Original Research Article: full paper

(2022), «EUREKA: Physics and Engineering»
Number 6

APPLICATION OF A PCA-BASED FAULT DETECTION AND DIAGNOSIS METHOD IN A POWER GENERATION SYSTEM WITH A 2 MW NATURAL GAS ENGINE

Yulineth Cardenas
Department of Energy
Grupo de Investigacion GIOPEN [1]

Gaylord Carrillo
Research Group on Alternative Energies and Fluids (EOLITO) [2]

Anibal Alviz✉
Grupo de Investigacion en Deterioro de Materiales, Transición Energética y Ciencia de datos DANT3
Universidad Señor de Sipan
Arquitectura y Urbanismo
Km 5 Via Pimentel, Chiclayo, Peru Chiclayo, Peru, 14820
alvizanibal@crece.uss.edu.pe

Antistio Alviz
Facultad de ciencias Farmaceutica
Grupo de Farmacologia y Terapeutica
Universidad de Cartagena
Carrera 6, Cl. de la Universidad #36-100, Cartagena de Indias, Bolivar, Colombia, 130014

Ivan Portnoy
Department of Productivity and Innovation Department [1]

Juan Fajardo
Research Group on Alternative Energies and Fluids (EOLITO) [2]

Eric Ocampo
Excellence Center in Renewable Energy and Energy Efficiency – EXCEN [3]

Edson Da-Costa
Excellence Center in Renewable Energy and Energy Efficiency – EXCEN [3]

[1]*Universidad de la Costa (CUC)*
str. 58 #55-66., Barranquilla, Colombia, 080002

[2]*Universidad Tecnologica de Bolívar (UTB)*
Parque Industrial y Tecnologico Carlos Velez Pombo Km 1 Via Turbaco., Cartagena, Colombia, 130002

[3]*Federal University of Itajuba – UNIFEI*
Av. BPS, 1303, Bairro Pinheirinho, Itajuba – MG, Brazil, 37500-000

✉**Corresponding Author**

Abstract

Based on increasing global energy demand, electric power generation from Internal Combustion Engines (ICE) has increased over the years. On this idea, the industries have adopted different methods and procedures to prevent failures in these engines, achieve an extension of the life cycle of the machines, improve their safety, and provide financial savings. For this reason, this work implements a methodology for detecting and identifying failures in a natural gas engine (JGS 612 GS-N. L), based on the integration of Principal Component Analysis (PCA) and alarm streak analysis.

A method used to describe a data set in terms of new uncorrelated variables or components. The components are ordered by the amount of original variance they describe, making the technique useful for reducing the dimensionality of a data set.

Technically, PCA searches for the projection according to which the data are best represented in terms of least squares, using the T^2 and Q statistics. In the initial stage, a PCA-based algorithm was developed to detect abnormal process trends and identify the variables of greater impact when these anomalies arise. In the next stage, an algorithm was developed and implemented, based on the analysis of alarm streaks, to identify the system's behavior and thus classify fluctuations into either normal operating condition drifts or system failures. The application of the proposed methodology with real operation data of the engine (JGS 612 GS-N. L) shows that the method outperforms operators in detecting and identifying faults, as it performs these tasks considerably earlier than operators.

Keywords: Principal Component Analysis, Fault Detection, Fault Diagnosis, Internal Combustion Engine.

DOI: 10.21303/2461-4262.2022.002701

1. Introduction

A fault is defined as an impermissible deviation in an observed variable from a nominal reference range [1]. Fault detection determines whether or not a fault has occurred, while fault diagnosis determines the type of fault that occurred [2].

Industrial plants in modern industries, including the automotive and ICE power generation industries, are large scale, highly complex, and operate with many variables in closed control loops [3]. Also, all processes are susceptible to experiencing inadequate operating conditions or failures. If these processes are not continuously monitored, faults will only be detectable once they have derived into major issues affecting the operation and generating economic losses [4]. The traditional approach to process monitoring is through control charts, which supervise each variable's behavior individually and establish operating limits for them, considering any violation of these limits as an atypical behavior or fault [5]. However, these techniques do not consider correlations between the different variables and their autocorrelations, so they do not allow processes with strongly correlated variables and non-linearities to be adequately monitored [6].

Given the inherent complexity of some processes or plants in the industry, the use of model-based fault detection approaches is impractical to implement, given that the accuracy of these models is compromised [7]. Multiple monitoring methods based on multivariate statistical analysis have been used for FDD [8]. These data-driven approaches have ensured that operation is safe and efficient in many industrial processes. Some of those methods are Principal Component Analysis (PCA) [9], Independent Component Analysis (ICA) [10], Neural Networks (ANN) [11], Fuzzy Logic (FL) [12], among others.

Specifically, in internal combustion engine research, a variety of multivariate techniques have been used for different purposes [13]. For example, principal component analysis (PCA) is popular as it allows a description of the data in terms of uncorrelated variables in a multivariate normal distribution to build predictive models of machine conditions [14]. PCA is advantageous because it simplifies a complex data set by reducing the dimensions of the original data into smaller sets of variables (principal components) that contain most of the information of the unique data set [15]. Some studies have used PCA to analyze internal combustion engines based on the evaluation of emissions produced by the engine [16].

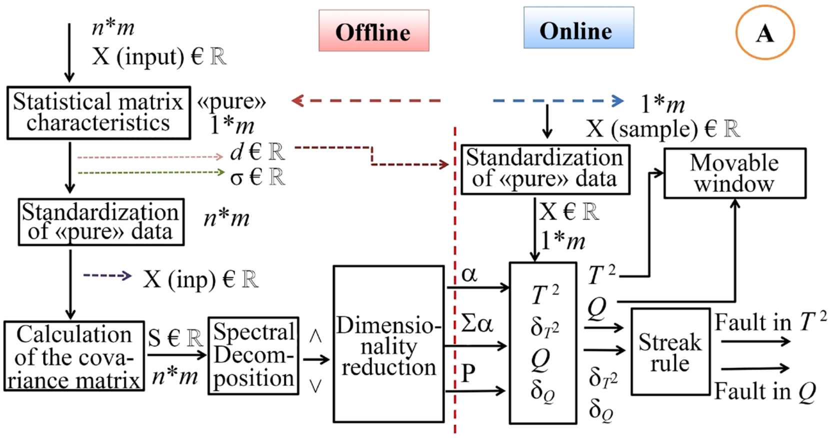
However, no literature investigates the correlation between combustion elements (such as spark plugs) and engine failures [17]. Nevertheless, there are researches, such as that of [18] which address the automatic detection and identification of process measurement equipment or sensor faults through a statistical approach, featuring PCA-based algorithms. Additionally, in their work [19] presented the combination of an exponentially weighted moving average control scheme with the PCA model to improve engine fault detection performance. PCA was used in this study to provide a modeling framework for the development fault detection algorithm, where the results show the effectiveness of the developed algorithm based on the moving average [20] briefly presented the application of the standard PCA technique for fault detection in ICE, omitting the study of fault sensitivity, where fitting and scaling faults are considered [21] showed the application of the standard PCA technique for the design of fault diagnosis systems developed under a failure isolation approach with the plausibility ratio test [22] proposed an improved formulation of PCA; it provides a dynamic and decentralized analysis of the main components (DDPCA) using a variable weighting method [23] developed an adaptive fault detection scheme based on recursive PCA analysis, which addressed false alarms due to regular changes in the actual process. Additionally, the study developed a fault isolation approach based on Generalized Probability. The ratio test (GLR) and Single Value Decomposition (SVD) are one of the general PCA techniques, in which the displacement and scaling error can be easily isolated with a fault model [24].

