



Diagnosis of defects by principal component analysis of a gas turbine

Fenghour Nadir¹  · Hadjadj Elias¹ · Bouakkaz Messaoud¹Received: 27 September 2019 / Accepted: 18 April 2020 / Published online: 28 April 2020
© Springer Nature Switzerland AG 2020

Abstract

This study examines the application of the Principal Component Analysis (PCA) technique to detect the failures in complex industrial processes such as gas turbines used for electric power generation. The early detection of failures in such complex processes is indeed paramount to prevent product deterioration, performance degradation, significant property damage and human health. We identified the PCA model by determining the optimal number of principal components retained in the PCA model, then we validated the PCA model by checking the evolution of measurements and estimated the two variables X2 and X8. Thereafter, the evolution of three detection indexes is illustrated highlighting that the filtered SPE index is the best suited one for our installation, and finally, we checked the efficiency of the linear PCA method from the filtered SPE detection index using real data of defects that may occur within the gas turbine. The results obtained will aid to confirm the performance of the linear PCA method in the field of early failure detection. Thus, the PCA method appears as an efficacious tool to monitor and diagnose complex installations.

Keywords Process monitoring · Fault detection · Linear principal component · Electric power production process

1 Introduction

In all industrial systems, breakdowns cause considerable economic losses. It is therefore essential to implement monitoring and diagnostic systems to avoid unexpected shutdowns, increase reliability, and ensure the safety of underlying systems. In industrial diagnosis, the detection process involves merely the detection of events that affect the evolution of the systems, and the assessment consists in comparing the actual functioning systems with those under the assumption of normal operations. Originally, the diagnosis was limited to high-risk industrial applications such as nuclear or aviation domains and advanced industries such as armament and aerospace [1–5]. Over the past three decades, the diagnosis has attracted much attention both in the industrial and in the scientific research world.

In the field of diagnostics, several methods based on the concept of redundancy of information have been developed. In previous works of other researchers, the

PCA, PLS, kernel PCA (kPCA) and kernel PLS (kPLS)-based generalized likelihood fault detection techniques have been developed, namely PCA, PLS, kPCA and kPLS have been implemented as a modeling framework for fault detection [6–9]. Their principle is generally based on a consistency test between an observed behavior of the process provided by sensors and an expected behavior provided by a mathematical representation of the process. Analytical redundancy methods, therefore, require a model of the system to be monitored. This model includes several parameters whose values are assumed to be known during normal operation. The comparison between the actual behavior of systems and their expected behaviors given by well-developed models provides a quantity, called a residue, which will be used to determine whether the system is in a failed state or not.

Multi-variable statistical methods are the most effective for treating the generation of residues. Among them, methods based on Principal Component Analysis (PCA).

✉ Fenghour Nadir, fen.nadir@gmail.com; Hadjadj Elias, hadjadjelias@yahoo.fr; Bouakkaz Messaoud, messaoud.bouakkaz@univ-annaba.org | ¹Department of Electromechanical, Badji Mokhtar University, Annaba, Algeria.



These will be effective for highlighting significant linear correlations between the variables of processes without explicitly formulating the models of their underlying systems. Thus, all the correlations between the different variables are taken into account in the PCA models.

In this paper, we study PCA models for the diagnosis (detection equipment or subassembly fault by precision) of the highly complicated system of a gas turbine [10], where the detection and localization of the fault has become a challenge requiring a profound diagnosis with specific equipment and a considerable time to determine the origin of the issue. However, with the method proposed (PCA), the issue detection will be faster and more efficient. Thus, the objective of this study is to aid in improving the diagnosis of a fault with a gas turbine by choosing the most efficient detection index for this installation, and then confirm the performance of the PCA method in this field of fault detection by the use of real data of a fault that already affected the machine during its lifetime. The plan of this article is as follows. In the first step of this study, we present the fundamental principles of the linear PCA. Secondly, we look at the existing methods of defect detection. Finally, we apply linear PCA to fault detection in an electrical energy generation process (gas turbine).

