the set of hyperparameters which allows to achieve a small error on a dataset that is independent from the training set.

Finally, we need to perform the EE phase of the optimal model with a separate sets of data $\mathcal{T}_t$ since the error that our model commits over $\mathcal{D}_n$ would be optimistically biased since $\mathcal{D}_n$ has been used to find f. For this reason we have to compute

$$M(\mathcal{A}_{\mathcal{H}^*, \mathcal{D}_n}, \mathcal{T}_t), \tag{10}$$

since $\mathcal{T}_t$ in independent from $\mathcal{D}_n$.

Note that, for supervised learning problems $\mathcal{L}_l^i, \mathcal{V}_v^i$ with $i \in \{1, \dots, r\}$, and $\mathcal{T}_t$ need to contain both inputs and outputs while for the weekly supervised learning problems $\mathcal{L}_l^i$ with $i \in \{1, \dots, r\}$ can contains just inputs and $\mathcal{V}_v^i$ with $i \in \{1, \dots, r\}$ and $\mathcal{T}_t$ needs to contain both inputs and outputs even if v and t can be very small (see Section 6).

5.4. The proposed approach

The problems described in Section 1 can be addressed in different ways using the methods presented in this section. It is worth to remark that the main aim of this work is to propose and validate a Weakly Supervised approach which is able to reduce the amount of labelled data to the minimum realistic requirement that can be readily available in real word operations.

As a first step it is possible to build a Digital Twin Cipollini et al. (2018b) able to estimate the performance variables $\boldsymbol{P}$ based on the control variables $\boldsymbol{C}$ exploiting regression models. In particular, we can exploit the dataset of normal (healthy) behaviour ($\text{DB}_{\text{Healthy}}^{\text{DM}}$ and $\text{DB}_{\text{Healthy}}^{\text{GM}}$) to train regression models to build a Digital Twin of the engine that can be used for fast simulation and performance drift detection problems. Note that for training, MS, and EE purposes of the Digital Twin (the regressors) the datasets of normal behaviour are enough.

If the Digital Twin is accurate enough it is possible to use it for simulating the behaviour of the engine and for checking if there is a large difference between the expected behaviour and the actual one (Coraddu et al., 2019). A large deviation may (should) indicate a malfunction. More formally, let us suppose to build a function f able to estimate $\boldsymbol{P}$ based on $\boldsymbol{C}$ such that

$$\hat{\boldsymbol{P}} = f(\boldsymbol{C}),$$

then the distance between $\hat{\boldsymbol{P}}$ and $\boldsymbol{P}$ measured with a proper distance (Adams and Fournier, 2003) $\|\cdot\|_p$ can be measured[3]

$$\Delta = \|\hat{\boldsymbol{P}} - \boldsymbol{P}\|_p,$$

and if this Δ is larger than a defined threshold δ^* then the behaviour can be classified as abnormal. $p \in [0, \infty)$ and $\Delta^* \in (0, \infty)$ are two hyperparameter to be tuned in the MS phase, respectively h_1 and h_2. Note that, for training the Digital Twin (the regressors) the datasets of healthy (normal) behaviour are enough but for tuning h_1 and h_2, we
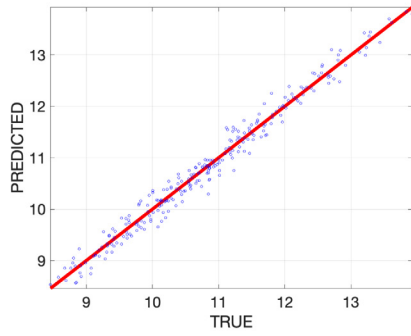
---

[3] p is an hyperparameter of the distance which allow to spam multiple notions of distance such as the Euclidean one with $p = 2$, or the Manhattan one with $p = 1$, or the Maximum Norm with $p \to \infty$.
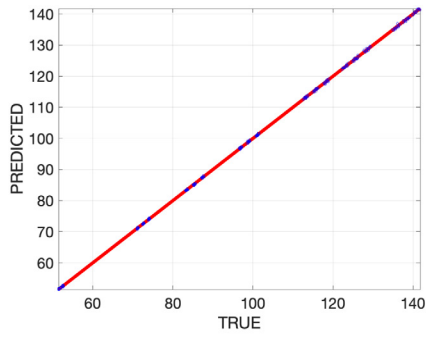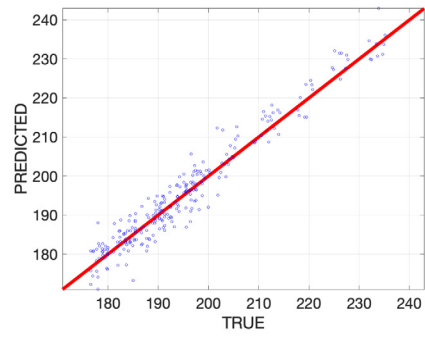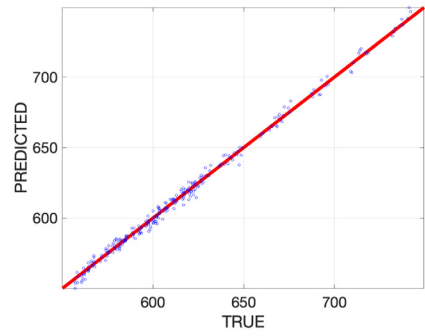
Fig. 3. Scenario 1 - Diesel Mode: scatter plot (true and predicted values), for the best performing model according to Table 4, of the main performance variables.

