JOURNAL OF MARINE ENGINEERING & TECHNOLOGY
2025, VOL. 24, NO. 1, 70–82
https://doi.org/10.1080/20464177.2024.2432777

Taylor & Francis
Taylor & Francis Group

Data-driven model for marine engine fault diagnosis using in-cylinder pressure signals

Chaitanya Patil, Gerasimos Theotokatos and Konstantinos Tsitsilonis

Maritime Safety Research Centre, Department of Naval Architecture Ocean and Marine Engineering, University of Strathclyde, Glasgow, UK

ABSTRACT

Effective diagnosis of marine engines that is crucial for safe and reliable ship operations requires fidel tools for the identification of critical faults. However, the unavailability of extensive measured data-sets corresponding to engine faulty conditions renders the development of such tools challenging. This study aims to develop a data-driven fault estimation model considering information extraction methods and regression techniques of low computational effort, namely multiple linear and polynomial regression. The required data-sets for healthy and faulty conditions for four critical faults and their combinations are generated by employing a calibrated zero-dimensional thermodynamic model representing a marine four stroke medium speed diesel engine, which is validated against engine shop trial measurements. Fourier analysis of the derived in-cylinder pressure profiles is employed to calculate the coefficients of the harmonic orders. Several harmonics coefficients sets are used as input to the regression models to estimate the severity of the four considered faults. The results demonstrate that initial 20 harmonics are sufficient to effectively estimate the severity for each fault, whereas polynomial regression is highly effective, exhibiting R^2 values greater than 98%. This study provides insights on the data-driven simultaneous faults severity estimation, and as such it impacts the advancement of cost-effective diagnostic methods for marine engines.

ARTICLE HISTORY
Received 4 September 2023
Accepted 19 September 2024

KEYWORDS
Fault diagnostics; marine engines; dimensionality reduction; fourier analysis; data-driven regression model

1. Introduction

1.1. Background

Forthcoming technological developments in maritime industry are driven by the 4th Industrial revolution (industry 4.0) concepts and tools. Several state-of-the-art technologies have been identified as top drivers including smart systems, advanced materials, big data analytics, robotics, sensors, and communications. Future developments in the shipping industry include the use artificial intelligence (AI) techniques and machine learning (ML) (Meier et al. 2019).

Fault diagnosis for marine engines and systems has progressed from approaches employing human expertise to smart and intelligent methods (Hou et al. 2020). The intelligent fault diagnosis typically relies on methods from Machine Learning (ML) and Artificial Intelligence (AI), which have been attracting interest from various industries including the shipping industry to support diagnostics and prognostics of systems in a sustainable way.

Machine learning methods include supervised and unsupervised learning of ship systems anomalies. Supervised learning methods for fault diagnosis employ classification models to categorise the faults and regression models to quantify the faults severity. A comprehensive review on the condition monitoring and fault diagnosis for diesel engines is reported in Jones and Li (2000). Rao et al. (2022) reviewed the real-time condition monitoring for cylinder liner piston rings for diesel engines. Various machine learning techniques are employed to detect failures in ship machinery (Elmdoost-gashti et al. 2023) and provide recommendations for appropriate maintenance activities (Karatuğ et al. 2023). Several studies focusing on faults or failure detection for engine subsystems, parts, actuators, sensors are published in the pertinent literature. The state of the art of fault diagnostic methods as well as their applications for detecting faults and degradation of ship mechanical and electrical systems is discussed in the next section.

1.2. Literature review

Data-driven techniques using supervised classification algorithms are employed for condition monitoring of engine subsystems. Hou et al. (2020) proposed a fault classification using Support Vector Machines (SVM) for the fuel oil supply system. A fault localisation algorithm using association rule mining for marine engine subsystems was developed by Cai et al. (2017) based on a historical faults database. Jung et al. (2018) reported a comparative study on the classification techniques employed to detect common faults with varying severity in a fuel system.

Artificial Neural Networks (ANN) are extensively employed for supervised fault classification. Mesbahi et al. (2005) introduced an ANN-based model to predict faults associated with the turbine and condenser in steam power plants. Raza and Liyanage (2009) developed an algorithm based on ANNs to clasify Faults related to oil pumps faults. Stoumpos and Theotokatos (2022) proposed an ANN-based approach to forecast sensor failures in dual fuel marine engines. Velasco-Gallego and Lazakis (2022) developed a methodology combining a first-order Markov chain with both ResNet50V2 and Convolutional Neural Networks (CNN) to detect anomalies in diesel generator sets through time series imaging. A methodology based on Recursive Neural Networks (RNN) was demonstrated by Senemmar and Zhang (2021) to simultaneously locate and classify faults in ship systems.

Ensemble models that integrate various machine learning techniques were developed for fault detection in marine engines. Tsaganos

CONTACT Chaitanya Patil chaitanya.patil@strath.ac.uk Maritime Safety Research Centre, Department of Naval Architecture Ocean and Marine Engineering, University of Strathclyde, Glasgow G4 0LZ, UK

et al. (2020) employed an ensemble (multi-technique) model to classify common faults, such as fuel and air leakage for marine two-stroke engines. An extreme learning ensemble was developed by Kowalski et al. (2017) to classify 14 faults for a four-stroke diesel engine. Wang et al. (2020) proposed a fault diagnosis framework combining both supervised and unsupervised algorithms to classify a limited number of engine faults at specific operating points. Perera (2016) demonstrated the effectiveness of a localised model with data point clustering using Gaussian mixture sub-models and an expectation-maximisation algorithm for fault detection in marine engines. Several techniques, including One Class Support Vector Machine (OCSVM), Support Vector Data Description (SVDD), Global k-nearest neighbors (GKNN), Local Outlier Factor (LOF), Isolation Factor (IF), and Angle Based Outlier Detection (ABOD), were investigated and validated for a marine machinery system by Tan et al. (2020). Furthermore, multi-class classification and simultaneous fault diagnosis using Binary Relevance (BR), Classifier Chains (CC), Multi-Label k-Nearest Neighbors (MLKNN), Binary Relevance k-Nearest Neighbors (BRKNN), and Multi-Label Twin Support Vector Machines (MLTSVM) were presented by Tan et al. (2021).

Machine learning models dealing with a single parameter or signal are frequently employed for identifying specific engine faults. These models require less computational power while providing more information through instantaneous data acquisition. Krogerus et al. (2018) analysed the common rail pressure signal in dual fuel engines to diagnose fuel injection faults. Coraddu et al. (2022) proposed a condition monitoring model based on the exhaust temperature prediction in various operating conditions. Vibration signals measured from engine subsystems and components also exhibit potential for diagnosing combustion engine faults. Xi et al. (2018) developed a classifier to reduce human errors using independent component analysis (ICA) on vibration data collected from multiple engine locations. Zabihi-Hesari et al. (2019) proposed a fault diagnosis method for 12-cylinder engines using vibration signature analysis and a neural network. Ayati et al. (2020) focused on classifying fuel injection faults based on vibration signatures. Feature extraction methods were developed to obtain required information for fault diagnosis from instantaneous signals, further reducing computational time. For instance, Zhang et al. (2018) developed a feature extraction method for instantaneous crankshaft speed to detect misfires in engine cylinders at three loads.

This study focuses on the extraction of pertinent information from acquired in-cylinder pressure signals, and identification of the following faults: blowby leakage, increased friction, injection issues, and charge air cooler fouling in marine four-stroke engines. Several levels of severity for these four faults as well as their combinations were considered at the whole operating range of the investigated marine engine.

1.3. Challenges and research gaps in fault diagnosis

The methods discussed in the preceding literature review require tremendous amount of data for training. Most of the ships (especially older) do not provide access to the engine manufacturer monitoring system rendering the data acquisition challenging. Thermodynamic models on the other hand are useful to generate the required realistic data-sets, which can subsequently be used to train data-driven diagnostic models (Altosole et al. 2022). The dimensionality reduction approach (Mohammad et al. 2022) can lower the number of required input/features, resulting in considerably reducing the training data amount and thus computational effort for data-driven models, without deteriorating their accuracy. When the in-cylinder pressure is used as input, the dimensions depend on the sampling rate of the pressure signal; typically, samples are acquired every $0.5 - 1$ deg CA. An effective method is required to reduce these signals dimensions and extract the required information for the engine health assessment.

Previous studies only focused on particular operating points of the investigated marine engines and systems to identify faults. However, the whole operating envelope of the investigated marine engine must be considered. The pertinent literature focused on detecting specific faults on marine engines and their systems. However, the capabilities of the engine diagnostics methods considering multiple faults combinations must be assessed. The development of fault classification models was the main focus of previous studies, whereas very few studies focused on quantifying the faults severity. Yet, the degree of fault severity is required to provide the marine engine health status. Computationally expensive ANNs were typically employed for estimating the combustion related faults by Patil et al. (2023). However, the development of shipboard fault diagnosis systems requires the use of computationally inexpensive methods, such as simple regression.

1.4. Research aim

To overcome the preceding challenges, this study aims at developing a data-driven fault estimation model for the following four critical faults and their combinations: charge air cooler (CAC) fouling, cylinders blowby, increased friction losses, and fuel injection timing. The model consists of three blocks, namely input dimensions reduction, data standardisation and regression, and estimates the engine faults and their severity. A large marine four-stroke engine is investigated. The in-cylinder pressures, the engine speed and load are the required parameters for developing and using the data-driven fault estimation model. This study proposes the processing of the in-cylinder pressure signals by employing Fourier analysis to reduce the input dimensions. Additionally, a feature standardisation method is employed, whereas the K-fold splitting technique is used to split the testing and validation data-sets. Two multivariate linear regression techniques are tested to select the one with the highest accuracy. Moreover, two parametric studies are performed; the former is employed to identify the number of Fourier coefficients that provide adequate regression model accuracy; the latter is employed to investigate the impact of the test-to-train ratio on the regression model accuracy.

