The Journal of Engineering and Exact Sciences - JCEC
ISSN: 2527-1075

Vol. 03 N. 07 (2017) 0920-0932

doi: 10.18540/jcecvl3iss7pp0920-0932
OPEN ACCESS

ALGORITHM FOR DETECTION AND DIAGNOSIS OF INDUSTRIAL FURNACE FAULTS APPLYING SINGULAR VALUE DECOMPOSITION AND GRAPHIC VISUALIZATION

ALGORITMO PARA DETECÇÃO E DIAGNÓSTICO DE FALHAS EM UM FORNO INDUSTRIAL APLICANDO DECOMPOSIÇÃO DOS VALORES SINGULARES E VISUALIZAÇÃO GRÁFICA

A. V. FONSECA[1], R. F. TEÓFILO[1], V. B. SOARES[2], D. de J. SILVA[1]

[1] Universidade Federal de Viçosa, Departamento de Química, Brasil
[2] Universidade Federal do Espírito Santo, Departamento de Engenharia Química, Brasil
E-mail: amandavfonseca12@gmail.com

a r t i c l e i n f o
Article history:
Received 2017-07-06
Accepted 2017-09-14
Available online 2017-11-20

KEYWORDS: Detector Failures; Fault Detection and Diagnosis; Principal Component Analysis; Singular Value Decomposition.

PALAVRAS-CHAVE: Análise de Componentes Principais; Decomposição dos Valores Singulares; Detecção e Diagnóstico de falhas; Falhas em detectores.

ABSTRACT: *The development of an algorithm based on singular value decomposition with graphic visualization to aid in the fault detection and diagnosis was the aim of this work. In order to test the algorithm, real temperature data from an industrial furnace of a natural gas processing unit were used. Nine temperature detectors were monitored over time. In addition to a real fault that caused the shutdown of the equipment, five simulated faults in the furnace detectors were also tested. Results showed that the algorithm was effective and has great potential to assist in the monitoring of industrial processes.*

RESUMO: *O desenvolvimento de um algoritmo baseado na decomposição dos valores singulares com visualização gráfica para auxiliar na detecção e no diagnóstico de falhas foi o objetivo deste trabalho. Para testar o algoritmo foram utilizados dados de temperatura reais de um forno industrial de uma unidade de processamento de gás natural. Nove detectores de temperatura foram monitorados no tempo. Além de uma falha real que gerou a parada do equipamento, foram também testadas cinco falhas simuladas nos detectores do forno. Os resultados mostraram que o algoritmo foi eficaz e possui grande potencial para auxiliar no monitoramento de processos industriais.*

1. INTRODUCTION

Progress in the field of process control has benefited several industrial segments such as petrochemical, cement, steel, among others, mainly in productivity and quality aspects. Advance occurred in a context of improvement of the computational capacity of complex processes, resulting in a higher level of automation of control systems. However, human activity is still present in a very important task within process control: the response to abnormal events. This task encompasses the detection of the abnormality, the correct diagnosis and taking actions to stop the fault (VENKATASUBRAMANIAN *et al.*, 2003a).

Several factors hinder human performance in the abnormality management. Among these are the difficulty of correct diagnosis, due to degradation, parameter drifts and equipment failures; the excess of information associated with a very large number of variables being measured and the measurements themselves which may be insufficient, incomplete or even unreliable due to malfunctions in measuring instruments. The direct consequence of these problems is the erroneous action taking by operators, thus impacting on economic, safety and environmental aspects of a company (VENKATASUBRAMANIAN *et al.*, 2003a).

According to Venkatasubramanian et al. (2003a), industrial statistics in 2003 indicated that 70% of accidents were caused by human error. Accidents at the Eurotunnel (1996), at the Texaco Refinery in Milford Haven (1994) and at the Three Mile Island Nuclear Power Plant (1979) are examples cited in the literature that show the difficulty of taking correct actions in the presence of abnormalities (SOARES, 2016). Although they have great repercussion, such accidents occur less frequently. On the other hand, deviations of smaller proportions are very common, causing great economic, safety, environmental and reliability losses (SOARES *et al.*, 2016). It is estimated that failures represent annually loss of billions of dollars for society (FEITAL, 2011; VENKATASUBRAMANIAN *et al.*, 2003a).

