Machine Learning and Data-Driven Fault Detection for Ship Systems Operations

Michail Cheliotis[1a], Iraklis Lazakis[b], Gerasimos Theotokatos[a]

[a]*Maritime Safety Research Centre, Dept. of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, Glasgow UK*

[b]*Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, Glasgow UK*

[1]Corresponding Author Email Address: Michail.cheliotis@strath.ac.uk (Michail Cheliotis)

Machine Learning and Data-Driven Fault Detection for Ship Systems Operations

Abstract

Well maintained vessels exhibit high reliability, safety and energy efficiency. Even though machinery failures are inevitable, their occurrence can be foreseen when predictive maintenance schemes are implemented. Predictive maintenance may be optimally applied through condition, performance, and process monitoring. Most importantly, it can include the detection of developing faults, which affect the performance of ship systems and hinder energy-efficient operations of ships. Under this viewpoint, this paper proposes a new data-driven fault detection methodology in a novel application for shipboard systems, by exploring the "learning potential" of recorded voyage data. The proposed methodology, combines the benefits of Expected Behaviour (EB) models, by selecting the optimal regression model, with the Exponentially Weighted Moving Average (EWMA) for fault detection, in novel ship applications. It is seen that a multiple polynomial ridge regression model, with testing R^2 score of nearly 0.96 and can accurately detect certain developing faults manifesting in both the Main Engine (ME) cylinder Exhaust Gas (EG) temperature and the ME scavenging air pressure. The early detection of developing faults can be used to supplement the daily monitoring of ship operations and enable the planning of pre-emptive rectifying actions by reducing sub-optimal machinery conditions.

Keywords: pre-processing, expected behaviour, machine learning, fault detection, ship machinery systems, ship operations

1. Introduction

The prosperity of global trade is closely tied with the performance of the global fleet (Stopford, 2018). A factor that influences the performance of the global fleet is the physical condition of its vessels and their systems and machinery. Detecting developing ship systems' faults, and taking rectifying actions before any major failure is an effective way of mitigating environmental and safety risks (Ančić et al., 2018).

Predictive maintenance represents one of the latest trends in maritime maintenance for ensuring the physical condition of ships (Mobley et al., 2008). It is applied through condition and process monitoring and it is best implemented in critical ship systems (CDNSWC, 2010).

The aim of condition and process monitoring is to use information from specific ship systems to assess the state of the examined system (e.g. failed, degraded, etc.) and detect developing faults (Kobbacy, 2008; Mohanty, 2015). The successful application of predictive maintenance is tied to the quality of the data and the applicability of the various algorithms, which can be employed for the systems condition assessment and Fault Detection (FD) (Cheliotis et al., 2019).

The early detection of faults can increase safety and reliability and reduce downtime, as it allows for pre-emptive rectifying actions and planning for the required maintenance (Zhang et al., 2019). As discussed by Beşikçi et al (2016), the energy-efficiency of ships can be improved and safeguarded by a variety of different factors, including practices in their management and operation. FD, as part of predictive maintenance, can have a substantially positive impact on the operation of ships and the management of their maintenance (Tan et al., 2019). Therefore, the use of FD can improve the energy-efficiency of ships in two ways: it can detect degraded machinery operations, and predict failures, avoiding the sub-optimal operations of machinery improving their energy-efficiency as well. In the same manner, FD can avoid inefficient operations resulting from the overloading and increased strain of other systems due to failures of specific machinery arrays as also discussed by Armstrong and Banks (2015). These benefits also translate to 10-35% more cost-efficient operations as discussed in Dikis and Lazakis (2019), when compared with reactive maintenance approaches. Similar cost-benefits have also been reported across different industries as seen in May and Thons (2015) and Kumar and Saini (2018).

1.1 Fault Detection Status Quo

The area of FD has been rapidly expanding and is currently facilitated using various condition describing signals, as reviewed and explained by shown in Martinez-Guerra and Luis Mata-Machuca (2013), Sari (2013) and Sayed-Mouchaweh (2018). As reviewed by Jardine et al (2006), FD includes a variety of different methods ranging from statistical to Machine Learning (ML), all aimed at identifying the presence of a fault in the examined system.

As discussed by Isermann (2006) the most basic statistical approach for FD is through limits checking (e.g. maximum and minimum values), whereas more advanced approaches are built around the identification of specific trends (e.g. cyclic patterns, rates of change, etc.). The use of the Exponential Weighted Moving Average (EWMA) for FD is an approach gaining popularity due to its versatility and accuracy and is based on the identification trends in the examined signal (Garoudja et al., 2017). EWMA-based FD models use selected signals and plot the signals' EWMA in a control chart, which contains upper and lower control limits for the detection of faults (Nounou et al., 2018). This type of FD creates easy to visualise models that can be used to detect various faults, with improved accuracy compared to traditional statistical control charts (Mukherjee et al., 2019; Shamsuzzaman et al., 2019). For instance, Harrou et al (2015) combined partial least squares with EWMA for the detection of faults in industrial processes. The effectiveness of this model was showcased through its ability to detect developing faults in distillation columns. Similarly, Badodkar and Dwarakanath (2017) developed a methodology based on EMWA for the detection of broken teeth in mechanical gearboxes. The EWMA analysed time-series acceleration signals, which showcased a good performance in detecting faults in their early stages. Awad et al (2018) developed a method for the detection of structural damage in buildings, based on Artificial Neural Networks (ANNs) and control charts. Nounou et al (2018) proposed a condition monitoring scheme for grid-connected photovoltaic panels. The scheme was based on the monitoring of environmental and performance parameters (voltage, current, and frequency) in an EWMA control chart. Finally, Adegoke et al (2019) proposed the use of an EWMA-based FD methodology for the manufacturing sector. The effectiveness of the methodology was showcased in an example of a continuous stirred tank reactor.

As discussed by Ma and Jiang (2011), FD based on ML approaches are gaining popularity in a variety of applications, including the manufacturing, nuclear, and offshore sectors. ML-based approaches for FD are traditionally based on classification algorithms, like Support Vector Machines (SVMs) and Logistic Regression, as reviewed by Liu et al (2018). However, these algorithms are restricted and not easily used in FD models trained on fault-free data only. Also, classification-based FD models are not easy to integrate with prognostic and diagnostic tasks (Hong et al., 2007).

FD based on Expected Behaviour (EB) models is an alternative approach to classification models (Hong et al., 2007). EB models are often used for FD tasks in efforts to improve safety in a variety of applications, including the offshore, automotive, nuclear, and manufacturing sectors. The use of such models offers several advantages. Notably, they replicate the normal behaviour of selected signals (target variables) based on appropriate inputs (predictor variable). These models leverage ML to assess any deviations from the normality to detect faults. Zaher et al (2009) examined the use of an ANN for the development of an EB model for FD in wind turbines, based on operational data. The ANN was trained with more than three months of operational data and was used to monitor the condition of the turbine's gearbox. Similarly, Schlechtingen and Ferreira Santos (2011) examined the application of ANNs and polynomial regression models for the development of EB-based FD for wind turbines. The examined models showed good performance in detecting faults in the stator and gearbox of a wind turbine, by modelling the power, speed, and various temperatures. The same authors developed an ANN-based EB model for the detection of a variety of faults in wind turbines based on operational data. The networks were trained by using more than 30 months of operational data (Schlechtingen et al., 2013; Schlechtingen and Santos, 2014). Lastly, Bangalore and Patriksson (2018) studied the topic of optimal maintenance planning for wind turbines by using an ANN-based EB model for the detection of faults in critical components. As with the previous cases, the developed models were trained on readily available operational data.