This study employs a zero-dimensional thermodynamic model to generate the required data for the engine healthy and faulty conditions. It must be noted that although healthy data can be acquired from measurements (shop trial and shipboard), it is not possible to obtain faulty data in the complete range of scenarios with multiple simultaneous faults and/or varying fault severity. This model is termed thermodynamic model henceforth to differentiate from the data-driven fault estimation model.

Thenovelty of this study stems from: (a) the prediction of the severity levels of the fault for each individual cylinder considering the complete operating envelope of the investigated marine engine; (b) the use of Fourier analysis as a dimensionality reduction technique, providing the harmonic orders coefficients as input of the regression model instead of the in-cylinder pressure data points; (c) the consideration of four critical faults and their combinations with varying severity; (d) the characterisation of the accuracy of the regression techniques and its dependence on the harmonic orders number.

2. Methodology

Figure 1 presents the methodology flow diagram for developing the data-driven model for each fault (top box), as well as its proposed

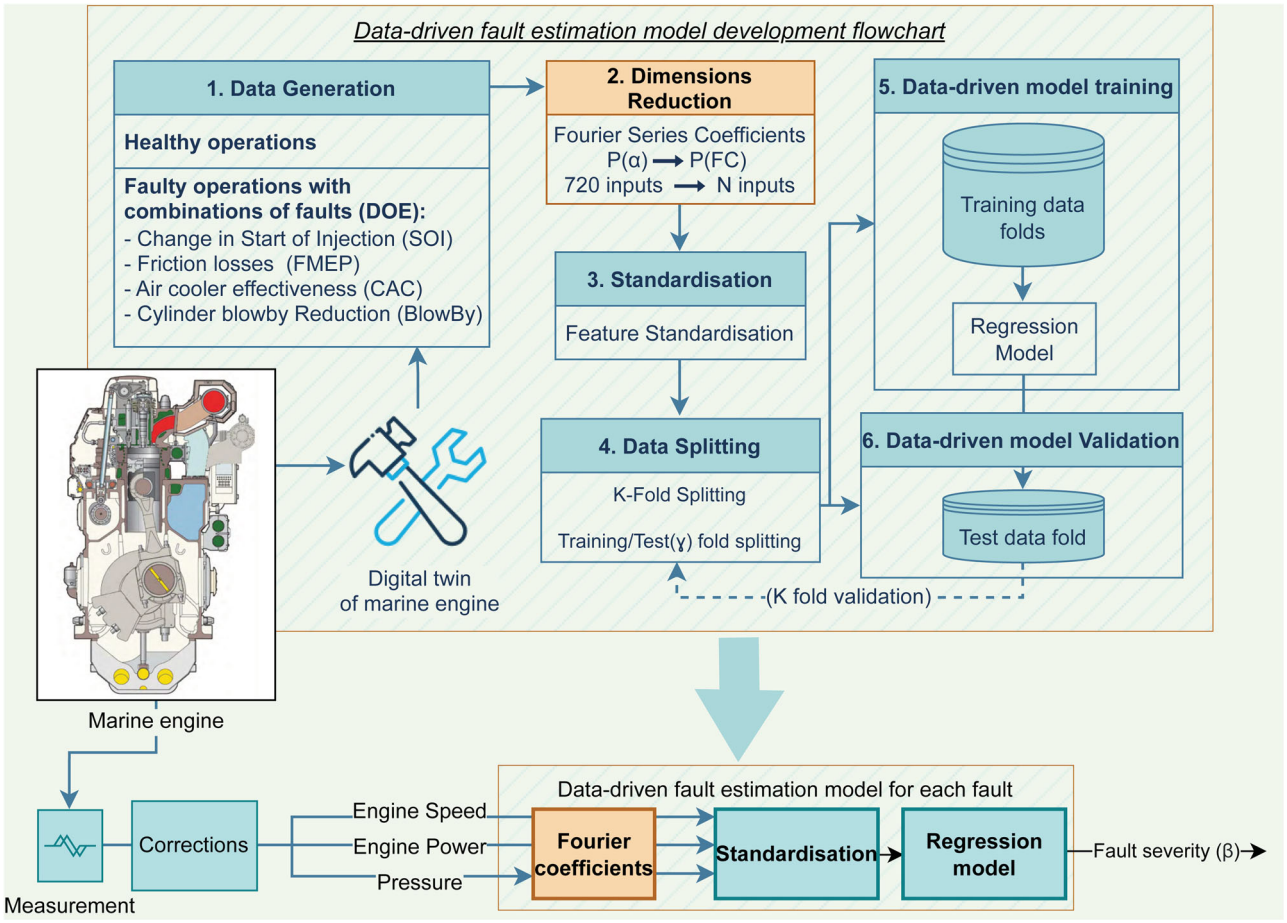


Figure 1. Methodology flowchart (top box) along with data-driven model application at the bottom.

application with shipboard measurements (bottom box). The generalised methodology to develop a fault estimation model for each fault consists of the following six steps:

Step 1 Data generation: A validated thermodynamic model of the investigated marine engine is employed to generate healthy and faulty data-sets as reported in Subsection 2.1. Several operating scenarios with combinations of four frequent faults of varying severity are considered. A design of experiment (DOE) approach is followed to perform simulation runs in the entire operating envelope and considering combinations of faults and their severity, thus generating the in-cylinder pressure profiles.

Step 2 Dimensional reduction: This step employs Fourier analysis to convert the in-cylinder pressure signals to harmonic-order coefficients. Considering the coefficients of the first N harmonic orders, the input dimensions reduce from 720 (considering 1^oCA sampling rate and four-stroke engine) to 2N+1 for each operating point. The detailed description of this method is provided in Subsection 2.2.

Step 3 Data standardisation: The standardisation of the generated data sets considering both the input and output parameters of the fault estimation model is carried out by centering to their mean values. This step is described in Subsection 2.3,

Step 4 Data splitting: Firstly, the data is separated in k-folds with equal distribution of samples for all faults. Secondly, the folds are classified into training and validation folds by the use of the test to train ratio (γ), which is provided as an input parameter. This step is described in Subsection 2.4.

Step 5 Regression model training: Multivariate linear and polynomial regression models are developed and trained by optimising the loss function described in 2.5 by employing the training data selected in step 4.

Step 6 Regression model validation: The trained model is validated against the validation data selected in Step 4.

Steps 4, 5 and 6 are repeated K times for cross-validation with one fold being used as the validation data set during training Section 2.5. The most accurate regression model is selected among the developed multivariate and polynomial regression models based on a parametric study with varying number of *Fourier series coefficients*. A second parametric study based on the test-to-train ratio (γ) is performed on the selected model to assess the impact on its accuracy.

2.1. Data-sets generation

2.1.1. Investigated marine engine

The engine used for this study is the four-stroke Wärtsilä 9L46C marine engine, which is a nine-cylinder turbocharged medium speed engine. In particular, this engine is installed in a RoPax ferry, the propulsion plant of which consists of two identical engines, each one driving a controllable pitch propeller and a power take-off generator through a gearbox. The main particulars of this engine are listed in Table 1 (Wartsila 2001).

The employed thermodynamic model uses semi-phenomenological models and widely acknowledged formulae to represent the engine processes. The state variables at cylinders and manifolds are calculated by employing the differential equations corresponding to

Table 1. Investigated marine engine specifications.

Maximum Continuous Rating point	9,450 kW @ 500 r/min
No. of Cylinders	9
Cylinder Bore	460 mm
Clutch-in Speed	300 r/min
Turbocharger	ABB TPL 77-A30

the mass and energy conservation, whereas the pressure is calculated by the ideal gas law. The Woschni-Anisits combustion model that is based on a single Wiebe function is employed for the heat release calculation. The Woschni equation is employed for calculating the heat transfer coefficient between the gas and the cylinder walls. The Chen-Flynn model (Chen and Flynn 1965) is used for the estimation for the friction mean effective pressure (FMEP). The turbocharger compressor and turbine are modeled by considering the digital forms of the compressor and turbine maps. The shaft speed of the engine and turbocharger is calculated by considering the respective angular momentum conservation equations. The detailed description of the employed thermodynamic model is provided in Tsitsilonis et al. (2021), Tsitsilonis and Theotokatos (2021) and Tsitsilonis et al. (2022).

The zero-dimensional thermodynamic DT for the investigated engine is calibrated and validated by considering the steady state measured parameters from the engine shop trials (factory acceptance tests). DT calibration involves the determination of the combustion and Friction Mean Effective Pressure (FMEP) parameters for the reference operating point, which is followed by the calibration of the Woschni-Anisits model constants considering all remaining shop test operating points. The employed DT calibration and validation processes for the investigated marine engine are described in Tsitsilonis et al. (2021), Tsitsilonis and Theotokatos (2021) and Tsitsilonis and Theotokatos (2022). This DT was further validated against in-cylinder pressure measurements acquired during the normal ship operation at five steady state operating points.

Subsequently, the validated DT is employed to generate the required data-sets considering the investigated engine operation at both healthy and faulty conditions; the latter are associated with several engine components faults and degradation. As mentioned in the methodology section, combinations of the previously identified four most frequently occurring faults/degradation (for which experimental data collection remains challenging (Lamaris and Hountalas 2010; Hountalas 2000; Rubio et al. 2018)) with varying severity are considered (Banisoleiman and Rattenbury 2006). The DT was employed to obtain the engine performance parameters including the in-cylinder pressure profiles. These four most common faults/degradations are as follows.

(1) **Fuel injection issues**:The advance and retard of the start of injection (SOI) occurs due to wear and tear of the fuel injection system (pumps and injectors), which results in deteriorating the engine performance. This study employs the SOI advance and retard to represent the fuel injection fault. Other injection issues, for example, worn injectors, were kept out of the scope of this study, as more detailed modelling approaches along with experimental data are required. Such cases can be investigated in future studies. The SOI is an input parameter of the thermodynamic model.

(2) **Increased friction losses**: Increased engine friction losses due to lubricating oil quality, or excessive bearing wear and tear result in reducing the engine brake power. This fault is induced in the DT by adjusting the Friction Mean Effective Pressure (FMEP).