Based on the above, it can be recognized the importance of using technologies that assist human operator in the task of detecting and diagnosing industrial failures. The area of detection and fault diagnosis relies on a large number of methods for this purpose. Some methods rely on modeling, while others use historical process data. The last ones, especially the multivariate, have gained industry attention given the accuracy and speed with which they warn about errors (MACGREGOR and CINAR, 2012). Principal Component Analysis (PCA) is a statistical method that performs data compression and makes visualization in multidimensional space easier. In this way the PCA can be applied in the monitoring of multiple process inputs, reducing the complexity of the systems (TEÓFILO, 2007; SARTORI *et al.*, 2012). PCA is based on the construction of a new reduced set of variables, the so-called principal components, which explain properly predominant trends in multivariate data. With PCA it is possible to reduce random measurement errors (MACGREGOR and CINAR, 2012). There are several algorithms to generate the principal components, being the singular value decomposition (SVD) a commonly used algorithm (TEÓFILO, 2007). Monitoring of principal components is accomplished in two steps. First the principal components model for process variables in normal operating condition is obtained. Afterwards, new observations are monitored through some statistics. Hotelling's T2 statistic is typically employed in the determination of abnormalities by monitoring the variability of principal components (QIN, 2012; MACGREGOR and CINAR, 2012; VENKATASUBRAMANIAN *et al.*, 2003b).

Chen (2010), Villegas *et al.* (2010) and Silva (2008) studies are examples of works that applied PCA and Hotelling's T2 statistic in process monitoring. The identification of abnormalities was performed through the construction of control charts. However, this kind of monitoring has as disadvantage the need of decomposing T2 statistic into its components to reveal the individual contribution of each variable and, consequently, to diagnose fault.

This work aimed to develop an algorithm to aid in the identification and diagnosis of failures based on singular value decomposition with graphic visualization. The algorithm was tested with real data from an industrial furnace of a natural gas processing unit. Besides a real fault event that led the equipment to shut down, five simulated faults in the furnace detectors were also tested.

2. MATERIALS AND METHODS

2.1. Thermal Oil Furnace System

In this paper a thermal oil furnace of a natural gas processing unit was monitored during its operation. The thermal oil furnace supplies the thermal load demanded by fractionation towers, therefore, it is a vital system for the gas processing. The monitored equipment has a vertical configuration with forced draft fans that compress the atmospheric air into the furnace in order to combust gas and draw it. Air flow through the furnace is controlled to maintain an optimum fuel gas / air ratio. The furnace has four (4) pipe passes that enter, run through the inside of the furnace and interconnect into a single outlet stream. Outlet temperature is controlled by regulating fuel gas flow rate to the furnace burners. Each pass has two temperature detectors with high temperature alarms (TAH) configured. These alarms share the same setpoint of 255 ° C.

Figure 1 shows the temperature detectors installed in the thermal oil furnace. Detectors are located at the outlet of the furnace (1 to 8) and at the interconnection header pipe (9).

