主機デジタルツイン技術を用いた船舶主機状態監視システムの開発

ボンダレンコ　オレクシー*，北川　泰士**，福田　哲吾*

Development of an Advanced Ship Engine Monitoring System Utilising Engine Digital Twin Technology

by

Oleksiy BONDARENKO*, Yasushi KITAGAWA**, Tetsugo FUKUDA*

Abstract

In response to increasingly stringent regulations, the shipping industry has witnessed significant technological advancements in ship and engine designs in recent years. These developments have resulted in engines that can accommodate a wide range of fuels, as well as electronic systems that offer flexible control and online tuning capabilities. Consequently, it has become increasingly important to simultaneously assess the performance and monitor the condition propulsion systems. To address these challenges, we have devised the concept of an "AI Chief Engineer" system which aims to combine the power of a digital twin with a comprehensive set of algorithms to provide in-depth analytics on the propulsion system and decision support.

The AI Chief Engineer system integrates the engine's digital twin, a virtual replica that emulates the engine's behaviour and characteristics in real-time, with advanced algorithms designed for performance monitoring, degradation analysis, failure detection, and more. By incorporating data from intelligent sensors and advanced data acquisition systems, the AI Chief Engineer system enables the application of complex analytics and machine learning algorithms to fulfill the engine's advanced monitoring requirements. Thus, the AI Chief Engineer system offers a robust platform interfacing the stream of raw data, the engine's digital twin, and a comprehensive suite of algorithms facilitating extensive analytics on the engine's state and decision support.

This report explores the architecture of the AI Chief Engineer concept and introduces how the integration of a digital twin with comprehensive algorithms can enhance data assimilation and the digitalisation of ships in service. The findings of this study contribute to the ongoing efforts in the shipping industry to enhance operational efficiency, reduce downtime, and ensure sustainability and competitiveness in a rapidly evolving regulatory landscape.

＊　環境・動力系
＊＊　流体性能評価系
原稿受付　令和 ５年 ７月 １４日
審査日　　令和 ５年 ９月　 ７日

Contents

1. Introduction

1.1 Background

Over recent years, concerns about environmental issues have spurred the maritime industry to undergo a major shift towards zero-emission shipping. In that respect, digitalisation is a primary driver in transforming the maritime industry into a more sustainable and efficient transportation sector. The propulsion systems of new-build ships are becoming increasingly complex and large in scale, accepting a diversity of energy sources, including fuel cells, hybrid systems, etc. [1, 2]. In this context, digital technologies are emerging as powerful enablers, offering innovative solutions to optimise vessel operations, enhance energy efficiency, and support the adoption of zero-emission technologies. The concept of digital shipping encompasses the integration of advanced technologies, data analytics, and connectivity throughout the entire shipping value chain. One of the key drivers behind the role of digitalisation in achieving zero-emission shipping is the ability to leverage real-time data and analytics. Through the installation of sensors, monitoring systems, and IoT devices [3], vast amounts of data on various operational parameters have become available for processing onboard or at on-shore data centres [4]. This real-time data provides valuable insights into energy usage patterns, identifies inefficiencies, and enables prompt corrective actions to optimise operations and reduce emissions.

Furthermore, digitalisation facilitates the implementation of advanced data analytics and machine learning algorithms to analyse the collected data and identify opportunities for energy savings and emission reductions [5, 6]. Predictive maintenance algorithms [7] can help detect and address potential issues before they lead to equipment failures, optimising maintenance schedules and minimising downtime. Advanced optimisation algorithms can also optimise route planning [8, 9], considering factors such as weather conditions, fuel consumption, and emissions, resulting in more efficient and eco-friendly voyages. To this end, the development of an integration platform which aims at interfacing the knowledge domain with advanced data analytics and machine learning is a key stride towards realizing digital shipping. In this respect, the concept of the 'AI Chief Engineer', combining the power of a digital twin with a comprehensive set of algorithms to provide analytics on the propulsion system and decision support, represents a promising approach. The present development stage focuses on the conceptual design and composition of the 'AI Chief Engineer' system, exploring the algorithms required to fulfil the system functionality.

1.2　System Architecture

The 'AI Chief Engineer' system is designed to support or replace the onboard ship Chief Engineer and thus can involve several components and layers to effectively combine real-time data streams, engine digital twin, and data analytics. The outline of the system architecture is depicted in Fig. 1.

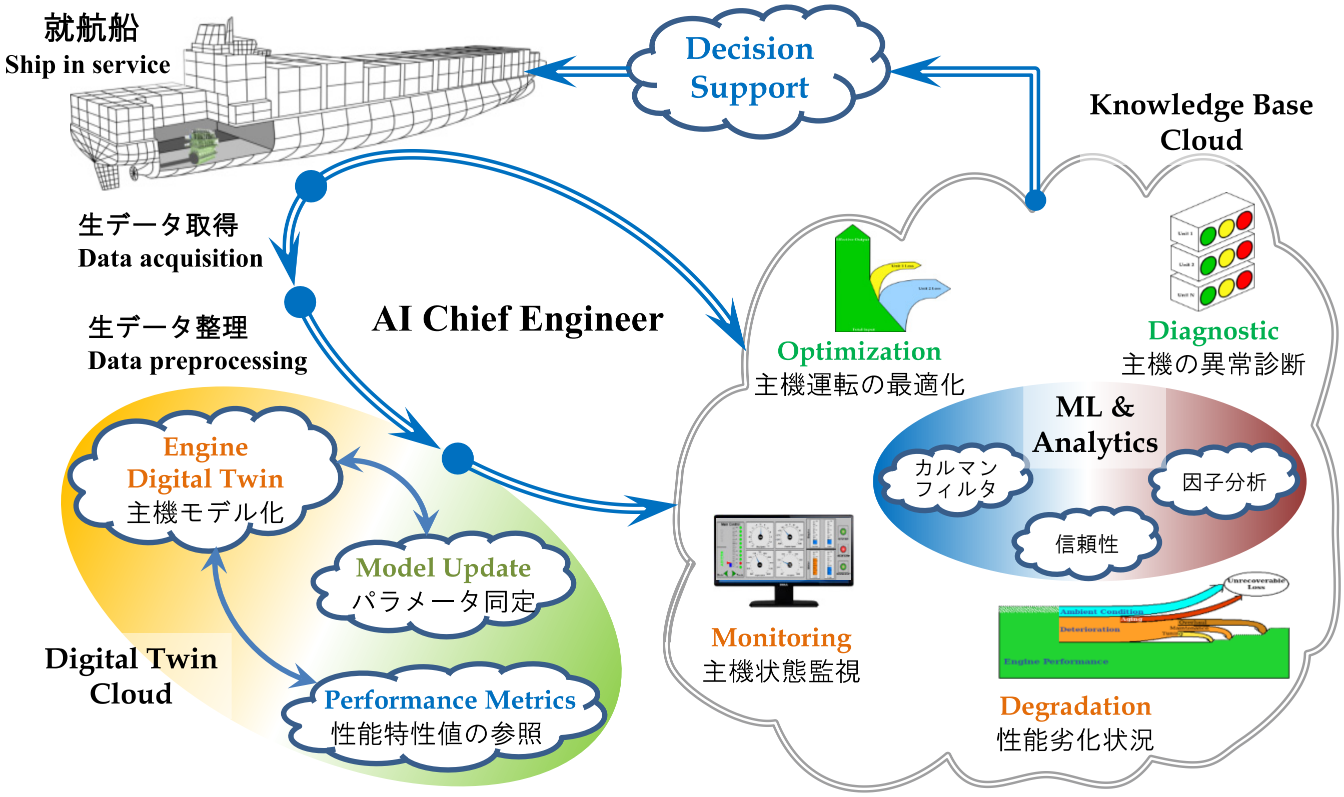


Fig.1 The outline of the ‘AI Chief Engineer’ concept

Although the data acquisition is not in the scope of the proposed system, it serves as an interface between a real asset and a twin, and thus, it cannot be overlooked. The data acquisition layer focuses on collecting real-time data from various sources on the ship, including sensors, meters, control systems and other relevant equipment. It involves data acquisition modules, such as IoT devices or data gateways, to gather data related to engine performance, operating conditions, environmental parameters, and other critical metrics [3].

Data preprocessing layer in which the acquired data undergoes information processing and integration to ensure its quality, consistency, and compatibility. These include data cleaning, filtering, normalisation, and synchronisation. The layer also involves the clustering of long-term historical data to create uniformly distributed data sets for the purpose of model parameters identification.

The digital twin cloud aims at developing and maintaining an accurate digital twin of the engine system. The engine's digital twin is a virtual representation that reflects the real-world ship's engine system, capturing its structure, components, behaviour, and performance characteristics. The cloud also consists of algorithms to manage the problem of parameters identification and adaptation and provide the reference performance metrics for a degradation analysis.

Knowledge base cloud consists of advanced analytics techniques, including machine learning algorithms, to analyse real-time and historical data. The analytics involves anomaly detection, degradation analysis, fault diagnosis, performance optimisation, and other relevant tasks. Machine learning models are trained using historical data and leverage the digital twin model to make accurate predictions and recommendations.

The decision support layer aims at conveying insights generated from the knowledge cloud to the users. This layer provides a user-friendly interface that displays real-time alerts, system health status, maintenance recommendations, and performance indicators. It can include interactive dashboards, reports, and visualisations to aid decision-making and facilitate proactive maintenance.

The outlined architecture aims to create an intelligent system that can monitor, analyse, and provide valuable insights for ship propulsion systems. It combines real-time data, engine digital twin, and analytics capabilities to support decision-making, improve operational efficiency, and enable proactive maintenance practices. The specific implementation and technologies used within each layer are presented and discussed in detail hereinafter.

2. Digital Twin Cloud

2.1 Outline

A digital twin is a vital piece of the digital transformation puzzle. It creates an accurate virtual replica of physical objects, assets, and systems to boost productivity, streamline operations and increase profits. The digital twin is a living-learning model that delivers valuable information by mapping the dynamic behaviour in real-time and ensuring that the twin is the exact replica of the asset. In such a way, using the digital twin to get insights, any aspect of the actual target can be explored through a digital interface. Specifically, simulation-based analysis of operational data from the physical counterpart can facilitate predicting the system performance and changes occurring over time, predicting the system response to safety-critical events and uncovering abnormalities by comparing predicted and actual responses. Thus, the digital twin binds an information/data source of the physical space, a set of dynamic models describing the physical counterpart in cyberspace, and a set of parameter values instantiated explicitly for the specific asset. Furthermore, the collection of parameters and, thus, models employed within this digital twin are continually updated based on the information/data from the physical twin. As such, the digital twin is envisaged as a holistic simulation platform and basement for analytics in a knowledge cloud.

