

Article

# Anomalistic Symptom Judgment Algorithm for Predictive Maintenance of Ship Propulsion Engine Using Machine Learning

Jinkyu Park  and Jungmo Oh \* 

Division of Marine System Engineering, Mokpo National Maritime University, Mokpo 58628, Republic of Korea; pjk2019@mmu.ac.kr

\* Correspondence: jmoh@mmu.ac.kr; Tel.: +82-61-240-7207

**Abstract:** Ships serve as crucial transporters of cargo and passengers in substantial volumes and operate for a long time; therefore, an efficient maintenance system is essential for economical and stable vessel operation. In this study, a machine learning based approach was developed that considers the rapidly changing load fluctuations on ships and large variability in normal operation data to apply predictive maintenance to the propulsion engines of ships. After acquiring propulsion engine data from the alarm monitoring system of a ship, data and maintenance items were analyzed to select the data that could determine the anomalistic symptoms of the propulsion engine. Further, the main engine condition criterion value was defined as the factor for anomalistic symptom prediction. An engine anomalistic symptom judgment algorithm that can be practically used for ship maintenance prediction was developed and verified using machine learning.

**Keywords:** ship propulsion engine; predictive maintenance; alarm monitoring system; machine learning; anomalistic symptom



**Citation:** Park, J.; Oh, J. Anomalistic Symptom Judgment Algorithm for Predictive Maintenance of Ship Propulsion Engine Using Machine Learning. *Appl. Sci.* **2023**, *13*, 11818. <https://doi.org/10.3390/app132111818>

Academic Editor: Koji Murai

Received: 22 September 2023

Revised: 26 October 2023

Accepted: 26 October 2023

Published: 29 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ships are a means of transporting cargo and passengers in large quantities and generally operate for a long time, 20–30 years, after construction. And the management costs (e.g., operating and maintenance costs) required until the ship is abandoned are higher than the cost of construction [1–3]. Depending on the type of ship, the maintenance costs for parts such as engines and machinery account for 10–30% of the total management costs of the ship [1–4]. In addition, the actual financial burden related to maintenance is greater when considering improper maintenance and unforeseen breakdowns and disruptions to ship operations due to repairs [4]. Therefore, an efficient maintenance process is essential for the economical and reliable operation of ships [3–10].

Predictive maintenance (PdM) has been studied for propulsion engines with high maintenance costs in ships. PdM detects the performance and condition of machines and systems to plan for cases of machine failure and maintenance, making it possible to maintain the best operation status and improve economic efficiency along with operational efficiency [11–14]. PdM has been utilized in public transportation, such as aircraft [15,16].

As shown in Table 1, although previous studies detected frequent load fluctuations, which rely on the operational characteristics and environmental conditions of ship engines, they could not detect abnormal operating conditions due to actual engine abnormalities [6,7,9]. Therefore, there are limitations in using the results of existing studies for PdM. Accordingly, a methodology is required that can intuitively detect any anomalistic symptoms that cause failures and defects during normal operating conditions, rather than making predictions immediately before the occurrence of failures, by considering various and rapidly changing load fluctuations on ships and the variability in their normal opera-

tion data. Additionally, to implement PdM through real time data acquisition on a ship, a method is required to efficiently determine the engine status with minimal factors.

**Table 1.** Literature review (source: own compilation based on the [6,7,9]).

Study	Content	Limitation
Park et al. [6,9]	<ul style="list-style-type: none"> <li>— Computed the correlation coefficients of propulsion engine scavenging and scavenging receiver temperature, piston cooling oil outlet temperature, main bearing oil temperature, fuel viscosity, and temperature sensor data acquired from the ship and simplified the data based on the data with high correlation coefficients.</li> <li>— Thereafter, an algorithm for detecting defect data was established through regression analysis, and an algorithm for predicting propulsion engine failure was proposed and verified through data analysis pre and post failure.</li> </ul>	<ul style="list-style-type: none"> <li>— The detected defect is a serious defect in the engine operation owing to piston ring failure and oil leakage of the propulsion engine, but it is not a defect with a high frequency of occurrence.</li> <li>— The sensor data utilized for defect detection cannot be used to readily assess the causality posed by the defect.</li> </ul>
Bae et al. [7]	<ul style="list-style-type: none"> <li>— Proposed a statistical method to support maintenance decisions by monitoring the anomalous status of the engine by classifying parameters that can cause abnormalities using a bootstrap based T2 multivariate chart of data acquired from the propulsion engines of ships.</li> <li>— It classifies the collected engine data into five groups with similar characteristics through statistical analysis, prepares normal standards for each group, sets an upper limit to detect an abnormal state of the engine, and monitors the engine condition.</li> </ul>	<ul style="list-style-type: none"> <li>— Ship engines bear a wide range of load fluctuations under various operating characteristics, producing a wide range of normal operating conditions of related data. Therefore, the proposed method may classify the acquired data as abnormal data owing to rapid temporary load fluctuations. Thus, normal operation data can be assessed to be abnormal.</li> <li>— Whether the data are beyond the upper bound of management cannot be easily determined owing to the failures or abnormal conditions of the actual engine through the statistical analysis results of the collected data.</li> </ul>

In this study, propulsion engine data such as revolutions per minute (RPM), power, engine load, exhaust gas temperature, lubricant and coolant temperature, cylinder pressure, and scavenge air temperature were acquired from the alarm monitoring system (AMS) of a ship and preprocessed based on the characteristics of the propulsion engine control mode. Further, the AMS data and maintenance items related to the propulsion engine were analyzed to select data that can determine the anomalistic symptoms of a propulsion engine. Thereafter, the selected cylinder exhaust gas temperature (CET) and maximum cylinder combustion pressure (CCP) for the intuitive detection of anomalistic symptoms in the propulsion engine were employed to define the main engine condition criterion value (MCCV). To improve the efficiency of anomalistic symptom detection, a propulsion engine anomalistic symptom judgment algorithm was established using the revised MCCV (RMCCV) reflecting engine operating conditions and average data based on the voyage. Subsequently, an algorithm for detecting anomalistic symptoms during normal propulsion engine operation was developed and verified using machine learning, affording a practical engine anomalistic symptom detection algorithm applicable to ship PdM.

## 2. Materials and Methods

### 2.1. Experimental Equipment

In this study, the ship under investigation was a 10,000 ton class training ship equipped with various technologies, such as a dynamic positioning system that uses a global positioning system to maintain the target position of the ship. Additionally, it features a selective catalytic reactor system to respond to environmental regulations. The ship was equipped with a controllable pitch propeller (CPP) via the propulsion engine load and the blade angle of the propeller to control speed, unlike general ships. The ship had a total length of 133 m,

a mold width of 19.4 m, and a sailing speed of 17.7 knot. The detailed ship specifications are listed in Table 2.

**Table 2.** Specifications of the ship used in this study (source: own compilation based on the ship manual).

Description	Specification
Gross tonnage	9196 ton
Length overall (L.O.A)	133.0 m
Breadth (Mlb)	19.4 m
Speed (design draft, 85% MCR with 15% S.M)	17.7 knot
Range of endurance	14,500 n.miles
Number of people on board	239 persons
Controllable pitch propeller	WARTSILA D 4000 MM × 4 blades

Ships are equipped with an AMS, loading and unloading systems, and engine room systems. The AMS collects and stores sensor data from each system. Therefore, ship operation data were acquired through the AMS.

Two stroke diesel engines are usually used as the propulsion engines of medium sized and large ships because they are advantageous in terms of output and weight considering the size of the ship. The ship to be studied was a medium sized ship with a two stroke diesel engine used as the propulsion engine. The propulsion engine was a six cylinder engine with a cylinder diameter of 400 mm and a cylinder stroke of 1770 mm, and the rated output was 6618 kW. The detailed specifications of the propulsion engine are presented in Table 3.

**Table 3.** Specifications of the propulsion engine used in this study [17].

Description	Specification
Type	MAN B&W 6S40ME-B9.5
Rated output	6618 kW
Diameter of the cylinder	400 mm
Stroke	1770 mm
Number of cylinders	6
Mean effective pressure	20 bar
Maximum cylinder pressure	185 bar
Turbo charger	HYUNDAI-ABB, 1 × A165L37

The propulsion engine of a ship is operated by an engine telegraph and controlled according to the engine control mode, as shown in Table 4, based on the RPM of the engine to control the ship speed. The control modes are divided into two major modes: the maneuvering mode of dead slow, slow, half, and full, four steps that can quickly control engine loads, such as arrival and departure, and a constant speed navigation mode (cruising mode) of navigation full. As described above, the ship was equipped with a CPP; thus, the pitch of the propeller was simultaneously controlled according to the engine mode.

The AMS installed in the engine room of the ship monitors factors such as temperature and pressure that can determine the operation status of mechanical systems, such as engines and pumps. In addition, the AMS is built to control the engine and mechanical systems. Engine monitoring and alarm items (lists) are composed of factors that directly and indirectly affect engine combustion and factors that affect mechanical defects.

**Table 4.** Engine telegraph control stage (source: own compilation based on the ship manual).

Steps		Stop	Dead Slow	Slow	Half	Full	Navigation Full
Speed (Knot)	Ahead	-	5.5	8.0	10.5	13.5	17.7
	Astern	-	3.6	5.3	7.3	8.7	9.8
RPM	Ahead	0	73	88	97	116	141
	Astern	0	110	121	127	135	141
Pitch (%)	Ahead	-	50	65	75	83	97
	Astern	-	-40	-50	-55	-62	-65

In total, 104 types of data were collected by checking numerical data, such as temperature and pressure, that can be analyzed; these data identify the operating condition of the propulsion engine in the AMS in the engine room. Considering changes in the marine environment and the influence of weather, data from eight voyages (21071, 21081, 21091, 21101, 21102, 21112, 21121, and 21122) acquired from the hot season (July) to the cold season (November) were utilized for this research. In addition, anomalistic symptom data (21123) were established based on the results of a previous study [18] on the implementation of abnormal symptoms using engine simulation and the analysis of abnormal operation items and maintenance determination factors of the engine to verify the anomalistic symptom judgment algorithm.

## 2.2. Experimental Methods

Machine learning is an artificial intelligence based system that learns from experience and improves performance through prediction and technology that develops and builds algorithms for it [19–21]. Therefore, machine learning was used to develop an anomalistic symptom prediction algorithm for propulsion engines.

The data must be free from abnormal data to ensure the reliability of results during data analysis and machine learning. The data classification that reflects the characteristics of the target to be analyzed, such as the engine, must be clear. In addition, the reliability of data processing is essential because machine learning results are derived in a different direction from the target if factors that are not related to the abnormal data or the analysis target are included. Therefore, the data collected by the data processing algorithm, considering the control characteristics of a propulsion engine constructed in a previous study [3], were preprocessed and then classified based on the engine telegraph control mode and used in the study [3]. The data preprocessing process is shown in Figure 1.