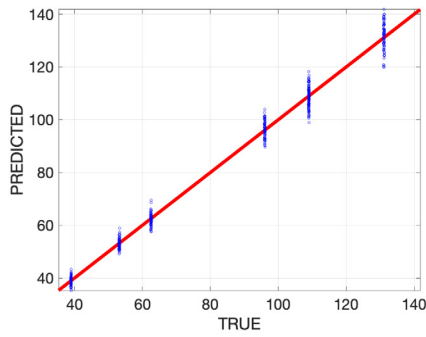(a) P_3 - Max Cylinder Pressure
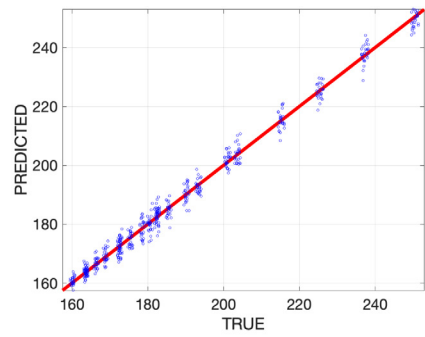
(b) P_{10} - BSFC
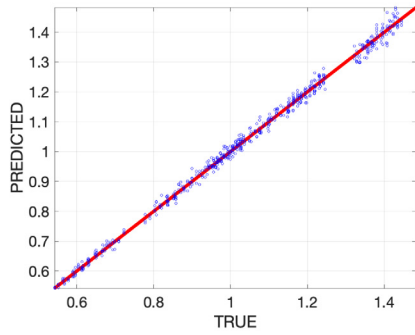
(c) P_8 - NO$_\text{X}$

(d) P_9 - CO$_2$

Fig. 4. Scenario 1 - Gas Mode: scatter plot (true and predicted values), for the best performing model according to Table 4, of the main performance variables.

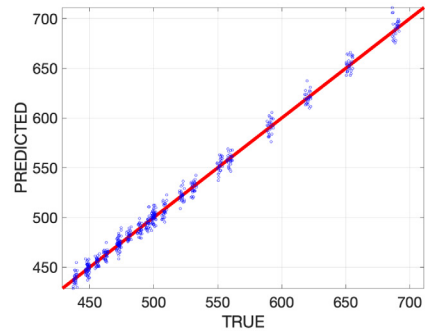(a) P_3 - Max Cylinder Pressure

(b) P_{10} - BSFC

(c) P_8 - NO$_\text{X}$

(d) P_9 - CO$_2$

need both a dataset of healthy behaviour (e.g. $DB_{Healthy}^{DM}$ and $DB_{Healthy}^{GM}$) and a dataset of faulty (abnormal) behaviour (e.g. DB_{Faulty}^{DM}).

If we have large datasets of normal and abnormal behaviour, instead of using the indirect approach (based on a Digital Twin), we can use a direct approach exploiting a binary classifier that based on P and C it has to predict if this combination refers to a healthy or faulty behaviour (Cipollini et al., 2018b). In this case for training, MS, and EE purposed a dataset of healthy behaviour and a dataset of faulty behaviour are needed.

More realistically, in practical situations, it is common to have a large datasets of healthy behaviour while the dataset of faulty behaviour is usually quite limited since engines work well for the majority of their lifetime (Cipollini et al., 2018b). In this case, it is then more appropriate and effective to use an anomaly detection method since, for these methods, the dataset of faulty behaviour is exploited just for MS and EE purposed and not for training the model (similarly to the indirect approach based on the Digital Twin). This allows to met or requirements: develop a model able to reduce the amount of labelled data to the minimum realistic requirement that can be readily available in real word operations.

6. Experimental results

In this section, we will present the results and the quality of the methodologies presented in Section 5 for solving the problem described in Section 1 (weekly supervised monitoring for marine DF engines targeting to reduce, as much as possible, the necessity of labelled data at least to a realistic level which is realistic to retrieve in the wild without compromising the recognition accuracy) by means of the data described in Section 4.

6.1. The scenarios

As described in Section 5, thanks to the use of data-driven methods, we will be able to face different scenarios. In particular, based on the available data it is possible to deal with

1. Fully Supervised Performance Estimation: In this setting it is available a dataset were both the control and the performance variables are available in healthy conditions of the engines. The scope is then to estimate the performance variables based on the control variables. This allows to build a Digital Twin of the engine that can be used for fast simulation and performance drift detection problems. As described in Section 4, in our study, it is available this dataset for both diesel and gas mode ($DB_{Healthy}^{DM}$ and $DB_{Healthy}^{GM}$).

2. Fully Supervised Health Status Estimation: In this setting it is available a dataset were both the control variables and the performance variables are available in healthy and faulty conditions. The scope is then to estimate the health status based on the control and performance variables. As described in Section 4, in our study, it is available this dataset for just the diesel mode ($DB_{Healthy}^{DM}$ and DB_{Faulty}^{DM}).

3. Weakly Supervised Health Status Estimation: In this setting, the most realistic one, it is available a dataset were both the control variables and the performance variables are available in healthy and faulty conditions, but the dataset referring to faulty status is very small (this is often the case in practice). The scope is then to estimate the health status based on the control and performance variables. As described in Section 4, in our study, we can simulate this scenario just for the diesel mode since we can use $DB_{Healthy}^{DM}$ and we can sub-sample the DB_{Faulty}^{DM}.

6.2. Experimental setting

The datasets considered in Section 4 were divided into different sets as reported in Section 5 based on the considered scenario (see Section 6.1)

- Scenario 1: we use 70% of the data for training the models and the remaining 30% for testing them for both $DB_{Healthy}^{DM}$ and $DB_{Healthy}^{GM}$;

- Scenario 2: for the direct approaches we merged the $DB_{Healthy}^{DM}$ and the DB_{Faulty}^{DM} assigning respectively the label Healthy and Faulty to the set of samples and we use 70% of the data for training the models and the remaining 30% for testing them. For the Digital Twin based methods we use 70% $DB_{Healthy}^{DM}$ to train the Digital Twin and the remaining data in $DB_{Healthy}^{DM}$ and DB_{Faulty}^{DM} to perform the other step of the methods (validation of the threshold and testing the model);

- Scenario 3: for both the direct approaches and the Digital Twin based methods we exploit the 70% $DB_{Healthy}^{DM}$ to train the model and the remaining data in $DB_{Healthy}^{DM}$ and DB_{Faulty}^{DM} to perform the other step of the methods (a very small amount for validation of the hyperparameters and the rest for testing the model).