The digital twin relies on a crucial demand: the models must precisely reflect the inherent attributes of a ship's propulsion system. This precision is vital to guarantee that the system's performance aligns seamlessly with its real-world counterpart. This synchronization should be held true across all operating conditions and be maintained in real-time [10].. The generic ship propulsion system may be considered to be made of three main components: an engine producing the torque, a shaft transmitting the torque from the engine to the propeller, and a propeller delivering thrust to a hull. Diesel engines remain an unavoidable part of the ship propulsion system, owing to the efficient conversion of a variety of fuel’s chemical energy into mechanical energy. Thus, the Diesel engines are considered a core part of the propulsion system’s digital twin and their mapping into cyberspace is of prime importance.

2.2 Engine’s Digital Twin

Diesel engine modelling has been evolving for decades since the development of computer simulation, and various model types can be classified depending on their degree of complication: transfer function models, quasi-steady mean value models and filling-emptying phenomenological models. Selecting the particular model is dictated by the requirements for the digital twin mentioned above – insight into the engine working process and real-time execution. Though, in the field of propulsion systems simulation, a cycle-mean value (CMV) engine modelling approach is widely used for evaluating the engine’s steady-state performance and transient response [11, 12]. In the CMV modelling approach, the engine is considered a series connection of throttles through which air and exhaust gas flow continuously, disregarding the intermittent nature of the engine cylinder processes. Standard thermodynamic equations are used to model the temporal evolution of pressures and temperatures in the control volumes, and there is no accumulation of mass and energy in the flow restrictions. The engine’s effective work due to fuel combustion is represented with the help of empirical functions. In this respect, the model provides the engine cycle-averaged temporal evolution of the engine operating parameters, neglecting the combustion process and in-cycle variation of parameters. The need for a more in-depth representation of the engine processes has led to the development of a Combined-CMV modelling approach [13], where a detailed model is used for representing the closed part of the engine combustion cycle, and the CMV model is used for the cycle-averaged representation of other engine components. As a result, the Combined-CMV approach is a suitable candidate for the digital twin of the engine in the framework of a propulsion system, and the details are discussed hereinafter.

The Diesel engine is a heterogeneous and complex system, and in order to develop its mapping into the cyberspace of digital twin, the method of system analysis is employed [14]. Following the method, the system under consideration is hierarchically decomposed into a number of lower-level entities. The information interface and physical variables are then determined to establish interconnections between the entities. Every entity is further decomposed into a finite number of components that are described by generic and reconfigurable mathematical models in terms of input/output relationships. The depth of decomposition depends on the required level and provided information. Thus, under the assumption used in the CMV model, the engine is decomposed to a finite number of elements that characterise the behaviour of the engine components. Specifically, as figure 2 shows, a cylinder element, air receiver and exhaust gas receiver elements, the turbocharger (TC) unit composed of the compressor and turbine elements, air cooler and exhaust ambient. The compressor and turbine elements are mechanically linked with the TC shaft element. An air cooler element is connected between the compressor and the air receiver. The fuel pump, i.e. fuel injector, is connected directly to the cylinder unit. The engine cylinder units are directly interfaced with the propeller by means of shaft rotational dynamics, expressed as:

$$2\pi I_e \frac{dn_e}{dt} = M_e - M_p \tag{1}$$

where M_e and M_p are the engine and propeller torques, respectively, I_e is the total moment of inertia, and n_e is the engine rotational speed.

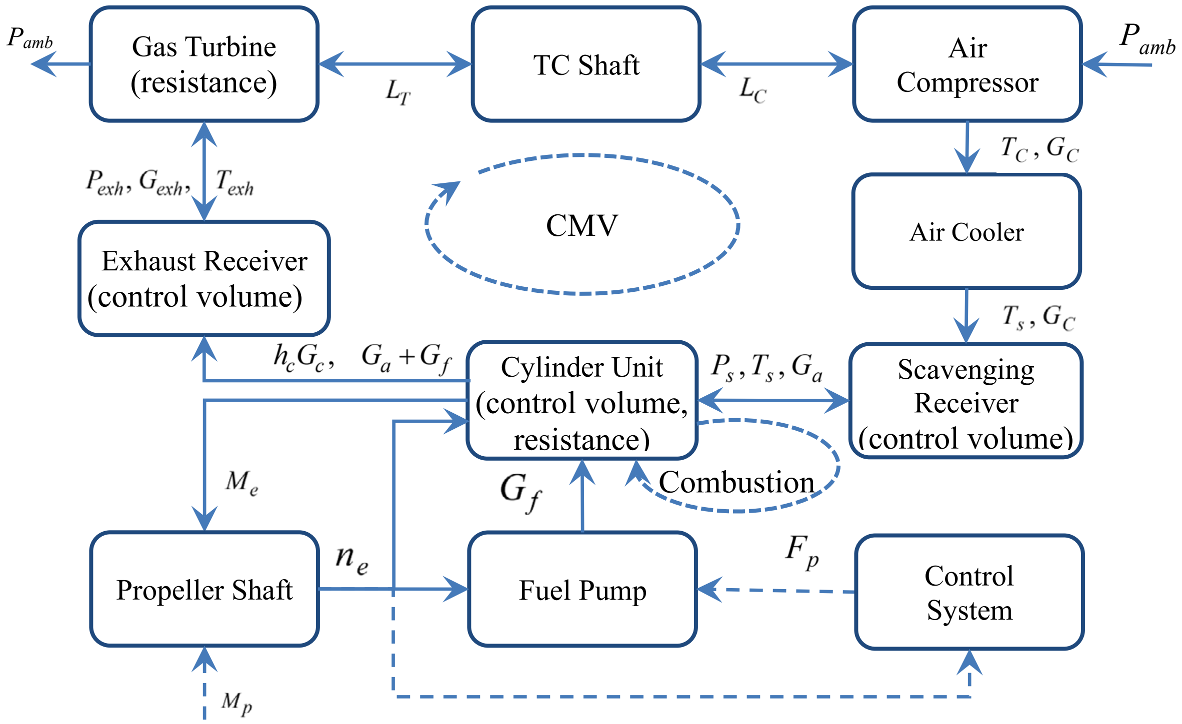


Fig. 2 System decomposition of the engine

The engine torque is the result of brake mean effective pressure (BMEP), P_b, which in turn is the difference between an indicated mean effective pressure (IMEP), P_i, developed in the cylinder volume V_s during one cycle and friction mean effective pressure (FMEP), P_f:

$$M_e = \frac{P_b V_s}{2\pi}, \quad P_b = P_i - P_f \tag{2}$$

From the above, it is clear that the objective of the engine model is to represent an external characteristic of the engine concerning the developed BMEP, which generally is a function of the engine states such as rotational speed, air mass flow and fuel mass flow. The complete details of the CMV engine model can be found in [14, 15], and brief details are given below.

The fundamental equations necessary to describe the temporal evolution of the state variables in the control volumes, V, can be obtained from the following mass and energy conservation laws along with the ideal gas equation:

$$\frac{dm}{dt} = \sum_i G_i \tag{3}$$

$$P = \frac{m}{V}\tilde{R}T \quad \Rightarrow \quad \frac{dP}{dt} = \frac{dm}{dt}\frac{\tilde{R}T}{V}, \quad \because T = \text{const} \tag{4}$$

$$c_v m \frac{dT}{dt} + u\frac{dm}{dt} = \sum_i h_i G_i \tag{5}$$

As was mentioned earlier, the central assumption in the CMV model is that the engine cylinder and turbine can be represented by the equivalent orifice, which produces the same mass flow rate for a given pressure ratio. This assumption allows employing a quasi-one-dimensional equation of flow through an orifice in the following form:

$$G = \mu\tilde{A}\frac{P_{in}}{\sqrt{RT_{in}}}\sqrt{\frac{2k}{k-1}\left[\left(\frac{P_{out}}{P_{in}}\right)^{\frac{2}{k}} - \left(\frac{P_{out}}{P_{in}}\right)^{\frac{k+1}{k}}\right]} \tag{6}$$

here, $\mu\tilde{A}$ is the effective equivalent area of the engine cylinder or turbine. The subscripts, in/out, stand for the inlet and outlet parameters of the considered element, correspondingly.

The energy flow rate, h_cG_c, exiting the engine cylinder element is calculated by taking into consideration the energy conservation equation, averaged over one engine cycle, thus:

$$h_cG_c = G_aC_{p.a}T_s + G_fH_U - W_i - Q_w \tag{7}$$

where W_i is the engine cylinder indicated power for one cycle, and Q_w is the corresponding heat loss rate. These two quantities are the result of combustion cycle calculation.

In the CMV modelling approach, the combustion simulation is substituted by the coefficient ζ_a, which denotes the proportion of the fuel chemical energy retained in the exhaust gas:

$$G_fH_U - W_i - Q_w = \xi_aG_fH_U \tag{8}$$

and such an approach allows for performing the fast calculation of engine states without explicit consideration of combustion.

Similarly, the simplification also applies to the engine's indicated power, W_i, which is considered proportional to the fuel pump index, F_p, modified by the combustion efficiency, η_c, to account for incomplete combustion:

$$W_i = F_pW_{i_0}\eta_c, \quad \because \eta_c = f\left(\frac{G_a}{G_f}\right) \quad \Rightarrow \quad P_i = \frac{W_i}{V_s} \tag{9}$$

In the meantime, the Combined-CMV modelling approach offers a more accurate estimation of the engine performance, in terms of W_i and h_cG_c, through the calculation of the combustion cycle. However, this comes at the cost of loss in the calculation speed. To address this concern, a fast calculation scheme of the engine cycle has been developed [15], and brief details are given hereinafter.