1.2 Fault Detection in the Maritime Industry

Following the above, the current literature on maritime FD is not as widespread as in other industrial sectors. For example, Capezza et al (2019) investigated the use of partial least squares regression in combination with the Hotelling's T^2 and the squared prediction error control charts, for FD under the scope of emissions monitoring. Also, the benefits of predictive maintenance have been addressed under a decision support framework using fuzzy–sets enhanced with Analytical Hierarchy Process (AHP) (Lazakis and Ölçer, 2015). Ahn et al (2017) also examined the use of a fuzzy-based Failure Mode Effects Analysis (FMEA) approach to study the risk profile of the gas turbine system of specialised tankers. Similarly, Cem Kuzu et al (2019) proposed the use of a fuzzy-based Fault Tree Analysis (FTA) to analyse the inherent risks of ship mooring operations. Dikis et al (2014) examined the use of data-driven dynamic Bayesian Networks (BN) for the maintenance prioritisation of multiple ship systems. Expanding on this, the coupling of BN with data mining and Markov Chains (MC) has been studied for the development of a predictive maintenance scheme of marine Main Engine (ME) and their supporting systems (Dikis and Lazakis, 2019). The application of a regularised feed-forward ANN classifier for the monitoring of the Exhaust Gas (EG) valve of a marine two-stroke engine, using acoustic emissions signals, has also been examined (Fog et al., 1999). Li et al (2011) also presented the use of a back-propagating ANN classifier for the condition monitoring a marine gearbox, based on the spectrum analysis of a vibration signal. Likewise, the use of a three-layer feed-forward ANN for the condition monitoring of the air intake and fuel injection system of a medium speed marine engine has been examined (Basurko and Uriondo, 2015). Raptodimos and Lazakis (2018) and Lazakis et al (2018) examined the application of ANNs and their combination with Self Organising Maps (SOM) and inter-clustering for the monitoring, prediction and healthiness assessment of a marine ME. Moreover, Lazakis et al (2018a) demonstrated the use of SVMs for the classification of faults and the development of a data-driven normality index for a marine generating engine. Zhan et al (2007a) (2007b) examined the use of a multi-class SVM for the fault diagnosis of marine ME cylinder covers, based on vibration analysis and Principal Components Analysis (PCA). The combination of data simulation, through physical modelling, with both supervised and unsupervised ML algorithms has been examined with application to system decay in naval vessels (Cipollini et al., 2018; Coraddu et al., 2016). Lastly, Begg et al (2018) developed a method for the detection of damaged mooring equipment by leveraging dynamic modelling.

1.3 Comparison, Gaps and Novelty

Having in mind the above-cited literature, the subsequent conclusions regarding maritime FD can be derived. Firstly, the characteristics of the available data (e.g. size and density) influence the developed methodology. As a result, the selection of the employed algorithm must be thoroughly investigated. Any selection must take into account the characteristics of the available data and demonstrate a good performance under case-specific constraints (e.g. lack of recorded faulty data, data with low sampling frequency).

Moreover, the majority of the EB models for FD are based on ANNs which, although they have good performance, they require large datasets for training, which are not always available within the maritime domain. In addition, the use of ANNs is related to a 'black-box' approach, which makes it difficult to impart domain knowledge. Consequently, the selection of the underlying approach for the EB models should serve the application and the data available. In cases of limited data and when accurate and fast results are required, regression-based EB models could be examined instead.

The selection of the EB model approach for the detection of faults is also beneficial when compared to the alternative classification approaches. Firstly, with EB models, there is more flexibility in the selection of the underlying algorithms used. With classification approaches, in the absence of observed faulty data, one-class SVM is the standard choice, with limited alternative algorithms. In contrast to classification, EB models have greater flexibility in the selection of the algorithm (e.g. ANNs, polynomial regression). Moreover, the output of the EB models (i.e. a time-series) is more interpretable and useful for future tasks (e.g. diagnostics), when compared to the output of classification approaches (i.e. decision space).

Summarising these critical reflections, the following gaps regarding maritime FD are revealed. Firstly, predictive maintenance in the maritime sector is at its infancy and is severely lagging compared to other sectors. A potential gap is also identified in the application and use of FD models addressing the particular needs and requirements of maritime predictive maintenance which in turn will lead to the assessment of energy-efficient operations of ship systems. The development of a novel methodology that addresses the previously discussed limitations and takes advantage of the benefits of regression-based EB modelling and the EWMA control chart, is a methodology that can address the above-mentioned gaps.

Therefore, this study proposes a novel methodology that integrates with a new way pre-processing analysis, regression-based EB modelling based on the investigation of the optimal regression model, and EWMA control charts. Moreover, the aim is to apply the novel methodology in a case study for FD in the main engine of a ship, which to the best of authors' knowledge has not been implemented before. The developed FD methodology includes the combination of data-driven models with engineering analysis, resulting in interpretable and effective models using real ship system data. Through this, the new methodology will increase safety and reliability, and establish energy-efficient operations by detecting developing faults and allowing for pre-emptive actions and corrective planning.

The application of the new methodology, presently missing from the maritime industry, will have the following impact: a) it will enable capturing of previously unseen anomalies based on an EB model, b) it will allow the examination of how signals evolve in real-time, based on contributing factors while uncoupled from variable operating conditions, due to EB modelling, c) it will enable the accurate detection of developing faults, due to the use of EWMA control chart, d) it will allow for the efficient filtering of noise due to the superior performance of EWMA control charts compared to traditional control charts and e) it will combine the above benefits in a single methodology aimed at improving the energy-efficient operations of ship systems.

Finally, the current paper is structured as follows: Section 2 discusses the details and novelty of the developed methodology, Section 3 presents the case study of a marine main engine together with the results and finally, Section 4 provides the overall conclusions and suggestions for future work.

2. Proposed Methodology

This study employs a fast and effective FD methodology that supports the requirements of the maritime industry, taking into account the limitations imposed by the used data and is aimed at maintaining energy-efficient ship operations. More specifically, the detection of developing faults can improve the energy-efficiency of the monitored ship system and can also safeguard the energy-efficiency of other systems by avoiding cascading failures and overloading patterns. The produced methodology uses a regression-based EB model, which provides accurate results, without the need to develop time-consuming physical models and complex 'black-box' approaches. In addition to that, the complexity of the suggested EB model can be controlled, without requiring very large training datasets, thus addressing the maritime domain limitation of small datasets available.

The necessary steps and processes to develop the novel FD methodology are presented in Figure 1:

1. Data collection: including the data gathering efforts by an onboard ship DAQ system.
2. Pre-processing: including form handling, unit checking, outlier detection, and further data filtering.
3. Model development: including the investigation for the optimal regression model, in terms of regression type and input.
4. Fault detection: including the estimation of the residuals and the implementation of the EWMA control charts.
5. Finally, the developed methodology is assessed during the model verification phase, to ensure that the required functionality is achieved.

As can be seen in Figure 1, the novel methodology initiates with the data collection step. During this step, two unique Process Monitoring (ProMon) datasets from a DAQ system are collected for use during the model development and methodology verification. The ProMon dataset used during model development contains historic information and represents "healthy" ship operation, as established by the ship's operators. Similarly, the ProMon dataset used during the FD and verifications steps includes incoming ProMon data. Once the data is collected, the pre-processing step follows, which includes the outlier detection and data filtering for the isolation of non-operational data points. It should be highlighted that the pre-processing step is used for all the data sets. Then, the model development step follows, which is based on pre-processed historic ProMon data and outputs the optimal EB model. During this step, the feasibility of different regression techniques and predictor variables is assessed in terms of their suitability for an EB model. Once this model is established, it is used to compare the recorded and the expected values of a target variable, generating the residuals. In detail, once the incoming ProMon data are pre-processed, they are used as input to the optimal EB model to produce the expected values of the target variables. These values are compared with the target variable from the incoming ProMon data. Thereupon, the obtained residuals are assessed in an EWMA control chart for FD, and the methodology gets verified.

2.1 Data collection

The data collection is the first step of the methodology and includes the required efforts to assemble the needed information. The output of the data collection step is the creation of an extended dataset, different segments of which are used for both model development and methodology verification purposes. The collected data may originate from a marine commercial DAQ system installed onboard a merchant navy vessel. Typically, DAQ signals for FD tasks of engineering systems include ProMon data, such as power output, rotational speed, injection and scavenging pressure, and exhaust gas temperature. Similarly, the data frequency usually ranges from one sample per second or per ten minutes and up to hourly intervals, depending on the application.

2.2 Pre-processing

Data pre-processing ensures that datasets reach their full knowledge-extracting potential. Pre-processing is a standard step in most data-driven research efforts (Martinez-Guerra and Luis Mata-Machuca, 2013; Sari, 2013; Sayed-Mouchaweh, 2018; Tanasa and Trousse, 2004). The output of this step is the creation of a processed dataset, ready for model development and methodology verification. The processes contained in this step include form handling, outliers detection, and data filtering.