This article proposes the design of an FDD method, implemented for a 2 MW natural gas engine, based on PCA integrated to a false alarm reduction method. With the methodology proposed in this research, the improvement of DDF is sought, through an alarm streak analysis, to identify faults or operation drifts in the process. For the analysis developed in this work, historical data of the operation of the Jenbacher JMS 612 GS-N engine, used to generate electricity in a company of the industrial sector, was used.

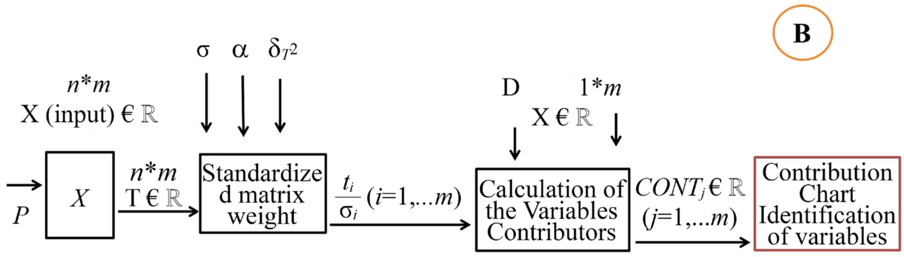
2. Materials and methods

2. 1. PCA algorithm

The PCA algorithm is used in this research as the fault detection technique. Moreover, the algorithm proposed by [25] is used to determine sample-wise contributions of the individual variables to arising faulty conditions as a means to identify those variables that are most likely affected by or causing the abnormal operating condition. The off-line stage of the PCA algorithm **Fig. 1**, which includes data pre-treatment and training. On the other hand, subsection 2.1.2. addresses the on-line stage, addressing the fault detection approach and presenting the contribution algorithm proposed by [25], which will be used to identify the signature variables (i.e., fault-related variables).



a



b

Fig. 1. Calculation algorithm sequence: *a* – Fault Detection; *b* – Diagnosis Algorithms

2. 1. 1. Off-line stage

Historical process data corresponding to normal operating conditions must be gathered and available prior to the implementation of this stage. This historical dataset, referred to as the training set, must be arranged into a matrix:

$$X_0 = \begin{bmatrix} X_{11} & X_{12} & X_{1m} \\ X_{21} & X_{22} & X_{2m} \\ \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & X_{nm} \end{bmatrix} \in R^{nXm}, \quad (1)$$

where n is the number of samples or observations gathered in the training set, while m is the number of variables.

Once the training set is created, the off-line stage is carried out. This stage comprises two sub-stages, accordingly: Data pre-processing and Training. These sub-stages are explained in detail in this subsection.

Data pre-processing. The first task in this sub-stage is to ensure that the number of observations, n is large enough to produce a detection threshold sufficiently closed to that obtained by assuming infinite data in the training set. Given a level of significance, α, and a dataset size (n, m), equation (2) [26] computes the relative error for the detection threshold:

$$\in = \frac{\left[\frac{m(n-1)(n+1)}{n(n-m)}\right] F_a\left(m,(n-m) - x_a^2(m)\right)}{x_a^2(m)}, \quad (2)$$

where F_a and x_a^2 are the Fisher and Chi square distributions, respectively.

Within the training set, there may be variables that do not provide information for process monitoring [27] due to null variance, signal loss issues, or reading errors, among other reasons. Therefore, these variables must be removed.

Similarly, measurements isolated from the rest of the group can be considered as out-of-control measurements, referred to as outliers. If not removed, the influence of these outliers on the estimation of the statistical parameters can be significant and affect the normal process behavior [28]. Thus, the data preprocessing described here must be carried out prior to the training stage, which is further explained in detail.

Training: consider the training dataset as defined in equation (1) after the data preprocessing has been carried out. This training set must be normalized to avoid biased measurements due to unevenly sized variables. This normalization comprises the subtraction, from each variable, of the corresponding mean and dividing it by its respective standard deviation. If the mean vector, d, is defined as a column array containing the means of all the variables, it can be computed as in equation (3):

$$d = \frac{1}{n}(x_0)TI_n = [\bar{x}_1 \quad \bar{x}_2 \quad \cdots \quad \bar{x}_3]^T, \left(I_n = [1 \quad 1 \quad \cdots \quad 1]^T \in R^n\right). \quad (3)$$

Furthermore, the standard deviations of the variables are stacked in a diagonal matrix, as shown in equation (4):

$$\sum = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{bmatrix} \in R^{mXm}. \quad (4)$$

Then, data standardization is carried out as follows:

$$X = \left(X_0 - I_n d^T\right)\sum{}^{-1} \in R^{nXm}. \quad (5)$$

The covariance matrix is now defined [29] as:

$$S = \frac{1}{n-1} X^T X \in R^{mXm}. \quad (6)$$

A spectral decomposition of S is carried out. The eigenvalues, λ, can be determined with the equation (11):

$$\det(\lambda I - S) = 0 \rightarrow \lambda_1, \lambda_2, ..., \lambda_m, \quad (7)$$

where I is the identity matrix. The eigenvalues must be sorted in decreasing order, i.e.:

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m. \quad (8)$$

The eigenvectors, v_1, are determined from equation (9):

$$(\lambda_1 I - S) \cdot v_i = 0, \; v_i \in R^m, \; i = 1,2,...,m. \quad (9)$$

Once the spectral decomposition is performed, it yields the eigenvalue matrix Λ and the eigenvector matrix V, as follows:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \in R^{mXm}, \quad (10)$$

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \in R^{mXm}. \quad (11)$$

Then, the covariance matrix, S can be re-written as follows:

$$S = V \Lambda V^T. \quad (12)$$

The reduction of variability is based on optimally capturing the data variability while avoiding the effect of noise. Therefore, a dimensionality reduction is required. The number of principal components, a, can be determined so that a percentage of the data's variability is retained (typically 95 %) [30], i.e.:

$$\frac{\sum_{j=1}^{a} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \geq 0.95 \rightarrow a. \quad (13)$$

The use of the eigenvalues as a proxy to the dataset's variability is allowed as the data has been transformed (by the PCA algorithm) into a space where these eigenvalues are directly related to the variability in the direction spanned by their corresponding eigenvectors.