2 Principal component analysis

Principal Component Analysis (PCA) is a part of the group of multidimensional descriptive methods called factorial methods. These methods, which appeared in the early 1930s, were mainly developed in France in the 1960s, in particular by Jean-Paul Benzécri, who made extensive use of geometric aspects and graphic representations.

It has been adapted for the detection and isolation of defects that may occur at the system level and categorized as a multivariate data analysis technique [11, 12].

This technique is based solely on the use and analysis of system input and output measurements. A data matrix is then constructed from these measurements. By decomposing the data matrix into singular values, the PCA divides this matrix into two distinct parts, one containing the dominant singular values, representing the relevant data, and the other one including the rest of the singular values assumed to be negligible and representing the noises [13].

Several studies have recently been published in the literature dealing with the use of the PCA in the field of defect detection [14–21].

Figure 1 represents the PCA algorithm illustrating all its main steps.

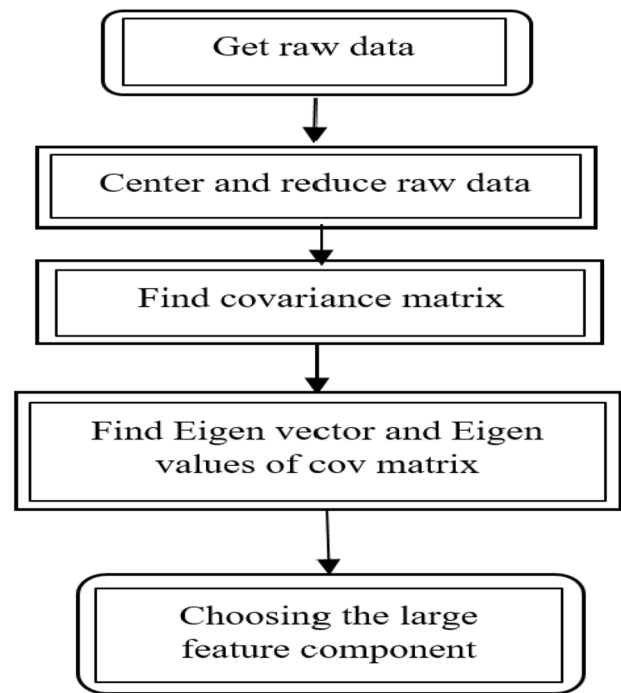


Fig. 1 PCA algorithm

To find a model based on the linear PCA, a database containing variables is required to track a set of measurements performed on the normal operation of the systems. Generally, the procedure for identifying the model involves, after the standardization of the data matrix, the estimation of the parameters of the model, the selection of a fixed structure and the validation of the model.

The basic idea of the PCA is to reduce the size of the data matrix. This reduction will only be possible if the initial variables are not independent and have nonzero correlation coefficients. These initial variables are transformed into new variables called main components.

The reconstruction error variance approach was chosen to determine the number of the key components to be retained. The Recon Error Variance (VNR) can be calculated as follows:

$$u_i = \text{var}\{\varepsilon^T(x - x_i)\} = \text{var}\{f_i\} = \frac{\tilde{\varepsilon}_i^T \Sigma \tilde{\varepsilon}_i}{(\tilde{\varepsilon}_i^T \tilde{\varepsilon}_i)^2} \tag{1}$$

$$u_i = \frac{\tilde{\varepsilon}_i^T \Sigma \tilde{\varepsilon}_i}{(\tilde{\varepsilon}_i^T \tilde{\varepsilon}_i)^2} = \frac{\|\tilde{\varepsilon}_i^0\|_{\Sigma}^2}{\|\tilde{\varepsilon}_i\|^2} \tag{2}$$

where f_i is an estimate of the magnitude of the defect f which measures the displacement in the direction ε_i , with Σ the correlation matrix and u_i , the variance of the

reconstruction error estimating x^* using x_j . Thus, the PCA method is employed as a modeling technique of the relationships between the different variables of the process. The estimation of the parameters of PCAL model is performed by estimating the eigenvalues and eigenvectors of the data matrix of correlations. However, for the determination of the model structure, it is necessary to determine the number of components to be retained in the model (number of eigenvectors).