Pre-processing starts with the form handling and unit checking of data. This is a simple yet important process, as it ensures that the data are placed in a fitting format for the next steps of the methodology.

In the next process, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is used to remove outliers and transient states of operation, which are out of scope in the present work. Outliers are sparse data points with significantly different values from the rest of the instances of the same variable. They are often caused by sensor errors and other instrumental faults and are not part of a fault indicative pattern. For instance, negative EG temperatures and power output above an engine's rated power are considered as outliers. Thus, outliers can be considered as data "anomalies" and if they are not removed they can have a negative impact on the developed models.

DBSCAN algorithm is very effective in detecting outliers and does not rely on domain knowledge, which offers several advantages. It works by examining each point in the dataset and identifying dense areas of points (clusters). DBSCAN requires the use of the user-defined $minP$ hyperparameter. The $minP$ defines the minimum number of points that are required to form a cluster. The $minP$ is simple to specify as it is a function of the dimensionality of the dataset. Larger values are preferable, with an exclusive global lower bound of 3 (Schubert et al., 2017). Lastly, the value of $minP$ should be close to the number of dimensions of the dataset (Chen and Li, 2011; Ester et al., 1996). As suggested in the literature, the $minP$ hyperparameter is selected by combining the above restrictions with domain knowledge (Schubert et al., 2017; Thang and Kim, 2011). Also, the ε hyperparameter is required, which defines the maximum distance between points for them to be considered to be in the same cluster. If ε is too small, the majority of the data points will be clustered as noise, whereas if it is too big all the data points will be in the same cluster. In general, smaller values are preferred. An approach for calculating the ε hyperparameter, is by considering the rate of change of the distance of each point to the nearest neighbour (k-nearest neighbour graph), as shown in the relevant literature (Gaonkar and Sawant, 2013; Rahmah and Sitanggang, 2016). However, this approach is not usable when using evenly spaced data (i.e time-series with constant sampling rate). As a result, the value of ε is obtained after iterative attempts.

Given these hyperparameters, the data are categorised in three groups. Core points are considered as data points with more than $minP$ points within a radius ε. Border points are defined as data points with fewer than $minP$ points within a radius ε. The remaining points are considered as outliers or noise. Moreover, point q is directly density-reachable from a point p, if p is a core point and q is within a radius ε from p. Assuming another point q_1 which is directly density-reachable from point q only, it is said that points p and q_1 are indirectly density-reachable (Çelik et al., 2011; Chen and Li, 2011; Thang and Kim, 2011). The working process of the DBSCAN algorithm, as used in this methodology is shown below:

1 Find all core points

2. Assign all points that are directly density-reachable and indirectly density-reachable in the same cluster
3. Mark any unassigned points as outliers.

Following the removal of the transients and outliers, further data filtering takes place. Specifically, the data are filtered to retain the points that represent operational periods. Since the data collection took place over an extended period, some points could have been recorded when the ship and its main engine were not operational. The data filtering is performed by using a value-based approach. Therefore, this process removes non-operational points whilst retaining the rest.

2.3 Model Development

The model development step follows the data pre-processing step and uses as input pre-processed historic ProMon data, as seen in Figure 2. The aim is the development of an EB model that can predict the ideal (expected) behaviour of a selected variable of a system based on appropriately selected inputs. EB models are often used for FD tasks, as they can model the expected behaviour of a variable subject to changing operating conditions. EB models are ideal in the absence of faulty labelled data, as they can detect developing faults by defining a range of normal operation. The output of this step is the developed EB model which is in turn used for the FD and verification steps, using the incoming ProMon data (Figure 2). It should be noted, that the model development step includes iterations for the identification of the optimal predictor variables. Throughout this step, the used data are divided into training, validation, and testing, based on empirical rules and common practices. The training sample of the data is used to fit the different ML models. It is said amongst practitioners, that the models “sees” and “learns” for the training data. The validation sample of the data is then used to tune the models’ hyperparameters. The validation set is withheld from the ML models during the training phase, but it can still affect the models’ performance, albeit in a more limited manner than training. Ultimately, the appropriate ML model is selected based on its performance on the validation set. Lastly, the test sample of the data is used to evaluate the overall performance of the selected model, after they are trained and their hyperparameters are tuned. The test sample is also withheld from the models during both training and validation.

2.3.1 Training and Validation

The training and validation process is used to fit and fine-tune the different ML models and is structured around the use of historic ProMon data, collected during the ship's operation. The aim is to use a training set to fit the different models and a validation set to fine-tune and ultimately select, the best performing model, before the evaluation of its generalisation capabilities in a test set. This process uses training and validation datasets, which are a portion of the historic ProMon data. Finally, it must be stressed, that the recorded ProMon data represent "healthy" ship operation, as established by the ship's operators.

Four ML regression models are generated including Ordinary Least Squares (OLS) single linear regression, multiple linear ridge regression, OLS single polynomial regression, and multiple polynomial ridge regression. As previously discussed, regression-based EB models do not depend on extremely large training datasets and offer greater flexibility in imparting domain knowledge. These ML models are used to produce an estimated output for a selected target variable, by relying on the use of appropriately selected inputs (predictor variables). The examined EB models are trained and validated, using the R^2 score. Linear and polynomial regression models are developed to examine the best type of fit, given the acquired data. There are several advantages for each type of model (linear and polynomial), however, the selection is application-specific (Müller and Guido, 2015). The suitability of linear or polynomial models is a function of the available data. Single-input OLS regression models, both linear and polynomial, are used as benchmark models (Assaf et al., 2019; Erto et al., 2015; Lepore et al., 2017; Naik et al., 2018). Ridge regression models are selected based in their overall accuracy and effectiveness, as discussed by Gkerekos et al (2019). Lastly, the specific inputs for the EB models are investigated separately.

The developed linear regression models have a form as shown in Equation 1, where $\hat{y}$ represents an estimate for the target variable, w_0 to w_p are the estimated regression coefficients, b is the estimated axis intercept and x_0 to x_p represent the p different predictors (inputs).

$$\hat{y} = w_0 x_0 + \cdots w_p x_p + b = \sum_{i=0}^{p} w_i x_i + b \quad \text{Equation 1}$$

The developed polynomial regression models have a form as shown in Equation 2, which is the general form of k^{th} order polynomial using two predictors (x_1, x_2). Equation 2 also includes interaction terms between the two predictors (Bowerman et al., 2015; Olive, 2005).

$$\hat{y} = w_0 x_1 + w_1 x_2 + w_2 x_1 x_2 + \cdots w_p x_1^k + w_p x_2^k + b \quad \text{Equation 2}$$

During the training phase, sets of known predictors (x_0 to x_p) and target variables (y) are used as input in Equation 1 and 2 to obtain $\hat{y}$. The aim of the training phase then becomes minimising the objective functions in either Equation 3 (OLS regression) or Equation 4 (ridge regression) to obtain the estimates for the coefficients (w) and intercept (b).

In OLS regression, the coefficients and intercept are estimated by minimising the sum of the squared difference between the predicted and the actual values of the target variable (residuals). The minimisation of this objective function is possible since both y and $\hat{y}$ are available during the training phase(Bowerman et al., 2015; Olive, 2005).

$$OLS: \|\hat{y} - y\|_2^2 \quad \text{Equation 3}$$

In ridge regression, the coefficients and intercept are estimated by minimizing an objective function similar to the OLS. In addition to the sum of the squared residuals, an additional term is included. The additional term is called L2 regularisation and limits the magnitude of the coefficients. L2 regularisation explicitly restricts the model to avoid overfitting. The limiting capability of the regularisation term is attributed by the user-specified hyperparameter, α. This hyperparameter limits the influence of the predictors to the target, given that α is appropriately selected. When $\alpha = 0$, the objective function becomes OLS and on the other hand, if α is very large the model will underfit the data. During this research effort, k-fold cross-validation was used to estimate the optimal α value (Bishop, 2006; Bowerman et al., 2015; Olive, 2005).

$$Ridge: \|\hat{y} - y\|_2^2 + a\|w\|_2^2 \quad with \quad a \in [0, \infty) \quad \text{Equation 4}$$

K-fold cross-validation iteratively trains and validates the examined models by using all the possible combinations of training and validating sets. The working principle of this process is demonstrated in Figure 3, which is a common example with k=3 folds. In essence, the k-fold cross-validation is used to evaluate the performance of the trained models and select the best performing approach for testing. This process trains and validates as many models as there are different combinations of model hyperparameters. The different hyperparameters included in this work are parameters of the learning methods (e.g. α regularization term), and the different inputs used (e.g. predictor variables). The k-fold cross-validation partitions the data in k different folds. Each fold is set aside once and the examined models are trained on the remaining k$-$1 folds. Then, the fold withheld from training is used, to obtain the validation score of the trained models. This sequence is repeated until every fold is used for validation once. For each model, the mean validation score is calculated and the model with the highest mean score is selected. Finally, the identified model is trained with all the training and validation data, before its generalisation capabilities are assessed in the test set.