On the one hand, the engine cylinder is considered an equivalent orifice, ensuring the gas flow from the air receiver to the exhaust receiver. On the other hand, the engine cylinder can be considered a closed thermodynamic system, excluding the gas exchange process. Additionally, the combustion can be considered a quasi-static thermodynamic process, which is why the thermodynamic properties of the cylinder can be calculated considering a zero-dimensional thermodynamic approach applied to a stationary system. In this case, only the law of conservation for energy is considered, assuming that the working medium is an ideal gas, the state of which is homogeneous in space, and time is only the independent variable (or its equivalent in crank angle). The internal energy change between two finite states, according to the 1st law of thermodynamics, can be written in an integrated form as follows:

$$U_2 - U_1 = -\int_{V_1}^{V_2} p_c dV + \Delta Q_f - \Delta Q_w \tag{10}$$

The heat loss, ΔQ_w, between the working medium and cylinder walls is taken into account by using the standard equation for convective heat transfer, assuming a constant mean cylinder wall temperature:

$$\Delta Q_w = \alpha A_w\left(T - T_w\right)\Delta\varphi \tag{11}$$

The heat addition due to fuel combustion, ΔQ_f, is:

$$\Delta Q_f = \eta_c H_U \Delta q_x, \quad \because \Delta q_\chi = m_{f.c}\Delta\chi_c \tag{12}$$

where the heat release rate, Δq_χ, is modelled according to the Wiebe function, which is widely used in the zero-dimensional approach due to its simplicity and versatility [16]. The Wiebe function for the non-dimensional increment of burn rate, $\Delta\chi_c$, can be written as:

$$\Delta\chi_c = C\left(a+1\right)\phi^a \exp\left(-C\phi^{a+1}\right)\Delta\varphi, \quad \because \phi = \frac{\varphi - \varphi_{soi} - \varphi_{id}}{\varphi_z} \tag{13}$$

where φ represents the normalised combustion duration.

The parameters of the Wiebe function, notably: $\{C, a, \varphi_z\}$, characterise the burn rate shape, which should be fitted for every operating point of the engine. However, for practical considerations, various correlations of constants with the engine operating conditions were proposed [16]. Additionally, to include the losses due to the incomplete fuel combustion, a coefficient, η_c, is introduced in Eq. (12).

The internal energy, U, can be evaluated from the average specific heat capacity, $\tilde{C}_v$, as a function of temperature. Furthermore, the working medium is considered a homogeneous mixture of charge air and stoichiometric combustion products, hence:

$$
\begin{aligned}
&U = m\,\tilde{C}_v\,T,\\
&\qquad \because \tilde{C}_v = r\,C_v\left(T\right)\Big|_{r=1} + \left(1-r\right)C_v\left(T\right)\Big|_{r=0}
\end{aligned}
\tag{14}
$$

where the specific heat capacities for air and combustion products can be represented by the published analytical expressions relating the temperature, T, and gas composition ratio, r [17].
Assuming a small change of gas state in the transition between two finite states, and applying a Taylor expansion for the nonlinear functions, the following quadratic equation with respect to the unknown temperature increment, ΔT, can be obtained:

$$
k_1\left(\Delta T\right)^2 + k_2\left(\Delta T\right) + k_3 = 0 \tag{15}
$$

Thus, starting from the initial state of gas at the exhaust valve closed position and advancing the crankshaft by the angle $\Delta\varphi$, the new state of the thermodynamic system can be evaluated from the solution of the above equation:

$$
\begin{aligned}
&T_{j+1} = T_j + \Delta T,\\
&p_{c_{j+1}} = \left(\frac{mT}{V}R\right)_{j+1}, \quad V_{j+1} = V_j + \Delta V\\
&W_{i_{j+1}} = \frac{p_{c_j} + p_{c_{j+1}}}{2}\Delta V
\end{aligned}
\tag{16}
$$

The presented calculation method preserves high accuracy at relatively large calculation steps. Furthermore, flexible adjustment of the calculation step for various stages of the engine cycle provides a fast and accurate calculation of the complete combustion cycle.

2.3 Model Validation and Verification

The developed engine model has been validated against an actual marine two-stroke Diesel engine. The data of the test engine in a variety of loads were provided by the Mitsui E&S Machinery. The specification of the test engine is given in Table 1. Figure 3 depicts the results of a steady-state simulation of the Combined-CMV model, superimposed with the experimental data of the test engine. As can be seen, the well-tuned model provides fairly well agreement of the selected engine states, such as air mass flow and pressure, turbine inlet and outlet temperatures, as well as specific fuel consumption. At the same time, the performance of the combustion model was also confirmed, as illustrated by the traces of in-cylinder pressure.

One of the aims of the digital twin is the online monitoring and performance prediction of its physical counterpart by integrating real-time data. Thus, the real-time dynamic performance of the model employed in the digital twin is of great importance. Therefore, following the model validation and verification at steady state conditions, the developed Combined-CMV model was also used for simulating the transient engine operation. As figure 4 shows, the model follows pretty well the acceleration/deceleration transition of the actual engine. Furthermore, the real-time performance of the engine model was quantified as the time step lagging/leading by execution on the real-time operating system (RTOS). The results, illustrated in Fig. 4, clearly prove the real-time performance of the developed model – the timing error is bounded in the range of +-0.2%.