3 Detection by the PCA method

PCA is defined as a linear transformation of the original correlated variables into a set of uncorrelated variables. However, the original variables can be represented by a small number of the main components due to the analytical redundancy between the variables, and therefore, instead of analyzing all the variables, the PCA analyzes only these components.

To detect such a change in the correlations between variables, PCA involves several detection indices that are used for the detection of abnormal PCA operations: Hotelling Statistics T^2 and the squared prediction error SPE.

3.1 SPE statistic (Q-statistic)

A typical statistic for detecting these unnatural conditions is the SPE statistic, also called Q (Squared Prediction Error) which is given by the equation:

$$SPE(k) = e^T(k)e(k) \quad (3)$$

To detect the presence of an anomaly, the SPE must check the following condition:

$$SPE(k) > \delta_\alpha^2 \quad (4)$$

or δ^2 is the threshold of confidence, Jackson and Mudholkar [19] that develop the expression of the confidence threshold for the SPE.

3.2 SPE filtered statistic

We have seen that the use of different detection tests can be affected by modelling errors that are capable of generating false alarms. The EWMA filter (exponentially weighted moving average) [22] is used to improve the detection for example, in the case of the SPE by filtering the residues to reduce the rate of false alarms. The general expression of this filter is given by;

$$\bar{e}(k) = (I - \beta)(\bar{e}(k - 1) + \beta e(k)) \quad (5)$$

where β is a diagonal matrix whose elements are the forgetting factors for the residues, I is the identity matrix.

The new SPE applied on the filtered residues noted by \overline{SPE} is given by:

$$\overline{SPE}(k) = \|\bar{e}(k)\|^2 \quad (6)$$

To detect the presence of an anomaly, the \overline{SPE} must verify the following condition:

$$\overline{SPE} > \delta_\alpha^2 \quad (7)$$

or

$$\delta_\alpha^2 = \frac{\gamma}{2 - \gamma} \delta_\alpha^2 \quad (8)$$

where γ is a factor of forgetfulness.

3.3 Hotelling T^2 Statistic

The T^2 statistic can be applied, on the first principal components [23]. Thus one obtains:

$$T^2(k) = \hat{t}^T(k) \Lambda_\ell^{-1} \hat{t}(k) = \sum_{i=1}^{\ell} \frac{t_i^2}{\lambda_i^2} \quad (9)$$

$$\Lambda_\ell = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_\ell) \quad (10)$$

Yet (7) is a diagonal matrix containing the ℓ largest eigenvalues of the correlation matrix, the statistic T^2 must check the following condition:

$$T^2(k) > \chi_{\ell, \alpha}^2 \quad (11)$$

where $\chi_{\ell, \alpha}^2$ the upper limit for a confidence level α .

We observed that the use of different detection tests can be affected by the modeling errors that could generate false alarms. The EWMA filter (exponentially weighted moving average) is then used to improve the detection.

4 Application of PCA on the gas turbine unit

To verify and illustrate the effectiveness of the PCA technique for defect detection, the data provided by a real process will be used. This part is dedicated to the application of the Linear Analysis of Main Components for fault detection in an industrial plant (single-cycle power plant gas turbine).

The group, as designed for most installations, consists of a single-cycle "single shaft" gas turbine designed for continuous operation and also to drive an alternator. The combustion of an air-fuel mixture produces the power

required to drive the compressor shaft, some auxiliaries and mainly the alternator.

Figure 2 represents an image of the HMI (Human Machine Interface) depicting the state of good functioning of the turbine for which we based to perform the modeling from the ACP, and moreover some essential variables (parameters) used in our study such as CPD, CDT, FTG, FPG2 and CSTGV (acronym description in Table 1) are put into use.