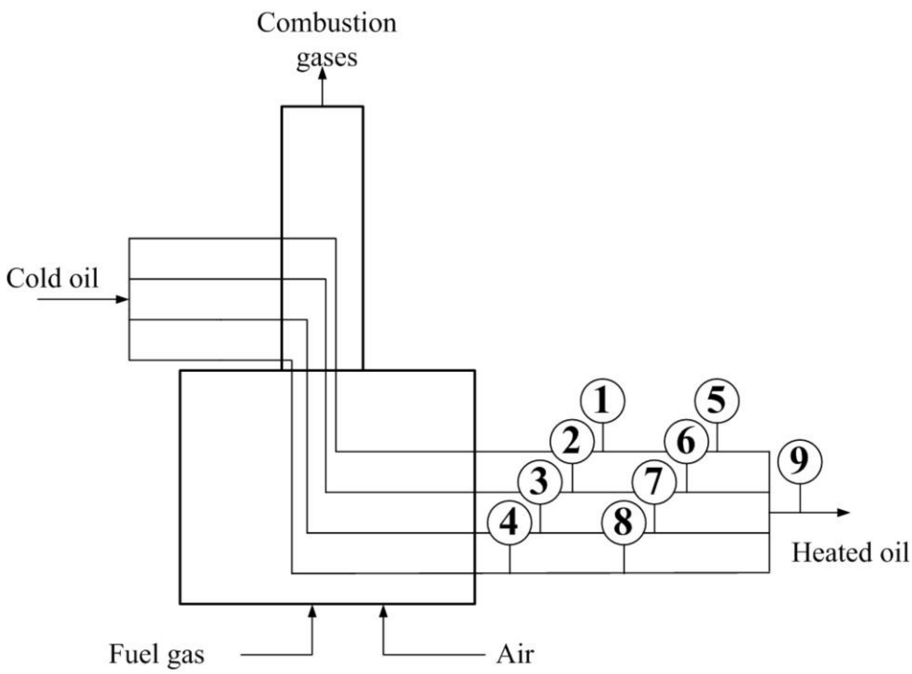

Figure 1 – Simplified diagram of the thermal oil furnace. Temperature detectors located at the furnace outlet (1-8) and at the interconnection header pipe (9).

2.2. Process Monitoring

An algorithm written as .m function was developed to monitor simultaneously the nine temperature detectors installed in the furnace. The algorithm was written in MATLAB version R2015a. In order to monitor the nine detectors simultaneously singular value decomposition and Hotelling's T^2 statistic were applied together. The algorithm was written with the purpose of identifying faults using graphical visualization.

Singular value decomposition (SVD): SVD is a mathematical decomposition used to eliminate multicollinearity of singular matrices. Since the new generated matrices by the SVD are orthogonal and relevant information is placed in descending and complementary order, SVD is used in multivariate statistical analyzes for visualization and resolution of complex linear systems. Equation 1 shows the decomposition of $\mathbf{X}_{np}$ matrix, where n represents the number of observations (samples) and p is the number of variables. In this equation, **U** represents the matrix of the original samples coordinates in the axis of the principal components (scores matrix), **V** is the matrix that contains relationship between the principal components and the original variables set (loadings matrix) and **S** contains in its diagonal the ordered singular values of matrix **X**, which allow to obtain the variance of each principal component (HERNANDEZ-VARGAS *et al.*, 2014; TEÓFILO, 2007; GEMPERLINE, 2006; MILLER e MILLER, 2005; VENKATASUBRAMANIAN *et al.*, 2003b).

$$\mathbf{X} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathrm{T}} \tag{1}$$

Reduction of dimensionality consists of the estimation of the matrix **X** by choosing only k principal components, which contain most of the system variance, namely in the truncation of the matrices **U, S** and **V**. Thus, only significant process information is retained, eliminating noises and redundancies (TEÓFILO, 2007; GEMPERLINE, 2006). According to Venkatasubramanian *et al.* (2003b), the first two or three components in general are sufficient to explain the variability of the original data set.

Hotelling's T^2 statistic: Hotelling's T^2 statistic is widely used for detection of process abnormalities. When applied to PCA models, it monitors the variability of the principal components, indicating if the new observations are in accordance with the model (MACGREGOR and CINAR, 2012; FEITAL, 2011; QIN 2003; VENKATASUBRAMANIAN *et al.*, 2003).