Among machine learning methods, the regression technique was used to derive the main factors and numerical data related to the improvement of the propulsion engine anomaly detection algorithm. The estimated regression model, through the regression algorithm, can determine the predicted value of the response variable, the linear relation between the response and explanatory variables, and the significance and influence of the explanatory variable correlated with the response variable [22,23]. Therefore, it was used to develop an anomalistic symptom judgment algorithm to examine the relationship among the variables of the propulsion engine.

In this study, a linear regression model was used among the regression techniques. The linear regression model is the most representative analysis method among regression algorithms, and it optimizes the straight regression line that minimizes the difference between the predicted and actual values. Using the collected data shown in Figure 2, the relation between the explanatory variable (feature) and the response variable (value) was estimated linearly, and the setting of the appropriate weight ( $W$ ) and bias ( $b$ ) was the key to linear regression [22,23].

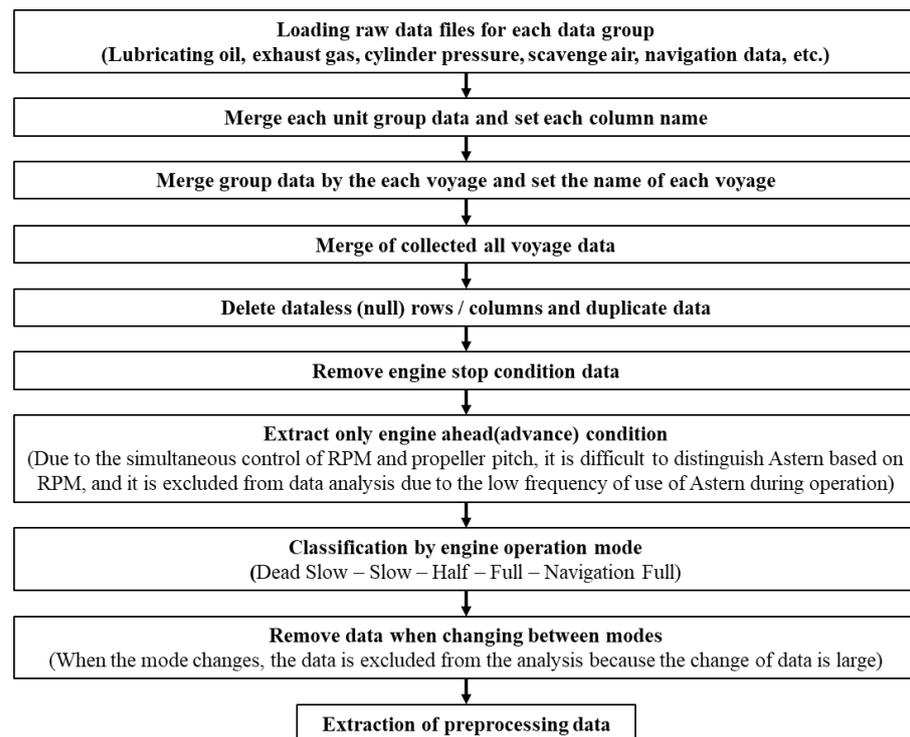


Figure 1. Data preprocessing algorithm for the propulsion engine [3].

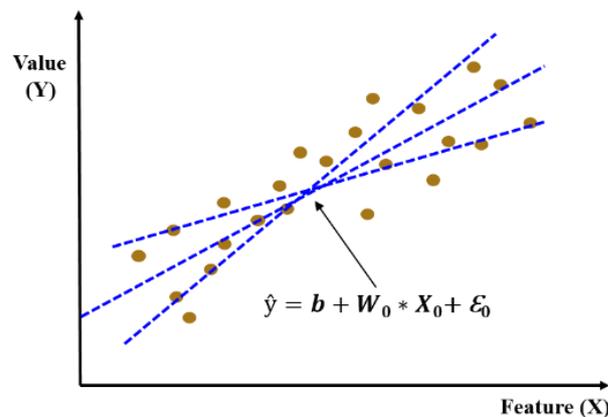


Figure 2. Simple linear regression model (source: own elaboration).

The mean absolute error (MAE),  $R^2$ , and  $R^2_{adj}$  of the loss functions were used, considering the various data scale ranges and normal operating ranges of the propulsion engine data, to evaluate the efficiency of the machine learning algorithm in implementing the collected data.

As shown in Equation (1), the MAE averages the errors between the collected and predicted values and expresses them as absolute values. It is less affected by outliers than other loss functions and shows the average error. The MAE depends on the data scale and is used as a regression performance evaluation index [24–29].

$R^2$  is a variance based evaluation index in the range of 0–1, as described in Equation (2). As it gets closer to 1, the accuracy improves and the influence of the data scale diminishes, making it possible to determine the relative performance. As the number of explanatory variables increases,  $R^2$  approaches 1 and is affected by the number of data points and response variables [24,30].

$R^2_{adj}$  is the value obtained by correcting  $R^2$  with the number of samples and the number of independent variables, as shown in Equation (3). It is smaller than or equal to

$R^2$ , and the prediction accuracy improves as it approaches 1. Moreover, if  $R^2_{adj}$  is negative, the regression algorithm cannot be used. When  $R^2_{adj}$  is considerably smaller than  $R^2$ , an unnecessary explanatory variable exists in the algorithm [24,30].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (1)$$

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

$$R^2_{adj} = 1 - \frac{(n-1)}{(n-p)(1-R^2)}, \quad (3)$$

where  $n$  is the sampling number,  $y_i$  is the collected value,  $\hat{y}_i$  is the predicted value,  $\bar{y}_i$  is the collected value average, and  $p$  is the explanatory variable number.

The machine learning algorithm was developed using Python™ 3. Python is designed to easily use various libraries necessary for data loading and visualization, image processing, and statistics. Few errors occur due to complex code configurations with a simple grammar configuration, and library linkage with other programming languages is easy; thus, the efficiency of the developed algorithm is high [19,31,32].

The utilization of the entire dataset collected by the AMS for detecting anomalistic symptoms and determining maintenance requires considerable time and effort, from algorithm processing to the drawing of results. Moreover, unnecessary data are included in the data used to determine anomalistic symptoms. Therefore, the factors used to determine anomalistic symptoms and maintenance must be identified by considering propulsion engine characteristics.

Abnormal operation issues that may occur during engine operation were analyzed by referring to propulsion engine related alarms and monitoring factors in the AMS, the propulsion engine related maintenance items of the planned management system, and the engine manual. Therefore, the item commonly considered to be the cause of abnormal operation was determined to be an exhaust system problem. An increase in the exhaust gas temperature is a major issue in abnormal engine operations, and the CET was selected for determining anomalistic symptoms of propulsion engines as the index that could directly determine the combustion state of the engine. In addition, the CCP was used to directly determine the combustion state in medium sized or large ships. Therefore, the CCP was also selected as a factor for determining the anomalistic symptoms of the propulsion engine.

The CET and CCP were derived as the criteria for determining anomalistic symptoms. The maintenance of the propulsion engine tends to increase CET and decrease CCP during abnormal combustion due to problems related to fuel systems and exhaust valves. Therefore, a reference factor capable of easily confirming the tendency of variation in the two factors is required [17,33,34]. The increase in reference factors for determining anomalistic symptoms makes the machine learning algorithm complicated, reducing the effectiveness of predicting anomalistic symptoms for ships in real time. Therefore, one factor that can intuitively grasp the trends of the two factors was defined to reflect the engine's status, and an anomalistic symptom judgment algorithm was constructed.

The MCCV that can determine the fluctuation trends of the two factors by analyzing the collected CET and maximum CCP data is described in Equation (4).

$$MCCV = CET - CCP, \quad (4)$$

The MCCV is the difference between the CET and the maximum CCP. The MCCV increases because of the increase in the CET or decrease in the maximum CCP or their simultaneous occurrence owing to an abnormal exhaust or fuel valve. After setting the

MCCV range in the normal operation state, it is possible to determine the anomalistic symptoms of the engine by understanding the trend of the MCCV.

The data on repetitive and continuous driving conditions are required for setting the MCCV for normal driving conditions to determine the anomalistic symptoms of the propulsion engine. Therefore, a normal operation status database was constructed by setting the conditions shown in Table 5 through engine data analysis.

**Table 5.** Criteria for constructing the normal operation state database for the propulsion engine (source: own elaboration).

Classification Criterion	Selection Reason
Navigation full mode  (ME ORDER RPM > 135 RPM)	The section where the engine is controlled at a constant target RPM, and the load fluctuation is small. Avoids the 135 RPM section, which is the set standard for full astern in the maneuvering mode.
CPP PITCH FEEDBACK ≥ 93	The section where constant speed navigation is performed at the target RPM, not the section where RPM increases to reach the target speed during the navigation full mode.  CPP pitch is >93% when ORDER RPM > 135 RPM in the navigation full mode.

As shown in Table 6, the criteria for determining engine anomalistic symptoms with the MCCV were set to concern (1) and abnormal (2) levels. The MCCV range of the normal operation state for each cylinder was set to the maximum value of the MCCV. The normal range of the MCCV was set to below the maximum value of the MCCV without a minimum standard because no minimum limit existed for the alarm and operating standard of the exhaust gas temperature in the AMS and engine manual.

**Table 6.** Criteria for MCCV abnormality judgment for the propulsion engine (source: own elaboration).

Level	Criterion
Concern (1)	Maximum MCCV value during normal operation < (1) < maximum MCCV value + 10
Abnormal (2)	Maximum MCCV value + 10 < (2)

The concern level involves the section from the time when the engine is out of the normal operation state to the time when it is determined to be in an abnormal operation state. It is a step to monitor whether abnormal operations are caused by temporary factors, such as emergency operations or marine environments, or actual exhaust valve problems. As shown in Table 6, the concern level was set for each cylinder from the maximum MCCV value in normal operation to the maximum MCCV value plus 10.

The abnormal level is a stage in which the cause of abnormal operation is diagnosed, and maintenance is deemed necessary by judging that the engine is operating abnormally beyond the concern level and showing anomalistic symptoms. As shown in Table 6, the interval exceeding the concern level was set as the abnormal level.

The algorithm for determining the anomalistic symptoms of the propulsion engine was established and verified based on the MCCV constructed as described above and the database of normal operating conditions.

The operating conditions of the propulsion engine rapidly change in real time based on the environment, such as sea routes and conditions. Accordingly, the MCCV changes rapidly. Therefore, MCCV correction for setting the normal range of a constant MCCV and data simplification for noise minimization were performed to determine anomalistic

symptoms. An improvement algorithm was constructed by deriving a revised MCCV for improving the performance of the anomalistic symptom determination algorithm.