For each supervised classification method described in Section 5, an MS procedure was performed, as described in Section 5.3. What follows is the list of hyperparameters tested during the MS with their respective intervals

- RF: the set of hyperparameters is $\mathcal{H}^{NN} = \{h_1, h_2\}$ and authors chose it in $\mathfrak{H}^{NN} = \{1000\} \times \{d^{1/8}, d^{1/4}, d^{1/2}, d^{3/4}, d^{7/8}\}$

- NN: the set of hyperparameters is $\mathcal{H}^{NN} = \{h_1, h_2\}$ and authors chose it in $\mathfrak{H}^{NN} = \{10, 10^{1.2} \cdots, 10^4\} \times \{10^{-6}, 10^{-5.8}, \ldots, 10^4\}$;

- KM: the set of hyperparameters is $\mathcal{H}^{KM} = \{h_1, h_2\}$ and authors chose it in $\mathfrak{H}^{KM} = \{10^{-6}, 10^{-5.8}, \ldots, 10^4\} \times \{10^{-6}, 10^{-5.8}, \ldots, 10^4\}$;

- OCSVM: the set of hyperparameters is $\mathcal{H}^{OCSVM} = \{h_1, h_2\}$ and authors chose it in $\mathfrak{H}^{OCSVM} = \{10^{-6}, 10^{-5.8}, \cdots, 10^4\} \times \{10^{-6}, 10^{-5.8}, \cdots, 10^4\}$;

- GKNN: the set of hyperparameters is $\mathcal{H}^{GKNN} = \{h_1\}$ and authors chose it in $\mathfrak{H}^{GKNN} = \{1, 3, 7, 13, 27, 51\}$;

- Approaches based on Digital Twin: the set of hyperparameters is $\mathcal{H}^{DigitalTwin} = \{h_1, h_2\}$ and authors chose it in $\mathfrak{H}^{DigitalTwin} = \{.001, .01, .1, 1, 10, 100, 1000\} \times \{10^{-6}, 10^{-5.999}, \cdots, 10^6\}$;

For the Scenario 3 the $\mathcal{V}_v^r$ cardinality was varied $v \in \{10, 20, 40\}$, in order to test the possibility of building an efficient model with a small number of labelled samples.

The performances of each model are measured according to the metrics described in Section 5.2. Each experiment was performed 30 times in order to obtain statistical relevant result, and the t-student 95% confidence interval is reported when space in the table was available without compromising their readability.

For RF the *randomForest* R package has been exploited.[4] For KM and GKNN a custom Matlab[5] implementation has been developed. For NN the Python *PyTorch* library has been exploited.[6] For OCSVM the *e1071* R package has been exploited.[7]

6.3. Scenario 1 : Fully supervised performance estimation

Table 4 reports the error of the three different models (RF, KM and NN) for both diesel and gas modes. Moreover, scatter plots for the best performing model (according to Table 4) are available in Figs. 3 and 4 for diesel and gas modes respectively. For both modalities, the authors report the estimations of the most relevant performance variables under analysis (P_3, P_8, P_9 and P_{10}).

---

[4] https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

[5] https://it.mathworks.com/products/matlab.html.

[6] https://pytorch.org/.

[7] https://cran.r-project.org/web/packages/e1071/e1071.pdf.

Table 4

Scenario 1: error, measured with the MAPE, of the different Models (RF, KM, and NN) in predicting the performance variables (see Section 4) for both diesel and gas mode.

Engine Model Perf. Variable	Diesel Mode			Gas Mode		
	RF	KM	NN	RF	KM	NN
P_1	0.97 ± 0.06	0.31 ± 0.02	1.66 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
P_2	1.54 ± 0.08	0.91 ± 0.05	2.49 ± 0.13	3.53 ± 0.27	2.67 ± 0.15	5.73 ± 0.34
P_3	0.71 ± 0.04	0.22 ± 0.01	0.88 ± 0.05	3.79 ± 0.23	3.56 ± 0.21	4.57 ± 0.24
P_4	0.85 ± 0.04	0.49 ± 0.03	1.27 ± 0.08	3.47 ± 0.18	2.97 ± 0.17	7.00 ± 0.40
P_5	1.62 ± 0.10	1.37 ± 0.07	2.69 ± 0.15	4.97 ± 0.26	4.20 ± 0.30	6.59 ± 0.36
P_6	1.48 ± 0.07	1.17 ± 0.06	2.45 ± 0.15	2.52 ± 0.14	2.38 ± 0.13	3.85 ± 0.21
P_7	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	4.08 ± 0.29	3.06 ± 0.18	5.45 ± 0.43
P_8	3.19 ± 0.18	1.77 ± 0.12	3.77 ± 0.23	2.02 ± 0.12	1.46 ± 0.08	2.10 ± 0.11
P_9	1.74 ± 0.09	0.55 ± 0.03	3.19 ± 0.21	1.49 ± 0.09	1.13 ± 0.07	1.57 ± 0.08
P_{10}	1.92 ± 0.10	1.71 ± 0.10	2.52 ± 0.16	1.49 ± 0.10	1.39 ± 0.08	1.98 ± 0.16

Table 5

Scenario 2: error (percentage of misclassification error) and confusion matrices (positive are faulty and negative are healthy) of the different models (RF, KM, and NN) with a Direct Approach or with a Digital Twin in predicting the health status conditions (see Section 4) of the engine in diesel mode.