(3) **Blowby**: Leakages of exhaust gases from the cylinder due to wear and tear in cylinder liner and piston rings cause the engine performance deterioration. This fault is modelled by using the blowby area (according to Table 2) and the nozzle flow equation to calculate the leakage gas mass flow rate (Perceau et al. 2022).

(4) **Charge Air Cooler (CAC) fouling**: This fault is attributed to the deposits and fouling in both sides (air, cooling water) of the CAC, and results in reducing the heat transfer rate from the air to the cooling water. It is modelled by reducing the charge air cooler effectiveness.

As faults and degradation are uncertain phenomena occurring in the systems and components, their combinations and varying severity may arise during the engine lifetime. To capture these combinations and generate the required data-sets for the fault estimation model, the full factorial design method reported in Cavazzuti (2013) is employed to model the four faults with five severity levels for the FMEP, CAC and blowby faults, and 9 levels for the SOI fault. The fault severity is represented by β_{fault}. The healthy conditions correspond to $\beta_{fault} = 0$, whereas the most severe fault corresponds to $\beta_{fault} = 0.6$ for the FMEP, CAC and blowby faults, as well as $\beta_{fault} = \pm 0.6$ for the most extreme SOI retard and advance, respectively.

Increments of (β_{fault}) by 0.15 were considered to represent the faulty conditions with several levels of severity. The increment of 0.15 may seem high for certain faults like FMEP and blowby, as these faults symptoms can be perceived easily by visual/audible observation. However, the preceding wider ranges are considered, as this study objective is to assess the potential of Fourier coefficients to diagnose specific faults and their combinations.

Table 2 lists the input parameters for the considered four faults/degradation, the equations to estimate these parameters, and the ranges of the corresponding β. In total, 1,125 ($5 \times 5 \times 5 \times 9$) combinations were generated with the full factorial design method for each operating point at steady state conditions for the investigated marine engine.

To retain the engine power output almost constant with varying faults severity, the engine injected fuel amount was adjusted by using a PID controller. It should be noted that this study considers all the engine cylinders being at the same health condition. The cases of different faults for different cylinders is out of this study scope. It was assumed that the acquired in-cylinder pressure will be pre-processed to remove offsets according to the method reported in Tsitsilonis and Theotokatos (2018), and subsequently corrected according to ISO3016 and manufacturers guidelines. Hence, the

Table 2. Faults, model input parameters, and corresponding severity.

Faults	Input parameter	Parameter Calculation	Range of β_{fault}
SOI	Start of injection (SOI)	$\theta_{SOI,\text{faulty}} = \theta_{SOI,\text{healthy}}(1 \pm \beta_{SOI})$	−0.6 – 0.6 (9 points)
FMEP	Friction Mean Effective Pressure (fmep)	$fmep_{\text{faulty}} = fmep_{\text{healthy}}(1 + \beta_{FMEP})$	0–0.6 (5 points)
Blowby	Blowby area (A_{BB})	$A_{BB} = \pi \dfrac{[D_{\text{cyl}}^2 - (D_{\text{cyl}} - 2t^{\text{a}})^2]}{4} \beta_{\text{BlowBy}}$	0–0.6 (5 points)
CAC	CAC effectiveness (η_{AC})	$\eta_{AC,\text{faulty}} = \eta_{AC,\text{healthy}}(1 - \beta_{CAC})$	0–0.6 (5 points)

[a]t: gap thickness.

ambient conditions impact on the engine performance parameters was not considered in this study.

To facilitate the fault/degradation analysis over the whole engine operating envelope, 50 operating points corresponding to steady state conditions were considered. These operating points are illustrated in Figure 2. Cumulatively, 56,250 (1125 × 50) simulation runs are performed for the considered fault combinations and severity levels, thus representing the conditions occurring on the investigated marine engine throughout its lifetime. The impact of the different faults combination on the engine performance parameters is presented and discussed in Subsection 3.1.

2.2. Dimensional reduction using Fourier coefficients

The in-cylinder pressure diagram conveys crucial information characterising the engine processes and their inter-dependencies, as reported in Tsitsilonis et al. (2021). To effectively capture the pressure diagram for a four stroke diesel engine with cycle duration of 720°CA, 720 points corresponding to 1°CA are typically required.

Fourier series (Shatkay 1995) is a method to represent continuous periodic signals through superposition of harmonic orders represented by sine and cosine terms. This method is useful to assess the complete in-cylinder pressure signal or its variation for different cycle phases (compression, expansion, gas exchange) through estimating the amplitudes for each harmonic order. Several efforts to reconstruct the in-cylinder pressure signals using Fourier analysis were reported in the pertinent literature (Johnsson 2006; Zeng and Assanis 2004; Taraza et al. 2005). Fourier series coefficients are proven useful to reduce dimensions in machine learning applications as reported in Manjunath Aradhya et al. (2008). Alternative methods including wavelet decomposition or local Fourier transform focus on signal specific windows. The computational power required for the above methods considering the whole signal or local windows does not vary significantly. Therefore, this study focuses on the Fourier analysis to extract information from the in-cylinder pressure in the complete engine cycle.

By employing Fourier analysis, the in-cylinder pressure signal can be approximated by the following equation (Zeng and Assanis 2004):

$$p(\phi) = \mathrm{A}_0 + \sum_{n=1}^{\infty} \mathrm{A}_n cos\left(\frac{2\pi\, n\phi}{T}\right) + \sum_{n=1}^{\infty} \mathrm{B}_n sin\left(\frac{2\pi\, n\phi}{T}\right) \quad (1)$$

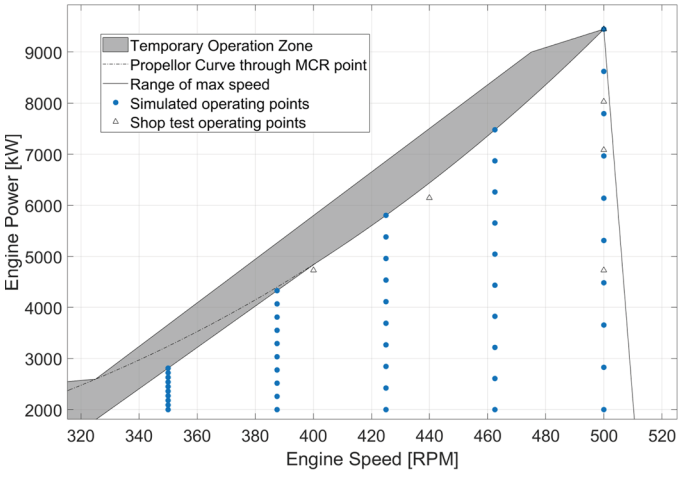


Figure 2. Engine load diagram (power/speed) and simulated operating points.

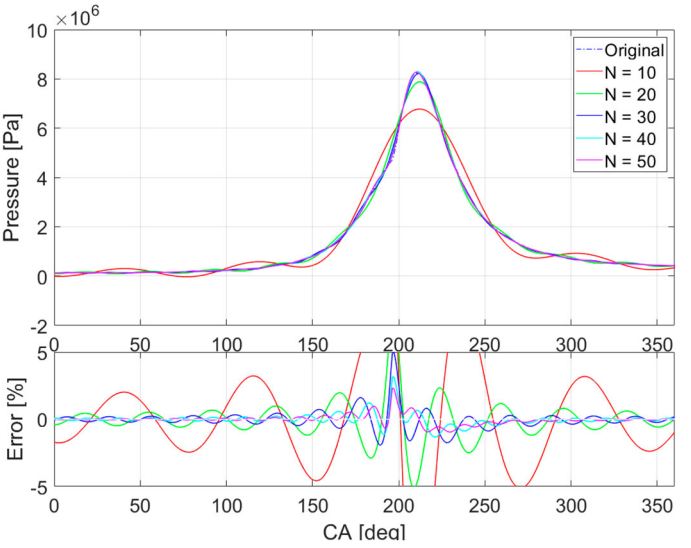


Figure 3. Reconstruction of pressure signal (first 360°*CA*) from N number of harmonic orders using 2N+1 Fourier coefficients.

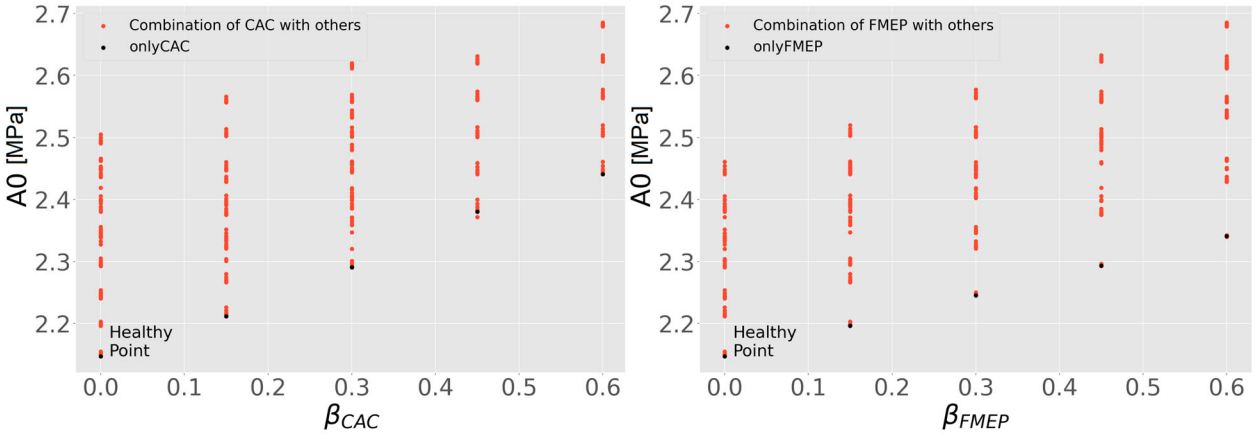
where T denotes the signal period equal to 720°CA for four-stroke engines, A_0 corresponds to the average value of the pressure signal, A_n and B_n represent the *cosine* and *sine* terms coefficients for the nth harmonic order; the amplitude of each harmonic order is $\sqrt{A_n^2 + B_n^2}$).