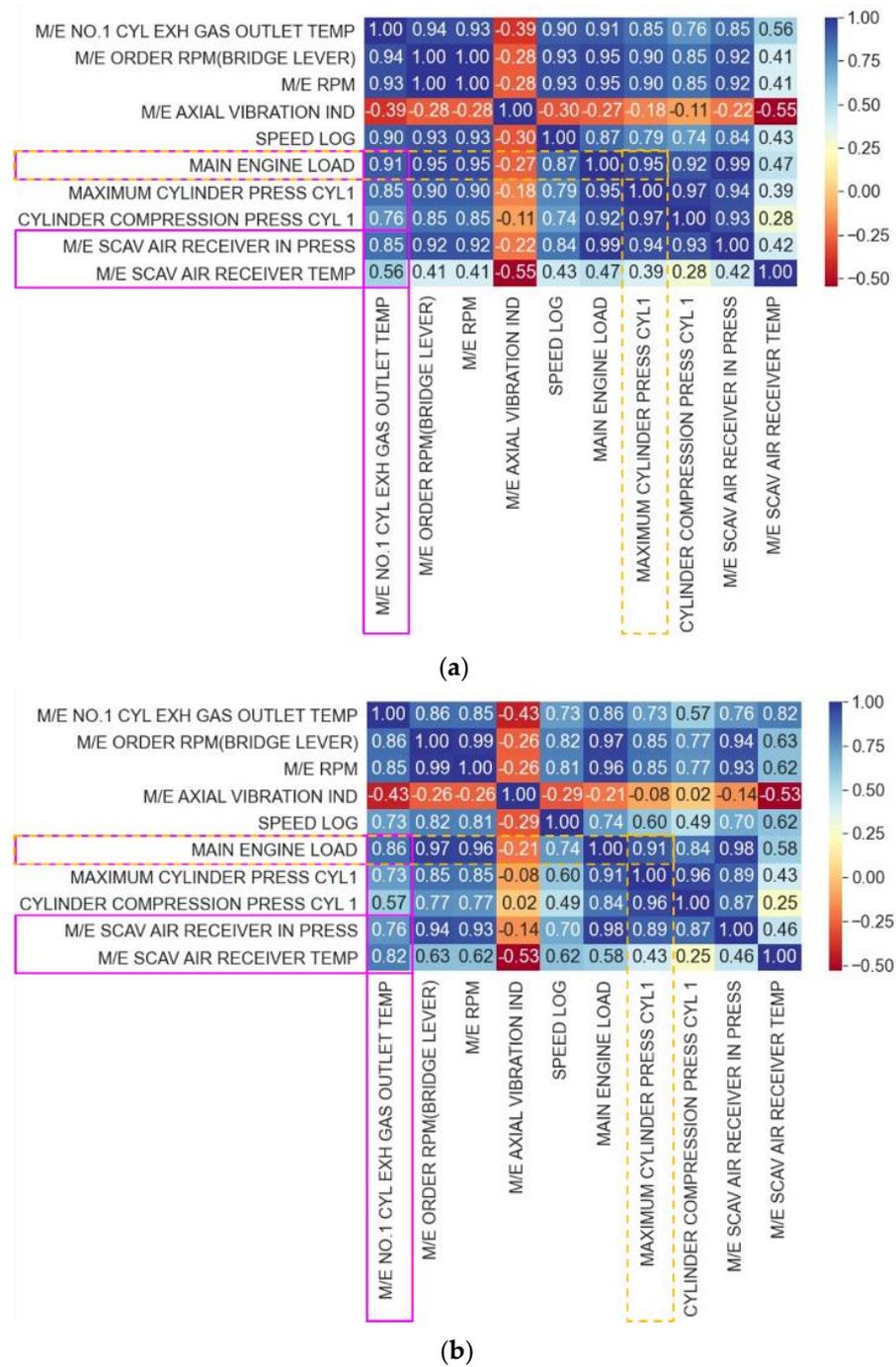
A correlation analysis of the acquired data was conducted to select the main factors necessary for MCCV correction. The Pearson correlation coefficient used in the correlation analysis is a quantification of the linear correlation between variables X and Y. It is obtained covariance of two variables divided by the product of each standard deviation, and it is a correlation coefficient commonly used in various categories [35,36]. Figure 3a shows the result of a repeated correlation analysis of the data acquired in all sections of the engine control telegraph based on cylinder 1, excluding factors that are not related to or have little correlation with engine combustion. The factors necessary for the MCCV correction were first selected by analyzing the correlation between the CET and the maximum CCP, which are the calculation factors for the MCCV, and other data. Then, a secondary correlation analysis was performed using only the data of the navigation full mode section, which is the criterion for determining anomalistic symptoms; the results are shown in Figure 3b. In the results of the correlation analysis in all operation modes, as a factor that drastically affects CET in the case of propulsion engine scavenging air temperature and pressure, the correlation coefficient of the scavenging air pressure was high at 0.85, and the correlation coefficient of the scavenging air temperature was low at 0.56. However, in the navigation full mode, the scavenging air temperature showed a high correlation of 0.82, and the scavenging air pressure was 0.76, which is lower than the result of the correlation analysis under all operation modes.

The scavenging air temperature (ME SCAV. AIR RECEIVER TEMP.) of the ship engine is configured to maintain a constant temperature through the air cooler. In general, an air cooler is used to reduce the high scavenging air temperature because the air supplied through the supercharger increases as the engine load increases. Therefore, because the temperature change in the engine is not large, the correlation of the scavenging air temperature is low, and the correlation of the scavenging air pressure is high in the analysis results of the operation modes.

The data in this study were acquired from a summer voyage in July to a winter voyage in December. In addition, navigation full is the mode in which engine load fluctuations are small, and constant speed navigation is performed at a high load above a certain level. There was no device that could increase the scavenging air temperature beyond a certain temperature, although the outside air temperature continued to decrease from July to December due to seasonal factors. Therefore, it is determined that the correlation between the CET and scavenging air temperature increases when the temperature of the scavenging air supplied to the engine decreases as the voyage progresses in the navigation full mode.

The engine load (ME LOAD), scavenging air pressure (ME SCAV. AIR RECEIVER TEMP.), and scavenging air temperature (ME SCAV. AIR RECEIVER IN PRESS.) were selected as factors for improving the MCCV correction anomalistic symptom judgment algorithm considering the correlation analysis results of all modes and the navigation full mode and their influence on engine combustion. Each correction factor was determined through the regression analysis of CET and the maximum CCP as the MCCV component factors and the correction factor selected as above, and the regression coefficient of each correction factor was derived for each cylinder and used for MCCV correction.

Data simplification and MCCV correction criteria of steady state databases were set through data row frequency analysis under the same conditions based on ME ORDER RPM and VOYAGE considering propulsion engine control and operation characteristics and used to improve the MCCV algorithm.



**Figure 3.** Correlation analysis of cylinder 1. (a) All modes and (b) navigation full mode (source: own elaboration, Python 3.8.3).

The correction equation for calculating RMCCV, which is the correction value of the MCCV, is described in Equation (5) using the regression coefficient and correction standard coefficient of each correction factor derived. An improvement algorithm was built, and the results were analyzed.

$$RMCCV = \left\{ + \left[ \begin{array}{l} CET \\ (ELR * (LC - CL)) \\ +(ESTR * (STC - CST)) \\ +(ESPR * (SPC - CSP)) \end{array} \right] \right\} - \{CCP + [CPLR * (LC - CL)]\} \quad (5)$$

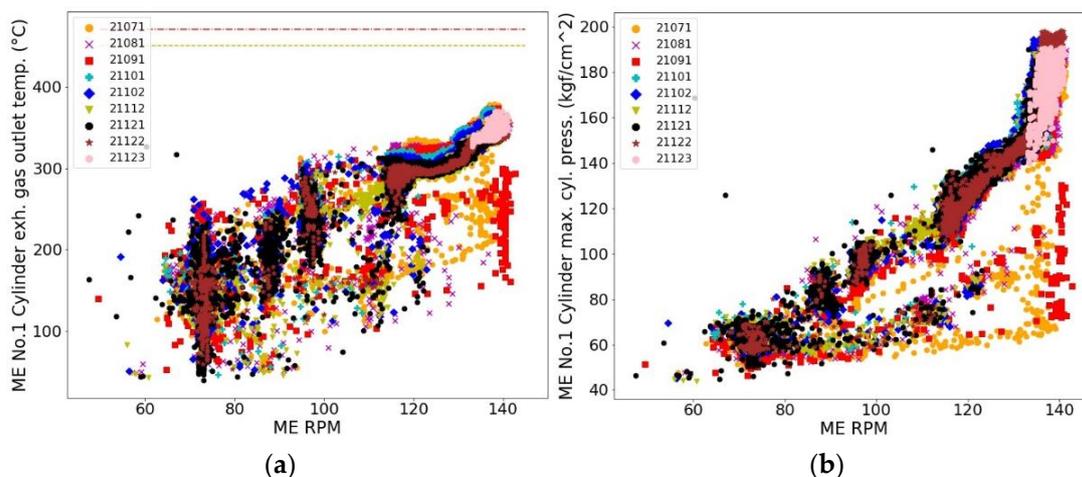
where LC is the load criterion, STC is the scavenge temperature criterion, SPC is the scavenge pressure criterion, CPLR is the combustion pressure/load regression coefficient, ELR is the exhaust temperature/load regression coefficient, ESTR is the exhaust temperature/scavenge temperature regression coefficient, ESPR is the exhaust temperature/scavenge pressure regression coefficient, CL is the collected load, CST is the collected scavenge temperature, and CSP is the collected scavenge pressure.

The RMCCV is calculated as the difference between the CET correction value and the maximum CCP correction value. The CET correction value multiplies the difference between the correction reference coefficient (LC, STC, and SPC) of each of the three selected correction factors (engine load, scavenging air pressure, and scavenging air temperature) and the collected correction factors (CL, CST, and CSP) by the calculated correction factor regression coefficient (ELR, ESTR, and ESPR). The three derived correction values are calculated by combining them with the CET. The maximum CCP correction value is calculated by multiplying the difference between the LC and the CL by the CPLR and then combining it with the obtained CCP.

The propulsion engine anomalistic symptom judgment algorithm built based on the MCCV and the RMCCV was verified using the entire dataset and average data, and the validity of the algorithm was confirmed by comparison.

### 3. Results

Figure 4 shows the exhaust gas temperature and maximum combustion pressure in all voyages, including the anomaly data of cylinder 1. It increases with increasing RPM of the engine, and the exhaust gas temperature and combustion pressure in the same RPM section change considerably depending on the operating environment, such as the engine load.