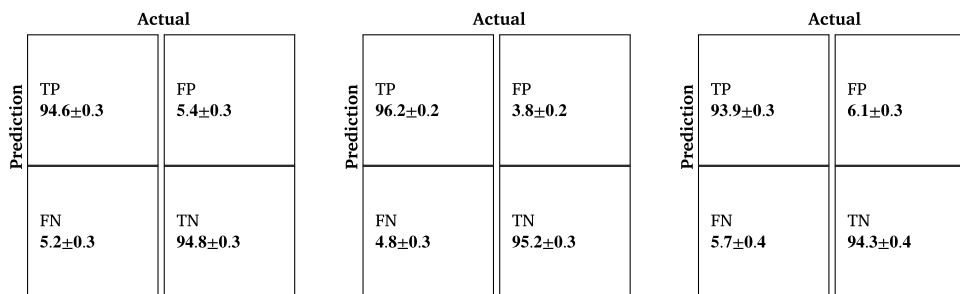| Engine	Diesel Mode					
Approach	Direct			Digital Twin		
Model	RF	KM	NN	RF	KM	NN
	2.1 ± 0.1	1.7 ± 0.1	2.7 ± 0.2	5.3 ± 0.3	4.3 ± 0.2	5.9 ± 0.4

(a) Misclassification error %
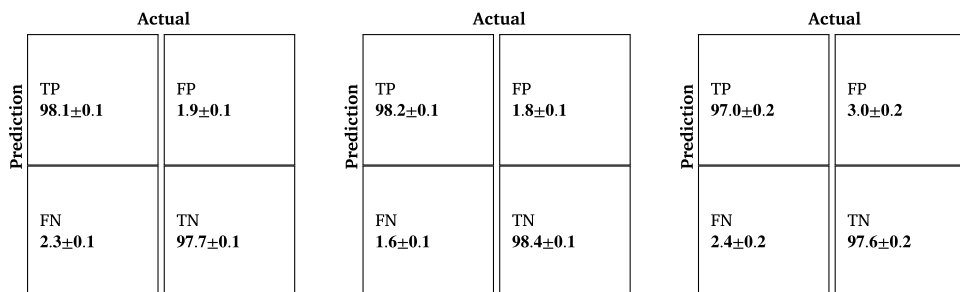
(b) Confusion matrix with RF for diesel mode using a Direct Approach.

(c) Confusion matrix with KM for diesel mode using a Direct Approach.

(d) Confusion matrix with NN for diesel mode using a Direct Approach.

(e) Confusion matrix with RF for diesel mode using a Digital Twin.

(f) Confusion matrix with KM for diesel mode using a Digital Twin.

(g) Confusion matrix with NN for diesel mode using a Digital Twin.

Observing the results KM outperforms the other two models (RF and NN) for all the performance variables. Moreover models seems to perform better for diesel model with respect to the gas mode. This results is reasonable given also the comments on the higher complexity of DF engines. From a practical point of view, the accuracies of the developed models is surely up to a level which is acceptable for their use in the wild. In fact, as reported in Table 4, the MAPE of the best performing model (KM) is always less than 4% (and in most case less than 2%) for all the performance variables, in the considered modalities.

6.4. Scenario 2 : Fully supervised health status estimation

In this section, the results of the engine's health status classification experiments are listed and analysed. As previously described in List 2,

Table 6

Scenario 3: error (percentage of misclassification error) and confusion matrices (positive are faulty and negative are healthy) of the different models with a Direct Approach (OCSVM and GKNN) or with a Digital Twin (based on RF, KM, and NN) in predicting the health status conditions (see Section 4) of the engine in diesel mode.

Engine		DM				
Approach		Direct		Digital Twin		
Model n_v		OCSVM	GKNN	RF	KM	NN
10		2.3±0.1	3.2±0.2	5.9±0.3	4.9±0.3	7.1±0.4
20		2.2±0.1	3.1±0.2	5.7±0.3	4.7±0.2	6.8±0.5
40		2.1±0.1	2.9±0.2	5.3±0.3	4.4±0.2	6.4±0.4

(a) Misclassification error %

	Actual	
Prediction	TP 97.6±0.1	FP 2.4±0.1
	FN 2.2±0.1	TN 97.8±0.1

(b) Confusion matrix with OCSVM for diesel mode using a Direct Approach ($n_v = 10$).

	Actual	
Prediction	TP 96.6±0.2	FP 3.4±0.2
	FN 3.0±0.2	TN 97.0±0.2

(c) Confusion matrix with GKNN for diesel mode using a Direct Approach ($n_v = 10$).

	Actual	
Prediction	TP 94.1±0.3	FP 5.9±0.3
	FN 5.9±0.3	TN 94.1±0.3

(d) Confusion matrix with RF for diesel mode using a Digital Twin ($n_v = 10$).

	Actual	
Prediction	TP 95.5±0.2	FP 4.5±0.2
	FN 5.3±0.3	TN 94.7±0.3

(e) Confusion matrix with KM for diesel mode using a Digital Twin ($n_v = 10$).

	Actual	
Prediction	TP 93.2±0.5	FP 6.8±0.5
	FN 7.4±0.4	TN 92.6±0.4

(f) Confusion matrix with NN for diesel mode using a Digital Twin ($n_v = 10$).

the experiments take into consideration only the diesel mode, merging in the same dataset both healthy and faulty engine's conditions. The three different models (RF, KM and NN) have been applied to distinguish between two possible engine's statuses: +1 for faulty condition and −1 for healthy condition. Table 5 reports the misclassification errors of the three models for the two different approaches under analysis (*Direct* and *Digital Twin*).