Once the number of principal components is determined, the eigenvector matrix, V, must be reduced by retaining only its first a column, becoming the so-called *loading matrix*, $P \in R^{mXn}$, as follows:

$$P = \begin{bmatrix} v_1 & v_1 & \cdots & v_a \end{bmatrix}. \quad (14)$$

Moreover, a matrix Γ_a R^{aXa} must be created containing the first a columns and rows of matrix $\Gamma \in R^{mXn}$, so that: $\Lambda = \Gamma^T \Gamma$, i.e., Γ is a diagonal matrix containing the squared roots of all the λ_j as (non-null, diagonal) entries.

Finally, the score matrix is defined as:

$$T = XP \in R^{nXa}. \quad (15)$$

2. 1. 2. On-line stage

In this stage, the process is monitored in real-time, receiving one observation per time sample. Two tasks are performed in this stage: Fault Detection and Signature Variables Identification.

Fault detection: given a current observation, $x \in R^m$, it must be normalized using the means and standard deviations obtained during the training (off-line) stage, and two statistics are to be computed to perform the fault detection task, accordingly. The Hoteling's T^2 [31] and the Q statistic. The T^2 statistic is computed as follows:

$$T^2 = x^T P \Gamma^{-2} P^T x, \tag{16}$$

where Γ_a contains the first a rows and columns of Γ.

The detection threshold for the T^2 statistic is defined by equation (17):

$$\delta_{T^2} = \frac{a(n-1)(n+1)}{n(n-a)} F_a(a, n-a). \tag{17}$$

The above threshold is a multivariate generalization of the Fisher's test statistic proposed by [32].

Any observation rendering a T^2 value greater than the detection threshold is considered an abnormal sample or fault with respect to the process's normal behavior in the principal space. Additionally, the behavior of the ($m-a$) eigenvalues, corresponding to the residual space, can be monitored using the Q statistic, also known as SPE (Squared Prediction Error) as shown in equation (18):

$$Q = r^T r, \tag{18}$$

where r is calculated as shown in equation (19):

$$r = \left(I - PP^T\right)x. \tag{19}$$

The detection threshold for the Q statistic is given by equation (20) [33]:

$$\delta_Q = \theta_1 \left[\frac{\theta_2 h(h-1)}{\theta_1^2} + \frac{c_a \sqrt{2\theta h^2}}{\theta_1} \right]^{1/h}, \tag{20}$$

where

$$\theta_1 = \sum_{i=r+1}^{n} \lambda_1,\ \theta_2 = \sum_{i=r+1}^{n} \lambda_i^2,\ \theta_3 = \sum_{i=r+1}^{n} \lambda_i^3 3, \tag{21}$$

$$h = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}. \tag{22}$$

And c_a is the normal deviate corresponding to the (1−α) percentile.

Whenever a threshold violation occurs for Q, it means the observation exhibits a residual space behavior different from that of the training set.

Signature Variables Identification: when either detection threshold $\left(\delta_{T^2}\right)$ or (δ_Q) is violated, the contribution of each variable to such an out-of-control event should be determined. The approach used in this paper, proposed by [34], is based on quantifying each process variable's contribution to the individual representative PCA scores.

Miller et al. proposed a sample-wise procedure to estimate the contribution of each variable x_j $\left(j = 1,...,m\right)$ to produce a threshold violation. The calculation of the contribution measure involves the following steps:

– check the standardized scores, $t_i/\sigma_i = \left(i = 1,...,m\right)$, where t is the column j-th column of the score matrix, T. Record the scores that meet the condition $\left(t_i/\sigma_i\right)^2 > \left(T_a^2\right)^{1/a}$;

– calculate the contribution of each variable, x_j, to scores out of control δ_{T^2}:

$$cont_{i,j} = \frac{t_i}{\sigma_i^2} P_{i,j}\left(x_j - \mu_j\right), \tag{23}$$

where $P_{i,j}$ are the loading matrix entries. If $cont_{i,j}$ is negative; it is considered as a zero value.

Finally, the overall contribution of the variable x_j is calculated as:

$$CONT_j = \sum_{i=1}^{r} cont_{i,j}. \tag{24}$$

Fig. 1, *a* shows an operation diagram of both the off-line and the on-line PCA stages.

2. 2. False alarm Reduction Approach for the T^2 and Q statistics

False alarm rates must be controlled for both the T^2 and Q statistics. Thus, a statistics-based false alarm reduction method is proposed to further control these false alarm rates. The proposed method is shown in **Fig. 2**.

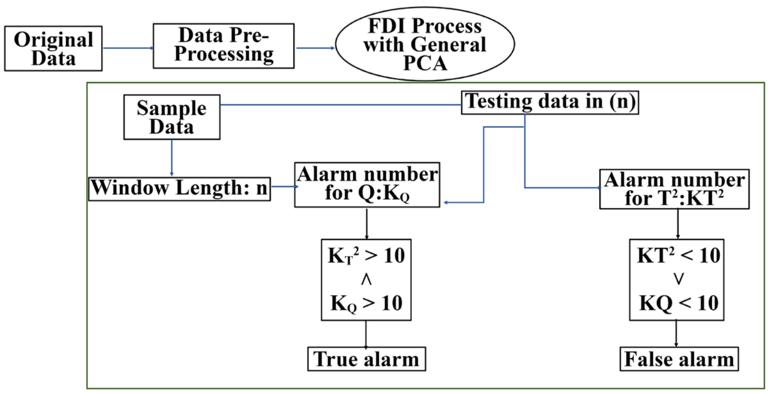


Fig. 2. Flow chart for the false alarm reduction method

The method works under the assumption that the false alarm rates for the T^2 and Q statistics should be around α (customary 5 %). The approach further considers an observation moving window with n samples. If k continuous violations take place, the algorithm considers such an event as a streak. When streaks occur simultaneously in both statistics, the algorithm triggers a 'true alarm,' otherwise, the algorithm dismisses the event as a 'false alarm.