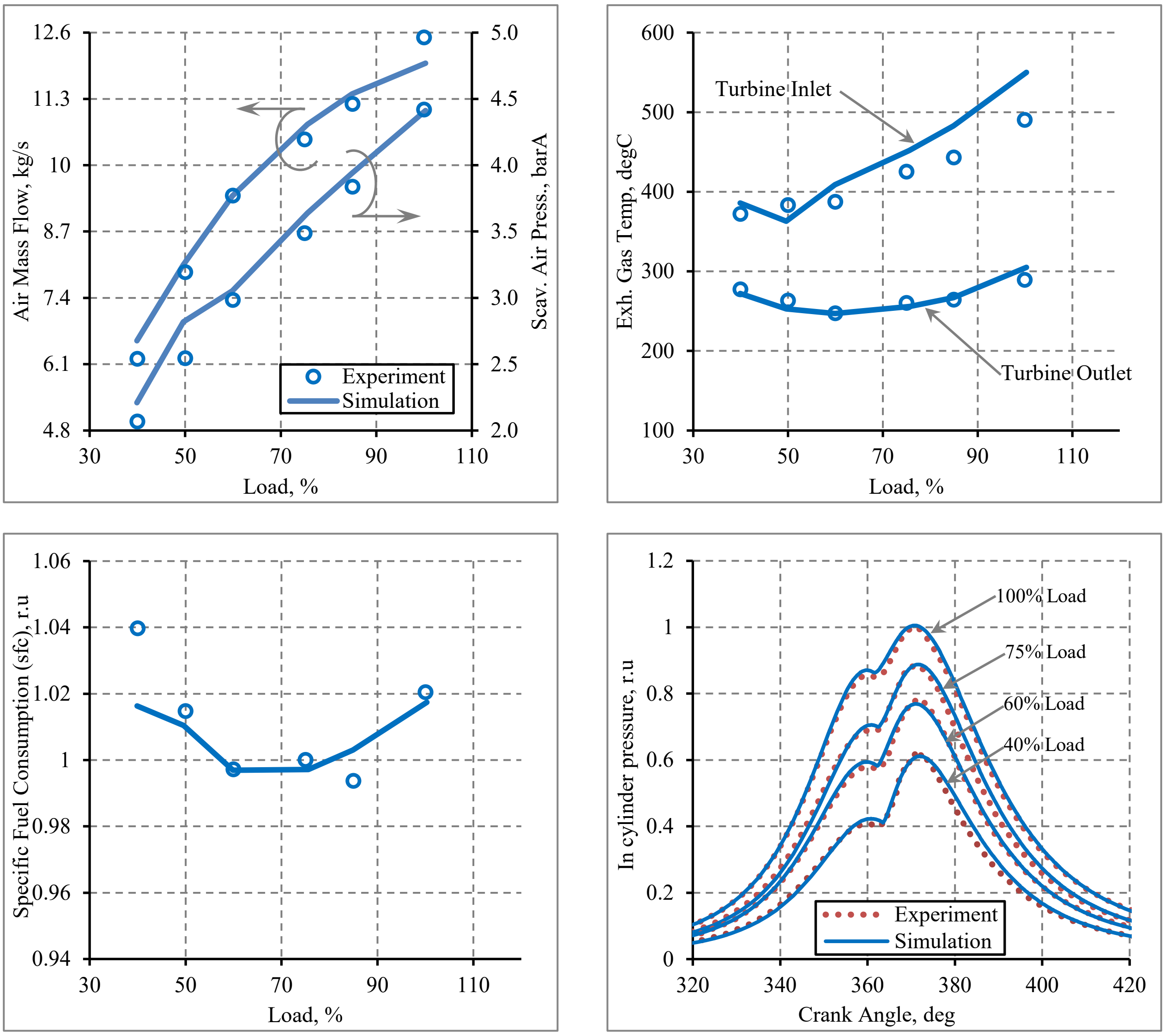


Fig. 2. Steady-state performance of the engine model

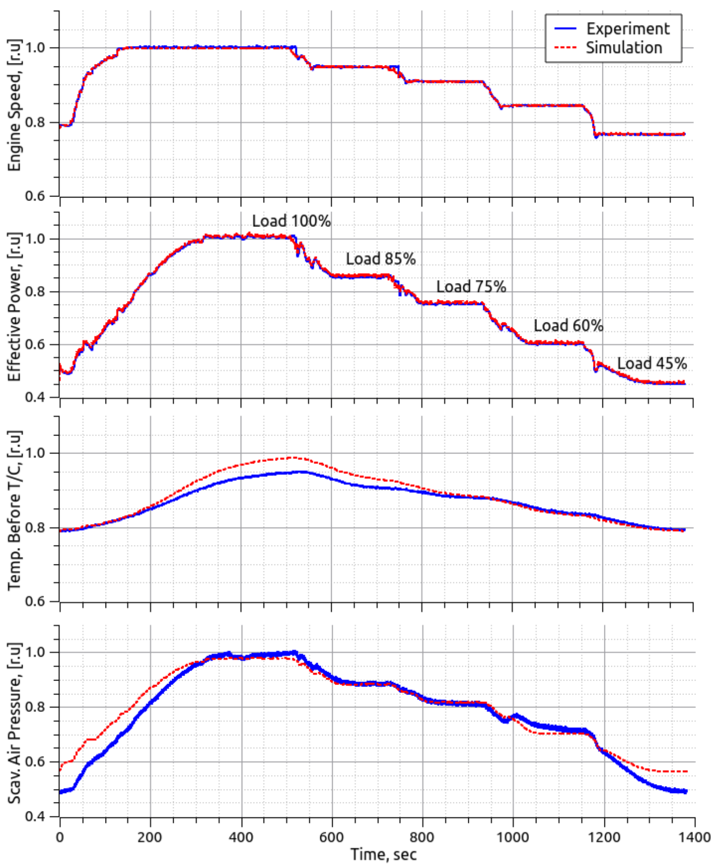


Fig. 3. Transient validation of the engine model

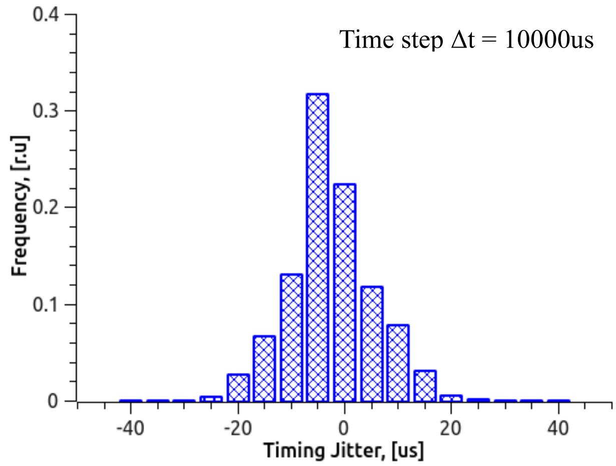


Fig. 4. Timing error of RTOS running the model

Table 1. Test engine specification @ MCR

Definition	Dimension	Value
Engine Type	[--]	Mitsui-MAN 4S50ME-T9
Cylinders	[--]	4 (in-line)
Bore/Stroke	[mm/mm]	500/2214
Power	[kW]	7120
Speed	[rpm]	117
IMEP	[bar]	22
Scav.Air Press.	[barA]	4.4

2.4　Model Parameters Identification and Adaptation

The preceding chapter's findings provide compelling evidence supporting the validity and effectiveness of the developed engine model. Additionally, as previously stated, the digital twin is a living-learning model that delivers valuable information by mapping the dynamic behaviour in real-time and ensuring that the twin is the exact replica of the asset. However, it is crucial to acknowledge that the model's performance relies on a set of tuning parameters and regression functions which are an integral part of the model's components. Therefore, it becomes imperative to address the problem of parameters identification and continuous adaptation to keep a model up to date with the actual operating conditions of the real ship. The approaches to complete the task mentioned above are diverse: nonlinear least squares and quasi-Newton, to name a few. However, the presence of noise and uncertainties associated with the actual operation data influence the robustness and accuracy of the parameters identification. In this respect, the task of parameters identification and adaptation can be viewed as a tracking filter problem consisting of the following key elements: 1) the underlying physics-based performance model of the engine's component, 2) the data-driven model representing the component's characteristic (either a simple regression or sophisticated neural network), 3) operation data from multiple sensors, and 4) an advanced algorithm for learning models to update the parameters in view of noise and uncertainties. The latter can be effectively tackled by utilising a family of stochastic recursive algorithms based on the Kalman Filter theory. Derivative-free implementation of the Kalman Filter, the Unscented Kalman Filter (UKF), is an efficient "second-order" technique for estimating the state of a dynamic system, and it can be easily reformulated to solve the parameters identification problem, though. This can be done by writing a "state-space" representation of the unit model with respect to unknown parameters:

$$\begin{aligned} &\boldsymbol{\theta}_{k+1}^{-} = \boldsymbol{\theta}_k + \boldsymbol{r}_k \\ &\boldsymbol{e}_k = \boldsymbol{Y}_k - \boldsymbol{F}\left(x_k, \theta_k\right) \\ &\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_{k+1}^{-} + K_k \boldsymbol{e}_k, \quad \therefore K_k = f\left(P_\theta\right) \end{aligned} \tag{17}$$

where the parameters set θ_k corresponds to a stationary process with identity state transition matrix, driven by process noise r_k, e_k is the error of the unit model output with respect to actual data.

The recursive update of parameters set is performed by the properly selected correction coefficient K_k (Kalman Gain), which in turn depends on the approximate error covariance matrix P_θ. The exact formulation of UKF is not displayed in this paper and can be found in the following reference [18].

The UKF algorithm is an intrinsically sequential procedure in that, for every recursion step, the parameters set is updated on the basis of one and only one data instance. However, in many training problems, and especially the considered problem in which input data sequence is heterogeneous, it is likely that the parameters be adapted unduly in favour of the currently presented training data. Furthermore, due to the operation profile of the ship, the data sequence may contain the data corresponding to only one operating mode. In contrast, the unit model should cover all possible operating modes. The remedy to this deficiency is to use a batch or, in other words, a multistream update algorithm [19]. It is based on the principle that each update step should simultaneously satisfy the demands from multiple sample points from the data set. In this respect, the extension of UKF is straightforward: the error vectors e_k of the unit models for every sample in the data set are concatenated to represent all the errors; at the same time, the parameter vector θ_k remains unchanged since all instances of the unit model share the same set of parameters:

$$\begin{matrix} \boldsymbol{e}_k^1 = \boldsymbol{Y}_k^1 - \boldsymbol{F}\left(x_k^1, \theta_k\right) \\ \vdots \\ \boldsymbol{e}_k^n = \boldsymbol{Y}_k^n - \boldsymbol{F}\left(x_k^n, \theta_k\right) \end{matrix} \Rightarrow \begin{bmatrix} \left\{-\boldsymbol{e}_k^1 -\right\}^T \\ \vdots \\ \left\{-\boldsymbol{e}_k^n -\right\}^T \end{bmatrix} \tag{18}$$

2.4.1 Service data processing

The ‘AI Chief Engineer’ system under development focuses on commercial ships, which operate in diverse marine environments and experience various ambient conditions. Furthermore, the presence of noise and uncertainties in the operational data can impact the robustness and accuracy of parameter identification. Therefore, it is necessary to preprocess the data to address these challenges. The initial step involves correcting the raw data in accordance with the relevant ISO standards. This correction ensures a standardised basis for comparison and model evaluation. Next, the variables are reduced to a non-dimensional form relative to the engine’s maximum continuous rating (MCR). Furthermore, the outliers and redundant data points due to measurement errors and other uncertain factors were labelled and removed from the data set based on the modified Z-score [20].

The ship’s operational profile has a significant impact on the accumulated data, such that the actual data is biased to its distribution and density of samples. An example of such an uneven and biased distribution is illustrated in Fig. 5. Consequently, when applying the identification algorithm to unevenly distributed data, there is a strong likelihood that the results will be skewed toward the group with a higher density of distribution. To address this issue, it is crucial to divide the data into subsets using cluster analysis. This approach aims to establish internal connections within each subset while ensuring clear separation between them. By doing so, we can mitigate the problem and improve the overall effectiveness of the identification process.

2.4.2 Adaptive clustering

In general, cluster analysis tries to subdivide a data set X into C subsets (clusters). The data in the same subset are more similar to each other than those in other subsets. In the Fuzzy c-means (FCM) clustering algorithm [21], the degree of similarity to either of the clusters is defined by introducing the membership function. The degree of memberships allows sampling the data in the close vicinity of the identified representative operating mode, thus equalising the distribution within one data set. For the purpose of this study, the Gustafson-Kessel FCM algorithm [22] has been adopted.

However, as an initial step, the algorithm requires an assumption on the number of clusters in the data and in many cases, the actual data does not reflect any particular structure of a set of clusters. Thus, it is proposed to adopt adaptive clustering. At first, an estimated number of clusters is overestimated, and after the initial clustering, a cluster merging algorithm is applied. Finally, the clusters are merged based on the similarity measure between them, defined as follows [23]:

$$FR_{ij} = \frac{dp_i + dp_j}{dv_{ij}} \tag{19}$$

where $d\upsilon_{ij}$ denotes the dissimilarity between two clusters C_i and C_j, i.e. Euclidian distance between clusters centres; In turn, a fuzzy dispersion dp_i is the fuzzy measure of the radius of cluster C_i, defined as follows:

$$dp_i = \sqrt{\tfrac{1}{n} \sum_{\boldsymbol{X} \in C_i} \boldsymbol{\mu}_i^m D_M^2\left(\boldsymbol{X}\right)}, \quad \therefore D_M\left(\boldsymbol{X}\right) = \sqrt{\left(\boldsymbol{X} - \boldsymbol{V}_i\right)^T S_i^{-1} \left(\boldsymbol{X} - \boldsymbol{V}_i\right)} \tag{20}$$

where $D_M(\boldsymbol{X})$ is the Mahalanobis distance, $\boldsymbol{S}_i$ is the cluster covariance matrix (available as the output of FCM algorithm), and $\boldsymbol{V}_i$ is the cluster prototype (centre).

Two clusters are considered similar if the corresponding similarity value, FR_{ij}, is the maximum and thus can be merged. The cluster merging process repeats until the similarity index drops below the threshold (usually < 1). Figure 6 illustrates the raw data and the results of adaptive clustering. As can be seen, the identified clusters fit well with the density distribution in Fig.5.