Looking at the percentages listed in Table 5a, the *Direct* approach outperforms the *Digital Twin* one in terms of misclassification errors for all the exploited models. Furthermore, KM demonstrates to perform better compared to RF and NN in accordance to what observed in Section 6.3.

In order to better represent the quality of the developed model, Tables 5b, 5c, 5d, 5e, 5f, and 5g report the confusion matrices for each of the considered models and for both the adopted approaches. By observing these confusion matrices it is possible to note that the misclassification errors are well distributed and models do not tend to predict more false positive than false negative. Also in this case the quality of the developed models is surely up to a level which is acceptable for their use in the wild with misclassification below the 5% (and in most case less than 3%). Note then that, switching from Scenario 1 to Scenario 2 does not compromise the ability to make accurate predictions.

6.5. Scenario 3 : Weakly supervised health status estimation

Table 6 reports the misclassification error percentages and the related confusion matrices for the third and last scenario. Diesel engine mode data and the Direct and Digital Twin approaches are exploited in this context, in accordance with previous scenario's experiments. In this case, the experimental setup is slightly different, as described in Section 6.1 by drastically reducing the need for labelled data.

For the Direct approach, OCSVM and GKNN models are taken into consideration and compared against the ones based on the Digital Twin (build with RF, KM and NN). In this setting, the validation set size (which is the only actual labelled one) is a parameter to be considered as another degree of freedom (which must be kept as small as possible to be able to apply this methodology in the wild).

Observing Table 6a, the Direct approach outperforms again the Digital Twin-based one. For the Direct approach, OCSVM model outperform the GKNN one. Meanwhile, for the Digital Twin-based approach, KM model confirms to outperform both RF and NN models. As expected, increasing the validation set size (the amount of labelled data), the models performance decrease. For the Direct approach the improvement is not so relevant, while for Digital Twin approach is more relevant. This confirm the higher ability of the Direct approach to deliver high performance with limited number of labelled samples.

Tables 6b, 6c, 6d, 6e and 6f report the confusion matrices of all the models under analysis for a validation set size equal to 10. Also in this case, observing these confusion matrices it is possible to note that the misclassification errors are well distributed and models do not tend to predict more false positive than false negative. Also in this case the quality of the developed models is surely up to a level which is acceptable for their use in the wild with misclassification below the 5% (and in most case less than 3%). Note then that, switching from Scenarios 1 and 2 to Scenarios 3 does not compromise the ability to make accurate predictions (the decrease in performance is less than 1%).

7. Conclusions

Marine dual fuel engines, running on both gaseous and liquid fuels, represent a viable way toward the reduction of emissions since maritime transportation accounts for around 80% of the world freight movements, remarkably contributing to the global environmental footprint. The side effect of this transition is in additional complexity in monitoring activities which are required to keep their performance always at the desired level. In fact, the degradation of marine engines performance decreases their operational efficiency, leading to a higher fuel consumption and consequently to an increase in greenhouse emissions. Moreover, losses in performance are often a precursor or signal for the necessity of maintenance activities.

In this paper, we focus on data-driven monitoring models to be employed in the wild. Unfortunately data-driven methods often require a large amount of labelled samples which are rarely available. For this reason, in this paper, we design and propose multiple alternatives toward the weakly supervised marine dual fuel engines data-driven monitoring. To this aim, we relied on a Digital Twin of the dual fuel engine or on novelty detection algorithms and compared them against state-of-the-art fully supervised approaches. The approach employed in this study includes three steps. The first step was characterised by a Fully Supervised Performance estimation scenario including the design of a Digital Twin, exploiting state-of-the-art supervised data-driven methods for enabling the prediction of the engine performance and emission parameters based on the control variables (e.g. engine load and engine speed), in healthy engine conditions. The second step consisted in the development of a Fully Supervised Health Status scenario, focusing on developing models capable of classifying the status of the engines as healthy or faulty, and it was accomplished by employing two approaches. The first one employed the Digital Twin developed in the first step to estimate the deviation (drift) of the parameters of the actual engine operation (based on the acquired data) from the respective Digital Twin predicted parameters. The second one exploited state-of-the-art supervised data-driven methods to classify the status of the investigated engine based on the control and performance parameters. Unfortunately, this approach required a large number of labelled samples to be implemented, which are rarely available in the wild. For this reason, in the third step a Weakly Supervised Health Status Estimation was proposed, focusing on reducing the amount of labelled data required to build the models developed in the second step by employing two approaches. The first one focused on the estimation of the engine performance parameters variation from the respective parameters calculated by employing the Digital Twin utilising a limited amount of labelled data for tuning the drift detection model. The second one, instead, exploited state-of-the-art unsupervised data-driven methods to detect abnormal conditions (anomalies) of the investigated engine by employing as input the considered control and performance parameters. The weakly supervised health status estimation step employed the models trained just on data acquired under the engine healthy conditions from the engine monitoring system. These models were subsequently fine tuned with a very small amount of labelled data.

Results on data generated from a real-data validated simulator of a marine dual fuel engine demonstrate that the proposed weakly supervised monitoring approaches lead to a negligible loss in accuracy compared with costly and often unfeasible fully supervised ones supporting the validity of the proposal for its application in the wild. In particular, in the Fully Supervised Performance estimation scenario, the error of the data-driven model is always less than 4% (and in most cases less than 2%) for all the performance variables in the considered modalities. This result is surely up to a level which is acceptable for the utilisation of data-driven models for dual fuel engine performance estimation. Considering the Fully Supervised Health Status scenario, the error of the fault detection models is always below the 5%, and in most cases less than 2% which is again suitable for real operational environment, but unfortunately it requires a number of labelled samples which is not realistic to obtain in the wild. Finally, in the Weakly Supervised Health Status Estimation scenario, we fill this gap by remarkably decreasing the amount of labelled samples necessary to train the model whilst obtaining an error below the 5% (and in most cases less than 3%) and not compromising the ability to make accurate predictions (the decrease in performance is less than 1%) for the use of this model in real operational conditions.