**Figure 4.** Conditions of cylinder 1. (a) EXH: exhaust gas temperature; (b) MAX: cylinder pressure (source: own elaboration, Python 3.8.3).

In addition, the high temperature alarm standard of the propulsion engine cylinder exhaust gas in the AMS is 450 °C (yellow dotted line in Figure 4a, the standard temperature for propulsion engine SHUT DOWN or SLOW DOWN is 470 °C (red dashed line in Figure 4a)). Therefore, if the temperature of the cylinder exhaust gas is less than 450 °C, it is determined as a normal operation state. As shown in Table 7, the average exhaust gas temperature of cylinder 1 during the voyages was 323.9 °C. The temperature range is from a minimum of 40.1 °C to a maximum of 378.6 °C. There is a big difference from the high temperature alarm standard, and the normal operating condition range is wide. Therefore, the data of VOYAGE (21123) with anomalistic symptoms are also included in the normal operating range, and criteria and algorithms for determining abnormal signs are required. Thus, a propulsion engine anomalistic symptom judgment algorithm was constructed using the MCCV via Equation (4).

**Table 7.** Condition data of cylinder 1 (source: own elaboration).

	CYL1 EXH. Gas Temp. (°C)	CYL1 MAX. PRESS. (kgf/cm <sup>2</sup> )
Mean	323.9	153.6
Std.	50.3	31.2
Min.	40.1	43.8
Max.	378.6	198.4

Tables 8 and 9 present the results of the CET and maximum CCP from the normal operating state database of the propulsion engine. When the propulsion engine is operated at a constant speed of 133 RPM or higher, the minimum/maximum difference in the CET is ~60 °C, depending on the operating environment of the engine, and the average exhaust gas temperature of each cylinder is different. The maximum CCP varies from cylinder to cylinder, with a minimum/maximum difference of ~60 kgf/cm<sup>2</sup>; the average pressure differs from cylinder to cylinder. Therefore, it is necessary to set the standard operating range of the MCCV for each cylinder.

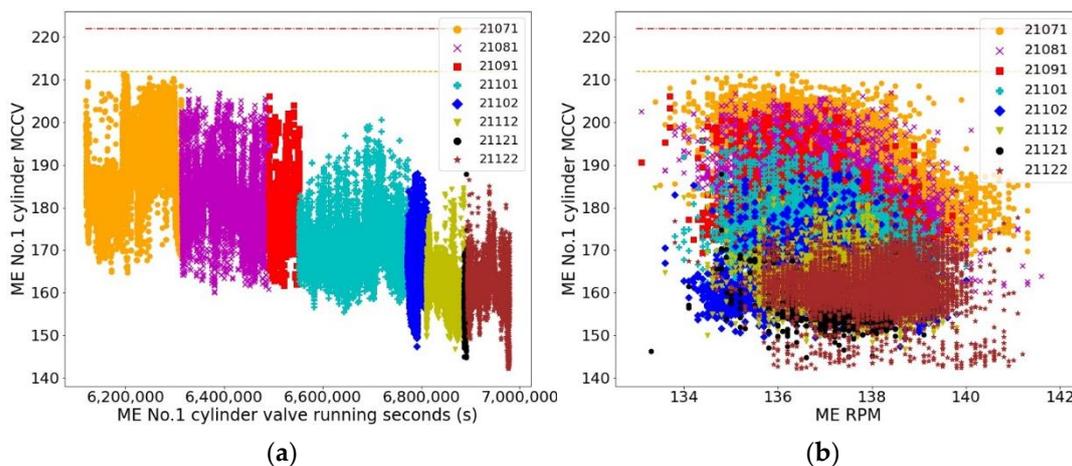
**Table 8.** Cylinder exhaust gas temperature data (source: own elaboration).

Cylinder	1	2	3	4	5	6
Mean (°C)	347.6	359.3	356.6	361.1	362.4	370.0
Std. (°C)	5.7	6.8	7.9	7.1	7.9	6.1
Min. (°C)	328.7	332.1	329.0	335.9	335.5	346.3
Max. (°C)	374.4	391.7	385.6	393.1	397.9	401.9

**Table 9.** Cylinder combustion pressure data (source: own elaboration).

Cylinder	1	2	3	4	5	6
Mean (kgf/cm <sup>2</sup> )	172.6	173.4	174.5	174.1	174.8	175.0
Std. (kgf/cm <sup>2</sup> )	8.8	8.7	8.5	8.7	9.0	8.7
Min. (kgf/cm <sup>2</sup> )	142.1	143.0	144.9	144.3	143.8	144.8
Max. (kgf/cm <sup>2</sup> )	198.4	198.4	198.1	200.4	200.2	201.2

Figure 5 shows the MCCV data of cylinder 1 in a normal operating state for each voyage based on the running time and RPM. The MCCV tended to decrease owing to changes in the navigating environment, such as changes in ME ORDER RPM and the lowering of the scavenge air temperature with the progression of the voyage. The yellow dotted line in Figure 5 is the Concern (1) level, and the red dashed line is the Abnormal (2) level.



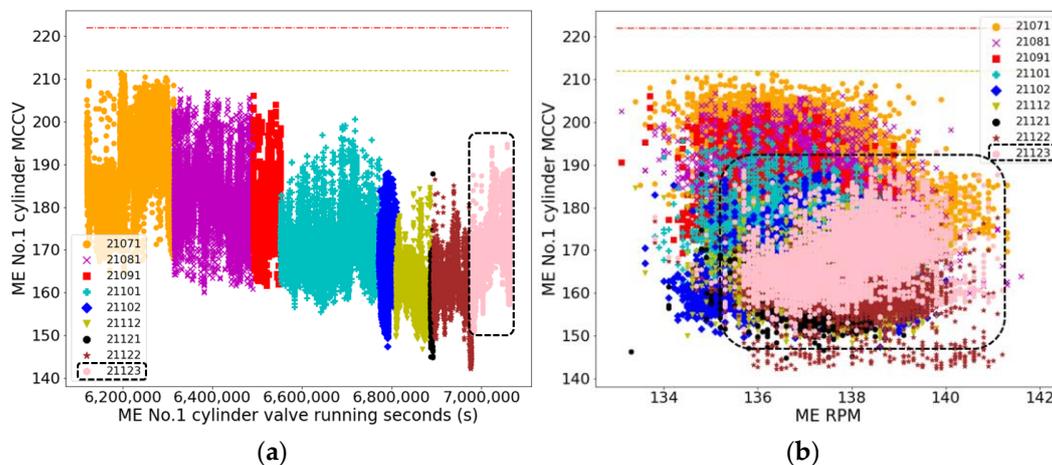
**Figure 5.** MCCV of the normal operating condition of cylinder 1 based on (a) running time and (b) power (source: own elaboration, Python 3.8.3).

Table 10 lists the MCCV data of cylinder 1 for all normal driving conditions. The average MCCV is 175.0, the minimum MCCV is 142.2, and the maximum MCCV is 211.5; thus, the maximum/minimum difference in the MCCV is 69.3, and the standard deviation is 11.6. The deviation resulted from the characteristics of engine control, where the load fluctuates in real time.

**Table 10.** MCCV data under the normal operating conditions of cylinder 1 (source: own elaboration).

Mean	Std.	Min.	25%	50%	75%	Max
175.0	11.6	142.2	166.2	174.7	182.5	211.5

Figure 6 shows the results of applying the entire dataset containing voyage 21123 data with anomalies to the MCCV based anomalistic symptom determination algorithm. Because the maximum MCCV for cylinder 1 was 211.5, the concern level was set to 212–222, and the abnormal level was set to more than 222. Figure 6a shows the results based on the running time of the propulsion engine. The MCCV of voyage 21123 with abnormalities tends to increase with the operating time. However, the abnormalities could not be determined from the results of this algorithm, as they tend to appear in the range below the concern level of the MCCV. Figure 6b shows the results based on RPM; the data of voyage 21123 with anomalies were located at the bottom of the steady state data group. In addition, determining anomalistic symptoms was impossible because of the difference between the concern level criteria of the MCCV. Anomalistic symptoms cannot be detected based on the MCCV set owing to changes in the normal operating range of the MCCV caused by major factors, such as weather conditions, the marine environment, and constant speed navigation target RPM, which differ for each voyage.



**Figure 6.** MCCV including anomaly symptom data of cylinder 1 based on (a) running duration and (b) power (source: own elaboration, Python 3.8.3).

The following presents the results of algorithm verification conducted using RMCCV, a correction value for MCCV that considers changes in the engine operating environment and conditions to improve the abnormality detection performance of the MCCV based algorithm. The data index frequency count analysis revealed the requirement for the RMCCV based algorithm under the same operating conditions, and the regression analysis presented the results for deriving the MCCV correction factor and its coefficient. The correction standard coefficients were set accordingly.

To minimize noise data in the steady state database, simplify the data, and set revision standards, temporary driving data before reaching the target RPM were excluded from the navigation full mode steady state driving data. Additionally, only the data with a data index (row) frequency count of  $\geq 1000$  EA based on the same ME ORDER RPM

and VOYAGE were extracted to ensure reliable machine learning results and anomaly detection for continuous and repetitive driving conditions. Table 10 shows the division into 12 ME ORDER RPMs, with partial overlapping existing for ME ORDER RPMs for each voyage, confirming their operation under the changing conditions of each voyage. Marine propulsion engines experience frequent load fluctuations depending on sea conditions, even at the same ME ORDER RPM, resulting in large fluctuations in the engine load, CET, CCP, and other data. However, determining anomalous symptoms in real time using all voyage data has limitations, necessitating improved algorithm reliability through reduced algorithm processing time and minimized noise data. Therefore, the average values of data with a frequency count of >1000 EA based on ME ORDER RPM and VOYAGE were derived (Table 11) and set as the steady state reference data. Furthermore, data analysis and correction were performed to set the normal range of consistent MCCVs according to factor changes.