The total number of harmonic orders required to exactly represent the pressure signal (resulting in zero error) tends to infinity. For reducing the pressure signal dimensions, a limited number of harmonics is sufficient, provided that the required information for estimating the engine faults is retained.

Figure 3 shows the reconstructed in-cylinder pressure profile considering N numbers of harmonic orders that require 2N + 1 Fourier coefficients. As N increases, the error in the reconstruction of the original pressure signal is reduced. A parametric study is also carried out using different values of N to assess the potential for information capture and accuracy of the final diagnostic model for predicting the engine faults, which is described in Section 3.

Figure 4 shows the variation of A_0 with the fault severity for the CAC and FMEP faults at full load and 462.5 rev/m. The points plotted in *black* represent the single fault conditions. The coefficient A_0 representing the average pressure value in the cylinder is correlated with the fault severity β_{fault}. This correlation extends to the other faults and coefficients derived from pressure profiles, showcasing a potential to map a regression for estimating β_{fault}. It is inferred from the derived results that the fault combinations compared to single faults affect the Fourier coefficients values. This further supports the selection of the Fourier coefficients as input for developing regression models to forecast the faults and their severity.

Hence, a separate fault estimation model is recommended to avoid complexity, as the Fourier coefficients ranges at different faults do not vary significantly (21–27 bar from Figure 4). The developed regression model for each fault employs the first N harmonics with 2N+1 *Fourier series coefficients* (A_0, A_n, B_N with $n = 1{-}N$) as input to estimate the fault severity parameter β_{fault}. The combined output of all fault models is displayed in an array format and are employed to detect the combination of faults at each engine operating point. Therefore, all the developed data-driven fault estimation models (for each fault) must be employed to provide the engine health status.



(a) CAC Fouling (b) Friction loss

Figure 4. A_0 coefficient (average in cylinder pressure) value for CAC and FMEP faults present alone (black) and in combination (orange) with other faults at 462.5 rev/m and full load. (a) CAC Fouling and (b) Friction loss.

2.3. Data standardisation

Considerable differences in the input data-sets ranges pose challenges for developing accurate machine learning models. To address this challenge, centering of the data-sets is required (Gal and Rubinfeld 2019). All numerical attributes the pertain to the input and output parameters of the developed regression models are standardised by removing their mean and scaling them to the unit variance. In this respect, each input attribute equally contributes to the objective function used for the model training. For a numerical attribute x, the standardised attribute x' is calculated according to the following equation:

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

where μ is its mean value, and σ denotes its standard deviation.

2.4. Data splitting

The optimisation of the regression model hyper parameters depends on the separation of training and validation/test data-sets. The stratified K-fold cross-validation technique (Kuhn and Johnson 2013) is used to verify that the regression model is not overfitted with training data. First, the data sets generated for each fault are divided into five folds (k = 5) with equal distributions of their severity levels. Figure 5 shows the distributions of fault severity for each fault for all folds. Each fold exhibits a similar distribution of the fault severity levels.

However, to effectively validate the developed regression models (specific to each fault), a parametric study for separation of the training and test or validation data-sets must be carried out considering the same data distribution ratio. Therefore, these folds are further divided into training and test (validation) folds by employing the test to train ratio (γ), which is defined by following equation:

$$\gamma_{Test/Train} = \frac{F_{test}}{F_{Train}} \tag{3}$$

where F denotes the respective data-sets folds.

Five iterations ($\gamma = 1/5$) are performed by selecting one fold as the validation/test data-set, and the remaining four folds as the training data-sets. For each iteration, the R^2 value is used to evaluate the developed regression models accuracy for each fault.

2.5. Regression model

The selection of the appropriate regression technique is essential in machine learning applications. Parametric models employ a finite set of hyperparameters (θ) that are obtained through training and error minimization using a training data set (D) with input x (Bishop and Nasrabadi 2006). The hyper parameters are used to predict the output based on new inputs unrelated to data-set D, according to the following equation:

$$P(x \mid \theta, D) = P(z \mid \theta) \tag{4}$$

Linear regression, artificial neural networks (ANN) and support vector regression (SVR) with a linear kernel are examples of parametric models with increasing complexity (Gkerekos et al. 2019). As the main focus of this study is on the reduction of dimensionality, its application to estimate the faults with simple form of the parametric regression techniques is used to map the fault severity. This study evaluates the following two linear regression techniques: multiple linear regression and polynomial regression. In both cases, the input parameters include the Fourier coefficients, along with the speed and power of the engine. The severity levels of the fault β_{fault} are predicted as output variables for each fault. The combination of faults can be inferred from the derived β_{fault} for each fault.

2.5.1. Multiple linear regression (MLR)

Linear regression is one of the simplest machine learning algorithms, which consider a linear combination of input parameters along with a bias (θ_0) to map the output parameters. The developed model governing equation is:

$$\begin{aligned}\beta_{fault,cyl}(C_{cyl}, \theta_{cyl}) &= \theta_{0,cyl} + \theta_{1,cyl}C_{0,cyl} + \theta_{2,cyl}C_{1,cyl} \\ &\quad + \cdots + \theta_{n,cyl}C_{n,cyl} \\ &= \theta_{0,cyl} + \sum_{j=1}^{n} \theta_{j,cyl}C_{j,cyl}\end{aligned} \tag{5}$$

where $C_{cyl} = [Speed, Power, A_0, A_1, \ldots, A_N, B_1, \ldots, B_N]_{cyl}$ consists of the engine speed and power along with the Fourier coefficients, whereas n denotes the C_{cyl} length, equal to $2N + 1 + 2$ (sum of length of A_0, A_N, B_N, engine speed and power).

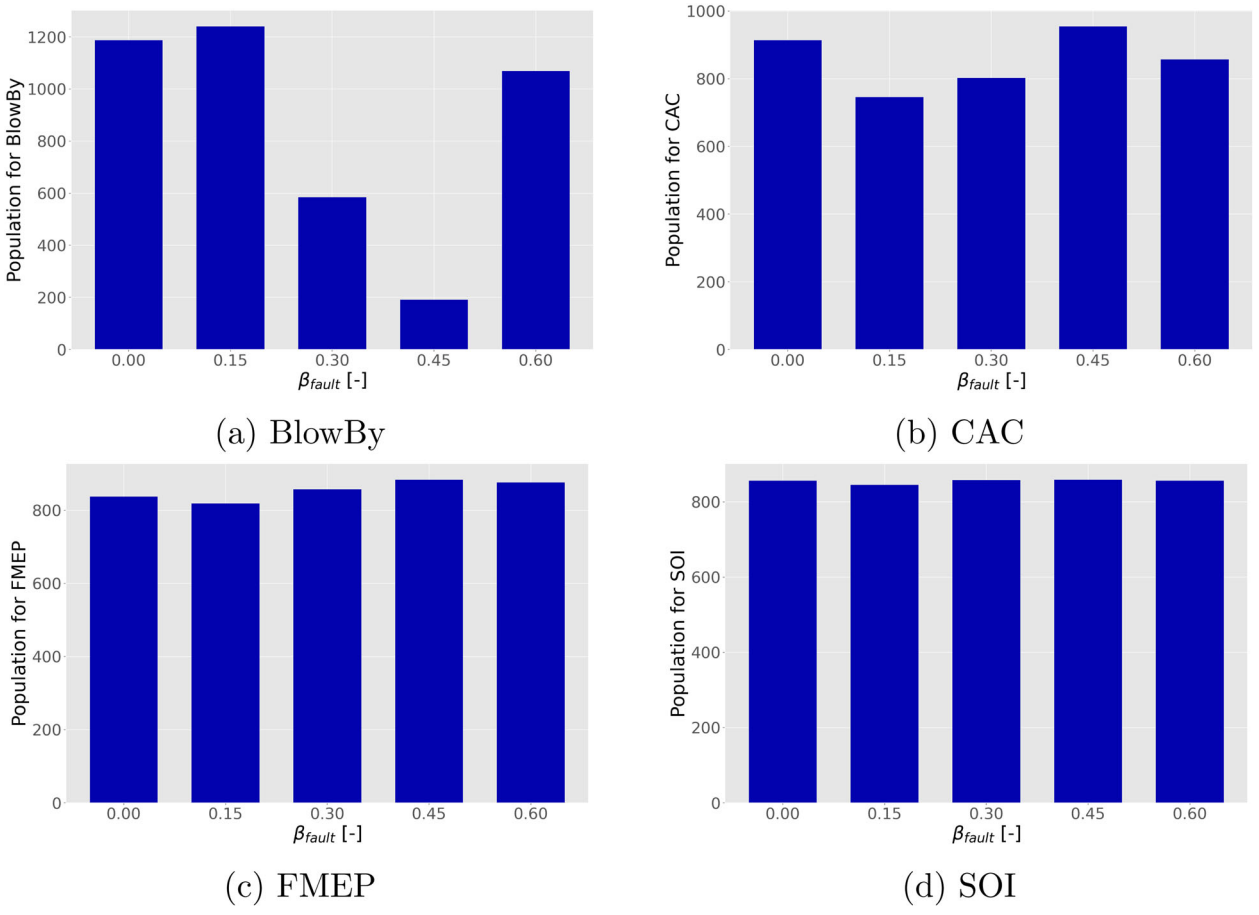


Figure 5. Fault severity distribution of data points for each fault from the simulation. (a) BlowBy. (b) CAC. (c) FMEP and (d) SOI

2.5.2. Polynomial regression (PR)

Polynomial regression (PR) is a form of linear regression known as a special case of multiple linear regression. It maps the relationship between input and output parameters as an nth degree polynomial (Gkerekos et al. 2019; Bishop and Nasrabadi 2006).