Moreover, the general working process that is followed as part of the development of the EB model is shown as an algorithm in Figure 4. The algorithm requires as input the predictor (X) and target (Y) variables. Also, it requires the number of folds (k) for the k-fold cross-validation and the size of the test set. Lastly, the set of the considered values for the model's hyperparameters is given. Figure 4 represents the generalised process for the development of the supervised model, including the optimisation of the α hyperparameter.

2.3.2 Testing

As mentioned above, four different types of ML regression models are used, namely OLS single linear regression, multiple linear ridge regression, OLS single polynomial regression, and multiple polynomial ridge regression. Also, during the training and validation, the value of α, and the different predictor variables are assessed in the required cases. After the k-fold cross-validation, the mean validation score for each of the examined model is obtained. The validation performance is assessed using the R^2 score and the model with the highest R^2 is selected for testing.

The model selected for testing is fully defined in terms of the regression type, the predictor variables, and the value of α. That model is then retrained using the training and validation datasets, and its testing performance is evaluated using the R^2 score. Once the testing performance of the selected model is analysed, the model is used to obtain the residuals.

2.4 Fault Detection

Following the full specification of the EB model, the fault detection step takes place. The output of the EB model is a prediction for the EB of a specifically selected variable of an engineering system. To facilitate fault detection, the aim is to monitor certain variables (y) and gauge any deviations from their expected value ($\hat{y}$), given the system's operational profile. As shown in Figure 1, the fault detection process has two inputs: a) incoming, previously unseen, ProMon data, and b) the EB estimate from the model for the monitored variable.

For each instance of the incoming database, the residuals (r) between the expected value and the recorded value is calculated according to Equation 5.

$$r_k = \widehat{y_k} - y_k$$
$$for \quad k = 1, \dots N$$

Equation 5

Analysing the residuals is an effective method for detecting faults in engineering systems, as the comparison between the actual and the expected behaviour can uncover developing faults. The residuals quantify the deviation of a variable from its expected value, given an operating profile (Harrou et al., 2015; Holmes and Mergen, 2000; Neubauer, 1997; Nounou et al., 2018)

After the residuals are calculated, the EWMA control chart is constructed (Hunter, 1986). In Equation 6, z refers to the EWMA statistic which is calculated for all of the k instances. For the special case of z_0, the mean value of the variable in the incoming data is used. The smoothing effect of the EWMA is attributed to the user-defined smoothing parameter, λ. The smoothing parameter is defined according to common practices. Lastly, the residual at each instance (r_k) is used.

$$z_k = \lambda r_k + (1 - \lambda) z_{k-1}$$
$$for \ k = 1, \dots N$$
$$and \ \lambda \in (0, 1]$$

Equation 6

A crucial component of the EWMA fault detection is the Upper Control Limit (UCL) and Lower Control Limit (LCL). These two limits provide the basis for the detection of faults, as any point above the UCL or below the LCL signifies faults. These limits are calculated according to Equation 7 and Equation 8.

$$UCL = \mu_0 + L\sigma \sqrt{\frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2i}]}$$

Equation 7

$$LCL = \mu_0 - L\sigma \sqrt{\frac{\lambda}{2-\lambda}[1 - (1-\lambda)^{2i}]}$$

Equation 8

In these equations, μ_0 is the mean value of the variable in the incoming data and σ is the standard deviation. Lastly, L represents the width of the control chart and its value is assigned based on the application. In essence, the UCL and LCL form the envelope of normal operations for the selected variable. As the choice of L affects this envelope, it must be appropriately selected so that it can correctly classify normal and faulty operating points. If the recorded data represent "healthy" operating points, as discussed in Section 2.3.1, the resulting UCL and LCL envelope, must fully encase all the data points. On the other hand, if a known fault exists in the data, the ULC or LCL must exceed at the point of the failure. If the resulting envelope does not exhibit this behaviour, the value of L must be altered. Consequently, assigning L its value can be an iterative process.

2.5 Methodology Verification

The last step of the proposed methodology is the verification of the developed methodology. This is arguably one of the toughest tasks, especially when access to the physical system is not a valid option. Similarly, the lack of access to a physical model of the examined system and the lack of recorded faulty data, increase the challenging nature of this step (Lazakis et al., 2018b). The main aim of this step is to ensure that the methodology performs as expected and that the developed models are fit for their purpose. This is achieved in two main ways. Initially, during the case study and while keeping the developed models constant, one predictor is altered and the outcome is observed. As an abnormal variation is given to the changed predictor, it is expected that the EB model and EWMA fault detection are able to identify this as a fault. By doing that, the ability of the model to capture abnormal values is evaluated. Also, the ability of the different regressors to achieve high R^2 score on previously unseen real data provides a practical approach to measure the models' fitness of purpose (Gkerekos and Lazakis, 2020).

3. Case Study

The FD methodology aims to establish a novel approach for the early detection of developing faults, to avoid sub-optimal ship operations and thus maintaining ship system energy-efficient operations. During this study, the considered faults are the result of gradual degradation and wear-and-tear. Consequently, sudden breakages and shock loads are not considered. The following sections showcase the effectiveness of the methodology and present its outcomes as obtained from a specific case study of a bulk carrier. The use of a specific case study does not reflect on the applicability of the developed model. On the contrary, the developed methodology has high transferability and can be applied to different cases (e.g. container ships, tanker, etc.). The pre-processing step is mainly based on the use of the DBSCAN for outlier detection, which does not rely on domain knowledge. The EB-models are not based on complex and time consuming physical modelling and as a result, they can easily capture the behaviour of a new target variable. However, this process is conditional on the availability of recorded voyage data. Lastly, the use of the EWMA is application-agnostic, which improves the model's flexibility.

3.1 Data Collection

The application of the novel methodology is examined through a specific case study, including data acquired by a DAQ system installed onboard a 64,000DWT Handymax bulk carrier. In this paper, the case study focuses on the monitoring and prediction of the Exhaust Gas (EG) temperature of a 5-cylinder two-stroke Main Engine (ME). This variable was selected due to its great importance in ship and system performance and process monitoring. Monitoring the ME cylinder EG temperature can help 1) control the ME's emissions, 2) understand the cylinders' combustion performance 2) identify underlying and developing faults. In more detail, faults in the air cooler, turbocharger, and gas passages of the ME can manifest through the ME EG temperature (HHI, 2010; MAN B&W, 20017; Woodyard, 2009). The schematic of the examined system is shown in Figure 5. The variables used during model development are shown in Table 1. As can be seen, along with the ME cylinder EG temperature (target), additional variables are used (predictors). In addition, the variables were recorded between 01st of January 2017 and 30th of March 2017 with a five minutes sampling rate, resulting in 25,627 points per variable. As mentioned in Section 2.3.1, the collected data represent fault-free operating conditions.

Table 1 Descriptive information of the variables in the ME system

Variable Name	Variable Description	Units	Count
ME CYL 1 EGT	ME cylinder 1 EG temperature	ºC	25,627
ME CYL 2 EGT	ME cylinder 2 EG temperature	ºC	25,627
ME CYL 3 EGT	ME cylinder 3 EG temperature	ºC	25,627
ME CYL 4 EGT	ME cylinder 4 EG temperature	ºC	25,627
ME CYL 5 EGT	ME cylinder 5 EG temperature	ºC	25,627
ME AVG EGT	ME mean cylinder temp	ºC	25,627

SCAV_AIR_TEMP	ME scavenging air temperature	°C	25,627
SCAV_AIR_PRESS	ME scavenging air pressure	bar	25,627
SHAFT_PWR	ME shaft power	kW	25,627
ME_RPM_TM	ME speed	rpm	25,627

3.2 Data Pre-processing

The pre-processing step, as described in Section 2.2, ensures that the data reach their full knowledge-extracting potential. All of the variables go through this step including the data used for training, validation, and testing. The first process of this step is to ensure that the units of the data are in the correct form and that the dataset is in a tabulated form. Following that, the DBSCAN algorithm is deployed to remove transient states of operation and outliers from each variable.