As presented by MacGregor and Cinar (2012), this statistic can be described by Equation 2, where $\mathbf{U}_k$ is the scores matrix of the k principal components and $\mathbf{S}_k$ is the diagonal matrix that contains the square root of the k eigenvalues of the covariance matrix of **X**.

$$\mathbf{t}^2 = \mathbf{U}_k\,\mathbf{S}_k^{-2}\,\mathbf{U}_k^{\mathrm{T}} \tag{2}$$

The upper control limit of the T^2 statistic (T^2_{max}) is defined by Equation 3. For this limit, n is the number of observations to generate the model, F is the percentage of the distribution

F with k and n-k degrees of freedom and tolerance α (MONTGOMERY, 2005).

$$T^2_{max} = \frac{k(n^2-1)}{n(n-k)} F_\alpha(k, n-k) \quad (3)$$

2.3. Normal operating condition

In order to test the algorithm, measurements during 2045 seconds of the temperature detectors were used, with a sampling rate of 5 seconds, following the same reading time of the regulatory control layer, totaling 410 observations. In that condition, thermal oil at approximately 142 ° C was heated to about 215 ° C. Normal operating condition of the thermal oil furnace was guaranteed by the tacit knowledge acquired by the operators over the years.

2.4. Tested fault scenarios

Several problems can occur during the operation of the thermal oil furnace, reducing its efficiency and, in more drastic situations, stopping the equipment with subsequent plant shutdown. Six particular cases were studied in this work: one event of a real fault in the equipment (Case A) and five simulations of faults in the detectors (Cases B to F).

<u>Case A:</u> In the first test, data from an actual fault event in the equipment were used, which caused the furnace to shut down. Temperatures in this situation fell gradually until approximately 160 ° C.

<u>Case B:</u> In the second case, it was tested the increase of all temperatures by 50 ° C, causing the process to operate in an abnormal region.

<u>Case C:</u> In this test, the malfunction of only one of the detectors was simulated. Fault caused the intermittent increase of its temperature in 5º C.

<u>Case D:</u> Simulated malfunction of two detectors was tested. Faults occurred with intermittent variation of the temperature of the detectors in 5º C.

<u>Case E:</u> In this situation, the abnormality tested was composed by both faults tested in cases B and C.

<u>Case F:</u> In the latter case, it was tested a 50 ° C temperature increase of one of the detectors along with an intermittent 5 ° C variation of the same detector, an intermittent variation of 5 ° C in another detector, and a constant sharp increase of 50 ° C in the temperatures of the other detectors.

3. RESULTS AND DISCUSSION

Figure 2 presents in a simplified form the algorithm developed in this work, using MATLAB software. In the routine, the matrix of data in normal operating condition and fault test data are initially concatenated. After that, singular value decomposition is applied to the normal data matrix plus the first observation of the test data set, obtaining the principal components model. Then, vector t2 and the limit for the Hotelling's T2 statistic (T2max) are calculated. If the maximum value of t2 is lower than T2max, graphs referring to the normal operating condition of the system are constructed. Otherwise, graphs for a fault condition are plotted. This procedure is repeated until all observations in the test data set have been incorporated into the principal component model. Normal condition is represented by two plots: one of the scores and another of the loadings of the first principal component as a function of the second one in the form of blue circles. On the score plot, a confidence ellipse based on the Hotelling's T2 statistic for data in normal operating condition is also inserted. Graphs for fault conditions are similar to those of the normal condition, with the difference that the scores are represented by a red "x", indicating the occurrence of faults.