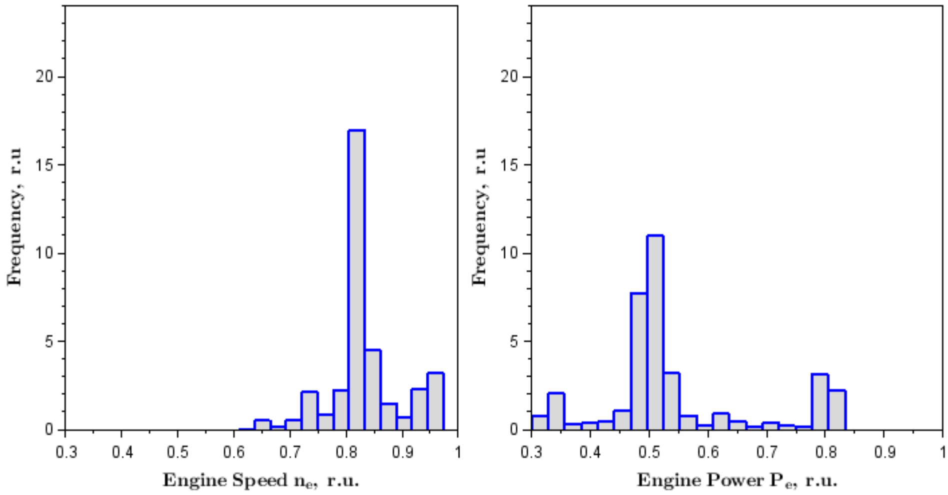


Fig. 5 Example of density distribution of engine power and shaft speed over one year of operation

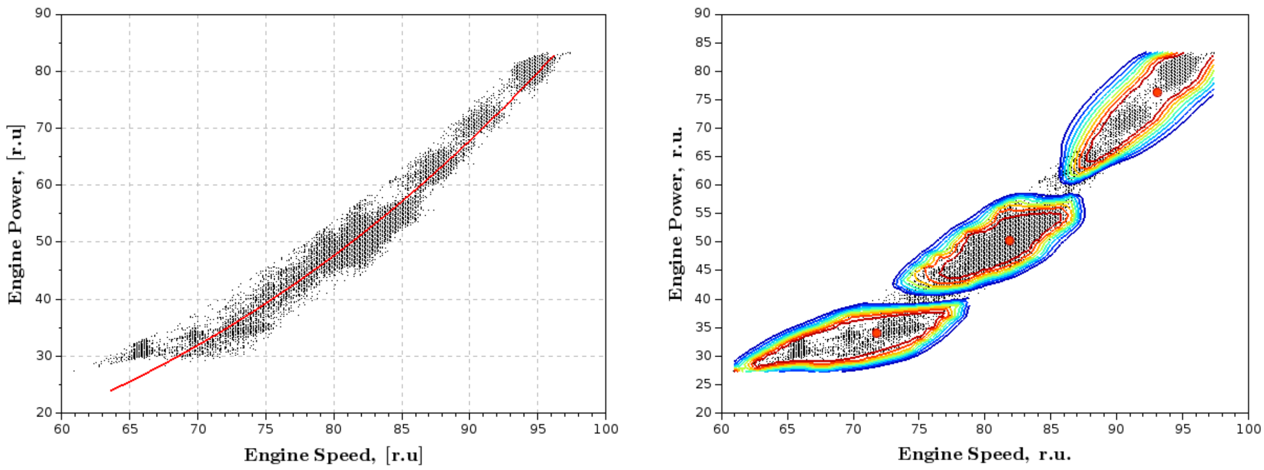


Fig. 6 Application of adaptive clustering to raw data of engine power – speed relation

2.4.3 Parameters Identification Example

In order to illustrate the developed framework of the propulsion engine digitalisation, for the sake of brevity, a simple yet informative case is considered hereinafter. The example shows the case of the turbocharger's turbine model parameters identification. The turbine performance affects the efficiency of fuel transformation to a great extent. The work done by the exhaust gas expansion in the turbine is transmitted to the compressor, which supplies air for combustion. However, the exact characteristics for efficiency, flow area, etc., are a priory unknown and have to be restored from the available in-service data.

At steady-state conditions, the power of the turbine is fully absorbed by the compressor, thus:

$$G_C L_C \frac{1}{\eta_{iC}} = G_T L_T \eta_{iT}\,, \quad \because L_C = C_{p.a} T_a \left(\pi_C^{\frac{k_a - 1}{k_a}} - 1 \right) \qquad \because L_T = C_{p.e} T_e \left(1 - \pi_T^{\frac{k_e - 1}{k_e}} \right) \tag{21}$$

The exhaust gas mass flow through the turbine is defined by Eq.(6), however, the effective flow area $\mu\tilde{A}$ is a function of the pressure ratio across the turbine:

$$\mu\tilde{A}_T = f\left(\pi_T\right) = \sum_{j=0}^{2} k_{Tj}\pi_T^j \tag{22}$$

Furthermore, the isentropic turbine efficiency, η_{iT}, holds a relation with the turbine characteristic speed in the following form:

$$\eta_{iT} = \eta_{iT_0}\left[u_c\left(k_0 - u_c\right) + k_1\right] \tag{23}$$

Finally, the exhaust gas temperature before the turbine requires the characteristic of the fuel chemical energy, ζ_a, retained in the exhaust gas in Eqs. (8), defined as follows:

$$\zeta_a = k_{\zeta_1} W_i + k_{\zeta_2} \tag{24}$$

Thus, the set of parameters to be identified includes the following coefficients:

$$\theta \in \left\{k_0, k_1, k_{T_0}, k_{T_1}, k_{T_2}, k_{\zeta_1}, k_{\zeta_2}\right\} \tag{25}$$

To this end, model parameters identification aims to restore the set of coefficients using the actual reading of available temperatures, pressures and mass flows. As a result, the identified turbine model provides a continuous estimation of exhaust gas temperatures, as illustrated in Figures 7a, b. The observed bias in the estimated exhaust gas temperature before the turbine is explained by the uncertainties in the measured exhaust gas temperatures and simplified representation of isentropic efficiency by Eq.(23).

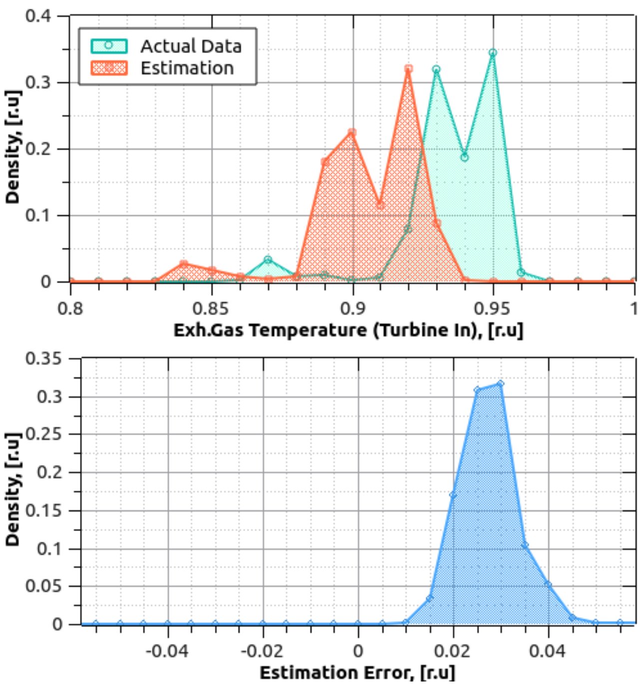


Fig. 7a Distribution of true value and error of estimated Exh.Gas temperature before turbine

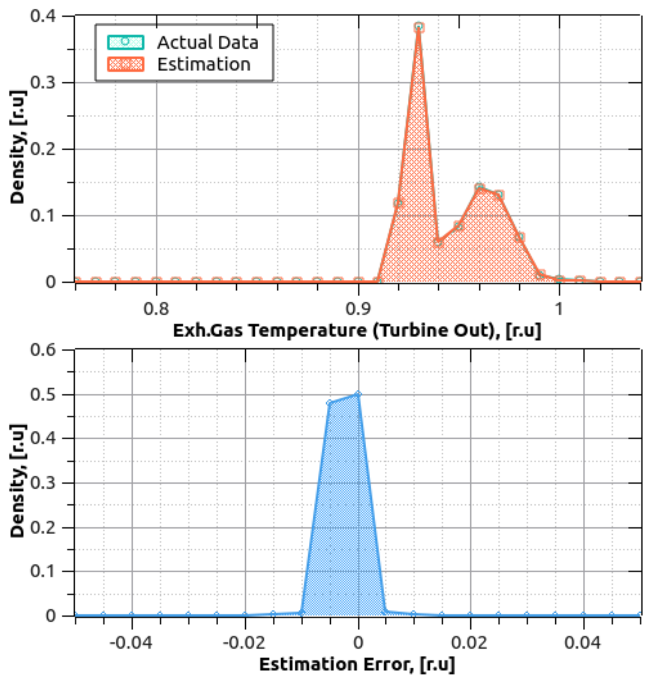


Fig. 7b Distribution of true value and error of estimated Exh.Gas temperature after turbine

Let's summarise the capabilities of the digital twin cloud, which is built upon the developed engine model using the Combined-CMV approach and a specialised cycle evaluation algorithm. This advancement offers enhanced detail and rapid execution. Additionally, the Tracking Filter concept integration enables seamless connectivity between the physical space's information/data source and the dynamic models representing the digital twin in cyberspace. As a result, the digital twin performs optimally, constantly aligning with its physical counterpart.

In conclusion, the digital twin cloud opens up various advanced monitoring capabilities for engines, such as detecting abnormalities or managing failure modes, tracking the impact of parameters or engine component degradation on thermal efficiency, and evaluating transient responses when control objectives are altered.

3. Knowledge Base Cloud

3.1 Outline

The 'AI Chief Engineer' system comprises two primary parts: the digital twin cloud and the knowledge base cloud. The latter acts as a repository of various algorithms and analytic tools, enabling advanced data analytics and decision capabilities within the considered advanced engine monitoring system. The interface between both parts plays a crucial role in extracting meaningful information from the digital twin's data and applying the appropriate algorithms for analysis. The digital twin cloud serves as a data source, providing real-time information on the propulsion system's reference state to the knowledge base cloud. Subsequently, the algorithms within the knowledge base cloud assist in identifying anomalies, predicting component's reliability degradation, and offering valuable recommendations for maintenance and operational optimisation. In particular, for the task of condition monitoring and anomaly classification, a combination of multidimensional statistical methods of factor analysis with methods from the fields of machine learning has been developed. Furthermore, for the task of degradation analysis, the method of quantification of the performance parameters departure from the reference based on the reliability function formulation has been developed. The brief details of both methods are given from now on.

3.2 Engine Condition Monitoring and Diagnostic

The application of the factor analysis (FA) method to the propulsion system condition monitoring is thoroughly examined within the knowledge base cloud. This method involves decomposing the covariance matrix of parameters' deviation from the reference state into a set of factor loading components. These components allow for the calculation of the relative contribution to the explained variance, which serves as a condition state indicator.

FA is the method designed to analyse interrelationships within a set of variables. The critical assumption of the method is that the measured variables can be correlated in such a way that their correlation may be reconstructed by a few (latent) common parameters, called factors. These factors, in turn, could represent the underlying structure and inherent interdependencies in a concise and interpretable form [24, 25]. Thus, the FA algorithm provides an explanatory model for the correlations among the data, which is highly suitable for engine diagnostics. The general factor model can be written as:

$$\boldsymbol{Y} = \boldsymbol{A}\,\boldsymbol{F} + \boldsymbol{\Psi}, \tag{26}$$

where $\boldsymbol{Y} = (y_1, y_2, ..., y_p)^{\mathrm{T}}$ is a vector of measured parameters, $\boldsymbol{F} = (f_1, f_2, ..., f_r)^{\mathrm{T}}$ is a vector of $r < p$ latent variables or factors, $\boldsymbol{A}$ is a (p x r) matrix of fixed coefficients (factor loadings) and $\boldsymbol{\Psi} = (u_1, u_2, ..., u_p)^{\mathrm{T}}$ is a vector of random error terms. The latter consists of errors of measurements, together with unique individual effects associated with each variable y_i.