The application of the DBSCAN algorithm requires the specification of the ε and $minP$ hyperparameters. The ε hyperparameter dictates the maximum distance between points for them to be considered in the same cluster. Also, the $minP$ hyperparameter controls the number of neighbouring points required to form a cluster. As discussed in Section 2.2, the value of ε is determined after iterative attempts. Different values are iteratively used until the transient states are removed, and consequently, the outliers are filtered out. Finally, the selected value was $\varepsilon = 0.25$. Considering the dataset used in the FD methodology has 10 dimensions, as seen in Table 1, and keeping in mind the restrictions suggested by Chen and Li (2011) the final value for $minP$ was determined to be 10. As the sampling rate of the data is 1 recording per 5 minutes and $minP = 10$ samples from 50 minutes are required to form a cluster. This is a realistic and reasonable time-frame for the operation of the ME when the ship in on voyage, as confirmed by the operators of the considered bulk carrier. The last process of the data checking step is to filter-out any points collected when the ship was not operational. For that purpose, a value-based filter was created, as seen in Equation 9.

$$SHAFT\_PWR > 10\ kW \quad \text{Equation 9}$$

As it can be seen in Figure 6, the pre-processing step is effective in filtering-out transient states and outliers. The spikes and dips, in the “Raw Data” graph of Figure 6, are filtered out by the DBSCAN clustering algorithm, while the flat-lines are removed by the value-based filter. For example, the dips and the spikes, recorded between the 1st of January 2017 and 15th of January 2017, are successfully removed.

3.3 Model Development

During the model development, the available historic data are divided into a training and validation set and a testing set. The former is used to fit the different models, tune the different hyperparameters, compare and ultimately select the best performing model. The best model is selected by primarily assessing the validation score and taking into account the standard deviation (σ) of the prediction errors. Once the best performing model is selected, its generalisation capabilities are assessed in the training set.

In total four different types of regression models are examined include OLS single linear regression, multiple linear ridge regression, OLS single polynomial regression and multiple polynomial ridge regression. Apart from the different types of models, different inputs (predictor variables) are considered for each model. The various resulting models are detailed in Table 2. For each of the ridge regression models, the a hyperparameter ranges from 0.1 to 0.6. Also, the examined polynomial models are assumed to be of 6th order. The value of a and the order of the polynomial models depend purely on each application and there is no standard guide for their selection. Instead, different ranges of a and different values for the polynomial order can be examined. In case none of the models exhibits good behaviour with the selected a and order, these values are changed and the analysis is repeated.

Table 2 List of the examined regression model, including single-input, multiple linear and multiple polynomial.

Model ID	ME Power	ME Speed	ME scavenging air temperature	ME scavenging air Pressure	Linear	Polynomial
M1	✓	✓	✓	✓	✓	
M2	✓		✓	✓	✓	
M3		✓	✓	✓	✓	
M4	✓	✓	✓		✓	
M5	✓	✓		✓	✓	
M6	✓			✓	✓	
M7			✓	✓	✓	
M8	✓		✓		✓	
M9		✓	✓		✓	
M10	✓	✓			✓	
M11		✓		✓	✓	
N1	✓	✓	✓	✓		✓
N2	✓		✓	✓		✓
N3		✓	✓	✓		✓
N4	✓	✓	✓			✓
N5	✓	✓		✓		✓
N6	✓			✓		✓
N7			✓	✓		✓
N8	✓		✓			✓
N9		✓	✓			✓

N10	✓	✓			✓
N11		✓	✓		✓
L	✓			✓	
P	✓				✓

3.3.1 Training and Validation

From the historic ProMon data, 80% of them are randomly selected for training and validation, according to empirical knowledge and common practices. The different models need a substantial amount of the collected data (e.g. 80%) to identify patterns and develop good generalisation capabilities. The random selection of the data controls the variance of the model and it enhances its generalisation capabilities, as during the training, data from the entirety of the operational profile are used (Kirk, 2017).

During this stage, k-fold cross-validation is used to train and validate the different models discussed above. For this work, the value $k = 7$ was assumed. Similarly with the previous hyperparameters (e.g. α), the selection of k was based on empirical knowledge, as its value is application-specific. Consequently, each model is trained and validated 7 times, each time using a different segment of the data for validation. The validation performance of each model is then assessed by averaging the 7 different validation scores resulting from the 7-fold cross-validation

Figure 7 shows the validation performance of the different models in terms of their average R^2 score during the k-fold cross-validation. This Figure examines different regression models with varying inputs and ranging α values. The upper bar chart examines the average score of each model (multiple linear ridge regression) shown in Table 2. To further evaluate the performance of the different models, the maximum score is shown as a dashed blue line. Similarly, the performance of the single OLS linear regression, using the ME power as an input, is shown as a solid black line. Models M11, M12 and M13 which use as inputs the ME power, speed, scavenging air temperature and pressure with $a = 0.1$, $a = 0.2$ and $a = 0.3$ respectively have the best validation score of nearly 0.93. The performance of the OLS single linear regression is lower than all the other multiple linear ridge regression models. The bottom bar chart of Figure 7 examines the average score of each model (multiple polynomial ridge regression) and its results are summarised in Table 3. The maximum score and the performance of the OLS single polynomial regression are also shown as a dashed and a solid line respectively. Models N53, N54, N55 and N56 which use as inputs ME power, speed, scavenging air pressure with $a = 0.3$, $a = 0.4$, $a = 0.5$ and $a = 0.6$ respectively have the best validation score of nearly 0.96. Interestingly, the performance of the OLS single polynomial regression is satisfactory and preferable to most of the multiple polynomial ridge regression models. Finally, it is observed that the polynomial models have a superior performance in terms of the mean validation score. As summarised in Table 3, the polynomial models have a higher mean validation score and a smaller score range compared to the linear models.

Table 3 Performance of linear and polynomial models in terms of R^2.

	Linear Models	**Polynomial Models**
Mean Validation Score	0.83	0.94
Validation Score Range	0.11	0.091

Figure 8 shows the validation performance of the different models in terms of the standard deviation (σ) of the average R^2 scores during the k-fold cross-validation. This Figure shows the results of the same models as Figure 8 and is supplementary for the evaluation of the different models. Similarly, the minimum σ of the average R^2 score is shown as a solid line and the σ of the R^2 scores from the OLS single linear and OLS single polynomial models are shown as dashed lines. The standard deviation of the validation performance describes the consistency of each model in their predictions during the k-fold cross-validation. It is observed that the linear models have a superior performance in terms of the σ of the mean validation score. As summarised in Table 4, the linear models have a lower mean σ and a smaller range compared to the polynomial models.

Table 4 Performance of linear and polynomial models in terms of σ

	Linear Models	**Polynomial Models**
σ of Mean Validation Score	0.046	0.07
Range of σ of Mean Validation Score	0.07	0.24

As previously discussed, the model development aims to identify and compare the models with the highest R^2 and lowest σ. As seen in Table 5, model N54 has the best mean validation R^2 and a σ comparable with the remaining models, and for these reasons is identified as the optimum choice. In summary ML model N54 uses multiple polynomial ridge regression has α=*0.4* and uses as input the ME Power, pressure, and speed.