4.1 The variables used to construct the linear PCA model

To build the PCA model, we chose 36 essential sensors. Measurements taken during normal gas turbine operation are shown in Table 1. Data are collected from a Distributed Control System (DCS) and sampled at a rate of 1 min with 1200 observations.

4.2 Determination of the optimal number of principal components

The principle of reconstruction represents an elimination of the effect of a default. In other words, this approach estimates the vector of amplitudes of such a defect.

We can determine the number of principal components to be retained by minimizing the reconstruction error (see Eq. 1). That is to say, we can find the number of principal components allowing to achieve the best reconstruction of the variables where u_i must necessarily have a minimum in the interval $[1, m]$.

In order to obtain an optimal number of main components used in the PCA model, we chose a reconstruction approach. After using this reconstruction method (structure of the PCA model). Figure 3 shows the evolution of

the reconstruction variance as a function of the number of main components. This result indicates that the seventh component corresponds to the minimum value of the VNR. Thereupon, the first seven main components are making it possible to achieve the best reconstruction of the model PCA.

4.3 Determination with the PCA evolution of the variables and their models

Figures 4 and 5 show, respectively, the measurements and their estimates for the inlet pressure of the compressor and its compression ratio using the PCA model already achieved.

4.4 The evolution of detection index case without defect

Figures 6, 7, and 8 illustrates the evolution of the three statistical tests; SPE, SPE filtered and T^2 , respectively. Data representative of normal operations are adapted to select detection thresholds for each test (red lines).

4.5 Fault detection with linear PCA

To test and verify the efficiency of the PCA model obtained for defect detection, we collected a database with a real defect to assign the gas turbine.

On the basis of the PCA model already obtained, Fig. 9 shows the evolution of the SPE filtered index detection in the presence of the defect:

Gas turbine trip by low gas pressure (fuel).

4.6 Discussion and interpretation of results

In Fig. 3, we obtained seven main components which represent the thirty-six initial variables given by the VNR reconstruction approach. This shows the importance of the PCA method in reducing the size of the initial data without losing their values.

In Figs. 4 and 5, we validated the PCA model obtained by comparing the evolution of the measured values of the two variables X2 and X8 with their modelled values, and we noted an impressive homogeneity.

By comparing Figs. 6, 7, and 8, the absence of modelling errors in the SPE filtered index is noticed, which signifies a better detection quality relative to the T^2 index and to the SPE index. Therefore, the SPE index is not an effective test for defect detection as it is a global test that accumulates all the error residues, and the T^2 index is not a reliable index of detection as the conditions of use are rarely checked [20]. Based on the results obtained, it is preferable