The FA algorithm analyses the correlations among the parameter deviations from the base (reference) level of monitored parameters available from the digital twin cloud. Factor analysis can be viewed as a particular case of the weighted principal components model, and thus principal components are used to estimate the factors. The correlation matrix for a set of p measured parameters during a time of length N is given as:

$$\boldsymbol{R} = \frac{\boldsymbol{Y}^T\boldsymbol{Y}}{N},$$

$$Y_{i,j} = \frac{X_{i,j} - \overline{X}^{\langle j \rangle}}{\sigma_{X^{\langle j \rangle}}}, \quad i = 1, \ldots, N,\ j = 1, \ldots, p \tag{27}$$

where $\boldsymbol{Y}$ is a matrix (N x p) of scaled parameter deviations from the base level, σ_x is the parameter standard deviation.

In turn, the (p x p) correlation matrix $\boldsymbol{R}$ can be expressed as the sum of p-eigenvalues multiplied by their eigenvectors and their transpose, that is a singular value decomposition (SVD) of matrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \boldsymbol{USV}^T, \tag{28}$$

Eventually, the estimator of the factor loadings matrix is defined as:

$$\boldsymbol{A} = \boldsymbol{U}\sqrt{\boldsymbol{S}}, \tag{29}$$

The principal components analysis gives a set of axes to which variables could be referred, and thus the factor loadings are the projections of the variables onto the factor axes. Although mathematically unique, a set of factors is the only one of an infinite number of sets that describes data configuration just as well. Still, it does not guarantee maximum statistical significance. Therefore, as the second step of the analysis, the maximum likelihood principle can be used to specify the elements of matrix ***A*** that best model the correlation structure of data ***Y***. In the first step towards refinement of the model, the factors ***F*** and data ***Y*** are assumed as a joint Gaussian distribution:

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{F} \end{bmatrix} \sim \aleph\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{A}\boldsymbol{A}^T + \boldsymbol{\Psi} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{I} \end{bmatrix}\right), \tag{30}$$

To maximise the parameters of the joint distribution above, the log-likelihood function of the parameters yields:

$$\ell(\boldsymbol{A},\boldsymbol{\Psi}) = \sum E \begin{bmatrix} -\frac{1}{2}\log\boldsymbol{\Psi} - \frac{N}{2}\log 2\pi ... \\ -\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{F})^T \boldsymbol{\Psi}^{-1}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{F}) \end{bmatrix}, \tag{31}$$

The maximisation of the likelihood function with respect to the matrix ***A,*** is performed by applying the iterative Expectation-Maximisation (EM) algorithm. For the sake of brevity, the details of the algorithm are omitted and can be found in [26].

Although, as was mentioned earlier, the factors contain the essential information related to the variance of measured parameters, the factors' loadings are considered more informative for the condition monitoring purpose. This is because the components of factor loadings matrix ***A*** reflect the degree of contribution of every parameter deviation onto the explained variance of every factor. Thus, a performance index of the equipment healthiness is defined as the relative share of the explained variance formed by the factors loadings of the measured parameters:

$$D^{\langle n\rangle} \equiv \sigma^2\left[F^{\langle n\rangle}\right] = \frac{\sum_{i=1}^{p} a_{i,n}^2}{\sum_{n=1}^{k}\sum_{i=1}^{p} a_{i,n}^2}, \quad n = 1,\ldots,k \tag{32}$$

As one may note, the number of performance indices is equal to the number of processed parameters p, and this fact does not facilitate efficient condition monitoring. Simultaneously, the eigenvalues of the correlation matrix ***R*** usually decay rapidly, which suggests that only a few first components of ***R*** and, thus, a few first factors can be considered as principal and containing a significant share of the explained variance. Therefore, k in Eq. (32) is limited to $k \leq 2$.

As mentioned above, the factor loadings indicate the relationship strength between parameters and factors variation. Therefore, this property can be considered as a feature that characterises every measured parameter's contribution to the variance of the performance index. Thus, the relative contribution of loadings to the variance of factors is given as:

$$\lambda_{i,k} = \frac{a_{i,k}^2}{\sum_{i=1}^{p} a_{i,k}^2}, \quad k = 1,2 \tag{33}$$

3.2.1 Case study of condition monitoring

This section presents a case study demonstrating the FA methodology implementation to the condition monitoring framework. Due to a lack of statistics concerning the failures of the actual engine, numerical simulation was used instead. With the support of the GT-Power engine simulation software, the Mitsui E&S Machinery provided modelling of the owned two-stroke marine diesel test engine. Table 1 summarises the specification of the test engine. The simulation condition was set to a constant speed and load corresponding to 75% of the Maximum Continuous Rating (MCR). The failures that were imitated in the simulation model are listed in Table 2. The data set for every failure consists of a steady-state simulation period followed by the response of engine state variables to the introduced incipient failures. The failures were introduced at T = 4200 sec and lasted for ΔT = 7200 sec. The failure development law is considered linear in time, however, due to the specificity of the software, only an exhaust valve blow-by is introduced as an abrupt event. The parameters listed in Table 3 were recorded every 1 sec.

Table 2. List of imitated failures

Item No.	Definition	Imitation method
1	Air cooler (AC) fouling	linear increase of AC outlet air temperature
2	Cylinder No.4 blow-by	linear increase of cylinder clearance
3	Exh. Valve No.4 burn out	orifice between the cylinder and exhaust gas receiver, initiated in an instant
4	Compressor filter clogging	linear increase of compressor suction pressure
5	Turbocharger degradation	linear decrease of efficiency multiplier

Table 3. List of recorded parameters

Engine Speed, [rpm]	Ne
TCH Speed, [rps]	Ntc
Scav. Air P., [kPa]	Ps
Scav. Air T., [K]	Ts
Exh. Gas P., [kPa]	Pe
Exh. Gas T., [K]	Te
Max. Compression P., [bar]	Pc
Max. Combustion P., [bar]	Pz
Indicated Mean Effective P.(IMEP), [bar]	Pi
Exhaust Valve Timing [deg]	τ_{exh}
Injection Timing Cyl. No.4 [deg]	τ_{inj}

Figure 8 shows an example of the incipient failures mapping in the domain of factor loadings. The sample number in Fig. 8 and hereinafter corresponds to a vector of measured parameters, each consisting of 180 timesteps. As can be observed, the factor loadings trends show different tendencies as different incipient failures progress with time. Therefore, every incipient failure or degradation of engine components is characterised by a unique set of factor loadings. In order to demonstrate the uniqueness of every set of factor loadings, a distance matrix based on the generalised Mahalanobis criteria is visualised in Fig. 9a. On the figure, the larger the circle, the better the distinction between data sets. In turn, Fig. 9b shows a dendrogram of hierarchical cluster analysis based on the distance matrix. The height of each junction represents the distance between the two data sets being connected. The analysis starts from the closely spaced data sets and progresses towards the outermost. Thus, data sets 2 and 5 are closely spaced, whereas data set 1 is the outermost of the others. The numbers correspond to Item No. In Table 2.

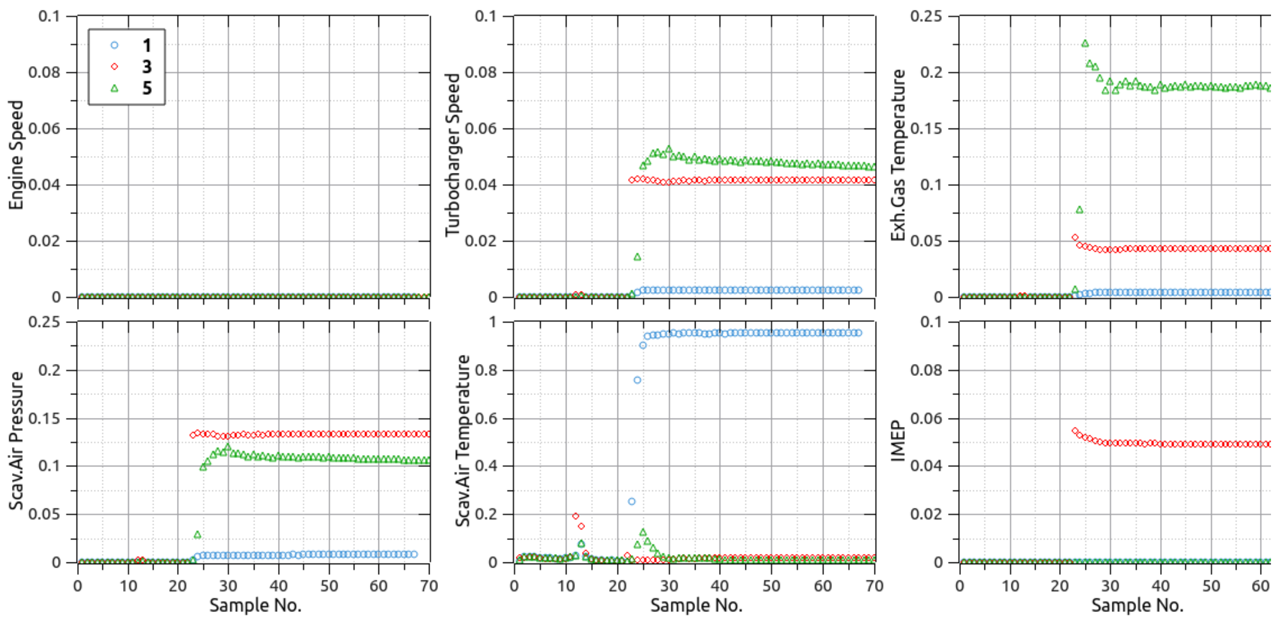
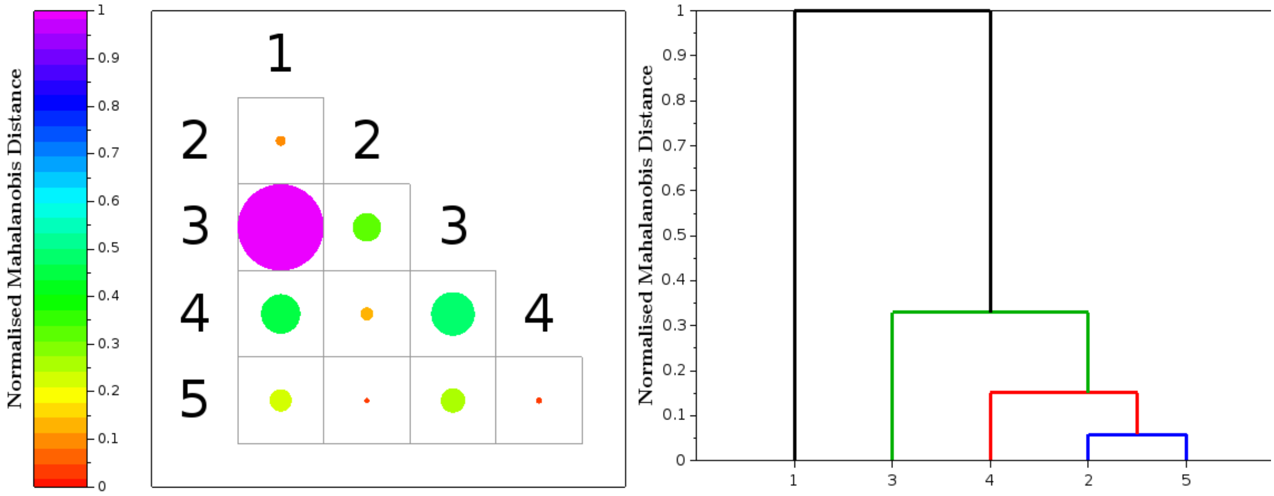