Table 5 Performance of models with the highest R^2 and performance of models with lowest σ

Models	Mean Validation R^2	σ of Mean Validation R^2
N54	0.96	0.03
M31	0.89	0.01
M32	0.89	0.01
M33	0.89	0.01
M34	0.89	0.01
M35	0.89	0.01
M36	0.89	0.01
M91	0.88	0.01
M92	0.88	0.01
M93	0.88	0.01
M94	0.88	0.01
M95	0.88	0.01
M96	0.88	0.01

Finally, Figure 9 shows the learning curves of model N54 having as target variable the EG temperature of the cylinders of the ME and the average EG temperature of all the cylinders of the ME. These curves show the training (red) and validation (black) scores for each case as a function of the number of folds in k-fold cross-validation. In effect, increasing the number of folds also increases the training data. Thus, the learning curves aim to evaluate if the model is either overfitting or underfitting the data. In other words, the learning curves are used to gauge the model's generalisation capabilities. In Figure 9, across all the graphs, as the number of fold increases the training performance reaches a plateau, indicating that the training performance can no longer improve by increasing the amount of training data. Similarly, the validation score increases indicating that the generalisation capabilities of the model are satisfactory. In general, the convergence of the training and validation learning curves indicate the presence of a model with a good fit on the data. For example, the upper left chart of Figure 9 shows the learning curves for the model predicting the average EG temperature of all the cylinders of the ME. As seen, the training score reaches a plateau of around 0.977, and the validation score reaches a maximum value of nearly 0.968.

3.3.2 Testing

Model N54 is identified as the optimal and fully defined choice for the prediction of the EG temperature of the cylinders of the ME. Following the completion of the training and validation process, model N54 is trained using the whole training set (no validation set is used). The trained model is then evaluated on the previously unseen test set. Figure 10 shows the training and testing scores of this process. It evaluates the model's capabilities in predicting the EG temperature of the cylinders of the ME and the average EG temperature of all the cylinders of the ME. As observed the testing performance is satisfactory as the R^2 ranges from 0.93 to 0.966. Also, the testing score is systematically lower than the training score, which is expected behaviour for this type of modelling.

3.4 Fault Detection and Verification

This step of the methodology uses the incoming ProMon data, which are pre-processed according to Section 2.2 and aims at detecting faults in the operation of the examined vessel. The identified and evaluated model (N54) is used to obtain the expected (predicted) values for the ME EG temperature of each cylinder of the examined vessel. Once these values are calculated, they are compared with the actual EG temperature from the incoming ProMon data, resulting in the residuals.

Once the calculation of the residuals is completed, the EWMA control chart is constructed, which requires the specification of the λ and L hyperparameters. The former is the smoothing parameter and is obtained according to common practices. In this work, it is assumed that $\lambda = 0.3$, according to Badodkarand and Dwarakanath (2017). On the other hand, the L hyperparameter controls the width of the control chart (distance between UCL and LCL) and its value is assigned after iterations. As previously discussed, the incoming data represent healthy operating conditions, as confirmed by the operator of the examined vessel. Therefore, the value of L was selected so that the residuals on the control chart do not exceed the UCL and LCL. Figure 11 shows the residuals of the average EG temperature of cylinders of the ME of the examined vessel, plotted in an EWMA control chart. The average EG temperature is used for simplicity reasons, as it summarises the behaviour of the individual cylinders. In this figure, the obtained residuals are shown with grey and the EWMA statistic for each residual is shown as blue. Lastly, the UCL, LCL, and the Center Line (CL) are also shown. In Figure 11, $L = 3$ was used, since the resulting EWMA statistics for the residuals lie between UCL and LCL. It should be specified that $L = 3$ is the first value correctly classifying the EWMA residuals.

The evaluation of the methodology transpires by using the developed EB model and analysing the residuals in an EWMA chart for fault detection. To examine the detection capabilities of the methodology, and by considering the fault-free nature of the available data, four different fault cases are examined through simulated data in the form of a sensitivity analysis (Law, 2009; Saltelli, 2004).

These four cases are presented in Table 6 and represent failure modes that can affect the target variable. In detail, according to domain knowledge and by considering the publications by Hountalas (2000) and Theotokatos et al (2015), the examined cases represent specific failure modes in the Turbocharger (TC), Air Cooler (AC) and gas passages of the ME that can affect the ME EG temperature. Table 7 provides a brief summary of the possible faults that included in the different cases of Table 6 (MAN B&W, 20017). Table 7 highlights the functionality of the methodology and provides practical suggestions, however, the faults are indicative and dedicated diagnostic efforts are required for the specification of their root-cause.

Table 6 Verification cases description

Case ID	Variable	Alteration	Limit	Value
Case 1	ME scavenging air	Increased	Upper	3.30
Case 2	ME scavenging air	Decreased	Lower	0.4 bar
Case 3	ME cylinder EG	Increased	Upper	420
Case 4	ME cylinder EG	Decreased	Lower	214

In addition to that, the limits presented in Table 6 are selected from the ME manufacturer guide (MAN B&W, 20017) and represent the alarm limits set by the manufacturer. In each case, the appropriate variables in the dataset used for verification are adjusted linearly, to reach and exceed the presented limits, simulating the faulty conditions. These adjustments take place, across all cases, from 10/01/2017 to 11/01/2017 and is assumed that after this period rectifying actions take place. Moreover, since the examined failures represent the result of gradual degradation, the faults are represented as a group of points exceeding the ULC or LCL. The identification of such patterns indicates the presence of a fault. Also, from a practical standpoint, the detection of a fault could be verified by inspecting the examined system, however, additional diagnostic efforts are required. Lastly, single points exceeding the limits of the EWMA control chart should be further investigated, even though they are out of the scope of this work. Figures 12 to 15 show the EWMA fault detection results for Case 1 to Case 4 respectively. All cases show the residuals of the average EG temperature across all the cylinders. This is due to simplicity reasons, as the examples of the examined faults affect the EG temperature of all the cylinders.

Table 7 Indicative possible faults organised in system-level and sub-system level, adapted from (MAN B&W, 20017).

System-Level Faults	Sub-System Level Faults
AC	Air-side fouling Water-side fouling
ME Gas Passages	Air filter fouling Corroded TC mechanical components TC fouling
Cylinder head	Leaking EG valve Blocked fuel valve or injector
Combustion Chamber	Blow-by

Figure 12 shows the EWMA smoothed residuals signal for case 1 which describes a typical example of a fault related to the overloading of the ME caused by the fouling of the ship's hull over time. As it can be observed, the points from the first few days have residuals with an error close to 0°C. However, from the 10/01/2017 the residuals begin to increase and reach more than 200°C. This surge is attributed to the increase of the ME scavenging air pressure to more than 3.20 bar. The residuals (as defined in Equation 5) increase, as the expected value of the target variable, increases with a higher rate. As it can be observed, the simulated fault is successfully detected as the EWMA exceeds the UCL.

Figure 13 shows the EWMA smoothed residuals signal for case 2. Similarly with the previous case, the residuals from the first few days fluctuate around 0°C. However, from 10/01/2017 the residuals begin to drop and reach a value of more than -350°C. This decrease is attributed to the controlled drop of the ME scavenging air pressure to nearly 0.20 bar. In this case, the residuals drop, as the expected value of the target variable declines with a sharper rate. As it can be observed, the simulated fault is successfully detected by the LCL. Such behaviour can be attributed to fouling and corrosion in the TC of the ship, and fouling and corrosion in the nozzle ring of the TC.

Figure 14 shows the EWMA smoothed residuals signal for case 3. As it can be observed, the points from the first few days fluctuate around 0°C. However, from 10/01/2017 the residuals begin to decrease and reach a value of approximately 150°C. This drop is attributed to the simulated rise in the ME cylinder EG temperature to nearly 420°C. The residuals in case 3 decline, as the actual value of the EG temperature drops decoupled from the expected value. As can be seen, the simulated fault is successfully detected, as the LCL is successfully exceeded. Typical examples of fault described in case 3 include the fouling of the main cooler of the ship and fouling in the AC of the ME (both air and water sides).

Figure 15 shows the EWMA smoothed residuals signal for case 4. In this case, the points from the first few days fluctuate around 0°C. However, from 10/01/2017 the residuals begin to increase and exceed the value of approximately 200°C. This surge is attributed to the simulated drop in the ME cylinder EG temperature to nearly 214°C. Following the same underlying reasoning as case 3, the decoupled decrease of the actual values of the target variable increases the residuals and the simulated fault is successfully detected by the UCL. A typical example of a fault described in case 4 includes the improperly maintained or improperly configured engine room conditions, resulting in obstructed air flows.