Each fault severity is derived according to the following governing equation of degree 2:

$$\begin{aligned}\beta_{fault,cyl}(C_{cyl},\theta_{cyl}) = {} & \theta_{0,cyl} + \theta_{1,cyl}C_{1,cyl} + \theta_{2,cyl}C_{2,cyl} \\ & + \cdots + \theta_{n,cyl}C_{n,cyl} + \theta_{n+1,cyl}C_{1,cyl}C_{2,cyl} \\ & + \theta_{n+2,cyl}C_{1,cyl}C_{3,cyl} + \cdots \\ & + \theta_{2n-1,cyl}C_{n-1,cyl}C_{n,cyl} + \theta_{2n,cyl}C_{1,cyl}^2 \\ & + \theta_{2n+1,cyl}C_{2,cyl}^2 + \cdots + \theta_{n^2+n,cyl}C_{n,cyl}^2 \end{aligned} \quad (6)$$

The error for both the employed linear regression methods (MLR and PR) is calculated according to the following equation:

$$e_{PR/MLR} = \underset{\theta_{cyl}}{argmin}\left\{\sum_{i=1}^{D}(\beta_{fault,cyl,i} - \beta_{fault,cyl}(C_{i,cyl},\theta_{cyl}))^2\right\} \quad (7)$$

Where $C_i \in [Speed, Power, A_0, \ldots, A_N, B_1, \ldots, B_N]$ represents the input corresponding to the ith sample of the training data for MLR and PR (degree 2), whereas D is total number of the training data samples.

The hyper parameters for both techniques (MLR and PR) are calibrated by minimising the loss function using the gradient descent method. The model accuracy is quantified using the R^2 value, which is calculated according to Equation (8), for both training and validation data. Only the R^2 values for the validation data are presented in the next section to assess the models accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y-\hat{y})^2}{\sum_{i=1}^{n}(y-\bar{y})^2} \quad (8)$$

Where, $\sum_{i=1}^{n}(y-\hat{y})^2$ is the sum of residuals and $\sum_{i=1}^{n}(y-\bar{y})^2$ is the total sum of squares (equal to variance) of the data with $\bar{y}$ being the mean value.

3. Results and discussion

First, the quantification of the faults effects on the critical engine parameters including the brake specific fuel consumption, the maximum in-cylinder pressure and the exhaust temperature are assessed in Section 3.1. Secondly, the test/validation results of the regression models for each fault are presented followed by the parametric studies on the number of Fourier coefficients (2N+1) as well as the test to train ratio (γ) to investigate the impact on the regression model accuracy R^2. This analysis highlights the amount of input dimensions required to train the regression models without loosing considerable information to predict each fault severity. Finally, the impact of the data-sets number used for training on the fault prediction accuracy considering the harmonic orders (N) is quantified.

3.1. Impact of faults on engine performance

The generated data-sets at healthy conditions and faults combinations are analysed to assess their impact on the engine performance

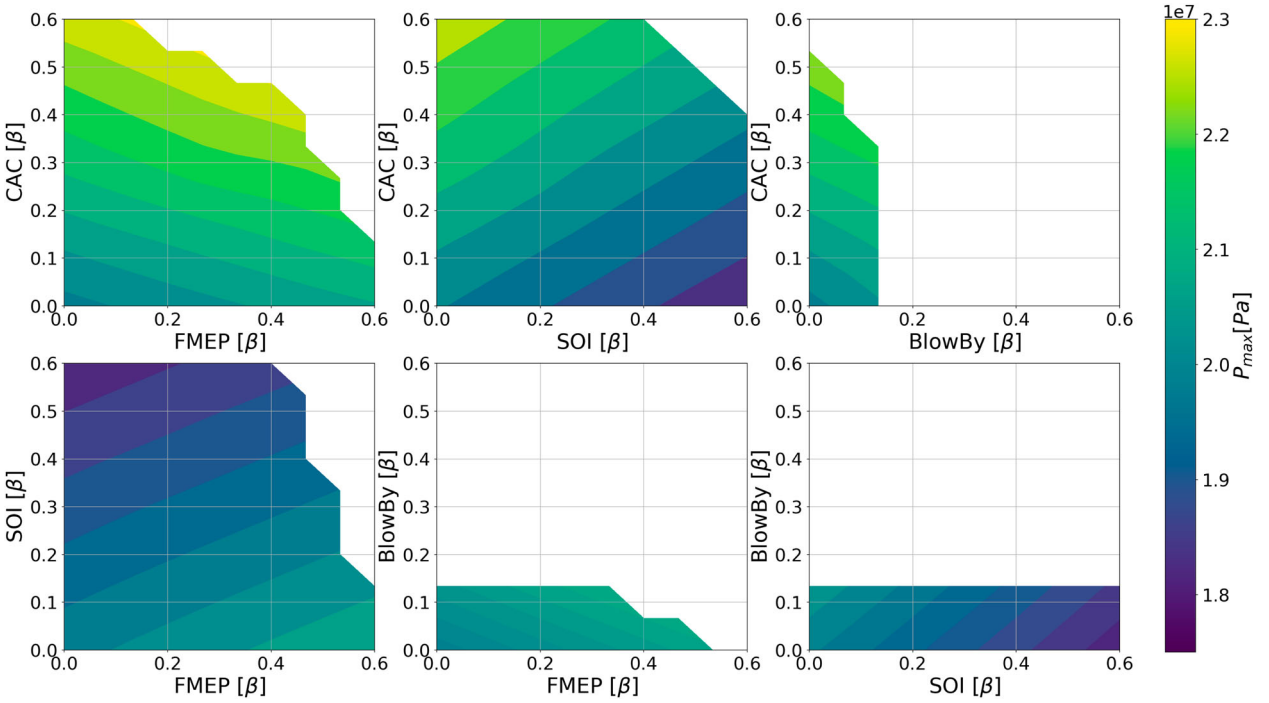


Figure 6. Impact of the fault pairs with different severity (β) on the maximum in-cylinder pressure at 100% load and 500 r/min.

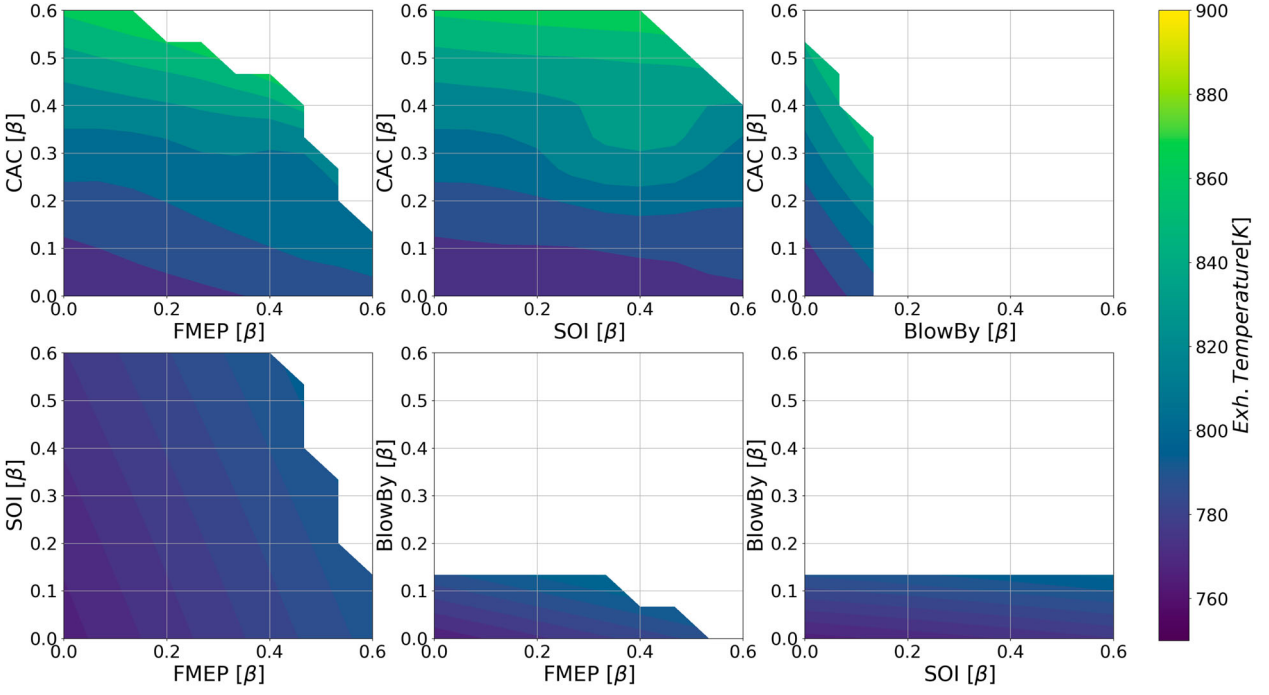


Figure 7. Impact of the fault pairs with different severity (β) on the exhaust temperature upstream turbine at 100% load and 500 r/min.

parameters. For brevity reasons, the results obtained at the operating point of 500 r/min and full load are only provided herein showcasing the impact on the maximum in-cylinder pressure Figure 6 and exhaust gas temperature Figure 7, considering pairs of the four different faults and their severity levels $\beta \in [0, 0.15, 0.5, 0.45, 0.6]$.

Thein-cylinder maximum (peak) pressure is a critical performance parameter affecting the engine BSFC and the engine components thermo-mechanical limits. The maximum in-cylinder pressure is mostly affected from β_{CAC} and β_{SOI}. High CAC fouling severity increases the in-cylinder pressure, whilst advancing the SOI reduces it. Simultaneous presence of these two faults (with same severity) only slightly affects the maximum in-cylinder pressure, as shown in top middle plot of Figure 6. Evidently, with more fuel injected, the energy content of the exhaust gas and thus the exhaust temperature before the turbine increases. Exhaust gas leak in the form of a blowby reduces the maximum cylinder pressure, however, the change is not considerable and requires more data to conclude on the effects of the blowby fault.

Table 3. Regression model training time for each fault with different number of harmonics (N) used as input with $\gamma = 0.2$.