**Table 11.** Average data of cylinder 1 for normal operating conditions (source: own elaboration).

ME Order RPM	Voyage	RPM	Load (%)	Scav. Air Temp. (°C)	Scav. Air Press. (kgf/cm <sup>2</sup> )	Exh. Gas Temp. (°C)	Cyl. Max. Press. (kgf/cm <sup>2</sup> )	Cyl. Comp. Press. (kgf/cm <sup>2</sup> )
136.1	21081	136.0	72.7	44.5	2.1	350.2	163.2	119.6
136.8	21091	136.7	74.1	43.3	2.2	347.1	168.8	126.1
136.9	21071	136.8	71.8	44.4	2.1	354.0	159.2	115.9
136.9	21101	136.8	76.2	42.5	2.4	345.5	176.4	138.2
137	21071	136.9	72.8	44.3	2.1	352.4	163.3	120.0
137	21101	136.9	76.3	42.7	2.4	346.9	176.7	138.7
137	21102	136.9	77.1	40.6	2.4	345.2	182.1	144.0
137.1	21081	137.0	74.0	45.5	2.2	349.2	168.1	124.9
137.1	21101	137.0	77.3	43.4	2.4	350.2	180.0	145.1
137.1	21122	137.0	76.9	34.6	2.2	339.5	179.0	143.3
137.3	21081	137.2	73.9	43.9	2.2	348.2	167.9	124.7
137.4	21081	137.3	73.8	43.8	2.2	348.0	167.3	124.2
137.5	21081	137.4	74.9	44.1	2.2	350.2	171.2	128.1
137.5	21101	137.4	78.3	43.0	2.5	350.7	181.3	147.0
137.5	21112	137.4	76.7	34.7	2.4	338.6	177.6	141.6
137.6	21071	137.5	75.2	45.3	2.2	353.4	171.9	128.3
137.8	21071	137.7	75.5	45.0	2.2	353.3	173.1	129.7
138.6	21122	138.5	79.0	35.7	2.5	343.9	183.6	148.5
138.8	21071	138.7	75.9	44.6	2.2	355.6	175.3	131.7
138.8	21122	138.7	79.3	36.4	2.5	346.2	183.7	149.4

The obtained results of the regression analysis of the CET and maximum CCP, which are the components of the MCCV and the selected MCCV correction factor, are as follows.

Figure 7a shows the results of the regression analysis of the exhaust gas temperature and MCCV correction factor of cylinder 1. The engine load (ME LOAD), scavenging air temperature (ME SCAV. AIR RECEIVER TEMP.), and scavenging air pressure (ME SCAV. AIR RECEIVER IN PRESS.) were determined as correction factors considering the explanatory power with CET and the influence of engine combustion. The  $R^2_{adj}$  of CET and three correction factors was 0.757, showing significant explanatory power. In addition, the  $p$ -value ( $p > |t|$ ), which is used to evaluate the effect of each correction factor on the CET, is <0.05 for all three correction factors. In addition, MAE, which is a predictive index of CET according to the correction factor, was 4.685, and the selected correction factor was found to be significant.

<b>Dep. Variable:</b> MEAN_ME_CYL1_EXH_TEMP <b>Model:</b> OLS <b>Method:</b> Least Squares <b>Date:</b> Tue, 25 Jul 2023 <b>Time:</b> 12:33:03 <b>No. Observations:</b> 20 <b>Df Residuals:</b> 16 <b>Df Model:</b> 3 <b>Covariance Type:</b> nonrobust	<b>R-squared:</b> 0.795 <b>Adj. R-squared:</b> 0.757 <b>F-statistic:</b> 20.72 <b>Prob (F-statistic):</b> 9.32e-06 <b>Log-Likelihood:</b> -42.113 <b>AIC:</b> 92.23 <b>BIC:</b> 96.21	<b>Dep. Variable:</b> MEAN_ME_CYL1_MAX_PRESS <b>Model:</b> OLS <b>Method:</b> Least Squares <b>Date:</b> Tue, 25 Jul 2023 <b>Time:</b> 12:33:02 <b>No. Observations:</b> 20 <b>Df Residuals:</b> 18 <b>Df Model:</b> 1 <b>Covariance Type:</b> nonrobust	<b>R-squared:</b> 0.970 <b>Adj. R-squared:</b> 0.968 <b>F-statistic:</b> 579.7 <b>Prob (F-statistic):</b> 3.83e-15 <b>Log-Likelihood:</b> -32.449 <b>AIC:</b> 68.90 <b>BIC:</b> 70.89																																																						
<table border="1"> <thead> <tr> <th></th> <th>coef</th> <th>std err</th> <th>t</th> <th>P&gt; t </th> <th>[0.025</th> <th>0.975]</th> </tr> </thead> <tbody> <tr> <td>const</td> <td>236.7577</td> <td>33.721</td> <td>7.021</td> <td>0.000</td> <td>165.271</td> <td>308.244</td> </tr> <tr> <td>MEAN_ME_LOAD</td> <td>1.8185</td> <td>0.762</td> <td>2.387</td> <td>0.030</td> <td>0.203</td> <td>3.433</td> </tr> <tr> <td>MEAN_ME_SCAV_TEMP</td> <td>0.9095</td> <td>0.228</td> <td>3.991</td> <td>0.001</td> <td>0.426</td> <td>1.393</td> </tr> <tr> <td>MEAN_ME_SCAV_PRESS</td> <td>-27.9974</td> <td>13.413</td> <td>-2.087</td> <td>0.053</td> <td>-56.433</td> <td>0.438</td> </tr> </tbody> </table>		coef	std err	t	P> t	[0.025	0.975]	const	236.7577	33.721	7.021	0.000	165.271	308.244	MEAN_ME_LOAD	1.8185	0.762	2.387	0.030	0.203	3.433	MEAN_ME_SCAV_TEMP	0.9095	0.228	3.991	0.001	0.426	1.393	MEAN_ME_SCAV_PRESS	-27.9974	13.413	-2.087	0.053	-56.433	0.438	<table border="1"> <thead> <tr> <th></th> <th>coef</th> <th>std err</th> <th>t</th> <th>P&gt; t </th> <th>[0.025</th> <th>0.975]</th> </tr> </thead> <tbody> <tr> <td>Intercept</td> <td>-82.7415</td> <td>10.646</td> <td>-7.772</td> <td>0.000</td> <td>-105.108</td> <td>-60.375</td> </tr> <tr> <td>MEAN_ME_LOAD</td> <td>3.3900</td> <td>0.141</td> <td>24.077</td> <td>0.000</td> <td>3.094</td> <td>3.686</td> </tr> </tbody> </table>		coef	std err	t	P> t	[0.025	0.975]	Intercept	-82.7415	10.646	-7.772	0.000	-105.108	-60.375	MEAN_ME_LOAD	3.3900	0.141	24.077	0.000	3.094	3.686
	coef	std err	t	P> t	[0.025	0.975]																																																			
const	236.7577	33.721	7.021	0.000	165.271	308.244																																																			
MEAN_ME_LOAD	1.8185	0.762	2.387	0.030	0.203	3.433																																																			
MEAN_ME_SCAV_TEMP	0.9095	0.228	3.991	0.001	0.426	1.393																																																			
MEAN_ME_SCAV_PRESS	-27.9974	13.413	-2.087	0.053	-56.433	0.438																																																			
	coef	std err	t	P> t	[0.025	0.975]																																																			
Intercept	-82.7415	10.646	-7.772	0.000	-105.108	-60.375																																																			
MEAN_ME_LOAD	3.3900	0.141	24.077	0.000	3.094	3.686																																																			
<b>Omnibus:</b> 2.429 <b>Durbin-Watson:</b> 2.574 <b>Prob(Omnibus):</b> 0.297 <b>Jarque-Bera (JB):</b> 1.667 <b>Skew:</b> 0.702 <b>Prob(JB):</b> 0.435 <b>Kurtosis:</b> 2.829 <b>Cond. No.</b> 5.99e+03	<b>Omnibus:</b> 7.213 <b>Durbin-Watson:</b> 2.010 <b>Prob(Omnibus):</b> 0.027 <b>Jarque-Bera (JB):</b> 5.294 <b>Skew:</b> 0.706 <b>Prob(JB):</b> 0.0709 <b>Kurtosis:</b> 5.088 <b>Cond. No.</b> 2.79e+03																																																								

(a)

(b)

**Figure 7.** Results of the regression analysis of correction factor of cylinder 1 based on (a) exhaust gas temperature and (b) maximum combustion pressure (source: own elaboration, Python 3.8.3).

Figure 7b shows the results of the regression analysis of the maximum combustion pressure of cylinder No. 1 and the MCCV correction factor. Based on the analysis of the relationship with engine combustion and the explanatory power with the maximum CCP,  $R^2_{adj}$  showed a very high explanatory power of 0.968 only with the engine load (ME LOAD). In addition, the  $p$ -value ( $p > |t|$ ) was analyzed as a significant factor at 0.00, and the engine load was selected as a correction factor for the maximum CCP. Also, the MAE of the CCP with the engine load was 1.375, showing excellent predictive performance.

The correction standard for calculating a constant MCCV in response to changes in factors such as engine load had the highest data frequency count of 13,470 EA among the classified data presented in Table 11, and considering engine operability, the 137RPM data of the 21101 voyage were selected as MCCV correction criteria. The MCCV correction criterion data are shown in Table 12.

**Table 12.** Correction criteria for the MCCV of cylinder 1 (source: own elaboration).

ME Order RPM	Voyage	Load (%)	Scav. Air Temp. (°C)	Scav. Air Press. (kgf/cm <sup>2</sup> )	Exh. Gas Temp. (°C)	Cyl. Max. Press. (kgf/cm <sup>2</sup> )
137	21101	76.3	42.7	2.4	346.9	176.7

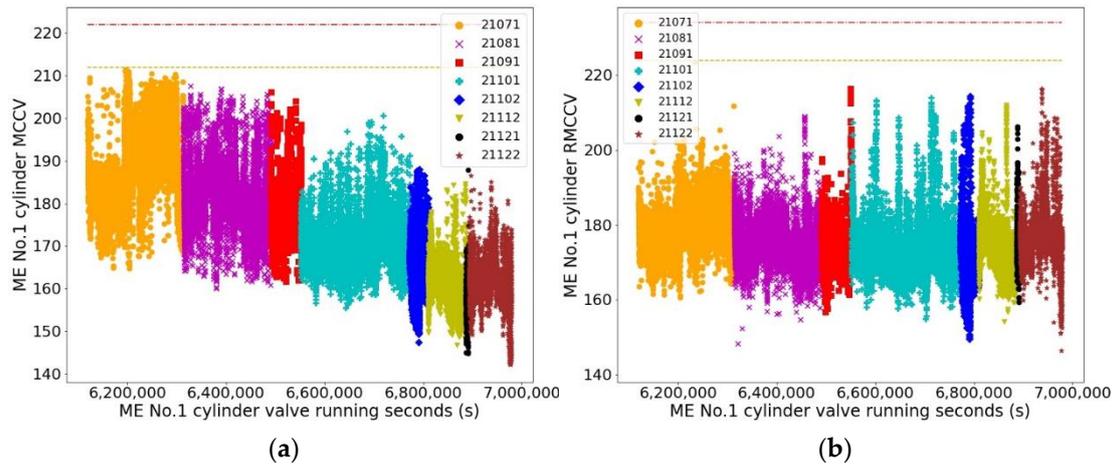
Table 13 shows the regression coefficient of each correction factor derived via the regression analysis of the CET, maximum CCP, and correction factor.