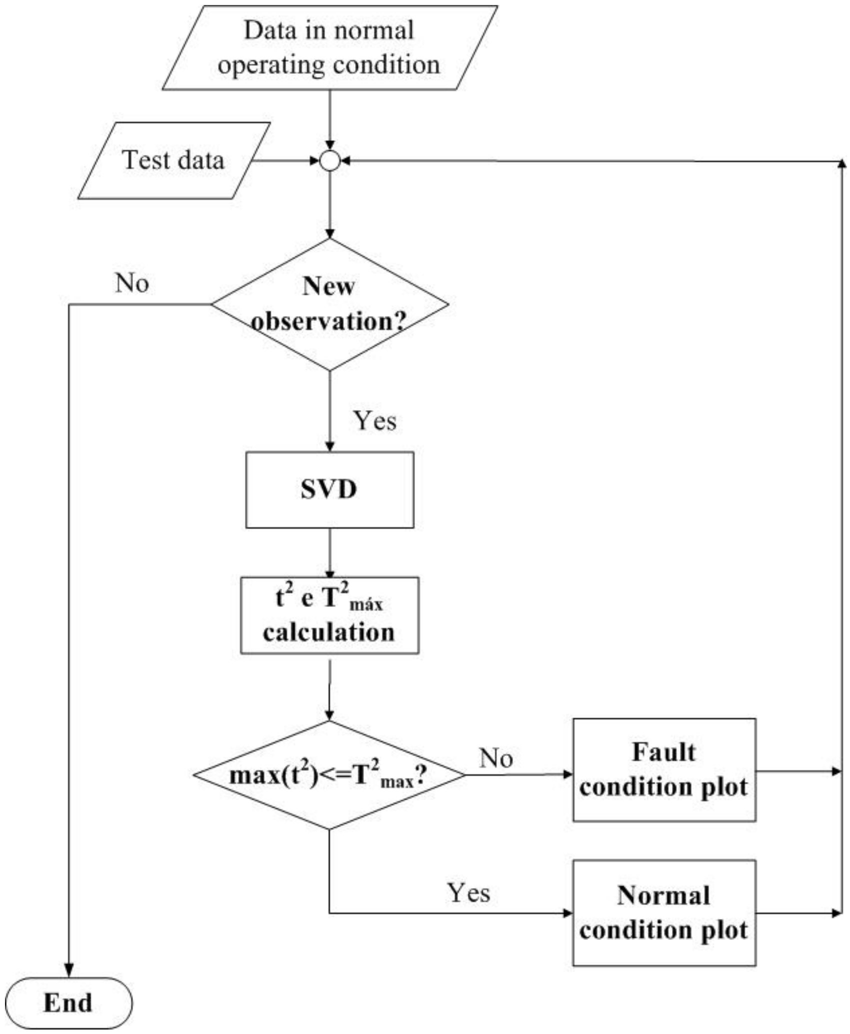

Figure 2 – Schematic flowchart of the developed algorithm.

Figure 3 shows the behavior of the temperature detectors before and after insertion of faults. Measured temperatures up to the time 2046 seconds consist of data in normal operating condition and are common for all fault cases studied. Temperatures after this time are the tested faulty data set. As observed, the faults occur at 2225 seconds for the actual fault (Case A) and at 2115 seconds for the simulated faults (Cases B-F).