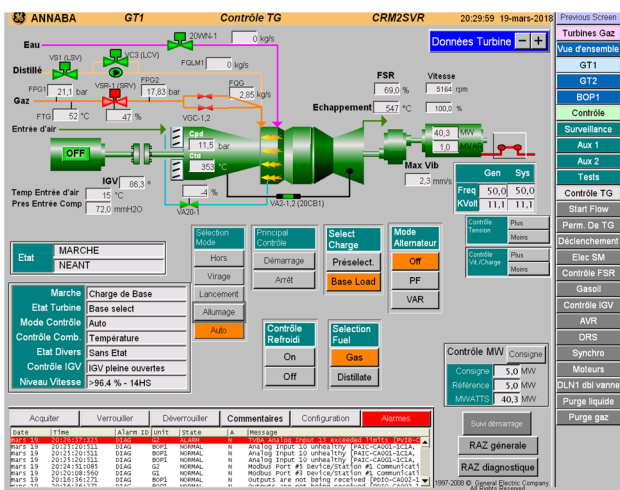


Fig. 2 Principle of operation of a gas turbine

Table 1 Description of the variables of the gas turbine

AAT	1	Temperature of the air main atomizations
AFPDB	2	Compressor inlet pressure
RFA	3	Mass flow air intake compressor
BBMAX	4	Maximum vibration turbo-alternator
BBMAX GT	5	Maximum vibration gas turbine
CMHUM	6	Humidity
CPD	7	Compressor outlet pressure
CPR	8	Compressor compression ratio
CSGV	9	IGV angle in degrees
CDT	10	Compressor discharge temperature
CTIM	11	Compressor inlet temperature
DF	12	Alternator frequency
CFO	13	Power factor
DV	14	Alternator voltage in %
PSUs	15	Reactive power alternator
DVX	16	Alternator voltage
DWATT	17	Active power alternator
FPG2	18	Inter-valve fuel pressure
PCF	19	Gas fuel mass flow rate
FTG	20	Gas fuel temperature
LTB1D1	21	Temperature purge lube oil bearing 1 turbine
LTB2D1	22	Temperature purge lubrication oil bearing 2 turbine
LTOT	23	Lubricating oil tank temperature
LTTH	24	Lubricating oil collector temperature
PAT	25	Fuel purge air temperature
GAP2	26	Collector pressure lubricating oil
TNH	27	Tree rotation speed
TNRCOR	28	Velocity shaft rotational %
TTWS1AO1	29	Inter-wheel temperature first stage Rear intake turbine
TTWS1FI1	30	Inter-wheel temperature first stage before intake turbine
TTWS2FI1	31	Inter-wheel temperature second stage before exhaust turbine
TTWS3AO1	32	Inter-wheel temperature third stage rear exhaust turbine
TTXM	33	Mid turbine exhaust temperature
TTXP1	34	First difference real control combustion
WTAD	35	Temperature evacuation cooling water
WTTL1	36	Temperature cooling water leg TG

to utilize the filtered SPE detection index to detect the gas turbine anomalies with higher efficiency.

To further illustrate this study, we made a comparison regarding other articles which use the main component analysis for defect detection. Namely, the work presented in [24] confronts the detection index SPE and T^2 from the data of a chemical reactor. The results obtained confirm that the detection index SPE is most relevant than the detection index T^2 for the detection of defects of a chemical reactor, which is in agreement with the present study comparing three detection indices SPE, SPE filtered and T^2 , which demonstrates that the filtered SPE index is the most suited for detecting faults of a gas turbine. Similarly, another work [25] discloses that the filtered detection

index D^2 is the most significant for detecting faults in an air quality control system compared to other detection indices such as the filtered SPE and the S.

We injected the data of a real defect (insufficient gas pressure which causes the gas turbine to trip) into the PCA model from the SPE filtered index and this confirms the presence of a fault indicated by a peak clearly exceeding the detection limit as depicted in Fig. 9.

On the basis of the above data, it should be noted that the detection rate (black curve) shows a large excess regarding the detection limit (red line) calculated by the PCA model for proper operation. Therefore, it can be concluded that we have an abnormal operating condition of the turbine which is ascertain. This confirms the efficiency

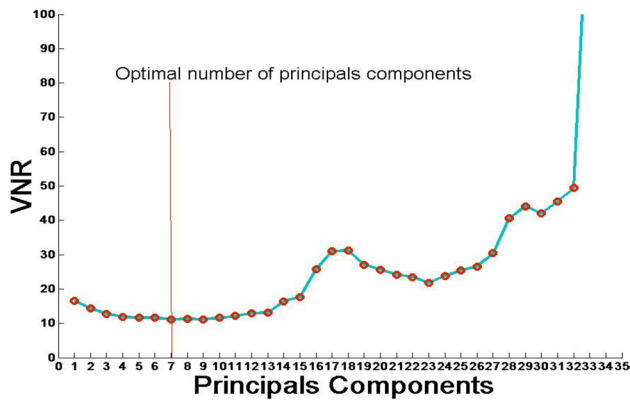


Fig. 3 Evolution of non-reconstruct variance VNR as a function of number of components