2. 3. Engine Description

The equipment selected is a Jenbacher JMS 612 GS-N engine, as shown in **Fig. 3.** It belongs to a 2 MW generation system operating with natural gas. The main specifications of the engine are shown in **Table 1.**

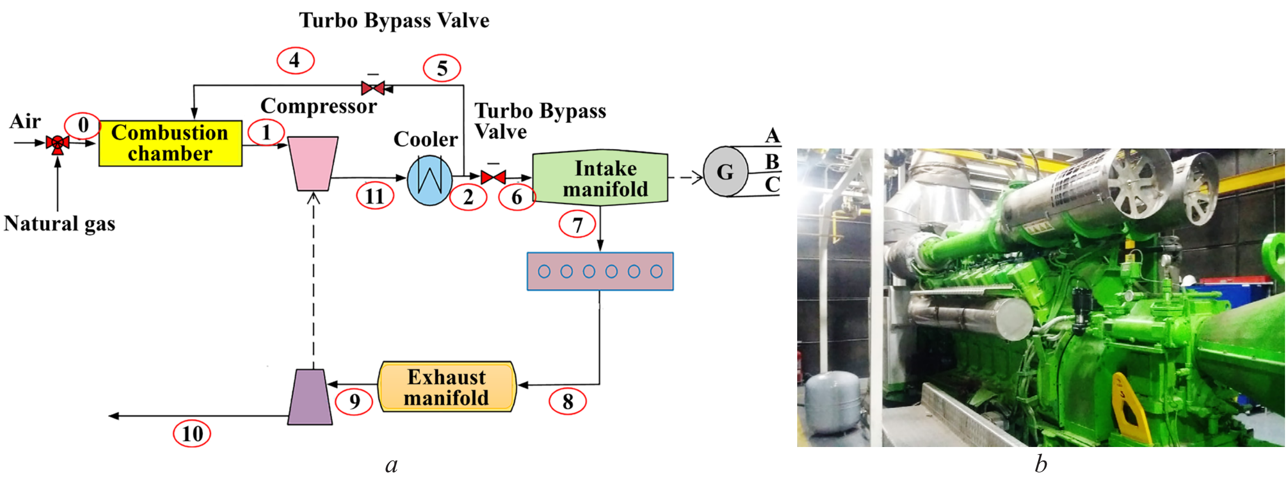


Fig. 3. Engine diagram: *a* – General flow diagram *b* – Physical structure of the motor for the engine generator system

Table 1
Characteristic of Jenbacher JMS 612 GS-N engine [13]

Parameter	Value
Volume (Liters)	74.852
Compression ratio	10.5
Number of Cylinders	12 V at 60°
Stroke length (mm*)*	220
Chamber diameter (mm)	190
Maximum torque (kN·m)	60.66
Maximum power (kW)	1820
Rotation speed (rpm)	1500

This engine has a three-phase electric generator working at a frequency of 60 Hz, with a power factor of 0.9 %, delivering a reactive power of 911 kvar, electrical power of 1975 kW, and apparent power of 2177 kvar. Also, it has an average voltage between lines of 13.264 V.

Fig. 3 shows the engine's configuration used in a plastic company in the industrial sector. The operation of this set begins at point (0), where the natural fuel gas (with the composition: 97.97 % CH_4, 1.5 % N_2, 0.25 % C_2H_6, and 0.16 % CO_2) is mixed with air (Data provided by plant operations staff). This mixture passes through the combustion chamber (0-1), where the process of releasing energy takes place, obtaining gases at high temperatures. Next, these gases are driven to the compressor (1-11), where pressure is increased and fed to the cooler (11-2). This component allows controlling the temperature of the gases which enter the engine's intake manifold (2-7). Finally, the inlet manifold is the component directing the combustion gases at high pressure and temperature to each of the engine's cylinders (7-8). The torque necessary to drive the electric generator (G) is produced at this stage of the process. Then, the gases that come out of each cylinder go to the engine's outlet manifold (8-9), where the gases are mixed and driven through the system's turbine (E9-10), where the gases are expanded to generate the torque required to ultimately drive the system's generator (G), generating electricity.

3. Results and discussion

3. 1. Selection of ICE measurement variables for the development of the PCA algorithm

The observation window of the engine's variables starts on 8/08/2019 at 00:58 AM and goes up to 08/21/2019 at 11:46:00 PM; the observations were obtained with a sampling period of 2 minutes for a total of 9,898 measurements for 50 variables, which are shown in **Table 2**.

Table 2
Selected variables for Principal Component Analysis

Variable (*s*)	Units	Total variables	Symbol (*s*)	Mean (*s*)
Voltage from 1 to 12	kV	12	(x_1–x_{12})	15.673
Compressor bypass	%	1	(x_{13})	48.773
Gas mixer 1 and 2	%	2	(x_{14}–x_{15})	183.342
Cylinder knocking noise from 1 to 12	mV	12	(x_{16}–x_{27})	81.082
Cylinder valve noise from 1 to 11	mV	11	(x_{28}–x_{38})	1377.668
Cylinder exhaust temperature from 1 to 12	°C	12	(x_{39}–x_{50})	592.493

The symbols provided in **Table 2** are a simplified way to summarize information. However, at the end of the analysis, the variables affected by or causing the assessed fault are described with their full tag/name, units, and the corresponding symbols as in **Table 2**.

An example of the behavior of some of the variables is shown below in **Fig. 4**.