**Table 13.** Regression coefficient of the revision factor of cylinder 1 (source: own elaboration).

Cylinder exhaust gas temperature		
Load	Scav. air temp,	Scav. air press.
1.8185	0.9095	-27.9974
Cylinder combustion pressure		
Load		
3.3900		

Using the MCCV correction factor regression coefficient, the correction criteria, and Equation (5), an RMCCV based anomalistic symptom judgment algorithm was constructed and verified as follows.

The results of the MCCV and RMCCV based on all normal operation status data of cylinder 1 are shown in Figure 8.



**Figure 8.** Comparison of MCCV and RMCCV of normal operating data of cylinder 1. (a) MCCV and (b) RMCCV (source: own elaboration, Python 3.8.3).

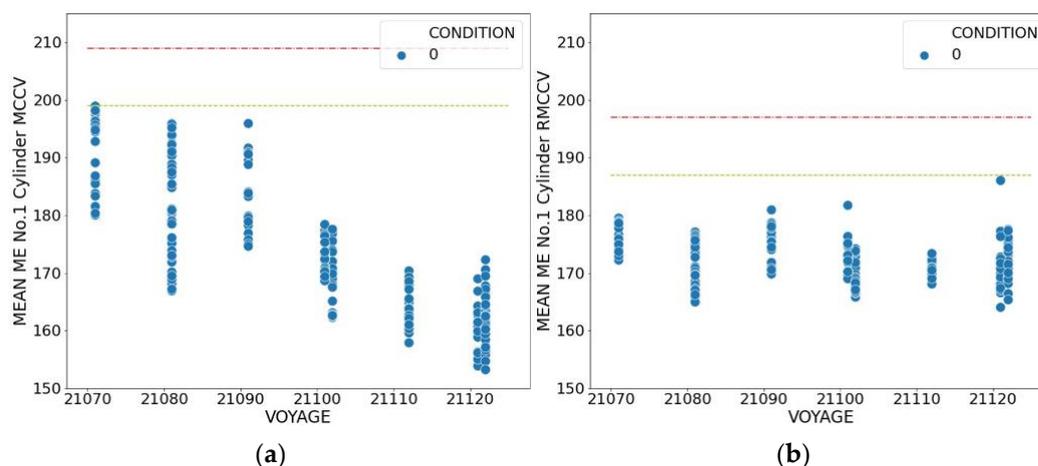
The MCCV showed a decreasing trend as the voyage progressed according to changes in the engine operating environment, such as target RPM. However, the RMCCV was corrected to respond to changes in engine operating conditions, and it remained relatively constant as the voyage progressed. However, some RMCCV data were overcorrected, resulting in abnormalities in noise data and the main data group, which are the criteria for determining the anomalistic symptoms of the MCCV.

Table 14 shows the MCCV and RMCCV results during normal operation. The mean value is the same, whereas the standard deviation improved to 11.6 for the MCCV and 6.1 for the RMCCV. However, the distribution range of the data was from 142.2 to 211.5 (69.3) in the case of the MCCV and from 146.5 to 223.3 (76.8) for the RMCCV owing to the noise data caused by overcorrection.

**Table 14.** Comparison of the MCCV and RMCCV analysis data of cylinder 1 for all normal operating data (source: own elaboration).

Item	MCCV	RMCCV
Mean	175.0	175.0
Std.	11.6	6.1
Min.	142.2	146.5
Max	211.5	223.3

Figure 9 shows the RMCCV and MCCV using the data average calculated based on the same ME ORDER RPM and VOYAGE to reduce the impact of noise data and improve the detection of anomalies. When the average values of the data are applied, the trends according to the voyage are similar for the MCCV and RMCCV (Figure 8), where all data are applied. However, the noise data of the RMCCV are confirmed to be reduced.



**Figure 9.** Comparison of MCCV and RMCCV using the average of normal operating data based on ME ORDER RPM and VOYAGE of cylinder 1. (a) MCCV (b) RMCCV (source: own elaboration, Python 3.8.3).

Table 15 shows the MCCV and RMCCV data in the normal operating state of the ME ORDER RPM and VOYAGE standard average values. The MCCV and RMCCV averages were lower than the overall results presented in Table 13. The data distribution range was also narrowed to 45.6 for the MCCV and 22.1 for the RMCCV, which was reduced by about 70%. As for the standard deviation, the MCCV increased by 0.8 compared to the total data. However, by reducing the RMCCV by 2.3, it was possible to reduce the range of normal operating conditions and the impact of noise data by utilizing the average data based on ME ORDER RPM and VOYAGE.

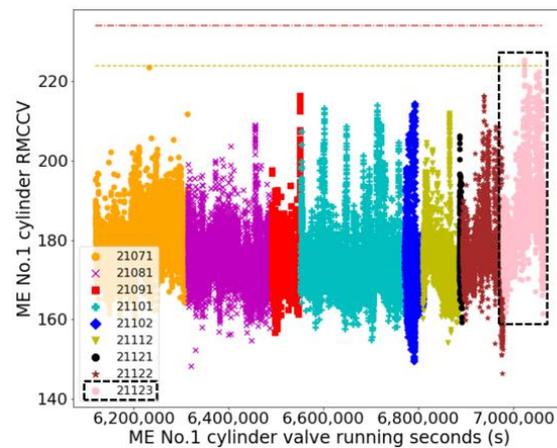
**Table 15.** Comparison of MCCV and RMCCV using the average normal operation data based on ME ORDER RPM and VOYAGE of cylinder 1 (source: own elaboration).

Item	MCCV	RMCCV
Mean	174.1	172.3
Std.	12.4	3.8
Min.	153.2	164.0
Max	198.8	186.1

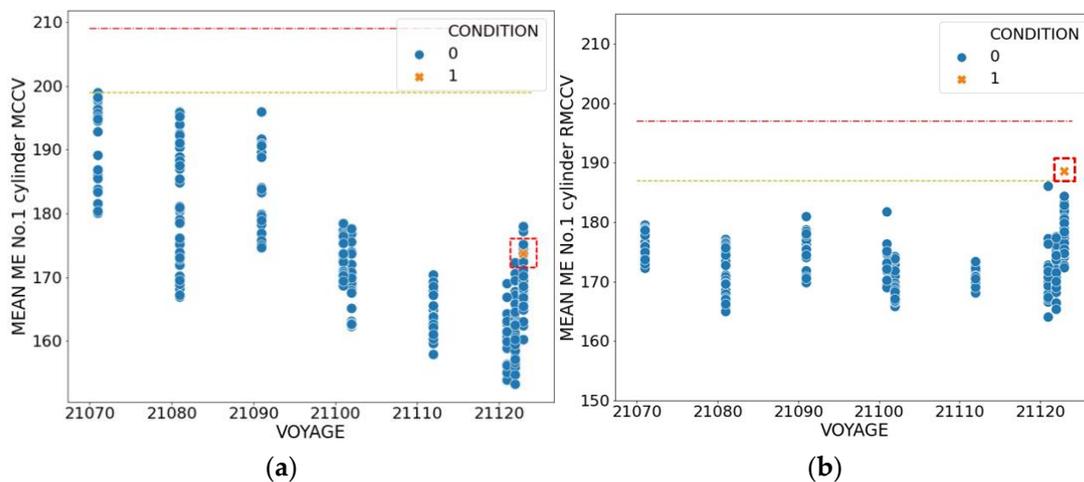
Based on the RMCCV of the mean values of ME ORDER RPM and VOYAGE for determining anomalistic symptoms, the concern level was set to 187–197, and the abnormal level was set to a section exceeding 197.

Figure 10 shows the RMCCV results of all data, including voyage 21123 data with anomalies. The RMCCV of voyage 21123 increases over time. However, clearly determining anomalistic symptoms is difficult due to the noise data.

Figure 11 shows the comparison results of the RMCCV of the average value data based on ME ORDER RPM and VOYAGE. Identifying anomalistic symptom trends with the MCCV in Figure 11a is difficult. In addition, the data of the concern level were included in the normal range, making it impossible to detect anomalistic symptoms. The RMCCV, as shown in Figure 11b, can detect anomalistic symptoms by confirming that the data of voyage 21123 rose beyond the normal data group to the concern level.



**Figure 10.** RMCCV including anomaly symptom data of cylinder 1 (source: own elaboration, Python 3.8.3).



**Figure 11.** RMCCV of the average of data including anomaly symptoms based on ME ORDER RPM and VOYAGE of cylinder 1. (a) MCCV and (b) RMCCV (source: own elaboration, Python 3.8.3).

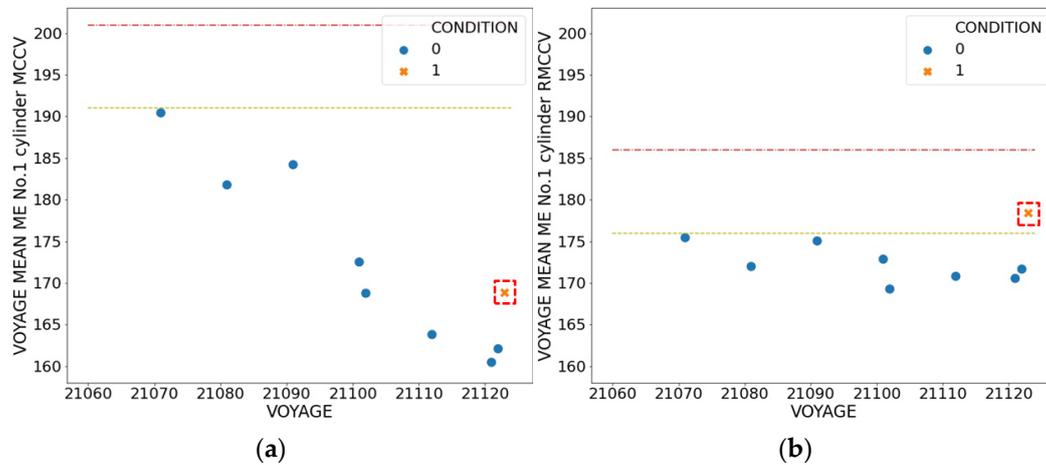
The verification results of the RMCCV algorithm using the average data of each VOYAGE were obtained to improve the detection of anomalistic symptoms. Table 16 presents the RMCCV results of the average normal operating data based on VOYAGE. The mean, standard deviation, and data distribution range improved compared to the average data analysis results based on ME ORDER RPM and VOYAGE. The concern level was set to 176–186, and the abnormal level was set to a section exceeding 186 based on the RMCCV of the VOYAGE averaged data for determining anomalistic symptoms.

**Table 16.** Comparison of MCCV and RMCCV data of average normal operating conditions based on VOYAGE of cylinder 1 (source: own elaboration).