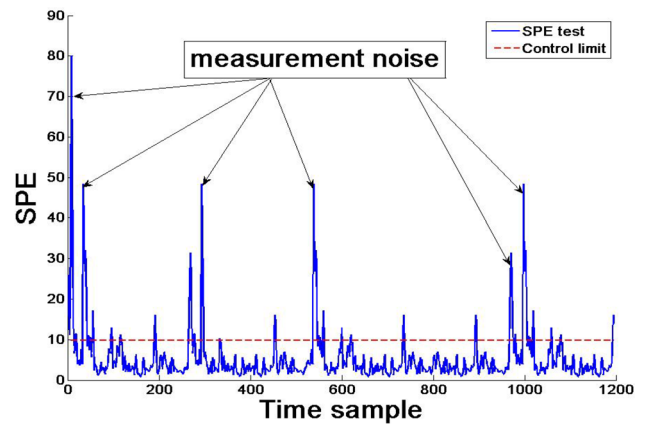


Fig. 6 Evolution of SPE “faultless case”

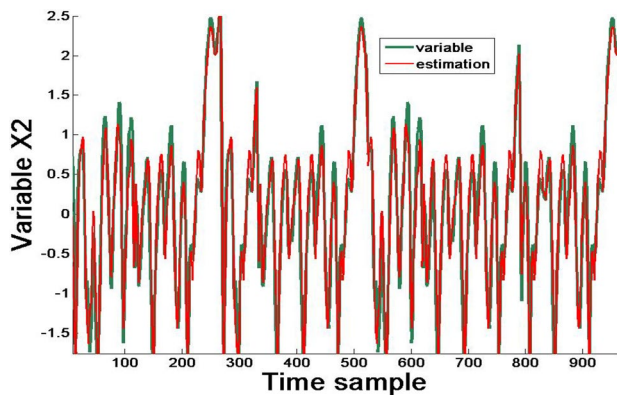


Fig. 4 Evolution of X2 and its estimate

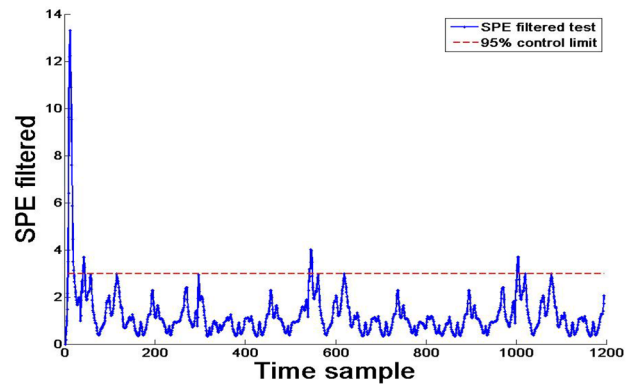


Fig. 7 Evolution of SPE filtered “faultless case”

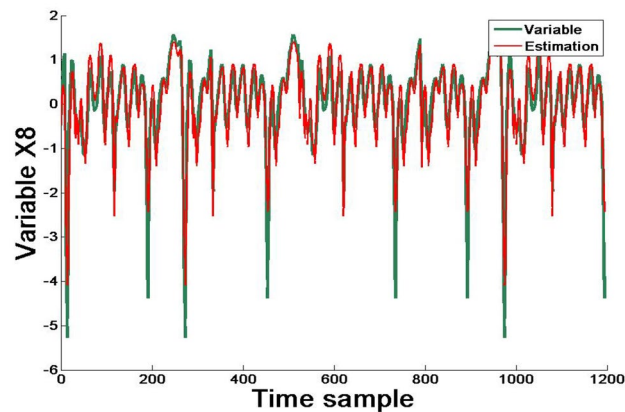


Fig. 5 Evolution of X8 and its estimate

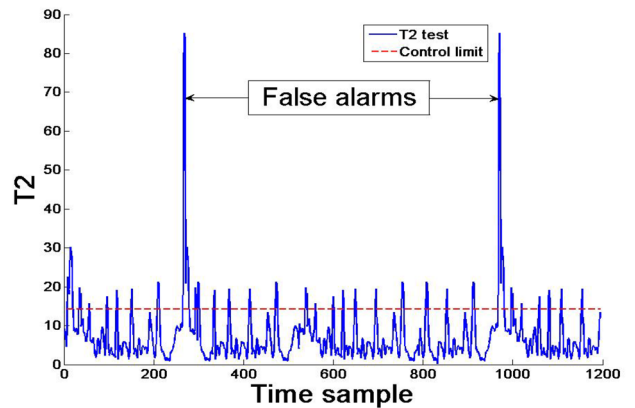


Fig. 8 Evolution of T^2 “faultless case”