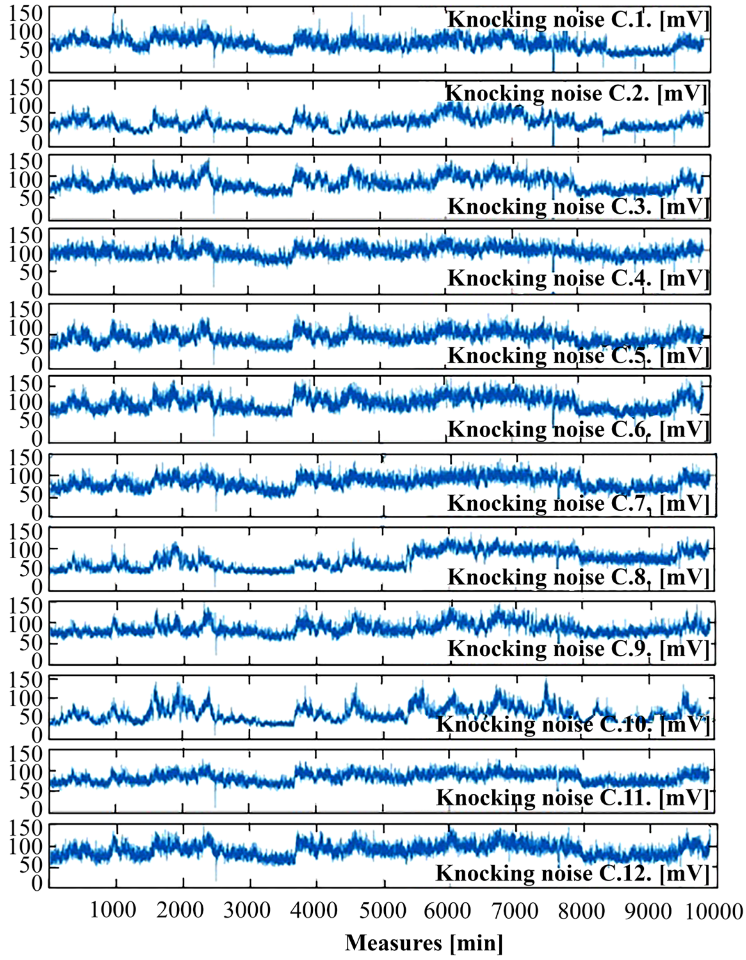


Fig. 4. Illustration of the behavior of some variables

The variables' behavior, taken as an example from the entire set of variables (shown in **Fig. 4**) shows disturbances taking place at different moments, revealing graphically that there was a change in the operation of the engine, false alarms, anomalies, and faults within the 9898 observations. In particular, the plots in **Fig. 4** show fewer disturbances up to observation 7000. However, fluctuations are stable throughout the measured window. This indicates that it is necessary to standardize and observe them to analyze the variables' real behavior over time.

3. 2. Application of the proposed approach

3. 2. 1. Off-line stage

3. 2. 1. 1. Data pre-processing

The study considered two datasets: a training set and a testing set. The raw training set comprised 5600 observations of 87 variables, while the raw testing set comprised 9898 observations of 87 variables as well. In order con carry out the data pre-processing, both datasets were merged. Thus, the entire raw dataset, $X_0^p \in R^{15498X87}$ comprised 15498 observations of 87 variables. After removing variables with null variability and signal loss or missing value issues, the pre-processed set, $X_0 \in R^{15498X50}$, comprises 15498 observations of 50 variables. After the pre-processing, the datasets were split again.

3. 2. 1. 2. Training

As explained in the Data-preprocessing subsection, the (pre-processed) training set comprises 5600 observations of 50 variables. This training set includes data obtained when the engine worked without reported failures or anomalies, according to the operators' knowledge. The PCA training is executed, such as explained in equations (3)–(15). **Fig. 5** shows the magnitude of all eigenvalues and graphically illustrates the selection of the number of principal components (NPC), symbolized with a in the notation used in this paper. The NPC selection is performed based on the 95 % variability capture, as previously shown in equation (13). The outcome of the NPC selection is $a = 30$ principal components.

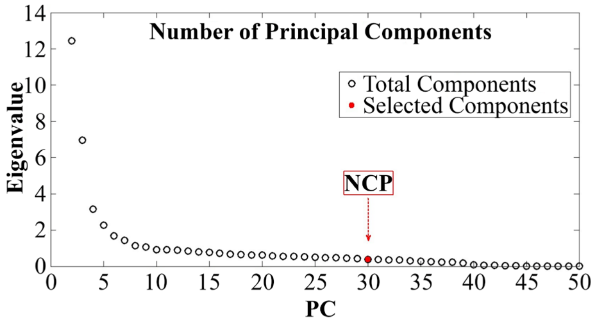


Fig. 5. Illustration of the selection of the NPC

Moreover, **Table 3** shows in detail the magnitude of all the principal components, along with the percentual variability they are associated with.

Table 3
Characterization of the PCs

a	*a* %	Eigenvalues
1	26.16	15.48
2	20.44	12.43
3	9.83	6.97
4	3.87	3.17
5	3.00	2.27
6	2.35	1.68
7	2.13	1.42
8	2.04	1.14
9	1.78	1.06
10	1.70	0.92
11	1.60	0.91
12	1.57	0.90
13	1.51	0.86
14	1.48	0.79
15	1.30	0.77
16	1.26	0.72
17	1.19	0.68
18	1.18	0.64
19	1.12	0.63
20	1.06	0.62
21	1.06	0.57
22	0.97	0.55
23	0.96	0.54
24	0.91	0.51
25	0.89	0.50
27	0.83	0.48
28	0.77	0.47
29	0.75	0.46
30	0.70	0.42

3. 2. 2. On-line Stage

Based on the T^2 and the Q statistics, this subsection addresses the on-line fault detection task, as well as the further identification of signature variables, i.e., those variables most affected by or causing the abnormal operating condition.

3. 2. 2. 1. Fault detection

Using equations (16)–(22), the fault analysis applied for this case study is made for the different normal operating conditions presented in the 9898 observations within the testing set. Results are displayed in **Fig. 6**.

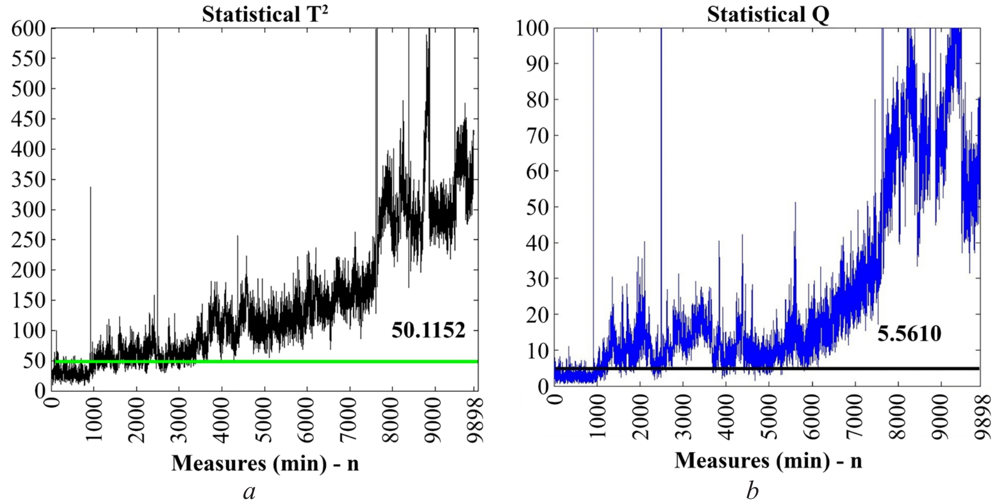


Fig. 6. Control charts: *a* – T^2; *b* – Q statistics

The normal operating condition was present in the first 1000 observations, indicating that no abnormal behavior arose. This level of regular service is consequent to the expected behavior of the variables in the control state. On the other hand, the thresholds violations indicate an operational change or a failure not reported or not detected by the operators. Additionally, the data shows a transient state, as shown in **Fig. 7**. The transient state identified by the fault detection algorithm as a continuous series of magnitude-increasing threshold violations is an early detection of an incipient fault, which would only be detected by the operators considerably long after.