Fig. 8 Incipient failures mapping in the domain of factor loadings



a) Distance matrix of factor loadings set

b) Hierarchical cluster analysis

Fig. 9 Factor loadings space clustering analysis

From the perspective of a demonstrated cluster separation, it becomes evident that the classification algorithms of ML are highly suitable for engine failure identification, and companion research [27] elaborated in detail on applying two algorithms: Self-Organising Map (SOM) and Support Vector Machine (SVM).

3.3 Analysis of the Engine Reliability Degradation

In general, the term degradation refers to an internal process in components of a system, where undesired performance departure takes place from its intended reference due to various circumstances. With respect to the propulsion engine, the performance change can be classified as change due to the ambient conditions, which are not defined as degradation, though, and the change sourced from the degradation of the engine components. In turn, the performance degradation of engine components can be classified as deterioration due to mechanical wear-out, fouling and ageing occurring over time-based. The degradation process consists of three key stages. The first stage, when the departure from the reference state is too small to influence the system significantly, represents a normal operating state. The second stage, when the departure reaches a predefined threshold, represents a state of failure where a significant reduction in machinery performance occurs. The last stage, when the departure exceeds the threshold, represents a faulty state where the system can no longer perform one or more required functions. Thus, the task of reliability monitoring can be formulated as a quantification of the degradation process, that is, the probability of performance metric departure from the reference state. There are three critical components in the problem of reliability monitoring: a machinery component information loop which binds available sensors measurements with the selected performance metric; a model of machinery component which provides reference characteristics of the performance metric; and a model, degradation process, which can quantify the metric departure.

3.3.1 Information loops of the propulsion engine

Following the system analysis method, the engine was decomposed into a set of functional elements in order to facilitate its mapping into the cyberspace of the digital twin. Every element is described by a mathematical model that relates input/output information preserving proper interface between components. Thus, a performance metric can be considered for every element based on physical principles and evaluating the available information. In order to illustrate the idea of information loops, simple yet informative cases are illustrated hereinafter. The turbocharger compressor provides air supply to the engine cylinder for the combustion of fuel and thus is crucial for the overall propulsion engine performance. The measurements of air temperatures, T_c, T_a, and pressures, P_s, P_a, across the compressor provide for the evaluation of the isentropic efficiency of the air compressor unit as a performance metric, expressed as follows:

$$\eta_{i_C} = \left[\left(\frac{P_s}{P_a}\right)^{\frac{k_a-1}{k_a}} - 1\right]\left(\frac{T_c}{T_a} - 1\right)^{-1} \tag{34}$$

Air cooler (AC) is used to keep the air density after the compressor as high as possible, and fouling of the air side or water side is the major cause of the performance degradation affecting the performance of the propulsion engine. The measurements of air temperatures, T_c, T_s, across the AC, together with the temperature of cooling water, T_w, provide for the evaluation of AC effectiveness as a performance metric, expressed as follows:

$$\eta_{AC} = (T_s - T_w)(T_c - T_w)^{-1} \quad (35)$$

The performance of the propulsion engine depends primarily on the fuel combustion efficiency in the cylinder unit. Therefore, as a rule, modern engines are equipped with in-cylinder pressure sensors and fuel flow meters to estimate indicating efficiency as a performance metric of cylinder unit. The indicating efficiency, η_i, is defined below as a ratio of indicating power output, $Z_c V_s n_e P_{imep}$, to the total fuel energy input, $LCV_f G_f$.

$$\eta_i = (Z_c V_s n_e P_{imep})(LCV_f G_f)^{-1} \quad (36)$$

Such data fusion for the purpose of performance metric evaluation is considered a part of the Digital Twin Cloud in the framework of The ‘AI Chief Engineer’ system. The information loops are not limited to the presented above and can be extended depending on the available information.

3.3.2 Quantification of the degradation process

A reliability function concept [28] is used as quantification metric to evaluate the degradation from the departure trend of the investigated engine’s system. The degradation of the machinery unit is manifested in the performance metric departure over time from the reference value toward the threshold. The threshold denotes the system operating limit (SOL) that ensures operation within acceptable reliability criteria. Thus, the operating limit is a deterministic value which is assigned by the machinery manufacturer or based on some consideration of reliability. On the other hand, the value of the performance metric, determined by the measured parameters, is a stochastic value affected by environmental factors, sensors noise, etc., and thereby it is characterised by some distribution. However, due to the operating profile of the ship, the parameters of the distribution, mean, and variance are variable, and what is more a priori unknown. Instead, the normal distribution generalising all the stochastic factors is bound with a limit value, the parameters of which can be fixed. In that respect, the degradation is quantified as a probability that the performance metric corresponds to the reference condition. Figure 10 illustrates the outlined concept. Thus, the reliability function can be defined as follows:

$$F_r = \int_{\Delta x}^{\Delta x_{\lim}} f(x_{\lim})\,dx = \Phi(\Delta x_{\lim}) - \Phi(\Delta x), \quad \therefore \Phi(x) = \frac{1}{2}\left[1 + erf\left(\frac{x - x_{\lim}}{\sigma_x \sqrt{2}}\right)\right] \quad (37)$$

Furthermore, the condition of normal operation can be defined as $\Delta x_c \le \Delta x_{\lim} - \Delta x$, where Δx_c denotes the state of failure, and hence the distribution to the right of $x_{\lim}$ is meaningless. In this way, the distribution, $f(x_{\lim})$, is truncated to a half: $f_{1/2}(x_{\lim}) = K f(x_{\lim})$, where K is the scale coefficient, and it can be shown that for the normal distribution $K = 2$. The parameters of limit value distribution are estimated from the assumption of a minimum reliability value when the parameter enters the failure state by solving the following equality:

$$erf\left(\frac{\Delta x_{\lim} - x_{\lim}}{\sigma_x \sqrt{2}}\right) - erf\left(\frac{\Delta x_{\max} - x_{\lim}}{\sigma_x \sqrt{2}}\right) = F_r^{\min}\Big|_{\sigma_x}, \quad \therefore \Delta x_{\max} = \Delta x_{\lim} - \Delta x_c \quad (38)$$

Figure 11 illustrates the parametrised reliability function for the following assumed parameters: the acceptable value of reliability, at the state of failure commence, is set to F_r = 85%, the maximum departure of the performance metric is set to $\Delta x_{\max}$ = 5%, and the limit value is set to $x_{\lim}$ = 10%.

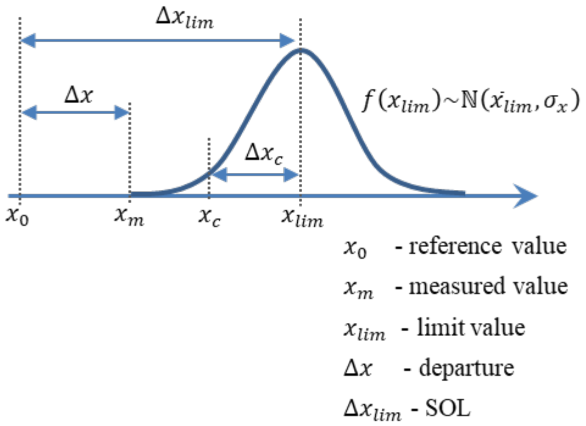


Fig. 10. Concept of the reliability quantification

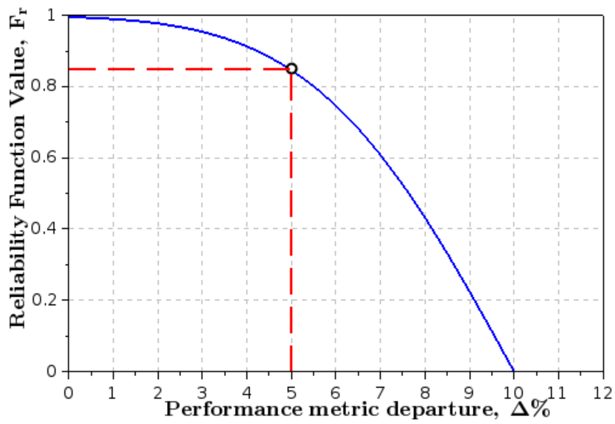


Fig. 11. Parametrised reliability function

3.3.3 Case study of degradation tracking

To demonstrate the applicability of the developed framework, the long-term service data was subjected to analysis and evaluation of the reliability function. The MOL shipping company provided the data set from one of the operated ships. The data are collected over five years of operation, starting from the vessel commissioning and consist of the records at the one-hour interval by the Voyage Data Recorder (VDR) and Engine Control System (ECS). The data are split into two periods by a dry-dock event. The reliability monitoring was applied to the engine cylinder information loop, those performance metric is expressed by Eq. (36). Figure 12 shows the results of the reliability function application according to Eq. (37) with the parameterisation as shown in figure 11. The results illustrate the average values on a monthly basis. As can be seen, there is some sign of degradation after commissioning, followed by a long period of normal operation which degrades again before the dry-docking. After the dry-docking, the performance was restored to the reference condition and through the second period of service, there were a few signs of degradation followed by restoration.