CRediT authorship contribution statement

Andrea Coraddu: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - reviewing and editing. **Luca Oneto:** Conceptualization, Methodology, Data curation, Writing - original draft, Writing - reviewing and editing, Software, Validation. **Davide Ilardi:** Data curation, Methodology, Writing - original draft. **Sokratis Stoumpos:** Data curation, Writing - original draft. **Gerasimos Theotokatos:** Supervision, Conceptualization, Reviewing, Writing - original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adams, R.A., Fournier, J.J.F., 2003. Sobolev Spaces. Elsevier.

Ahmed, R., El Sayed, M., Gadsden, S.A., Tjong, J., Habibi, S., 2015. Automotive internal-combustion-engine fault detection and classification using artificial neural network techniques. IEEE Trans. Veh. Technol. 64 (1), 21–33.

Antony, J., 2014. Design of Experiments for Engineers and Scientists. Elsevier.

Bae, C., Kim, J., 2017. Alternative fuels for internal combustion engines. Proc. Combust. Institut. 36 (3), 3389–3413.

Baldi, F., Theotokatos, G., Andersson, K., 2015. Development of a combined mean value-zero dimensional model and application for a large marine four-stroke Diesel engine simulation. Appl. Energy 154, 402–415.

Boretti, A., 2019. Advantages and disadvantages of diesel single and dual-fuel engines. Front. Mech. Eng. 5, 64.

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

Brzozowska, L., Brzozowski, K., Nowakowski, J., 2005. An application of artificial neural network to diesel engine modelling. In: IEEE Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications.

Cai, C., Weng, X., Zhang, C., 2017. A novel approach for marine diesel engine fault diagnosis. Cluster Comput. 20 (2), 1691–1702.

CAT, 2019. DICARE - system for diesel engine diagnostic and predicitive maintenance. www.t.ly/uqrq.

Christen, C., Brand, D., 2013. IMO tier 3: gas and dual fuel engines as a clean and efficient solution. In: Conseil International Des Machines a Combustion Congress.

Cipollini, F., Oneto, L., Coraddu, A., Murphy, A.J., Anguita, D., 2018a. Condition-based maintenance of naval propulsion systems: Data analysis with minimal feedback. Reliab. Eng. Syst. Saf. 177, 12–23.

Cipollini, F., Oneto, L., Coraddu, A., Murphy, A.J., Anguita, D., 2018b. Condition-based maintenance of naval propulsion systems with supervised data analysis. Ocean Eng. 149, 268–278.

Coraddu, A., Oneto, L., Baldi, F., Cipollini, F., Atlar, M., Savio, S., 2019. Data-driven ship digital twin for estimating the speed loss caused by the marine fouling. Ocean Eng. 186, 106063.

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.

Gamma-Technologies, 2019. GT-SUITE Manual. https://www.gtisoft.com/.

Geertsma, R.D., Negenborn, R.R., Visser, K., Loonstijn, M.A., Hopman, J.J., 2017. Pitch control for ships with diesel mechanical and hybrid propulsion: Modelling, validation and performance quantification. Appl. Energy 206, 1609–1631.

Geertsma, R.D., Visser, K., Negenborn, R.R., 2018. Adaptive pitch control for ships with diesel mechanical and hybrid propulsion. Appl. Energy 228, 2490–2509.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Gratsos, G.A., Psaraftis, H.N., Zachariadis, P., 2009. Life cycle cost of maintaining the effectiveness of a ship's structure and environmental impact of ship design parameters: an update. In: RINA Conference on the Design and Operation of Bulk Carriers, Athens, Greece.

Grimmelius, H., Mesbahi, E., Schulten, P., Stapersma, D., 2007. The use of diesel engine simulation models in ship propulsion plant design and operation. In: Conseil International Des Machines a Combustion Congress.

Hanson, R.K., Salimian, S., 1984. Survey of rate constants in the n/h/o system. In: Combustion Chemistry.

Hansson, J., Månsson, S., Brynolf, S., Grahn, M., 2019. Alternative marine fuels: Prospects based on multi-criteria decision analysis involving Swedish stakeholders. Biomass Bioenergy 126, 159–173.

He, Y., Rutland, C.J., 2004. Application of artificial neural networks in engine modelling. Int. J. Engine Res. 5 (4), 281–296.

Herceg, M., Raff, T., Findeisen, R., Allgowe, F., 2006. Nonlinear model predictive control of a turbocharged diesel engine. In: IEEE Conference on Computer Aided Control System Design, IEEE International Conference on Control Applications, IEEE International Symposium on Intelligent Control.

Heywood, J.B., 1988. Combustion Engine Fundamentals. McGraw-Hill Education.

Hsieh, M.F., Wang, J., 2009. Diesel engine selective catalytic reduction (SCR) ammonia surface coverage control using a computationally-efficient model predictive control assisted method. In: ASME Dynamic Systems and Control Conference.

Jiaqiang, E., Zhao, X., Liu, G., Zhang, B., Zuo, Q., Wei, K., Li, H., Han, D., Gong, J., 2019. Effects analysis on optimal microwave energy consumption in the heating process of composite regeneration for the diesel particulate filter. Appl. Energy 254, 113736.

Joshi, A.A., James, S., Meckl, P., King, G., Jennings, K., 2009. Assessment of charge-air cooler health in diesel engines using nonlinear time series analysis of intake manifold temperature. J. Dynam. Syst. Measur. Control 131 (4).

Karim, G.A., 2015. Dual-fuel diesel engines. CRC Press.