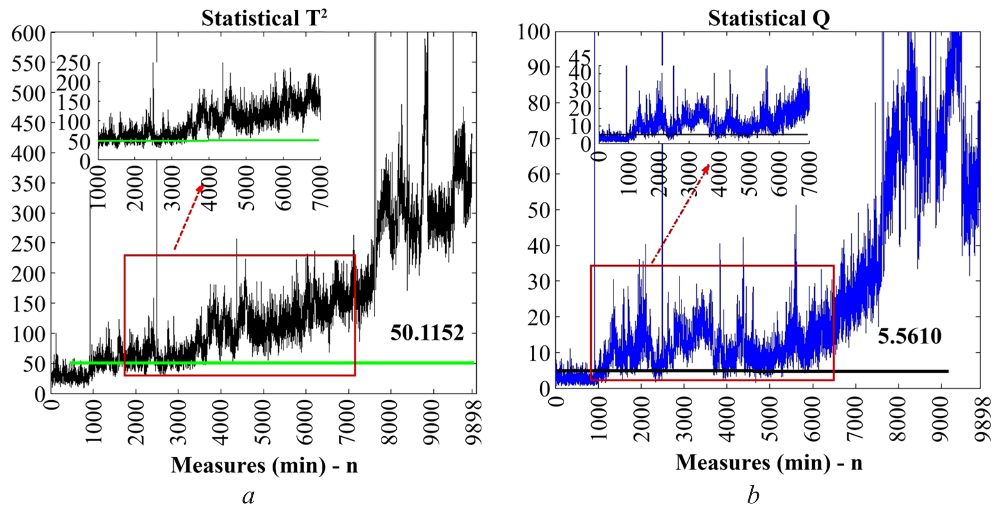


Fig. 7. Illustration of the transient behavior of the engine:
a – T^2; *b* – Q statistics

Fig. 8 illustrates, within the control charts for T^2 and Q, the moment when the plan operator reported an engine malfunction. The algorithm outperformed the operators regarding the detection capacity, anticipating them by roughly 7500 samples (i.e., roughly 10 days).

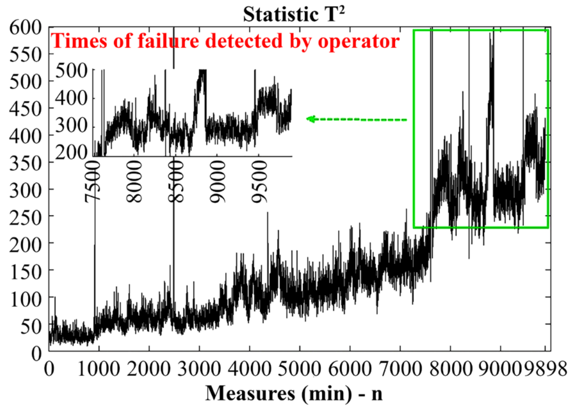


Fig. 8. Fault detection time by operators

The operators' lack of detection capacity is normal as they are humans, yet it led them to stop the engine only when the consequences of the failure had caused damage, generating losses due to prolonged idle times and corrective maintenance.

In addition to this, the stop is extended by ignorance of the variable or variables that could create the failure. Therefore, identifying the fault-related variable(s) is an important task, which will be addressed now.

3. 2. 2. 2. Signature variables identification

Using equations (23), (24), the variables' contributions to the out-of-control operation are determined. The Cylinder Noise 2 (x_{17}) and Cylinder Voltage 8 (x_8) exhibit an average contribution, as computed by the approach proposed by [34] of 50 and 30, respectively, when the fault is fully developed. **Fig. 9** shows a heatmap displaying these contributions, along with their time evolution. Let's notice that the operators do not detect the anomaly until one of these variables represents 50 % or more contribution. Operators report defects from observation 7600 onwards when the criterion is met.

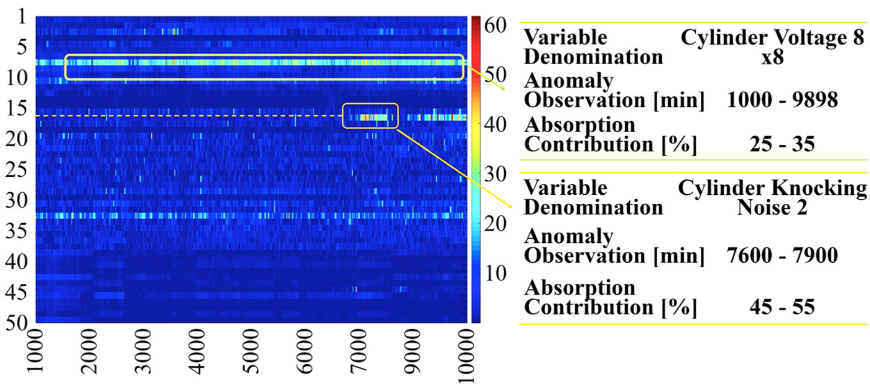


Fig. 9. Contribution heatmap

Results prove it relevant to identify the variable(*s*) most associated with the fault from the moment the ICE presents an anomaly, as these can be used as signatures to characterize and further diagnose this fault. In this case, the fault starts developing from observation 1000. **Fig. 10** summarizes the main findings with the fault detection and signature variable identification approaches.