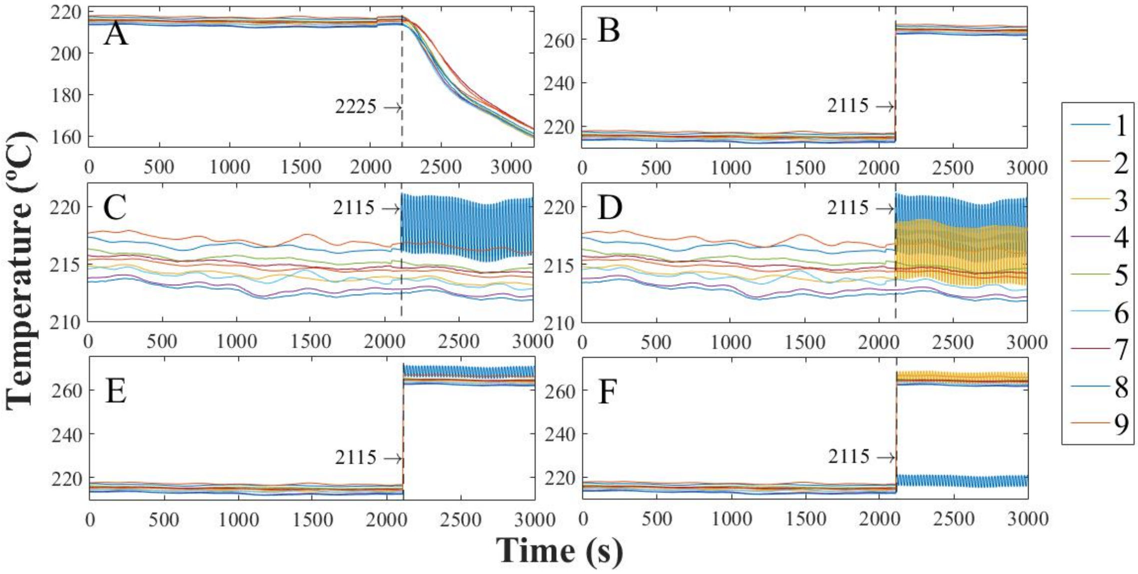

Figure 3 – Temperature detectors behavior for different fault scenarios (A-F).

Figure 4 shows the scores plot on the left, and the loadings plot on the right. The first principal component (PC1) was plotted against the second (PC2) at the time of 2045 seconds, that is, before insertion of the set of faults and is common for all cases studied. In this condition, more than 90% of the system variance can be explained by only these two components. It is possible to observe that all scores are within the confidence ellipse T^2 and that, according to this statistic, they are in a normal operating condition. It is possible to notice that there is a slight proximity between loadings of pairs of detectors (pairs 1 and 5, 2 and 6, 3 and 7 and 8 and 4), which is consistent with the fact that these pairs are located in the same pipes.

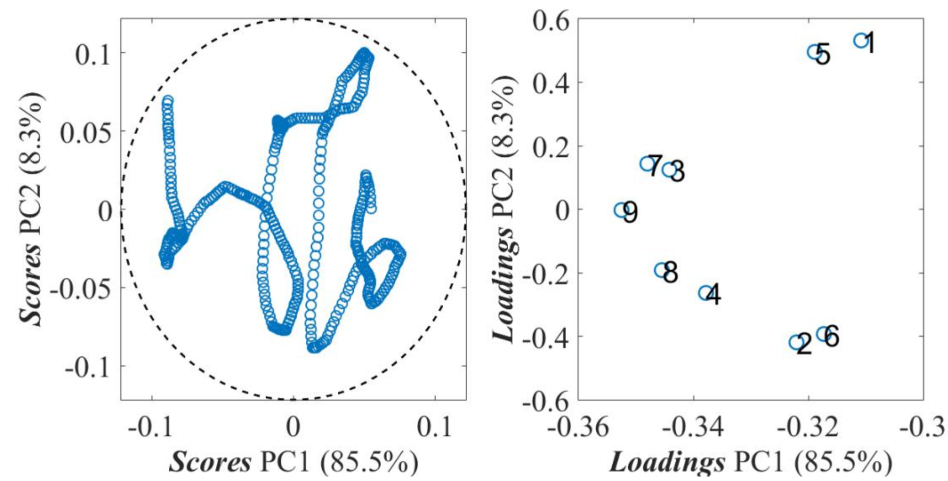

Figura 4 – S*cores* (left) and loadings (right) plots at 2045 seconds, *i.e*, before faulty data set insertion.

Evolution of the first faulty condition tested (Case A) is shown by Figures 5 and 6, at times, in seconds, 2350 (intermediate time after fault insertion) and 3160 (end of fault), respectively (Figure 3A). It is possible to visualize the reduction of the ellipse as the fault occurs. The algorithm in this case was able to identify the failure in time 2265 seconds. This fact was due to sensitivity of the statistics used. Studies of Chen (2010) and Yue and Qin (2001) showed the use of statistics that could reduce time to identify failures. Some of the scores considered normal are presented out of the confidence ellipse due to the influence that the incorporation of the fault data had on the principal components model.

As there is no clear separation between loadings of one or more detectors, it can be concluded that all of them failed, which is in agreement with the proposed scenario.