Karlsson, M., Ekholm, K., Strandh, P., Johansson, R., Tunestål, P., 2010. Multiple-input multiple-output model predictive control of a diesel engine. IFAC Proc. Vol. 43 (7), 131–136.

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

Kowalski, J., Krawczyk, B., Woźniak, M., 2017. Fault diagnosis of marine 4-stroke diesel engines using a one-vs-one extreme learning ensemble. Eng. Appl. Artif. Intell. 57, 134–141.

Lavoie, G.A., Heywood, J.B., Keck, J.C., 1970. Experimental and theoretical study of nitric oxide formation in internal combustion engines. Combust. Sci. Technol. 1 (4), 313–326.

Lebedevas, S., Pukalskas, S., Daukšys, V., 2020. Mathematical modelling of indicative process parameters of dual-fuel engines with conventional fuel injection system. Transport 35 (1), 57–67.

Li, Z., Yan, X., Guo, Z., Liu, P., Yuan, C., Peng, Z., 2012. A new intelligent fusion method of multi-dimensional sensors and its application to tribo-system fault diagnosis of marine diesel engines. Tribol. Lett. 47 (1), 1–15.

Liu, Z., Zuo, Q., Wu, G., Li, Y., 2018. An artificial neural network developed for predicting of performance and emissions of a spark ignition engine fueled with butanol-gasoline blends. Adv. Mech. Eng. 10 (1), 1687814017748438.

Lloyd, A.C., Cackette, T.A., 2001. Diesel engines: environmental impact and control. J. Air Waste Manage. Assoc. 51 (6), 809–847.

Louis, J., 2001. Well-to-wheel energy use and greenhouse gas emissions for various vehicle technologies. SAE Trans. 1682–1689.

MAN, 2019. Cocos EDS - engine diagnostics system. www.t.ly/KW5Q.

Mangan, J., 2017. Future of the sea: trends in the transport of goods by sea. Futur. Sea.

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., Bezirtzoglou, E., 2020. Environmental and health impacts of air pollution: A review. Front. Publ. Health 8.

Merker, G.P., Schwarz, C., Stiesch, G., Otto, F., 2005. Simulating Combustion: Simulation of Combustion and Pollutant Formation for Engine-Development. Springer Science & Business Media.

Muchiri, P.N., Pintelon, L., Martin, H., Chemweno, P., 2014. Modelling maintenance effects on manufacturing equipment performance: results from simulation analysis. Int. J. Prod. Res. 52 (11), 3287–3302.

Müller-Steinhagen, H., 1999. Cooling-water fouling in heat exchangers. In: Advances in Heat Transfer.

Munim, Z.H., Dushenko, M., Jimenez, V.J., Shakil, M.H., Imset, M., 2020. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Marit. Policy Manage. 47 (5), 577–597.

Nixon, S., Weichel, R., Reichard, K., Kozlowski, J., 2018. A machine learning approach to diesel engine health prognostics using engine controller data. In: Annual Conference of the PHM Society.

Oneto, L., 2020. Model Selection and Error Estimation in a Nutshell. Springer.

Özener, O., Yüksek, L., Özkan, M., 2013. Artificial neural network approach to predicting engine-out emissions and performance parameters of a turbo charged diesel engine. Therm. Sci. 17 (1), 153–166.

Pavlenko, N., Comer, B., Zhou, Y., Clark, N., Rutherford, D., 2020. The climate implications of using LNG as a marine fuel. In: International Council on Clean Transportation, Washington DC.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. 2 (1), 37–63.

Qian, Y., Wang, J., Li, Z., Jiang, C., He, Z., Yu, L., Lu, X., 2020. Improvement of combustion performance and emissions in a gasoline direct injection (GDI) engine by modulation of fuel volatility. Fuel 268, 117369.

Rakopoulos, C.D., Giakoumis, E.G., 2007. Prediction of friction development during transient diesel engine operation using a detailed model. Int. J. Veh. Des. 44 (1–2), 143–166.

Ramaswamy, S., Rastogi, R., Shim, K., 2000. Efficient algorithms for mining outliers from large data sets. In: ACM Sigmod Record.

Reitz, R.D., Ogawa, H., Payri, R., Fansler, T., Kokjohn, S., Moriyoshi, Y., Agarwal, A., Arcoumanis, D., Assanis, D., Bae, C., et al., 2020. IJER Editorial: The Future of the Internal Combustion Engine. SAGE Publications Sage UK, London, England.

Reitz, R.D., Rutland, C.J., 1995. Development and testing of diesel engine CFD models. Prog. Energy Combust. Sci. 21 (2), 173–196.

Scholkopf, B., 2001. The kernel trick for distances. In: Advances in Neural Information Processing Systems.

Shalev-Shwartz, S., Ben-David, S., 2014. Cambridge university press.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge university press.

Shayler, P.J., Goodman, M., Ma, T., 2000. The exploitation of neural networks in automotive engine management systems. Eng. Appl. Artif. Intell. 13 (2), 147–157.

Shin, S., Lee, Y., Kim, M., Park, J., Lee, S., Min, K., 2020. Deep neural network model with Bayesian hyperparameter optimization for prediction of NOx at transient conditions in a diesel engine. Eng. Appl. Artif. Intell. 94, 103761.

Sitkei, G., 2013. Kraftstoffaufbereitung und Verbrennung bei Dieselmotoren. Springer-Verlag.

Sixel, E.J., Hiltner, J., Rickert, C., 2016. Use of 1-D simulation tools with a physical combustion model for the development of Diesel-Gas or DF engines. In: Conseil International Des Machines a Combustion Congress.

Song, R., Huang, L., Xue, Y., Vanthienen, J., 2018. A data-driven fault prediction method for LNG engine city buses. In: International Conference on Logistics, Informatics and Service Sciences.

Stewart, G., Borrelli, F., 2008. A model predictive control framework for industrial turbodiesel engine control. In: IEEE Conference on Decision and Control.