4. Conclusions

The present paper provides a novel ML and data-driven FD methodology, based EB modelling and EWMA control charts, and its application on ship systems. The ultimate goal of the FD is to allow for pre-emptive actions and scheduling, reducing downtime and improving safety and energy efficiency and supplement the daily monitoring of ship operations. The early detection of developing faults can reduce the sub-optimal operations of the monitored system, safeguarding its energy-efficient operations. Similarly, as failures can cascade to other systems, avoiding the failure of certain machinery can have a positive effect on the energy-efficiency of a wider spectrum of systems. More specifically, it can avoid increased strain and overloading on other systems, maintaining their energy-efficient operations. The EB modelling approach is used to predict the EB of the ME cylinder EG temperature. The ME EG temperature parameter is selected, as several faults in the ME's supporting systems can manifest through this variable. Also, the data used in all the aspects of the EB model are pre-processed using the DBSCAN algorithm and a value-based filter. Then, the residuals from the incoming and expected ME cylinder EG temperature are analysed in an EWMA control chart to detect developing faults through the UCL and LCL. During the development of the EB model, the optimal ML model is identified by examining the performance of several ML regression types with varying inputs. The main outcomes of this paper are summarised below:

- The developed methodology can successfully detect imminent faults by analysing the difference between the recorded and expected (predicted) ME cylinder EG temperature.
- By analysing the residuals of the ME cylinder EG temperature, it is possible to detect faults manifesting in both the ME cylinder EG temperature and the ME scavenging air pressure.

- The optimal EB model is based on multiple polynomial ridge regression and has as inputs the ME power, ME speed and ME scavenging air pressure.
- The optimal EB model can accurately predict the target variable, having a test R^2 score of more than 0.96. This is advantageous, especially when considering that the EB model is not based on time-consuming and complex physical models.
- It is possible to obtain accurate prediction, without resorting to "black-box" approaches while keeping the ability to both interpolate and extrapolate the value of a continuous, target variable.
- The EWMA smooths the residuals signal and gives a dynamic aspect to the methodology, as previous points are taken into account.
- The pre-processing step is essential for the maximisation of the knowledge-extracting capabilities of the used data.
- The developed methodology is highly transferable and can be applied in a variety of different cases (e.g. tankers, container ships). The various steps of the methodology are not reliant on domain knowledge (pre-processing), are not based on complex application-specific models (EB-models) and employ application-agnostic tools (EWMA control chart).

The proposed novel methodology was verified in a case study, where simulated faults, selected from domain knowledge, were fed into the model. Through this, it was observed that all of the simulated faults were detected. Even though the selected target variable can provide the means for efficient monitoring of ship systems, future steps of this work could include the development of an independent diagnostic system. For instance, the results from the FD could be aggregated and used as input in diagnostic networks fault localisation in the supporting systems of the ME.

Disclosure Statement

No potential conflict of interest was reported by the authors

Funding

The work presented in this paper is partially funded by the Integrated Ship Energy and Maintenance Management System (ISEMMS) project. ISEMMS project has received research funding from the Innovate UK Programme. This publication reflects only the authors' views and Innovate UK is not liable for any use that may be made of the information contained within.

References

Adegoke, N.A., Abbasi, S.A., Dawod, A.B.A., Pawley, M.D.M., 2019. Enhancing the performance of the EWMA control chart for monitoring the process mean using auxiliary information. Qual. Reliab. Eng. Int. 35, 920–933.

Ahn, J., Noh, Y., Park, S.H., Choi, B. Il, Chang, D., 2017. Fuzzy-based failure mode and effect analysis (FMEA) of a hybrid molten carbonate fuel cell (MCFC) and gas turbine system for marine propulsion. J. Power Sources 364, 226–233.

Ančić, I., Theotokatos, G., Vladimir, N., 2018. Towards improving energy efficiency regulations of bulk carriers. Ocean Eng. 148, 193–201.

Armstrong, V.N., Banks, C., 2015. Integrated approach to vessel energy efficiency. Ocean Eng. 110, 39–48.

Assaf, A.G., Tsionas, M., Tasiopoulos, A., 2019. Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression. Tour. Manag. 71, 1–8.

Awad, M.I., AlHamaydeh, M., Faris, A., 2018. Fault detection via nonlinear profile monitoring using artificial neural networks. Qual. Reliab. Eng. Int. 34, 1195–1210.

Badodkar, D.N., Dwarakanath, T.A., 2017. Machines, Mechanism and Robotics. In: Badodkar, D.N., Dwarakanath, T.A. (Eds.), Proceeding of INACOMM 2017. Springer, Mumbai, p. 841.

Bangalore, P., Patriksson, M., 2018. Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines. Renew. Energy 115, 521–532.

Basurko, O.C., Uriondo, Z., 2015. Condition-based maintenance for medium speed diesel engines used in vessels in operation. Appl. Therm. Eng. 80, 404–412.

Begg, S.E., Fowkes, N., Stemler, T., Cheng, L., 2018. Fault detection in vibration systems: Identifying damaged moorings. Ocean Eng. 164, 577–589.

Beşikçi, E.B., Kececi, T., Arslan, O., Turan, O., 2016. An application of fuzzy-AHP to ship operational energy efficiency measures. Ocean Eng. 121, 392–402.

Bishop, C.M., 2006. Pattern Recognition and Machiner Learning, 1st ed, Information Science and Statistics. Springer, Singapore.

Bowerman, B.L., O'Connell, R.T., Murphree, E.S., 2015. Regression Analysis Unified Concepts, Practical Applications, and Computer Implementation.

Capezza, C., Coleman, S., Lepore, A., Palumbo, B., Vitiello, L., 2019. Ship fuel consumption monitoring and fault detection via partial least squares and control charts of navigation data. Transp. Res. Part D Transp. Environ. 67, 375–387.

CDNSWC, 2010. Handbook of Reliability Prediction Procedures for Mechanical Equipment. Naval Surface Warfare Center Carderock Division, West Bethesda.

Çelik, M., Dadaşer-Çelik, F., Dokuz, A.Ş., 2011. Anomaly detection in temperature data using DBSCAN algorithm. INISTA 2011 - 2011 Int. Symp. Innov. Intell. Syst. Appl. 91–95.

Cem Kuzu, A., Akyuz, E., Arslan, O., 2019. Application of Fuzzy Fault Tree Analysis (FFTA) to maritime industry: A risk analysing of ship mooring operation. Ocean Eng. 179, 128–134.

Cheliotis, M., Gkerekos, C., Lazakis, I., Theotokatos, G., 2019. A novel data condition and performance hybrid imputation method for energy efficient operations of marine systems. Ocean Eng. 188.

Chen, Z., Li, Y.F., 2011. Anomaly Detection Based on Enhanced DBScan Algorithm. Procedia Eng. 15, 178–182.

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

Coraddu, A., Oneto, L., Ghio, A., Savio, S., Anguita, D., Figari, M., 2016. Machine learning approaches for improving condition-based maintenance of naval propulsion plants. Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ. 230, 136–153.

Dikis, K., Lazakis, I., 2019. Dynamic predictive reliability assessment of ship systems. Int. J. Nav. Archit. Ocean Eng.

Dikis, K., Lazakis, I., Turan, O., 2014. Probabilistic Risk Assessment of Condition Monitoring of Marine Diesel Engines. Int. Conf. Marit. Technol. 2014, 7-9 July 2014, Glas. United Kingdom 1–9.

Erto, P., Lepore, A., Palumbo, B., Vitiello, L., 2015. A Procedure for Predicting and Controlling the Ship Fuel Consumption: Its Implementation and Test. Qual. Reliab. Eng. Int. 31, 1177–1184.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

Fog, T.L., Hansen, L.K., Larsen, J., Hansen, H.S., Madsen, L.B., Sorensen, P., Hansen, E.R., Pedersen, P.S., 1999. On condition monitoring of exhaust valves in marine diesel engines. In: Neural Networks for Signal Processing - Proceedings of the IEEE Workshop. IEEE, pp. 554–564.

Gaonkar, M., Sawant, K., 2013. Auto Eps DBSCAN: DBSCAN with Eps automatic for large dataset. Int. J. Adv. Comput. Theory Eng. 2, 11–16.

Garoudja, E., Harrou, F., Sun, Y., Kara, K., Chouder, A., Silvestre, S., 2017. Statistical fault detection in photovoltaic systems. Sol. Energy 150, 485–499.