Harmonics number (N)	Training time (s)			
	Blowby	CAC	FMEP	SOI
	Linear Regression			
10	0.02	0.02	0.02	0.017
20	0.05	0.05	0.0.35	0.031
30	0.07	0.09	0.07	0.08
40	0.11	0.1	0.1	0.09
50	0.16	0.15	0.14	0.14
	Polynomial Regression			
10	0.36	0.36	0.35	0.34
20	1.92	1.9	2.1	1.83
30	8.10	7.9	8	7.5
40	27.5	27.5	27.25	26.1
50	79.07	77	77.5	74

The exhaust gas temperature upstream of the turbocharger turbine is also an important parameter from the point of view of thermo-mechanical limits of engine components materials, and it determines the loss of exhaust gas energy as well as the performance of the turbine (Galindo et al. 2018). The exhaust gas temperature (varying from 760 to 880 K) greatly depends on the charge air cooler fouling severity as shown in the top plots of Figure 7. The combination CAC with other faults (except SOI) increases the exhaust temperature with lower values of β. The SOI fault, on the other hand, exhibits slight effects on the exhaust temperature, unless it is combined with friction and blowby faults, which cause an increase in temperature with a slight severity β_{SOI}.

From the above discussion, it is evident that, combination of faults with less severity exhibits a greater impact on the engine performance parameters compared to the presence of a single fault. Therefore, a data-driven fault estimation model is developed to predict the severity of each fault.

3.2. Validation of regression model for each fault

Effect of *Fourier coefficients* on the regression models accuracy: The time required for training of MLR and PR models for each fault with varying number of Fourier coefficients used as input is showcased in Table 3. As expected, the MLR model required shorter training time compared to the polynomial regression models (PR). In addition, the training time further increases with increased number of inputs dimensions. The model accuracy (R^2) variations for each

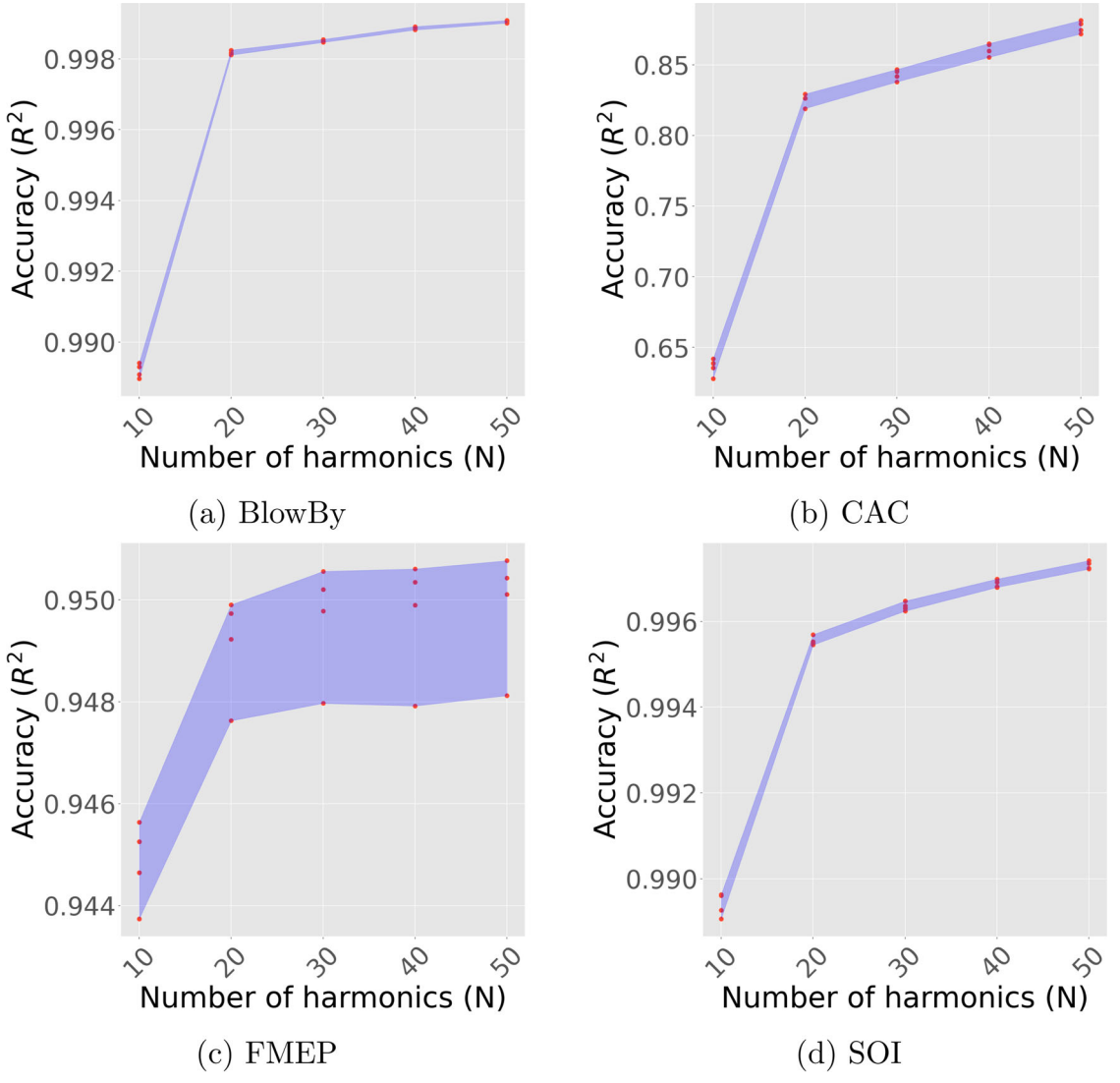


Figure 8. MLR models accuracy (R^2) variation on validation data with the initial N number of harmonics considered by using 2N+1 Fourier coefficients as inputs to regression model. (a) BlowBy. (b) CAC. (c) FMEP and (d) SOI.

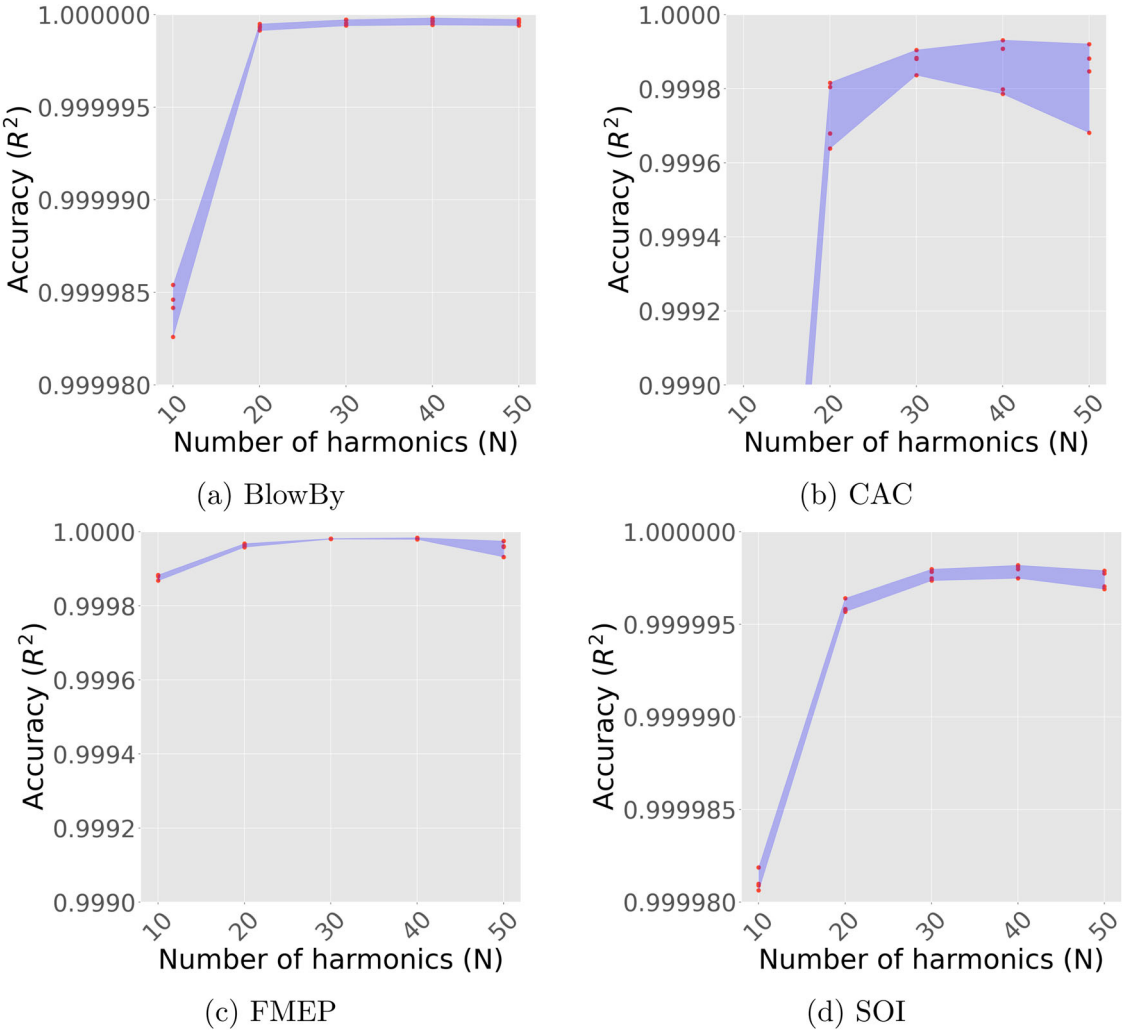


Figure 9. PR models accuracy (R^2) variation on validation data with the initial N number of harmonics considered by using 2N+1 Fourier coefficients as inputs to regression model. (a) BlowBy. (b) CAC. (c) FMEP and (d) SOI.

fault on validation data-sets as function of the harmonics number (N) are presented for MLR and PR in Figures 8 and 9, respectively.