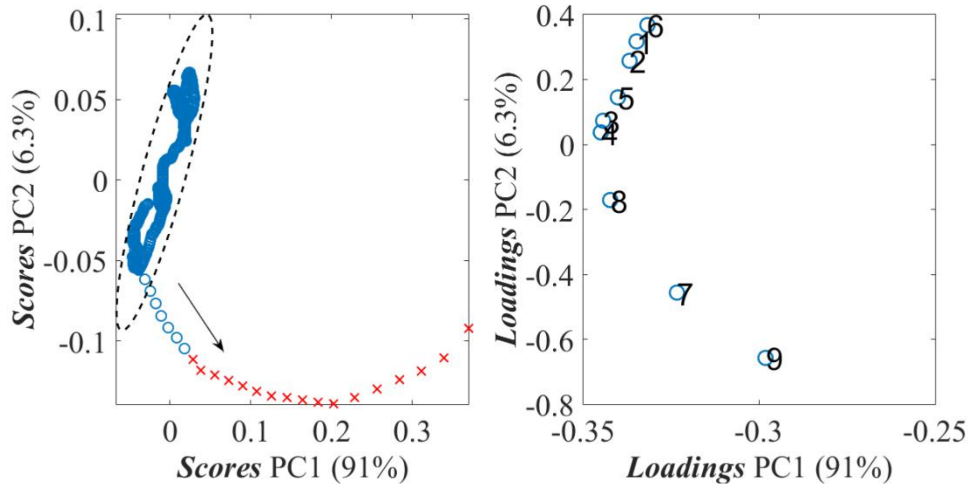

Figure 5 – Scores (left) and loadings (right) plots at 2350 seconds for the real faulty condition (A).

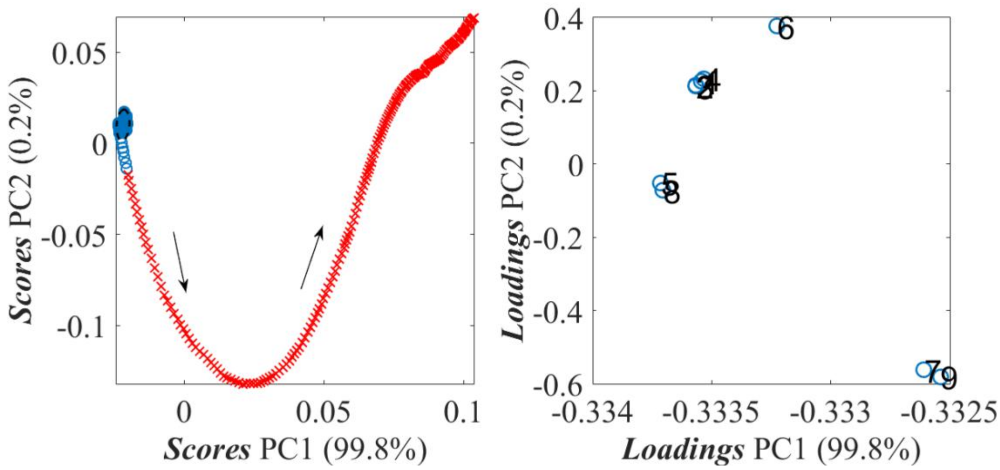

Figure 6 – Scores (left) and loadings (right) plots at 3160 seconds for the real faulty condition (A).

Figure 7 shows the plots obtained for the second fault tested in Case B at 3000 seconds (Figure 3B). In this test, the algorithm was able to identify the fault at the exact moment of its occurrence (2115 seconds). Similarly to case A, no clear separation was observed in the loadings between one or more detectors, indicating that all of them failed in accordance with the behavior of the tested fault.