and the reliability of the PCA method with SPE filtered detection index in the field of defect detection, noticing that the data used correspond to the defect that already actually produces in the life of the machine (gas turbine).

To display the advantage of using this method in complex processes such as a gas turbine, when an identification trigger of the failure trigger is difficult, but the use of the main component analysis with the SPE filtered

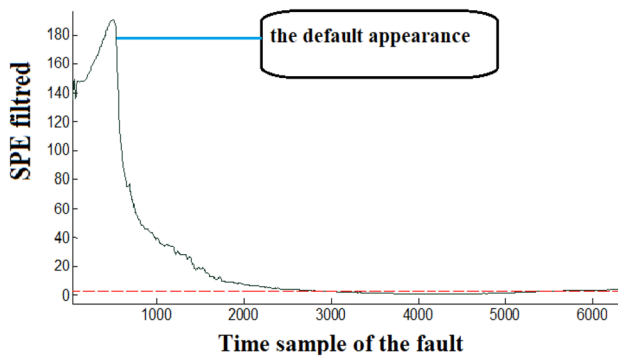


Fig. 9 Detection defect lowering gas pressure by PCA

detection index facilitates the identification of the defective element. This gives a huge advantage in terms of minimizing the time of diagnosis of a defect and thus the time of repair, and considerably reduces the cost of the intervention.

5 Conclusion

The work presented in this article focuses on the diagnosis of defects through the analysis of major components of the PCA, which is used as a technique for modelling the relationships between different process variables to obtain a compact representation of these variables. Therefore, to obtain the PCA model, the parameters of the PCA model are first acquired by estimating the own values and vectors of the data correlation matrix, and then by determining the structure of the model, and finally by finding the number of components to keep in the model (number of clean vectors).

Once the model has been identified, the fault detection procedure can be performed by generating the fault indicators (residuals) comparing the observed behavior of the process given by the measured variables with the expected behavior supplied by the PCA model. Most detection methods based on the PCA approach employ the SPE (squared error statistics) with the use of the EWMA filter to improve the detection and the T^2 statistic.

Finally, we presented the application of PCA linear models for fault detection in a gas turbine electric power generation process. In this study, we compared three detection indices (SPE, filtered SPE and T^2). The results show that the filtered SPE index has the advantage of being less sensitive to the modeling errors and more sensitive to the defects than the SPE index. Lastly, to complete this work, we used the data of the actual defects affecting the machine, the purpose of which was to verify the efficiency and the performance of the linear PCA model

of the detected defects. The results obtained by the PCA model validate the presence of defects, which confirm the performance and efficiency of the PCAL method in the field of defect detection of a gas turbine. The benefit of this method is mainly to reduce the diagnostic time. In the context of the perspectives, we propose to extend it to the localization of the defects. Another prospect would consist in studying the nonlinear PCA which extract linear and nonlinear relations with the dynamic model. With regard to the application being processed, some proposals are made to improve the quality of the PCA model, a proposal consists in using different PCA models to choose and identify the best fit to the gas turbine, and since the different variables exhibit a certain behavior cyclically, they can be modeled by a PCA using batches of data where each batch corresponds to a period of operation. Another proposal is to exploit the multi-model form, and for modeling using multi-PCA as an extension to the non-linear model.

Compliance with ethical standards

Conflict of interest They have no conflict of interest.