Gkerekos, C., Lazakis, I., 2020. A novel, data-driven heuristic framework for vessel weather routing. Ocean Eng. 197, 106887.

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.

Harrou, F., Nounou, M.N., Nounou, H.N., Madakyaru, M., 2015. PLS-based EWMA fault detection strategy for process monitoring. J. Loss Prev. Process Ind. 36, 108–119.

HHI, 2010. Insrtuctions Hyundai-MAN Diesel Engines Operation. Oper. Man.

Holmes, D.S., Mergen, A.E., 2000. Exponentially weighted moving average acceptance charts. Qual. Reliab. Eng. Int. 16, 139–142.

Hong, Y.S., Cho, Y.M., Hameed, Z., Song, C.K., Ahn, S.H., 2007. Condition monitoring and fault detection of wind turbines and related algorithms: A review. Renew. Sustain. Energy Rev. 13, 1–39.

Hountalas, D.T., 2000. Prediction of marine diesel engine performance under fault conditions. Appl. Therm. Eng. 20, 1753–1783.

Hunter, J.S., 1986. The Exponentially Weighted Moving Average. J. Qual. Technol. 18, 203–210.

Isermann, R., 2006. Fault-Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance. Springer, Darmstadt.

Jardine, A.K.S., Lin, D., Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process.

Kirk, M., 2017. Thoughtful machine learning. O'Reily, Sebastopol.

Kobbacy, K.A.., 2008. Complex System Maintenance Handbook, 1st ed. Springer, New Jersey.

Kumar, A., Saini, M., 2018. Stochastic modeling and cost-benefit analysis of computing device with fault detection subject to expert repair facility. Int. J. Inf. Technol. 10, 391–401.

Law, A.M., 2009. How to build valid and credible simulation models. Proc. - Winter Simul. Conf. 24–33.

Lazakis, I., Gkerekos, C., Theotokatos, G., 2018a. Investigating an SVM-driven, one-class approach to estimating ship systems condition. Ships Offshore Struct. 5302.

Lazakis, I., Gkerekos, C., Theotokatos, G., 2018b. Investigating an SVM-driven, one-class approach to estimating ship systems condition. Ships Offshore Struct. 14, 432–441.

Lazakis, I., Ölçer, A., 2015. Selection of the best maintenance approach in the maritime industry under fuzzy multiple attributive group decision-making environment. Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ. 230, 297–309.

Lepore, A., dos Reis, M.S., Palumbo, B., Rendall, R., Capezza, C., 2017. A comparison of advanced regression techniques for predicting ship CO 2 emissions. Qual. Reliab. Eng. Int. 33, 1281–1292.

Li, Z., Yan, X., Yuan, C., Zhao, J., Peng, Z., 2011. Fault detection and diagnosis of a gearbox in marine propulsion systems using bispectrum analysis and artificial neural networks. J. Mar. Sci. Appl. 10, 17–24.

Liu, R., Yang, B., Zio, E., Chen, X., 2018. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mech. Syst. Signal Process. 108, 33–47.

Ma, J., Jiang, J., 2011. Applications of fault detection and diagnosis methods in nuclear power plants: A review. Prog. Nucl. Energy 53, 255–266.

MAN B&W, 20017. Operational and Maintenance Manual.

Martinez-Guerra, R., Luis Mata-Machuca, J., 2013. Understanding Complex Systems Fault Detection and Diagnosis in Nonlinear Systems A Differential and Algebraic Viewpoint. Springer, London.

May, A., Thons, S., 2015. Integrating Structural Health and Condition Monitoring: A Cost Benefit Analysis For Offshore Wind Energy. In: OMAE (Ed.), Proceedings of the ASME 2015 34th International Conference on Ocean, Offshore and Arctic Engineering. ASME, Newfoundland.

Mobley, K., Higgins, L., Wikoff, D., 2008. Maintenance Engineering Handbook, 7th ed. McGraw Hill.

Mohanty, A.R., 2015. Machinery Condtion Monitoring, 1st ed. Taylor & Francis.

Mukherjee, A., Chong, Z.L., Khoo, M.B.C., 2019. Comparisons of some distribution-free CUSUM and EWMA schemes and their applications in monitoring impurity in mining process flotation. Comput. Ind. Eng. 137, 106059.

Müller, A.C., Guido, S., 2015. Introduction to Machine Learning with Python and Scikit-Learn, 1st ed, O'Reilly Media, Inc. O'Reily, Sebastopol.

Naik, J., Bisoi, R., Dash, P.K., 2018. Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression. Renew. Energy 129, 357–383.

Neubauer, A.S., 1997. The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation.

Nounou, M., Nounou, H., Mansouri, M., Harkat, M.F., Al-khazraji, A., Hajji, M., 2018. Wavelet optimized EWMA for fault detection and application to photovoltaic systems. Sol. Energy 167, 125–136.

Olive, D.J., 2005. Linear Regression, 1st ed, Springer Nature. Springer, Cham.

Rahmah, N., Sitanggang, I.S., 2016. Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra. In: IOP Conference Series: Earth and Environmental Science. p. 12012.

Raptodimos, Y., Lazakis, I., 2018. Using artificial neural network-self-organising map for data clustering of marine engine condition monitoring applications. Ships Offshore Struct. 13, 649–656.

Saltelli, A., 2004. Sensitivity analysis in practice: a guide to assessing scientific models (Google eBook).

Sari, A.H.A., 2013. Data-driven design of fault diagnosis systems, Data-Driven Design of Fault Diagnosis Systems. Springer, Rostock.

Sayed-Mouchaweh, M., 2018. Fault diagnosis of hybrid dynamic and complex systems, Fault Diagnosis of Hybrid Dynamic and Complex Systems. Springer, Douai.

Schlechtingen, M., Ferreira Santos, I., 2011. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mech. Syst. Signal Process. 25, 1849–1875.

Schlechtingen, M., Santos, I.F., 2014. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: Application examples. Appl. Soft Comput. J. 14, 447–460.

Schlechtingen, M., Santos, I.F., Achiche, S., 2013. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description. Appl. Soft Comput. J. 13, 259–270.

Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Trans. Database Syst. 42.

Shamsuzzaman, M., Haridy, S., Maged, A., Alsyouf, I., 2019. Design and application of dual-EWMA scheme for anomaly detection in injection moulding process. Comput. Ind. Eng. 138, 106132.

Stopford, M., 2018. Maritime Economics, 3rd ed. Taylor & Francis, Oxon.

Tan, Y., Niu, C., Tian, H., Hou, L., Zhang, J., 2019. A one-class SVM based approach for condition-based maintenance of a naval propulsion plant with limited labeled data. Ocean Eng. 193, 106592.

Tanasa, D., Trousse, B., 2004. Advanced data preprocessing for intersites Web usage mining. Intell. Syst. IEEE 19, 59–65.

Thang, T.M., Kim, J., 2011. The anomaly detection by using DBSCAN clustering with multiple parameters. 2011 Int. Conf. Inf. Sci. Appl. ICISA 2011 1–5.

Theotokatos, G., Tzelepis, V., 2015. A computational study on the performance and emission parameters mapping of a ship propulsion system. Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ. 229, 58–76.

Woodyard, D., 2009. Manire Diesel Engines and Gas Turbines, 9th ed. Elsevier, Oxford.

Zaher, A., McArthur, S.D.J., Infield, D.G., Patel, Y., 2009. Online wind turbine fault detection through automated SCADA data analysis. Wind Energy 12, 574–593.

Zhan, Y., Shi, Z., Liu, M., 2007a. The application of support vector machines (SVM) to fault diagnosis of marine main engine cylinder cover. In: IECON Proceedings (Industrial Electronics Conference). IEEE, pp. 3018–3022.

Zhan, Y., Shi, Z., Liu, M., 2007b. The Application of Support Vector Machines (SVM) to Fault Diagnosis of Marine Main Engine Cylinder Cover. In: IECON 2007 - 33rd Annual Conference of the IEEE Industrial Electronics Society. IEEE, pp. 3018–3022.

Zhang, C., Zhang, D., Zhang, M., Mao, W., 2019. Data-driven ship energy efficiency analysis and optimization model for route planning in ice-covered Arctic waters. Ocean Eng. 186, 106071.