The higher number of harmonic orders retain higher amount of information from the in-cylinder pressure signal showcasing potential to estimate the fault severity using linear regression model. However, the regression models accuracy also depends on the particular fault type and the employed training data. The MLR model estimating faults severity related to Blowby, friction and injection exhibited R^2 values greater than 0.94 with only 10 harmonics. Lower number of data points with high severity levels for the Blowby fault resulted in very high accuracy of fault estimation. The R^2 values in Figure 9(a) for Blowby faults are close to 0.99 with only 10 harmonics used as input. The cooler fouling fault severity estimation using the same regression type with 10 harmonics exhibited R^2 close to 0.65. Therefore, although it is inferred that the first 10 harmonic orders contain sufficient information to diagnose combustion chamber related faults, the faults not directly related to combustion chamber require higher number of harmonics.

However, the PR models developed for each fault exhibited greater accuracy. Even in the case where only 10 harmonic orders were used as input, the calculated R^2 exceeded 0.99 for the CAC and friction faults, whereas R^2 for the blowby and injection faults was even higher. These results indicate that the instantaneous in-cylinder pressure conveys essential information for estimating faults and their severity. This information can be derived using traditional dimensionality reduction techniques and subsequently used for fault diagnosis without the need for complex regression models. The noise from the measurement of pressure signal must be removed prior of using it as input for the proposed diagnostic model. The model does not consider uncertainties involved in in-cylinder pressure measurements, which may require alternative regression methods depending on the error type.

Models comparison: Although, the PR model exhibits high accuracy on the considered faults severity prediction, it is imperative to check its over-fitting tendency compared to the linear regression model. Statistical methods based on generalised degree of freedom, such as Akaike Information Criterion (AIC), are useful to compare the model accuracy, however the cross-validation technique is found to be more robust and effective to diverse distributions according to Hauenstein et al. (2018). Figures 8 and 9 illustrate the k-fold cross validation results for the linear and polynomial models, respectively. The R^2 values are reduced for the PR model compared to the LR model for the same number of harmonics.

The PR model with the first 20 harmonic orders (41 Fourier coefficients as input) has the potential to attain R^2 values greater than 0.95 for all the considered faults. From the cross validation (4 cases where validation/test fold was different) for the injection and friction faults, it was inferred that the models seem to over-fit the training data-sets

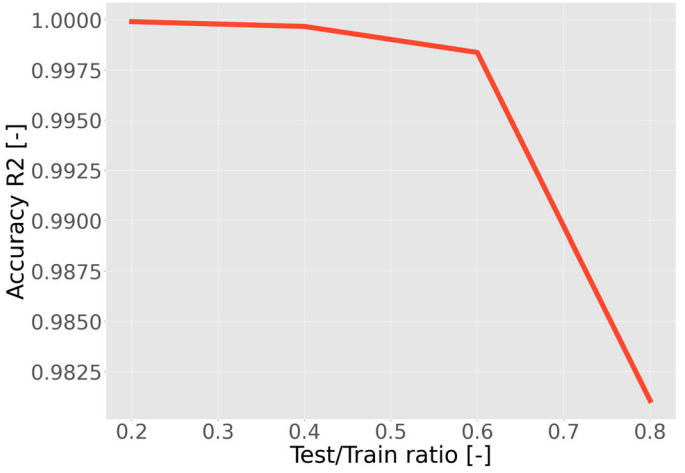


Figure 10. PR models (R^2) variations versus the test to train ratio for CAC fault. Higher test to train ratio means that less data-sets were used for training.

for the cases where more than 40 harmonic orders were employed as input. This reduces the accuracy of the PR models in specific validation folds. Based on the preceding discussion, it can be inferred that the PR models are effective and hence, they are selected for conducting the parametric study to assess the impact of data-sets amount used for training on their accuracy.

Effect of training data-set amount on accuracy: Figure 10 presents the accuracy of the polynomial regression models for the four considered faults with respect to the test-to-train ratio of data. As the test-to-train ratio increases, the amount of training samples used for training gets reduced, which lowers the accuracy of the models on the validation data sets. The R^2 drops greatly for values of test-to-train ration above 0.6. This indicates that up to 60% of the data-sets can be used for the training of the PR models without loosing accuracy.

4. Study implications

The data-driven model proposed in this study (consisting of blocks of dimensionality reduction (by Fourier analysis), standardisation, and regression models) provides a tool for performing marine engine diagnostics based on measure parameters (in-cylinder pressure, speed, and power). Therefore, it is beneficial to support operations on smart and future autonomous ships. The advantage of low computational power and memory requirement makes their easy shipboard implementation (edge computing) that supports predictive maintenance of the machinery and contributes to reduced operational costs. Moreover, due to their dependency on the collected data used for training, this data-driven model can be updated to self-adapt the current operating conditions of the monitored marine engine. The model can be extended to other internal combustion engine types including propulsion, electricity generation, providing that in-cylinder pressure signals can be acquired.

The input parameters (in-cylinder pressure signal) required for the proposed data-driven model can be obtained by direct measurement on each cylinder (which is common in advanced ships) or intermittent measurement using portable devices. It is also expected that the development of systems and methods to estimate in-cylinder pressure signals, for example, using the measured instantaneous engine torque as reported in Tsitsilonis and Theotokatos (2022)), will further help develop diagnostic methods and will be benefited by their integration with the proposed data-driven model.

5. Conclusions

This study focuses on the development of a data-driven fault estimation model using a novel method for the extraction of information from in-cylinder pressure signals. The proposed data-driven model facilitates the fault detection (classification model) and the quantification of the detected faults severity. The main findings of this study are summarised as follows.

- Each fault severity can be identified by determining changes in engine parameters like maximum cylinder pressure, and exhaust temperature. The proposed model however can accurately identify each fault severity in presence of simultaneous faults.
- The Fourier coefficients derived from the in-cylinder pressure signals convey the required information pertinent to the faults and their severity, and therefore, their use can facilitate the engine fault diagnosis process.
- The first 10–20 harmonics are sufficient to identify and estimate the faults severity, which reduce the input dimensions to maximum 43 (41 coefficients, engine speed, load) per operating point, instead of providing whole signal.
- Considerable improvement of the accuracy for both the multivariate linear and polynomial regression techniques was exhibited when 20 harmonics were employed, compared to the case of 10 harmonics. The polynomial regression model considering the first 20 harmonic orders (41 Fourier coefficients) exhibited accuracy more than 0.99 for all faults.
- The estimation of the charge air cooler fouling requires the use of at least 20 harmonics and polynomial regression to achieve acceptable accuracy.
- The polynomial regression model training required around 40% of the total data-sets (56,250) to obtain R^2 above 0.98.

This study considers the same faults/degradation for all the cylinders of a multi-cylinder engine, whereas the employed data-sets were numerically generated. The developed regression models are trained by employing the in-cylinder pressure signals generated by using the engine digital twin without accounting for any measurement errors, sensor uncertainties and faults, or human interventions. A process is required to remove noise and errors from the measured in-cylinder pressure signals prior to feeding them as input to the proposed model. Future studies could consider combinations of faulty cylinders, investigate the sensitivity and impact of the measured parameters uncertainties, as well as validate the proposed models in shipboard operations.

Acknowledgments

The authors affiliated with the MSRC also greatly acknowledge the funding from DNV AS and RCCL for the MSRC establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of Innovate UK, DNV AS and RCCL.

Disclosure statement

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

Funding

This study was carried out in the framework of the i-HEATS project, which is funded by the Innovate UK Smart Grants under the grant agreement No 99958.

Symbols

R^2	Coefficient of determination [-]
N	Selected number of Fourier series cosine (A_N) and sine (B_N) coefficients

cyl	Cylinder
fault	Selected Fault
t	gap thickness

Greek Symbols

β	Engine fault severity [-]
$\gamma_{Test/Train}$	Test to Train ratio [-]
ϕ	Crank angle [degrees]
θ	Hyper parameters of regression model
F	number of folds of the data-set

Subscripts

faulty	Faulty engine conditions
healthy	Healthy engine conditions

Abbreviations

CA	Crank Angle
BlowBy	Blowby fault
CAC	Charge Air Cooler fault
FMEP	Friction Mean Effective Pressure fault
MLR	Multiple Linear Regression
PR	Plynomial Regression
SOI	start of injection fault
SVR	Support Vector Regression

ORCID

Chaitanya Patil http://orcid.org/0000-0001-8139-1514

References

Altosole M, Balsamo F, Acanfora M, Mocerino L, Campora U, Perra F 2022. A digital twin approach to the diagnostic analysis of a marine diesel engine. IOS Press.

Ayati M, Shirazi FA, Ansari-Rad S, Zabihihesari A. 2020. Classification-based fuel injection fault detection of a trainset diesel engine using vibration signature analysis. J Dyn Syst Meas Control. 142:051003. doi: 10.1115/1.4046270.

Banisoleiman K, Rattenbury N. 2006. Reliability trends, operating issues and acceptance criteria related to exhaust gas turbochargers used in the marine industry-a classification society view. In: Institution of Mechanical Engineers: 8th International Conference on Turbochargers and Turbocharging. United Kingdom. p. 289–304.

Bishop CM, Nasrabadi NM. 2006. Pattern recognition and machine learning. Vol. 4. Springer.

Cai C, Weng X, Zhang C. 2017. A novel approach for marine diesel engine fault diagnosis. Cluster Comput. 20:1691–1702. doi: 10.1007/s10586-017-0748-0.

Cavazzuti M. 2013. Design of experiments. In: Optimization methods. Springer. p. 13–42.

Chen SK, Flynn PF. 1965. Development of a single cylinder compression ignition research engine. SAE paper.

Coraddu A, Oneto L, Cipollini F, Kalikatzarakis M, Meijn GJ, Geertsma R. 2022. Physical, data-driven and hybrid approaches to model engine exhaust gas temperatures in operational conditions. Sh Offshore Struct. 17:1360–1381. doi: 10.1080/17445302.2021.1920095.