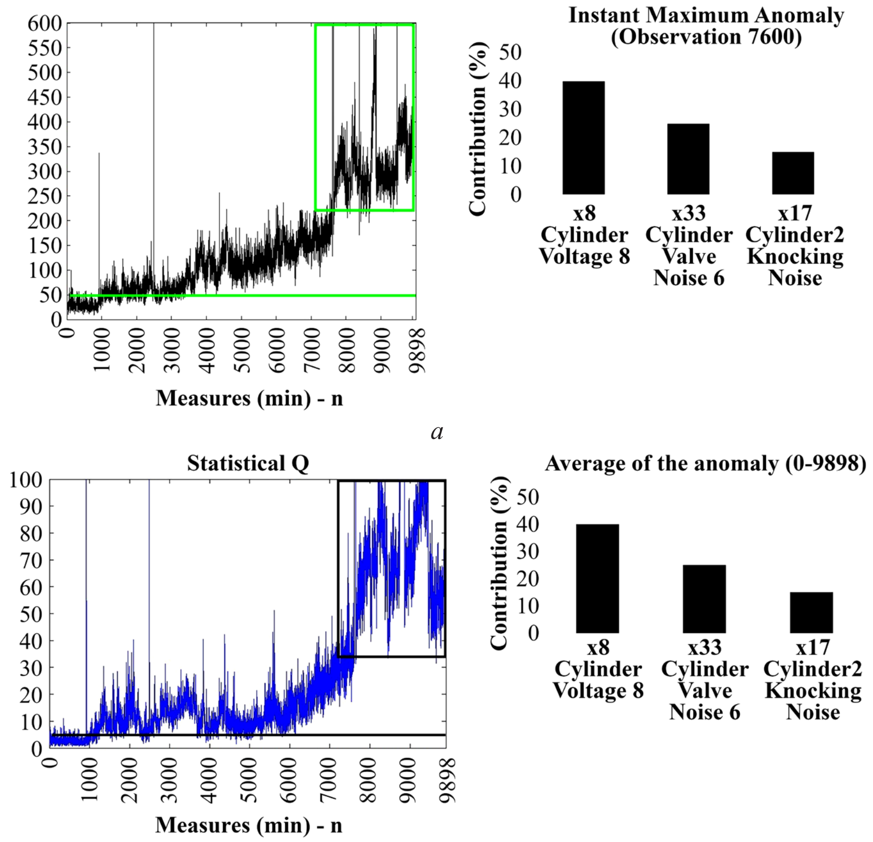


Fig. 10. Summary of findings:
a – instant maximum anomaly; *b* – average of the anomaly

It is important to mention that one of the limitations of this study corresponds to the variables that were not taken into account due to the lack of measurements and processing of the sensors of the system. Aspects should be taken into account when trying to replicate this study in other combustion engines. However, the variables taken into account for this study correspond to **Table 2**, and most of them correspond to determining factors in the combustion of the Jenbacher JMS 612 GS-N engine equipment, which allowed this study to be carried out efficiently. However, the inclusion of other variables could be an interesting aspect to be evaluated in future studies.

4. Conclusions

The main result of this study is the validation of the mathematical model proposed for the detection of failures in internal combustion engines, this model was validated in a Jenbacher JMS 612 GS-N engine. The model was built based on principal component analysis, a statistical tool that allows to assessment the variability or failure component in a series of normless operating parameters. Thirty (30) principal components representing at least 95 % of the engine variability were obtained.

The explanation of this result is focused on the analysis related to the T^2 and Q stipends that allow the quantitative assessment of the behavior in a time series. The method and its application become successful to the extent that human operators can detect engine failures and identify related variables up to ten days in advance. It will allow the industry that uses this type of engines to have greater control of the operations minute by minute avoiding any anomaly of the process.

Thus, this research was able to demonstrate that the proposed methodology allows the early detection of failures, controlling the false alarm rate, identifying the variables that contribute to or explain the out-of-control state. Future research should consider the evaluation of the reduction of the false alarm rate achieved with the application of the proposed methods.

Conflict of interest

The authors declare that there is no conflict of interest in relation to this paper, as well as the published research results, including the financial aspects of conducting the research, obtaining and using its results, as well as any non-financial personal relationships.

Funding

The University Señor de Sipán provided the funds to cover the APC of the paper.

Data availability

Data cannot be made available for reasons disclosed in the data availability statement.

Acknowledgments

Acknowledgments to Universidad del Atlantico. The authors wish to thank the National Agency of Petroleum, Natural Gas and Biofuels (ANP) [grant number PRH-ANP n°46.1].

References

[1] van Schrick, D. (1997). Remarks on Terminology in the Field of Supervision, Fault Detection and Diagnosis. IFAC Proceedings Volumes, 30 (18), 959–964. doi: https://doi.org/10.1016/s1474-6670(17)42524-9

[2] Quiñones-Grueiro, M., Prieto-Moreno, A., Verde, C., Llanes-Santiago, O. (2019). Data-driven monitoring of multimode continuous processes: A review. Chemometrics and Intelligent Laboratory Systems, 189, 56–71. doi: https://doi.org/10.1016/j.chemolab.2019.03.012

[3] Coussement, A., Gicquel, O., Parente, A. (2013). MG-local-PCA method for reduced order combustion modeling. Proceedings of the Combustion Institute, 34 (1), 1117–1123. doi: https://doi.org/10.1016/j.proci.2012.05.073

[4] Jung, D., Ng, K. Y., Frisk, E., Krysander, M. (2018). Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation. Control Engineering Practice, 80, 146–156. doi: https://doi.org/10.1016/j.conengprac.2018.08.013

[5] Haanchumpol, T., Sudasna-na-Ayudthya, P., Singhtaun, C. (2020). Modern multivariate control chart using spatial signed rank for non-normal process. Engineering Science and Technology, an International Journal, 23 (4), 859–869. doi: https://doi.org/10.1016/j.jestch.2019.12.001

[6] Isermann, R. (2005). Model-based fault-detection and diagnosis – status and applications. Annual Reviews in Control, 29 (1), 71–85. doi: https://doi.org/10.1016/j.arcontrol.2004.12.002

[7] Jafarian, K., Mobin, M., Jafari-Marandi, R., Rabiei, E. (2018). Misfire and valve clearance faults detection in the combustion engines based on a multi-sensor vibration signal monitoring. Measurement, 128, 527–536. doi: https://doi.org/10.1016/j.measurement.2018.04.062

[8] Portnoy, I., Melendez, K., Pinzon, H., Sanjuan, M. (2016). An improved weighted recursive PCA algorithm for adaptive fault detection. Control Engineering Practice, 50, 69–83. doi: https://doi.org/10.1016/j.conengprac.2016.02.010

[9] Niu, G., Xiong, L., Qin, X., Pecht, M. (2019). Fault detection isolation and diagnosis of multi-axle speed sensors for high-speed trains. Mechanical Systems and Signal Processing, 131, 183–198. doi: https://doi.org/10.1016/j.ymssp.2019.05.053

[10] Albarbar, A., Gu, F., Ball, A. D. (2010). Diesel engine fuel injection monitoring using acoustic measurements and independent component analysis. Measurement, 43 (10), 1376–1386. doi: https://doi.org/10.1016/j.measurement.2010.08.003

[11] Shahnazari, H. (2020). Fault diagnosis of nonlinear systems using recurrent neural networks. Chemical Engineering Research and Design, 153, 233–245. doi: https://doi.org/10.1016/j.cherd.2019.09.026

[12] Ahmadi, H., Gholamzadeh, M., Shahmoradi, L., Nilashi, M., Rashvand, P. (2018). Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review. Computer Methods and Programs in Biomedicine, 161, 145–172. doi: https://doi.org/10.1016/j.cmpb.2018.04.013

[13] Cardenas, Y. (2019). Fallas en bujías para motores de generación a gas. (Tesis de maestría). Universidad del Atantico.