Stoumpos, S., Theotokatos, G., Boulougouris, E., Vassalos, D., Lazakis, I., Livanos, G., 2018. Marine dual fuel engine modelling and parametric investigation of engine settings effect on performance-emissions trade-offs. Ocean Eng. 157, 376–386.

Stoumpos, S., Theotokatos, G., Mavrelos, C., Boulougouris, E., 2020. Towards marine dual fuel engines digital twins-integrated modelling of thermodynamic processes and control system functions. J. Mar. Sci. Eng. 8 (3), 200.

Swersky, L., Marques, H.O., Sander, J., Campello, R.J.G.B., Zimek, A., 2016. On the evaluation of outlier detection and one-class classification methods. In: IEEE International Conference on Data Science and Advanced Analytics.

Syed, J., Baig, R.U., Algarni, S., Murthy, S., Masood, M., Inamurrahman, M., 2017. Artificial neural network modeling of a hydrogen dual fueled diesel engine characteristics: An experiment approach. Int. J. Hydrogen Energy 42 (21), 14750–14774.

Talaat, M., Gobran, M.H., Wasfi, M., 2018. A hybrid model of an artificial neural network with thermodynamic model for system diagnosis of electrical power plant gas turbine. Eng. Appl. Artif. Intell. 68, 222–235.

Theotokatos, G., Stoumpos, S., Bolbot, V., Boulougouris, E., 2020. Simulation-based investigation of a marine dual-fuel engine. J. Mar. Eng. Technol. 19, 5–16.

Thomson, H., Corbett, J.J., Winebrake, J.J., 2015. Natural gas as a marine fuel. Energy Policy 87, 153–167.

Turkson, R.F., Yan, F., Ali, M.K.A., Hu, J., 2016. Artificial neural network applications in the calibration of spark-ignition engines: An overview. Eng. Sci. Technol. Int. J. 19 (3), 1346–1359.

Wainberg, M., Alipanahi, B., Frey, B.J., 2016. Are random forests truly the best classifiers? J. Mach. Learn. Res. 17 (1), 3837–3841.

Wakode, V.R., Kanase-Patil, A.B., 2017. Regression analysis and optimization of diesel engine performance for change in fuel injection pressure and compression ratio. Appl. Therm. Eng. 113, 322–333.

Wang, H., Kolmanovsky, I., Sun, J., 2016. Set-membership condition monitoring framework for dual fuel engines. In: American Control Conference.

Wang, Y.S., Liu, N.N., Guo, H., Wang, X.L., 2020. An engine-fault-diagnosis system based on sound intensity analysis and wavelet packet pre-processing neural network. Eng. Appl. Artif. Intell. 94, 103765.

Wärtsilä, 2012. Wärtsilä 50DF Product Guide. http://cdn.wartsila.com/docs/default-source/product-files/engines/dfengine/product-guide-oe-w50df.pdf.

Wolpert, D.H., 1996. The lack of a priori distinctions between learning algorithms. Neural Comput. 8 (7), 1341–1390.

Woschni, G., 1967. A universally applicable equation for the instantaneous heat transfer coefficient in the internal combustion engine. In: SAE Technical Paper.

Xi, W., Li, Z., Tian, Z., Duan, Z., 2018. A feature extraction and visualization method for fault detection of marine diesel engines. Measurement 116, 429–437.

Xiang, L., Theotokatos, G., Ding, Y., 2019. Investigation on gaseous fuels interchangeability with an extended zero-dimensional engine model. Energy Convers. Manage. 183, 500–514.

Xu, S., Anderson, D., Hoffman, M., Prucka, R., Filipi, Z., 2017. A phenomenological combustion analysis of a dual-fuel natural-gas diesel engine. Proc. Inst. Mech. Eng. D 231 (1), 66–83.

Xu, K., Tang, L.C., Xie, M., Ho, S.L., Zhu, M.L., 2002. Fuzzy assessment of FMEA for engine systems. Reliab. Eng. Syst. Saf. 75 (1), 17–29.

Zabihi-Hesari, A., Ansari-Rad, S., Shirazi, F.A., Ayati, M., 2019. Fault detection and diagnosis of a 12-cylinder trainset diesel engine based on vibration signature analysis and neural network. Proc. Institut. Mech. Eng. C 233 (6), 1910–1923.

Zhai, Y.J., Yu, D.L., 2009. Neural network model-based automotive engine air/fuel ratio control and robustness evaluation. Eng. Appl. Artif. Intell. 22 (2), 171–180.

Zhao, D., Liu, C., Stobart, R., Deng, J., Winward, E., Dong, G., 2013. An explicit model predictive control framework for turbocharged diesel engines. IEEE Trans. Ind. Electron. 61 (7), 3540–3552.

Zhong, W., Pachiannan, T., Liu, Q., Li, Z., Qian, Y., Zhang, Y., He, Z., Lu, X., 2020. Experimental study the effect of injection strategies on combustion and emission characteristics in gasoline compression ignition engines using gasoline/hydrogenated catalytic biodiesel blends. Fuel 278, 118156.

Zuo, Q., Xie, Y., Guan, Q., Zhu, G., Jiaqiang, E., Zhu, X., Tang, Y., Wang, Z., Chen, W., 2020a. Effect of critical dual-carrier structure parameters on performance enhancement of a dual-carrier catalytic converter and the gasoline engine system. Energy Convers. Manage. 204, 112325.

Zuo, Q., Xie, Y., Jiaqiang, E., Zhu, X., Zhang, B., Tang, Y., Zhu, G., Wang, Z., Zhang, J., 2020b. Effect of different exhaust parameters on NO conversion efficiency enhancement of a dual-carrier catalytic converter in the gasoline engine. Energy 191, 116521.