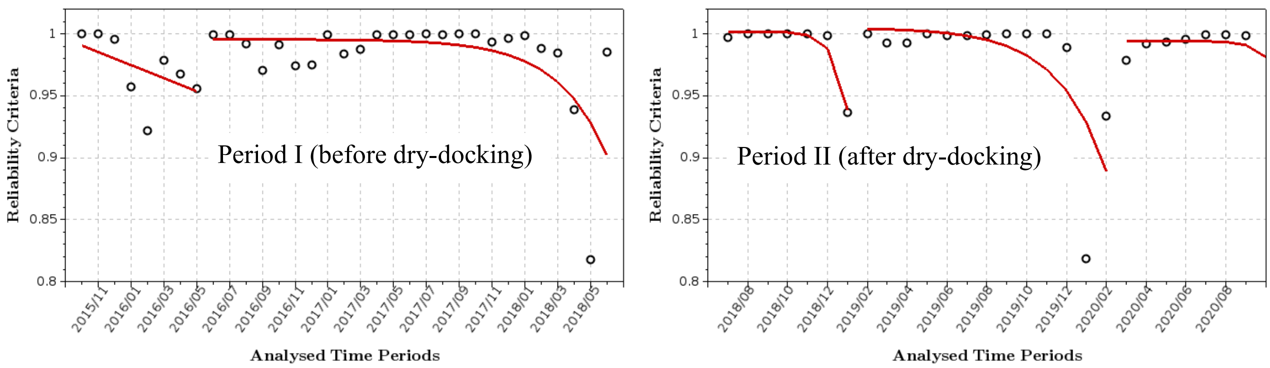


Fig. 12. Propulsion engine reliability analysis throughout the service period

In conclusion, the developed reliability monitoring framework, integrated within the knowledge cloud and operating in conjunction with the engine digital twin and component performance metrics, enables real-time monitoring and assessment of the engine's health, thereby facilitating proactive maintenance and predictive decision-making. This integration allows for a comprehensive understanding of the engine's behaviour and the identification of potential issues before they escalate into critical problems.

4. Conclusions

In conclusion, the ongoing development of the 'AI Chief Engineer' system, with the introduction of the digital twin cloud and the knowledge base cloud, represents a significant advancement in the field of propulsion system digitalisation. This integrated system combines the power of digital twins, advanced analytics algorithms, and real-time data to optimise the performance and reliability of propulsion engines. The digital twin cloud forms the foundation of the 'AI Chief Engineer' system, incorporating an engine digital twin, a parameters identification algorithm, and a collection of engine component performance metrics. The engine's digital twin serves as a virtual replica of the physical engine, capturing its behaviour, characteristics, and performance in real-time. Owing to the developed Combined-CMV approach with the special algorithm of the cycle evaluation, great detailing and fast execution are now provided. At the same time, the concept of a tracking filter is proposed. This allows interfacing the information/data source of the physical space with the set of dynamic models describing the physical counterpart in cyberspace, thus providing the ultimate performance of the continuously matched digital twin. This digital representation enables comprehensive monitoring, analysis, and prediction of the engine's operational status. Complementing the digital twin cloud, the knowledge base cloud is a collection of advanced analytics algorithms based on actual data and digital twin predictions. This cloud harnesses the power of data analytics to extract meaningful insights, identify patterns, and detect anomalies. There are two algorithms have been elaborated within the knowledge base cloud. These are the condition monitoring algorithm based on the factor analysis method and the degradation tracking algorithm based on reliability function formulation. The condition monitoring algorithm utilises factor analysis to identify hidden variables and patterns within the engine's performance data. By analysing these factors, the algorithm can assess the current condition of the engine, detect early signs of potential issues, and trigger timely maintenance actions. This proactive approach helps to minimise downtime, optimise maintenance schedules, and prevent costly failures.

The degradation tracking algorithm focuses on the long-term performance and reliability of the engine. It leverages time-series data, combined with insights from the digital twin, to quantify the probability of the engine's performance departure from the reference providing an estimate of its remaining useful life.

Overall, the ongoing development of the 'AI Chief Engineer' system, incorporating the digital twin cloud and the knowledge base cloud, demonstrates the potential for transformative advancements in shipping digitalisation. By combining real-time data, digital twin technology, and advanced analytics algorithms, this integrated system empowers organisations to enhance their engines' performance, reliability, and efficiency. As further refinements and advancements continue, the 'AI Chief Engineer' system is poised to revolutionise the field, enabling proactive and data-driven decision-making in engineering operations.

Acknowledgements

The authors wish to express sincere gratitude to Mitsui E&S Machinery for granting access to their test engine for the extensive tests and sharing their extensive experience on engine performance and failure modes. Furthermore, the authors wish to extend appreciation to MOL Shipping for providing the necessary data on ships in service for our analysis.

The authors are truly grateful to both Mitsui E&S Machinery and MOL Shipping for their kind assistance, which has significantly contributed to the success of our project.

References

1) Sato E., Amano M.: Introduction of Hybrid Propulsion Plant STaGE for LNG Carrier and Application to Offshore Power Plant, Journal of Japan Society of Marine Engineering, Vol.54, No.2 (2019), pp.6-11
2) Hirata K., Kiyokawa K., et.al: Efforts to develop a hydrogen fuel cell ship, Journal of the Japan Society of Marine Engineering, Vol.54, No.2 (2019), pp.34- 37
3) Shibata J., Mimura Y., Ando H.: Ship Data Collection Platform SIMS and Data Utilization Initiatives, Journal of the Japan Society of Marine Engineering, Vol.54, No.2 (2019), pp.121-125
4) Ikeda Y.: Latest Information on IoS Open Platform Consortium and ShipDC (Ship Data Center), Journal of Japan Society of Marine Engineering, Vol.54, No.2 (2019) pp.67-75
5) Omine M.: "SOPass" a ship operation management support system using IoT technology, Journal of the Japan Society of Marine Engineering, Vol.54, No.2 (2019), pp.126-128
6) Murakami M., Hatta T.: Ship Operation Support Service "Fleet Monitor" - Configuration, Examples of Use and Introduction of Next Generation Service "Fleet Transfer", Journal of Japan Society of Marine Engineering, Vol.54, No.2 (2019), pp.129-134
7) Coraddu A., Oneto L., Ghio A., et.al.: Machine learning approaches for improving condition-based maintenance of naval propulsion plants, Proc. IMechE, Part M: J. of Engineering for the Maritime Environment, Vol. 230(1) (2014), pp.136-15.
8) Mizuryu M.: Introduction of Optimal Operation Support Tool "ClassNK-NAPA GREEN", Journal of Japan Society of Marine Engineering, Vol.54, No.2 (2019), pp.117-120
9) Figari M., Zaccone R., Boeris G. and Martelli M.: Ship Voyage Optimisation: a Comparison between Forecast and Real Route Selection, Proc. The 2nd International Conference on Modelling and Optimization of Ship Energy Systems (MOSES2019), Glasgow, (2019), pp.199-205.
10) GE Digital Twin. Analytic Engine for the Digital Power Plant, Available from: https://www.ge.com/digital/applications/digital-twin
11) Hendrics E.: Mean value modelling of large turbocharged two-stroke diesel engines. SAE technical paper no 890564 (1989)
12) Theotokatos G.: On the cycle mean value modelling of a large two-stroke marine diesel engine. Proc. IMechE Part M: J. Engineering for the Maritime Environment, Vol. 224, (2010), pp.193-205.

13) Baldi F, Theotokatos G, Andersson K.: Development of a combined mean value zero-dimensional model and application for a large marine four-stroke diesel engine simulation. Appl. Energy (2015), 154:402–15
14) Bondarenko O., Fukuda T.: Assessment of Propulsion System Integration via Process Modelling, Papers of National Maritime Research Institute， Vol.17, No.4 (2018), pp.527-546
15) Bondarenko O., Fukuda T., 2020, Development of a Diesel engine's Digital Twin for the prediction of propulsion system dynamics, Energy 196
16) Ghojel, J.I.: Review of the development and applications of the Wiebe function: a tribute to the contribution of Ivan Wiebe to engine research. International Journal of Engine Research. Vol.11 (2010), 297-312
17) Grigor'eva, V.A., Zorina, V.M.: Theoretical foundations of thermal engineering, Energoatomizdat, Moscow, Vol. 2 (1988)
18) Haykin S., Kalman Filtering and Neural Networks, Wiley (2002).
19) Puskorius G.V., and Feldkamp L.A.: Multistream extended Kalman filter training for static and dynamic neural networks, IEEE International Conference on Systems, Man, and Cybernetics, Vol.3 (1997), pp. 2006-2011
20) Iglewicz B., Hoaglin D.: The ASQC Basic References in Quality Control: Statistical Techniques, Editor: Edward F. Mykytka, Volume 16: How to Detect and Handle Outliers (1993)
21) Bezdek J.C., Ehrlich R., Full W.: FCM: The Fuzzy c-Means Clustering Algorithm, In: Computer & Geosciences, Vol.10, Issue.2-3 (1984), pp.191-203
22) Gustafson D.E., Kessel, W.C.: Fuzzy clustering with a covariance matrix, In: Proc. IEEE CDC, Vol. 2 (1979), pp.761-766.
23) Xiong X., Chan K.L., Tan K.L.: Similarity-Driven Cluster Merging Method for Unsupervised Fuzzy Clustering, In: Proc. of the 20th Conference on Uncertainty in Artificial Intelligence (UAI2004) (2004), pp. 611-618.
24) Harman H.H.: Modern Factor Analysis. Second Edition. The University Chicago Press. (1968)
25) Cureton E.E., D'Agostino R.B.: Factor Analysis, An Applied Approach. Psychology Press. (1983)
26) Basilevsky A.: Statistical Factor Analysis and Related Methods, Theory and Application. A Wiley-Interscience Publication. (1993)
27) Bondarenko O., Fukuda T., and Miyachi K. The Outlook for Machine Learning Application in the Problem of Propulsion Engine Monitoring and Diagnostics. In Proc.: The 10th Intern. Conf. COMODIA, (2022)
28) Bondarenko O., Fukuda. T., and Miura S. Assessment of Propulsion Engine Reliability Degradation Through the Continuous Analysis of Service Data. In Proc.: The 1st Intern. Conf. G-NAOE, (2022)