Item	MCCV	RMCCV
Mean	173	172.2
Std.	11.2	2.2
Min.	160.5	169.3
Max	190.4	175.4

Figure 12 shows the results of anomalistic symptom detection using the average data based on the VOYAGE. The MCCV was unable to detect anomalies, as indicated by the

results presented in Figures 6 and 11. However, the RMCCV could detect anomalistic symptoms in the engine more clearly than the results using the mean data based on ME ORDER RPM, VOYAGE, and all data.



**Figure 12.** RMCCV of average data including anomaly symptoms based on VOYAGE of cylinder 1. (a) MCCV and (b) RMCCV (source: own elaboration, Python 3.8.3).

Table 17 summarizes the MCCV correction factor analysis results and revised coefficients of each cylinder for calculating the RMCCVs of the cylinders.

**Table 17.** Revision factors and results of all cylinder MCCV correction factors for the abnormal symptom prediction of the propulsion engine (source: own elaboration).

Cylinder	1	2	3	4	5	6	
Cylinder exhaust temperature							
R <sup>2</sup>	0.795	0.867	0.793	0.880	0.845	0.869	
Adj. R <sup>2</sup> .	0.757	0.842	0.754	0.857	0.817	0.844	
<i>p</i> >  t	Load	0.03	0.014	0.260	0.555	0.234	0.001
	Scav. temp,	0.001	0.000	0.001	0.000	0.000	0.000
	Scav. press.	0.053	0.127	0.259	0.581	0.220	0.013
	MAE	4.685	5.541	7.411	4.946	5.332	4.095
Regression (revision) coefficient							
	ELR	1.8184	2.1837	1.2959	0.4746	1.1577	2.7299
	ESTR	0.9095	1.5505	1.3895	1.5084	1.4077	1.1866
	ESPR	−27.9974	−22.3134	−22.8538	−7.8111	−21.0387	−32.8586
Cylinder MAX combustion pressure							
R <sup>2</sup>	0.970	0.961	0.955	0.959	0.964	0.966	
Adj. R <sup>2</sup> .	0.968	0.959	0.952	0.957	0.962	0.965	
<i>p</i> >  t	Load	0.000	0.000	0.000	0.000	0.000	0.000
	MAE	1.375	1.721	1.867	1.688	1.756	1.188
Regression (revision) coefficient							
	CPLR	3.3900	3.3459	3.1987	3.3506	3.4613	3.3541

R<sup>2</sup><sub>adj</sub> between CET and the correction factor showed a high explanatory power of 0.754–0.857, depending on the cylinder. Moreover, R<sup>2</sup><sub>adj</sub> between the CCP and the correction factor showed a very high explanatory power as the range of 0.952–0.968.

The *p*-values between the CET and the engine load and scavenging air pressure correction factors are relatively high for cylinders 3, 4, and 5. This is due to the high correlation between the engine load and scavenging air pressure.  $R^2_{adj}$  and *p*-value are indicators that identify the relation between the explanatory and response variables [19,37]. Therefore, the reference data used for determining anomalistic symptoms affected the *p*-value owing to the use of a database with a limited data range of  $\geq 135$  RPM based on ME ORDER RPM [19,37].

The MAE of the CET according to the correction factor was in the range of 4.095–7.411, exhibiting good prediction performance compared to the exhaust gas temperature scale. The correction coefficients (ELR, ESTR, and ESPR) of each CET were effectively derived considering  $R^2_{adj}$ , *p*-value, and MAE. The MAE range of the entire CCP was 1.188–1.867, and the prediction result was highly reliable. In addition, considering  $R^2_{adj}$  and the *p*-value, the CCP correction factor (CPLR) was highly reliable, and a valid RMCCV was calculated.

Table 18 presents the RMCCV results of each cylinder using the average data based on ORDER RPM and VOYAGE. Table 19 lists the RMCCV results of each cylinder using the average data based on VOYAGE.

**Table 18.** RMCCV of average data based on ME ORDER RPM and VOYAGE (source: own elaboration).

Cylinder	1	2	3	4	5	6
Mean	172.3	183.9	179.2	183.8	184.3	193.1
Std.	3.8	4.1	4.8	4.0	4.3	4.4
Min.	164.0	174.4	171.8	176.6	175.5	185.0
Max.	186.0	194.8	191.7	202.4	206.8	204.5

**Table 19.** RMCCV of average data based on VOYAGE (source: own elaboration).

Cylinder	1	2	3	4	5	6
Mean	172.2	183.5	178.8	183.6	183.9	192.8
Std.	2.2	2.7	2.7	2.6	2.9	2.5
Min.	169.3	179.1	175.4	179.1	179.0	189.7
Max.	175.4	186.4	183.7	185.8	187.4	196.9

As a result of analyzing the RMCCV of the VOYAGE based average data, the RMCCV standard deviation of each cylinder was improved by about 38% on average compared to the VOYAGE and ME ORDER RPM based average data. In addition, the data distribution range was reduced by 68% on average, making it possible to detect clear anomalistic symptoms.

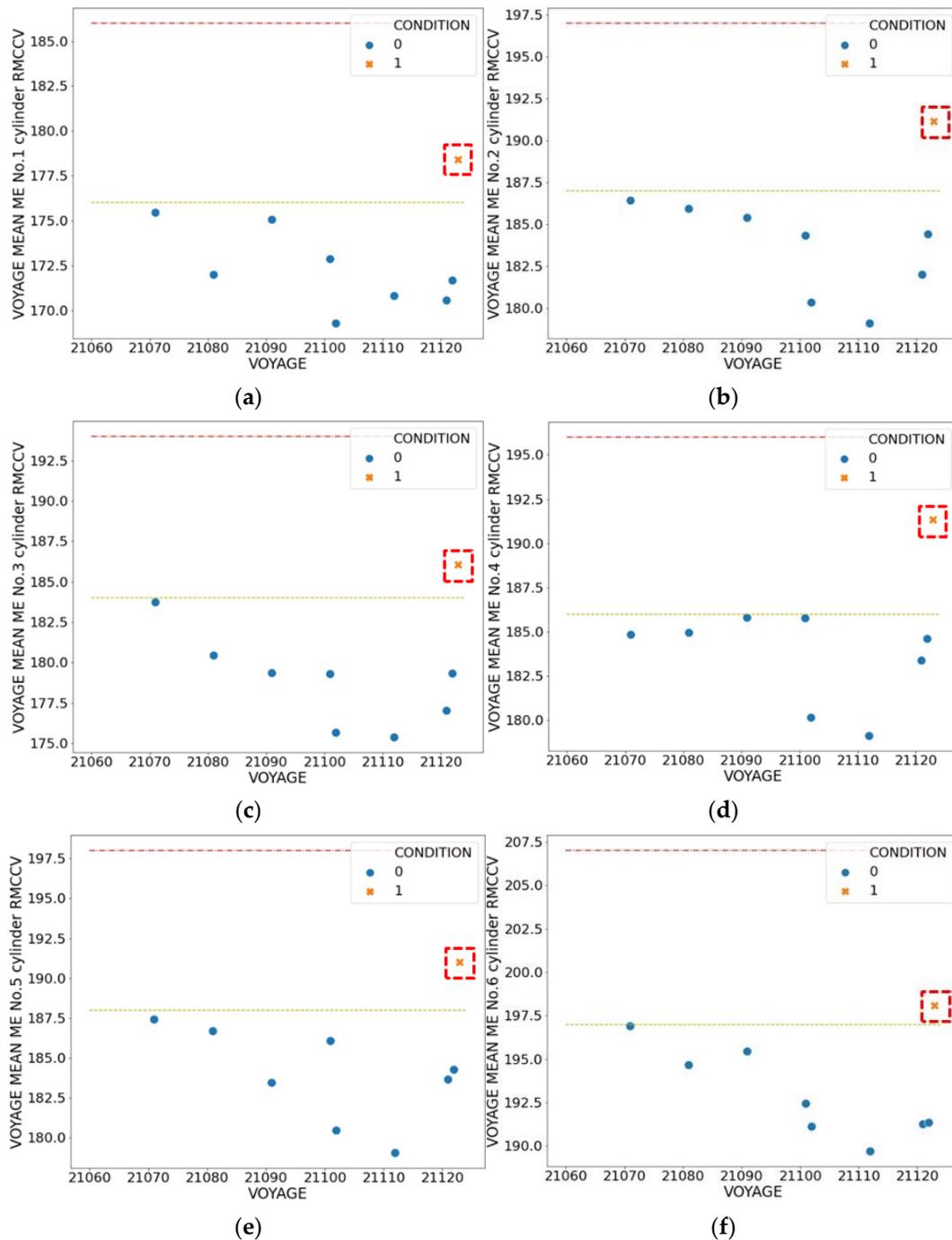
Table 20 lists the criteria for determining the anomalistic symptoms of the propulsion engine established by the RMCCV analysis in the normal operating state of the VOYAGE based average data for each cylinder.

**Table 20.** RMCCV criteria for the abnormality determination of the propulsion engine (source: own elaboration).

Cylinder	1	2	3	4	5	6
Concern (1)	176 < (1) < 186	187 < (1) < 197	184 < (1) < 194	186 < (1) < 196	188 < (1) < 198	197 < (1) < 207
Abnormal (2)	186 < (2)	197 < (2)	194 < (2)	196 < (2)	198 < (2)	207 < (2)

Figure 13 presents the detection results of anomalistic symptoms in all cylinders based on VOYAGE averaged data. Voyage 21123 data displayed anomalistic symptoms in all

cylinders. This confirms the efficacy of the RMCCV based algorithm in determining engine anomalistic symptoms and validates the effectiveness of the constructed algorithm.



**Figure 13.** RMCCV of average data including anomalistic symptoms based on VOYAGE of all cylinders. (a–f) Cylinders 1–6 (source: own elaboration, Python 3.8.3).

#### 4. Discussion

The CET and maximum CCP were identified as anomaly determination factors through the analysis of engine data and manuals. Thereafter, to intuitively determine the engine abnormalities, the MCCV was defined as a judgment factor of abnormality symptoms. However, the indicators of engine abnormalities could not be clearly detected owing to the frequently varying driving environment. Thus, to ensure the detection of engine abnormalities, the RMCCV was constructed to reflect the fluctuating conditions

of engine operations. The RMCCV corrected based on the varying engine load conditions, scavenging air temperature, and the operating environment was derived using the CET and maximum CCP of the engine cylinder. In summary, a reliable algorithm for detecting propulsion engine anomalistic symptoms was developed using average data for each voyage to improve anomalistic symptom detection by minimizing noise data and shortening the algorithm operation. Therefore, a foundation that can be used for ship PdM was established.