Elmdoost-gashti M, Shafiee M, Bozorgi-Amiri A. 2023. Enhancing resilience in marine propulsion systems by adopting machine learning technology for predicting failures and prioritising maintenance activities. J Mar Eng Technol. 23(1):1–15.

Gal MS, Rubinfeld DL. 2019. Data standardization. NYUL Rev. 94:737.

Galindo J, Climent H, Varnier O, Patil C. 2018. Effect of boosting system architecture and thermomechanical limits on diesel engine performance: part-i – steady-state operation. Int J Engine Res. 19:854–872. doi: 10.1177/1468087417731654.

Gkerekos C, Lazakis I, Theotokatos G. 2019. Machine learning models for predicting ship main engine fuel oil consumption: a comparative study. Ocean Eng. 188:106282. doi: 10.1016/j.oceaneng.2019.106282.

Hauenstein S, Wood SN, Dormann CF. 2018. Computing aic for black-box models using generalized degrees of freedom: a comparison with cross-validation. Commun Stat-Simul Comput. 47:1382–1396. doi: 10.1080/03610918.2017.1315728.

Hou L, Zhang J, Du B. 2020. A fault diagnosis model of marine diesel engine fuel oil supply system using pca and optimized svm. In: Journal of Physics: Conference Series. Vol. 1576. Institute of Physics Publishing. doi: 10.1088/1742-6596/1576/1/012045.

Hountalas DT. 2000. Prediction of marine diesel engine performance under fault conditions. Appl Therm Eng. 20:1753–1783. doi: 10.1016/S1359-4311(00)00006-5.

Johnsson R. 2006. Cylinder pressure reconstruction based on complex radial basis function networks from vibration and speed signals. Mech Syst Signal Process. 20:1923–1940. doi: 10.1016/j.ymssp.2005.09.003.

Jones N, Li YH. 2000. A review of condition monitoring and fault diagnosis for diesel engines. Tribotest. 6:267–291. doi: 10.1002/tt.v6:3.

Jung M, Niculita O, Skaf Z. 2018. Comparison of different classification algorithms for fault detection and fault isolation in complex systems. In: Procedia Manufacturing. Vol. 19. Elsevier B.V.. p. 111–118. doi: 10.1016/j.promfg.2018.01.016.

Karatuğ Ç, Arslanoğlu Y, Soares CG. 2023. Review of maintenance strategies for ship machinery systems. J Mar Eng Technol. 22(5):1–15.

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. doi: 10.1016/j.engappai.2016.10.015.

Krogerus T, Hyvönen M, Huhtala K. 2018. Analysis of common rail pressure signal of dual-fuel large industrial engine for identification of injection duration of pilot diesel injectors. Fuel. 216:1–9. doi: 10.1016/j.fuel.2017.11.152.

Kuhn M, Johnson K. 2013. Applied predictive modeling. Vol. 26. Springer.

Lamaris V, Hountalas D. 2010. A general purpose diagnostic technique for marine diesel engines–application on the main propulsion and auxiliary diesel units of a marine vessel. Energy Convers Manag. 51:740–753. doi: 10.1016/j.enconman.2009.10.031.

Manjunath Aradhya V, Hemantha Kumar G, Noushath S. 2008. Multilingual ocr system for south indian scripts and english documents: an approach based on fourier transform and principal component analysis. Eng Appl Artif Intell. 21:658–668. doi: 10.1016/j.engappai.2007.05.009.

Meier M, Sudwoj G, Theodossopoulos P, Tzanos E, Karakas I. 2019. A real time comprehensive analysis of the main engine and ship data for creating value to ship operators. In: Proceedings of the 29th CIMAC World Congress on Combustion Engine Technology, Vancouver (BC), Canada. p. 10–14.

Mesbahi E, Genrup M, Assadi M. 2005. Fault prediction/diagnosis and sensor validation technique for a steam power plant. J Mar Eng Technol. 4:33–40. doi: 10.1080/20464177.2005.11020187.

Mohammad A, Rezaei R, Hayduk C, Delebinski T, Shahpouri S, Shahbakhti M. 2022. Physical-oriented and machine learning-based emission modeling in a diesel compression ignition engine: dimensionality reduction and regression. Int J Engine Res. 24(3):14680874211070736.

Patil C, Theotokatos G, Milioulis K. 2023. Marine engines combustion diagnostics employing fourier series and ann. In: 11th European Combustion Meeting. France.

Perceau M, Guibert P, Clenci A, Iorga-Simăn V, Niculae M, Guilain S. 2022. Investigation of the aerodynamic performance of the miller cycle from transparent engine experiments and cfd simulations. Machines. 10:467. doi: 10.3390/machines10060467.

Perera LP. 2016. Marine engine centered localized models for sensor fault detection under ship performance monitoring. In: IFAC-PapersOnLine. Vol. 49. Elsevier B.V.. p. 91–96. doi: 10.1016/j.ifacol.2016.11.016.

Rao X, Sheng C, Guo Z, Yuan C. 2022. A review of online condition monitoring and maintenance strategy for cylinder liner-piston rings of diesel engines. Mech Syst Signal Process. 165:108385. doi: 10.1016/j.ymssp.2021.108385.

Raza J, Liyanage JP. 2009. Application of intelligent technique to identify hidden abnormalities in a system: a case study from oil export pumps from an offshore oil production facility. J Quality Mainten Eng. 15:221–235. doi: 10.1108/13552510910961156.

Rubio JAP, Vera-García F, Grau JH, Cámara JM, Hernandez DA. 2018. Marine diesel engine failure simulator based on thermodynamic model. Appl Therm Eng. 144:982–995. doi: 10.1016/j.applthermaleng.2018.08.096.

Senemmar S, Zhang J. 2021. Deep learning-based fault detection, classification, and locating in shipboard power systems. In: IEEE Electric Ship Technologies Symposium. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ESTS49166.2021.9512342.

Shatkay H. 1995. The fourier transform-a primer. Brown University. p. 1–20.

Stoumpos S, Theotokatos G. 2022. A novel methodology for marine dual fuel engines sensors diagnostics and health management. Int J Engine Res. 23:974–994. doi: 10.1177/1468087421998635.

Tan Y, Tian H, Jiang R, Lin Y, Zhang J. 2020. A comparative investigation of data-driven approaches based on one-class classifiers for condition monitoring of marine machinery system. Ocean Eng. 201:107174. doi: 10.1016/j.oceaneng.2020.107174.

Tan Y, Zhang J, Tian H, Jiang D, Guo L, Wang G, Lin Y. 2021. Multi-label classification for simultaneous fault diagnosis of marine machinery: a comparative study. Ocean Eng. 239:109723. doi: 10.1016/j.oceaneng.2021.109723.

Taraza D, Henein NA, Gade MJ, Bryzik W. 2005. Cylinder pressure reconstruction from crankshaft speed measurement in a four-stroke single cylinder diesel engine. In: ASME Internal Combustion Engine Division Spring Technical Conference. Chicago, Illinois, USA. p. 387–395. doi: 10.1115/ICES2005-1023.

Tsaganos G, Nikitakos N, Dalaklis D, Ölcer A, Papachristos D. 2020. Machine learning algorithms in shipping: improving engine fault detection and diagnosis via ensemble methods. WMU J Marit Aff. 19:51–72. doi: 10.1007/s13437-019-00192-w.

Tsitsilonis K, Theotokatos G. 2021. Engine malfunctioning conditions identification through instantaneous crankshaft torque measurement analysis. Appl Sci. 11:3522. doi: 10.3390/app11083522.

Tsitsilonis KM, Theotokatos G. 2018. A novel systematic methodology for ship propulsion engines energy management. J Clean Prod. 204:212–236. doi: 10.1016/j.jclepro.2018.08.154.

Tsitsilonis KM, Theotokatos G. 2022. A novel method for in-cylinder pressure prediction using the engine instantaneous crankshaft torque. Proc Inst Mech Eng Pt M J Eng Marit Environ. 236:131–149.

Tsitsilonis KM, Theotokatos G, Coventry A. 2021. Development of an intelligent marine engines health assessment system based on digital-twins and data-driven models. In: Annual Marine Technology Conference. Athens, Greece. p. 1–12.

Tsitsilonis KM, Theotokatos G, Patil C, Coraddu A. 2022. Health assessment framework of marine engines enabled by digital twins. Int J Eng Res. 24(7):14680874221146835.

Velasco-Gallego C, Lazakis I. 2022. Development of a time series imaging approach for fault classification of marine systems. Ocean Eng. 263:112297. doi: 10.1016/j.oceaneng.2022.112297.

Wang S, Wang J, Wang R. 2020. A novel scheme for intelligent fault diagnosis of marine diesel engine using the multi-information fusion technology. In: IOP Conference Series: Materials Science and Engineering. Vol. 782. Institute of Physics Publishing. doi: 10.1088/1757-899X/782/3/032022.

Wartsila . 2001. Project guide for marine applications: Wartsila 46c. [accessed 2022 October 06].

Xi W, Li Z, Tian Z, Duan Z. 2018. A feature extraction and visualization method for fault detection of marine diesel engines. J Int Meas Confed. 116:429–437. doi: 10.1016/j.measurement.2017.11.035.

Zabihi-Hesari A, Ansari-Rad S, Shirazi FA, Ayati M. 2019. Fault detection and diagnosis of a 12-cylinder trainset diesel engine based on vibration signature analysis and neural network. Proc Inst Mech Eng Pt C J Mech Eng Sci. 233:1910–1923. doi: 10.1177/0954406218778313.

Zeng P, Assanis DN. 2004. Cylinder pressure reconstruction and its application to heat transfer analysis. SAE paper: 0922.

Zhang M, Zi Y, Niu L, Xi S, Li Y. 2018. Intelligent diagnosis of v-type marine diesel engines based on multifeatures extracted from instantaneous crankshaft speed. IEEE Trans Instrum Meas. 68:722–740. doi: 10.1109/TIM.2018.2857018.