References

1. Ashish S, Dale E (2006) Evaluation of pattern matching method for the Tennessee Eastman challenge process. *J Process Control* 16:601–613
2. Choi SW, Park JH, In-Beum L (2004) Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Comput Chem Eng* 28:1377–1387
3. Jolliffe IT (1986) *Principal component analysis*. Springer, Berlin
4. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* 2:37–52
5. Kresta JV, MacGregor JF, Marlin TE (1991) Multivariate statistical monitoring of process operating performance. *Can J Chem Eng* 69:37–47
6. Botre C, Mansouri M, Nounou M, Karim MN (2016) Kernel PLS based GLRT method for fault detection of chemical processes. *J Loss Prev Process Ind* 43:212–224
7. Mansouri M, Nounou M, Nounou H, Nazmul K (2016) Kernel PCA-based GLRT for nonlinear fault detection of chemical processes. *J Loss Prev Process Ind* 26(1):129–139
8. Mansouri M, Sheriff MZ, Baklouti R, Nounou M, Nounou H, Hamida AB, Karim N (2016) Statistical fault detection of chemical process-comparative studies. *J Chem Eng Process Technol* 7(1):282–291
9. Sheriff MZ, Mansouri M, Karim MN, Nounou H, Nounou M (2017) Fault detection using multiscale PCA-based moving window GLTR. *J. Process Control* 54:47–64
10. Navi M, Davoodi MR, Meskin N (2015) Sensor fault detection and isolation of an industrial gas turbine using partial kernel PCA. *IFAC-Papers*
11. Quin SJ, Dunia R (2000) Determining the number of principal components for best reconstruction. *J Process Control* 10:254–2050

12. Jaffel I, Taouali O, Harkat MF, Messaoud H (2015) Online process monitoring using a new PCMD index. *Int J Adv Manuf Technol* 80:947–957
13. Ricker NL (1994) Optimal steady-state operation of the Tennessee Eastman challenge process. *Comput Chem Eng* 19:949–959
14. Tamura M, Tsujita S (2007) Une étude sur le nombre de composants principaux et de détection des défauts utilisant l'ACP. *Comput Chem Eng* 31:1035–1046
15. Weihua L, Sirish S (2002) Structured residual vector-based approach to sensor fault detection and isolation. *J Process Control* 12:147–160
16. Westerhuis JA, Gurden SP (2000) Generalized contribution plots in multivariate statistical process monitoring of residual in MSPC. *J Intell Lab Syst* 51:95–114
17. Wuse BM (1991) Adapting multivariate analysis for monitoring and modeling of dynamic systems. Ph.D. Dissertation. University of Washington, Seattle
18. Zwingelstein G (1995) Diagnostic des défaillances, théorie et pratique pour les systèmes industriels. livre. Edition Hermès
19. Jakson JE, Mudholkar GS (1979) Control procedures for residuals associated with principal component analysis. *Technometrics* 21(3):341–349
20. Chaouch H, Najeh T, Nabli L (2016) Compression de données de processus multivariable et isolation des défauts à l'aide de l'APC en ondelettes et d'un algorithme génétique. *Int J Adv Manuf Technol* 91:869–878
21. Garoudja E, Harrou F, Sun Y, Kara K, Chouder A, Silvestre S (2017) Statistical fault detection in photovoltaic systems. *Sol Energy* 150:485–499
22. Kourti T, Macgregor JF (1996) Recent developments in multivariate SPC methods for monitoring a diagnosing process and product performance. *J Qual Technol* 28(4):409–428
23. Harket MF, Mourot G, Ragot J (2001) Sensor failure detection and isolation of air quality monitoring network. In: 4th international conference on acoustical and vibratory surveillance methods and diagnostic techniques, Compiègne, France
24. Taouali O, Jaffel I, Lahdhiri H, Harkat MF (2015) New fault detection method based on reduced kernel principal component analysis. *Int J Adv Manuf Technol*. <https://doi.org/10.1007/s00170-015-8059-1>
25. Harkat MF, Mourot G, Ragot J (2006) An improved PCA scheme for sensor FDI: application to an air quality monitoring network. *J Process Control* 16:625–634

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.