The established PdM algorithm intuitively assesses the operating state of the engine with factors related to its combustion state. Consequently, abnormal operating conditions can be preemptively identified by detecting anomalistic symptoms. However, the detailed maintenance items responsible for the anomalistic symptoms are difficult to determine; e.g., it is difficult to identify whether they originate from an issue related to the exhaust gas system or the fuel system. Furthermore, the practical real time application of the constructed PdM algorithm to operational ships requires careful evaluation. Accordingly, the developed PdM algorithm can be effectively applied to operational ships for improvement research, such as algorithm verification, reliability improvement, and developing functionalities to identify detailed maintenance items.

## 5. Conclusions

In this study, the following conclusions were inferred by developing and verifying an engine anomalistic symptom judgment algorithm using operational ship data and machine learning to employ PdM to ship propulsion engines.

1. The CET and maximum CCP of the cylinder were selected as the data that can directly determine the engine condition by analyzing troubleshooting data, maintenance items, and collected data of the engine.
2. To intuitively detect the anomalistic symptoms of the propulsion engine, the MCCV was defined based on the CET and maximum CCP data.
3. The MCCV based anomalistic symptom judgment algorithm failed to determine anomalistic symptoms due to changing operating conditions, such as RPM and the marine environment.
4. To improve anomalistic symptom detection, the RMCCV, which reflects changing operating conditions and environments, was established.
5.  $R^2_{\text{adj}}$  between the CET of the RMCCV and correction factors was in the range of 0.754–0.857 in all cylinders.  $R^2_{\text{adj}}$  between the maximum CCP and correction factors was found to be in the range of 0.952–0.968, showing high explanatory power.
6. The MAE range of the CET according to the correction factor was 4.095–7.411, and the MAE of the maximum CCP was in the range of 1.188–1.867, showing excellent predictive performance compared to the data scale. Therefore, the developed RMCCV based anomalistic symptom prediction algorithm was reliable.
7. To reduce the impact of noise data and improve the predictability of anomalistic symptoms, the average data based on the VOYAGE were applied to the RMCCV based anomalistic symptom judgment algorithm, resulting in the effective anomalistic symptom determination of the propulsion engine.

As described above, an algorithm for effectively determining propulsion engine anomalistic symptoms, suitable for ship PdM, was successfully developed. Given the diverse voyage standards for different routes and operational characteristics of ships, it is crucial to set a certain period, considering the operating characteristics of the target ship and engine, and integrate it into the RMCCV based algorithm.

The future plan involves validating the developed anomalistic symptom judgment algorithm by applying it to operational ships to ensure its reliability. In addition, there are intentions to develop a PdM platform capable of real time anomaly detection in the engine and a function to identify corresponding maintenance items based on detected anomalistic symptoms. These initiatives will contribute to further improving ship maintenance practices and optimizing the PdM system for engines.

**Author Contributions:** Conceptualization, J.P.; methodology, J.O.; investigation, J.P.; data curation, J.P.; writing—original draft preparation, J.P.; writing—review and editing, J.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NO. 2020R111A2073426).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This paper was developed from the Ph.D. thesis of the first author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Drewry Shipping Consultants. *Post-Panamax Containerships—The Next Generation*; Drewry Shipping Consultants: London, UK, 2001.
2. Notteboom, T.E. The time factor in liner shipping services. *Marit. Econ. Logist.* **2006**, *8*, 19–39. [[CrossRef](#)]
3. Park, J.; Oh, J. Analysis of collected data and establishment of an abnormal data detection algorithm using principal component analysis and K-nearest neighbors for predictive maintenance of ship propulsion engine. *Processes* **2022**, *10*, 2392. [[CrossRef](#)]
4. Lazakis, I.; Ölçer, A. Selection of the best maintenance approach in the maritime industry under fuzzy multiple attributive group decision-making environment. *Proc. Inst. Mech. Eng. M* **2016**, *230*, 297–309. [[CrossRef](#)]
5. Michala, A.L.; Lazakis, I.; Theotokatos, G. Predictive Maintenance Decision Support System for Enhanced Energy Efficiency of Ship Machinery. In Proceedings of the International Conference on Shipping in Changing Climates, Glasgow, UK, 24 November 2015.
6. Park, J.-H.; Jang, M.; Lee, G.; Oh, E.; Hur, S. Forecasting algorithm for vessel engine failure. *J. Korean Inst. Inf. Technol.* **2016**, *14*, 109–117. [[CrossRef](#)]
7. Bae, Y.-M.; Kim, M.; Kim, K.; Jun, C.; Byeon, S.; Park, K. A case study on the establishment of upper control limit to detect vessel's main engine failures using multivariate control chart. *J. Soc. Nav. Archit. Korea* **2018**, *55*, 505–513. [[CrossRef](#)]
8. Göksu, B.; Erginer, K.E. Prediction of ship main engine failures by artificial neural networks. *J. ETA Marit. Sci.* **2020**, *8*, 98–113. [[CrossRef](#)]
9. Park, J.H.; Oh, E.; Jang, M.; Seo, Y.; Hur, S. Improved forecasting algorithm for vessel engine failure. *J. Korean Inst. Inf. Technol.* **2017**, *15*, 175–185. [[CrossRef](#)]
10. Zhou, R.; Cao, J.; Zhang, G.; Yang, X.; Wang, X. Heat load forecasting of marine diesel engine based on long short-term memory network. *Appl. Sci.* **2023**, *13*, 1099. [[CrossRef](#)]
11. Youn, I.; Park, J.; Oh, J. A Study on the concept of a ship predictive maintenance model reflection ship operation characteristics. *J. Korean Soc. Mar. Environ. Saf.* **2021**, *27*, 53–59. [[CrossRef](#)]
12. Mehairjan, R.; Osborne, M.; Smit, J. Substation maintenance strategies. *Substations* **2019**, *12*, 997–1009.
13. Alsyouf, I. Maintenance practices in Swedish industries: Survey results. *Int. J. Prod. Econ.* **2009**, *121*, 212–223. [[CrossRef](#)]
14. Rojek, I.; Jasiulewicz-Kaczmarek, M.; Piechowski, M.; Mikołajewski, D. An artificial intelligence approach for improving maintenance to supervise machine failures and support their repair. *Appl. Sci.* **2023**, *13*, 4971. [[CrossRef](#)]
15. Diana, T. Has market concentration fostered on-time performance? A case study of seventy-two U.S. airports. *J. Air Transp. Manag.* **2017**, *58*, 1–8. [[CrossRef](#)]
16. Bala, A.; Ismail, I.; Ibrahim, R.; Sait, S.M.; Oliva, D. An improved grasshopper optimization algorithm based echo state network for predicting faults in airplane engines. *IEEE Access* **2020**, *8*, 159773–159789. [[CrossRef](#)]
17. HYUNDAI Heavy Industries Co., Ltd. *Man B&W S4 OME-B9.5 Diesel Engine Manual*; HYUNDAI Heavy Industries Co., Ltd.: Ulsan, Republic of Korea, 2019.
18. Kim, J.; Jinkyu, P.; Seunghun, L.; Jungmo, O. A study on the use of engine simulation to secure collect machine learning data for failure diagnosis and predictive maintenance in ship engine of ships. In Proceedings of the KSAE 2021 Annual Autumn Conference Y Exhibition, Yeosu, Republic of Korea, 17–20 November 2021; p. 1083.
19. Uyanik, G.K.; Güler, N.; Müller, A.C.; Guido, S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2016.
20. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; Francisco, R.d.P.; Basto, J.P.; Alcalá, S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [[CrossRef](#)]
21. Kecici, A.; Yildirak, A.; Ozyazici, K.; Ayluctarhan, G.; Agbulut, O.; Zincir, I. Implementation of machine learning algorithms for gait recognition. *Eng. Sci. Technol. Int. J.* **2020**, *23*, 931–937. [[CrossRef](#)]
22. Maulud, D.; Abdulazeez, A.M. A review on linear regression comprehensive in machine learning. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 140–147. [[CrossRef](#)]
23. Uyanık, G.K.; Güler, N. A study on multiple linear regression analysis. *Procedia Soc. Behav. Sci.* **2013**, *106*, 234–240. [[CrossRef](#)]

24. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [[CrossRef](#)]
25. Despotovic, M.; Nedic, V.; Despotovic, D.; Cvetanovic, S. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renew. Sustain. Energy Rev.* **2016**, *56*, 246–260. [[CrossRef](#)]
26. Lee, Y.T. A comparison of machine learning models in photovoltaic power generation forecasting. *J. Korean Inst. Ind. Eng.* **2021**, *47*, 444–458.
27. Botchkarev, A. A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdiscip. J. Inf. Knowl. Manag.* **2019**, *14*, 45–76. [[CrossRef](#)] [[PubMed](#)]
28. Birgen, C.; Magnanelli, E.; Carlsson, P.; Skreiberg, Ø.; Mosby, J.; Becidan, M. Machine learning based modelling for lower heating value prediction of municipal solid waste. *Fuel* **2021**, *283*, 118906. [[CrossRef](#)]
29. Pervez, M.N.; Yeo, W.S.; Shafiq, F.; Jilani, M.M.; Sarwar, Z.; Riza, M.; Lin, L.; Xiong, X.; Naddeo, V.; Cai, Y. Sustainable fashion: Design of the experiment assisted machine learning for the environmental-friendly resin finishing of cotton fabric. *Heliyon* **2023**, *9*, e12883. [[CrossRef](#)] [[PubMed](#)]
30. Karch, J. Improving on adjusted R-squared. *Collabra Psychol.* **2020**, *6*, 45. [[CrossRef](#)]
31. Van Rossum, G. Python programming language. In Proceedings of the USENIX Annual Technical Conference, Santa Clara, CA, USA, 17–22 June 2007; pp. 1–36.
32. Srinath, K.R. Python—the fastest growing programming language. *Int. Res. J. Eng. Technol.* **2017**, *4*, 354–357.
33. HYUNDAI Global Service Co., Ltd. *Hyundai HIMSEN H21/32 Diesel Engine Instruction Book*; HYUNDAI Global Service Co., Ltd.: Ulsan, Republic of Korea, 2019.
34. Kowalski, J. An experimental study of emission and combustion characteristics of marine diesel engine in case of cylinder valves leakage. *Pol. Marit. Res.* **2015**, *22*, 90–98. [[CrossRef](#)]
35. Cohen, I. *Pearson Correlation Coefficient. Noise Reduction in Speech Processing*; Springer Science & Business Media: Berlin, Germany, 2009; pp. 1–4.
36. Adler, J.; Parmryd, I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry A* **2010**, *77*, 733–742. [[CrossRef](#)]
37. Lu, M.; Ishwaran, H. A Machine Learning Alternative to p-Values. *arXiv* **2017**, arXiv:1701.04944.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.