[14] Camacho, J., Pérez-Villegas, A., García-Teodoro, P., Maciá-Fernández, G. (2016). PCA-based multivariate statistical network monitoring for anomaly detection. Computers & Security, 59, 118–137. doi: https://doi.org/10.1016/j.cose.2016.02.008

[15] Meglen, R. R. (1992). Examining large databases: a chemometric approach using principal component analysis. Marine Chemistry, 39 (1-3), 217–237. doi: https://doi.org/10.1016/0304-4203(92)90103-h

[16] Aversano, G., Parra-Alvarez, J. C., Isaac, B. J., Smith, S. T., Coussement, A., Gicquel, O., Parente, A. (2019). PCA and Kriging for the efficient exploration of consistency regions in Uncertainty Quantification. Proceedings of the Combustion Institute, 37 (4), 4461–4469. doi: https://doi.org/10.1016/j.proci.2018.07.040

[17] Li, Z., Yan, X., Yuan, C., Peng, Z., Li, L. (2011). Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. Mechanical Systems and Signal Processing, 25 (7), 2589–2607. doi: https://doi.org/10.1016/j.ymssp.2011.02.017

[18] D. Rosković, A., Grbić, R., Slišković (2011). Fault tolerant system in a process measurement system based on the pca method. MIPRO, 2011 Proceedings of the 34th International Convention, 1646–1651.

[19] Harrou, F., Nounou, M., Nounou, H. (2013). A statistical fault detection strategy using PCA based EWMA control schemes. 2013 9th Asian Control Conference (ASCC). doi: https://doi.org/10.1109/ascc.2013.6606311

[20] Ding, S., Zhang, P., Ding, E., Naik, A., Deng, P., Gui, W. (2010). On the application of PCA technique to fault diagnosis. Tsinghua Science and Technology, 15 (2), 138–144. doi: https://doi.org/10.1016/s1007-0214(10)70043-2

[21] Yin, S., Steven, X. D., Naik, A., Deng, P., Haghani, A. (2010). On PCA-based fault diagnosis techniques. 2010 Conference on Control and Fault-Tolerant Systems (SysTol). doi: https://doi.org/10.1109/systol.2010.5676031

[22] Tong, C., Lan, T., Shi, X. (2017). Fault detection and diagnosis of dynamic processes using weighted dynamic decentralized PCA approach. Chemometrics and Intelligent Laboratory Systems, 161, 34–42. doi: https://doi.org/10.1016/j.chemolab.2016.11.015

[23] Hu, Z., Chen, Z., Gui, W., Jiang, B. (2014). Adaptive PCA based fault diagnosis scheme in imperial smelting process. ISA Transactions, 53 (5), 1446–1455. doi: https://doi.org/10.1016/j.isatra.2013.12.018

[24] Huang, Y., Shen, L., Liu, H. (2019). Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. Journal of Cleaner Production, 209, 415–423. doi: https://doi.org/10.1016/j.jclepro.2018.10.128

[25] Miller, P., Swanson, R. E., Heckler, C. E. (1998). Contribution plots: A missing link in multivariate quality control. Applied mathematics and computer science, 8 (4), 775–792.

[26] Oliveira, J. C. M., Pontes, K. V., Sartori, I., Embiruçu, M. (2017). Fault Detection and Diagnosis in dynamic systems using Weightless Neural Networks. Expert Systems with Applications, 84, 200–219. doi: https://doi.org/10.1016/j.eswa.2017.05.020

[27] Mårtensson, J., Hjalmarsson, H. (2009). Variance-error quantification for identified poles and zeros. Automatica, 45 (11), 2512–2525. doi: https://doi.org/10.1016/j.automatica.2009.08.001

[28] Wu, X. (2015). Study on mean-standard deviation shortest path problem in stochastic and time-dependent networks: A stochastic dominance based approach. Transportation Research Part B: Methodological, 80, 275–290. doi: https://doi.org/10.1016/j.trb.2015.07.009

[29] Boutellaa, E., Kerdjidj, O., Ghanem, K. (2019). Covariance matrix based fall detection from multiple wearable sensors. Journal of Biomedical Informatics, 94, 103189. doi: https://doi.org/10.1016/j.jbi.2019.103189

[30] Yang, H., Li, S., Li, K. (2012). Order estimation of multivariable ill-conditioned processes based on PCA method. Journal of Process Control, 22 (7), 1397–1403. doi: https://doi.org/10.1016/j.jprocont.2012.06.013

[31] Zumoffen, D. (2008). Desarrollo de Sistemas de Diagnóstico de Fallas Integrado al Diseño de Control Tolerante a Fallas en Procesos Químicos.

[32] Lane, S., Martin, E. B., Morris, A. J., Gower, P. (2003). Application of exponentially weighted principal component analysis for the monitoring of a polymer film manufacturing process. Transactions of the Institute of Measurement and Control, 25 (1), 17–35. doi: https://doi.org/10.1191/0142331203tm071oa

[33] Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., Yin, K. (2003). A review of process fault detection and diagnosis Part III: Process history based methods. Computers and Chemical Engineering, 27, 327–346.

Received date 22.08.2022
Accepted date 30.10.2022
Published date 29.11.2022

How to cite: *Cardenas, Y., Carrillo, G., Alviz, A., Alviz, A., Portnoy, I., Fajardo, J., Ocampo, E., Da-Costa, E. (2022). Application of a PCA-based fault detection and diagnosis method in a power generation system with a 2 MW natural gas engine. EUREKA: Physics and Engineering, 6, 84–98. doi: http://doi.org/10.21303/2461-4262.2022.002701*