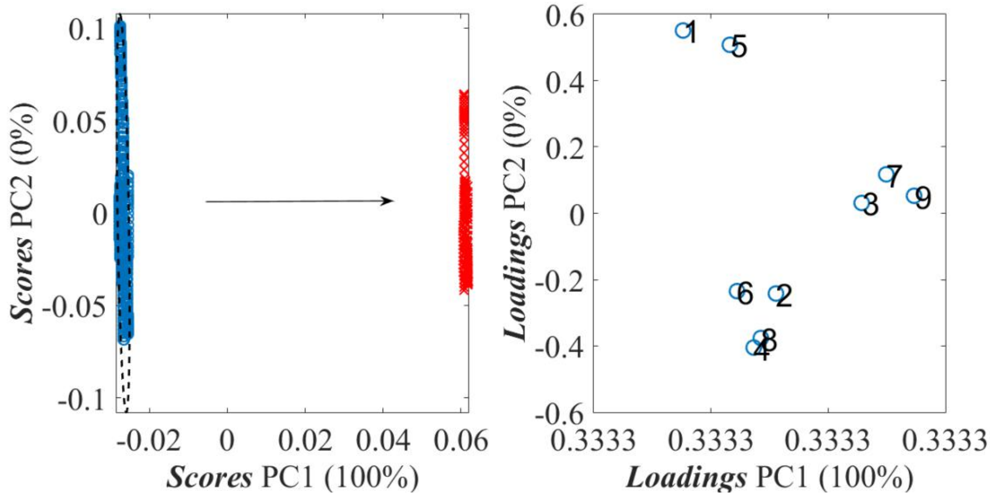

Figure 7 – S*cores* (left) and *loadings* (right) plots at 3000 seconds for the simulated faulty condition B.

Algorithm plots for the tested fault C are shown in Figure 8. The algorithm in this case was efficient in both fault detection and diagnosis. Time of fault detection agreed with the moment of its occurrence. In the scores plot it is observed that part of these is inside the confidence ellipse, due to the return of the detector temperature to normal condition. Since there is a quite clear separation of the loadings of the detector 1 from the other ones, it is presumed that the scores separation was due to a fault in that instrument.

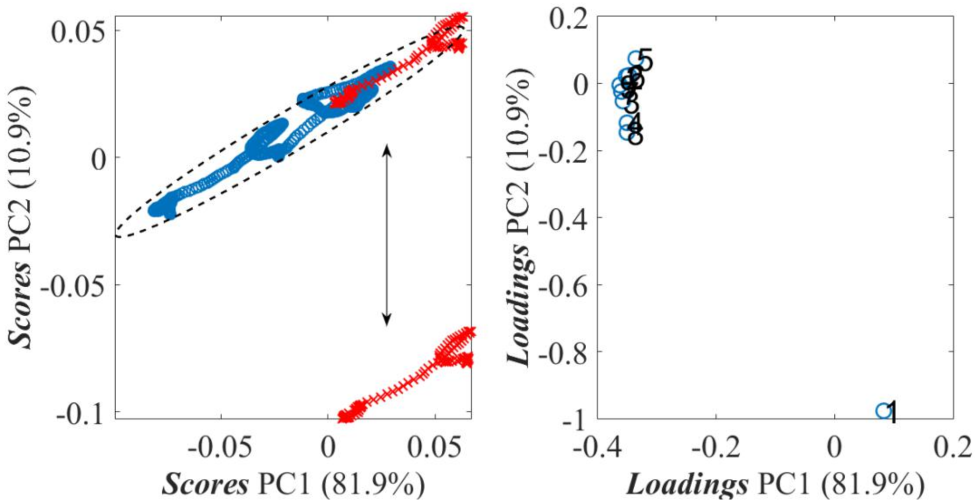

Figure 8 – S*cores* (left) and *loadings* (right) plots at 3000 seconds for the simulated faulty condition C.

Test of the algorithm with the faulty scenario D resulted in Figure 9. A similar result to case C was obtained, in that the algorithm was able to identify the fault at the exact moment of its occurrence and allowed the identification of which detectors caused the abnormalities, *i.e.* detectors 1 and 3, due to the separation of their loadings from the others.

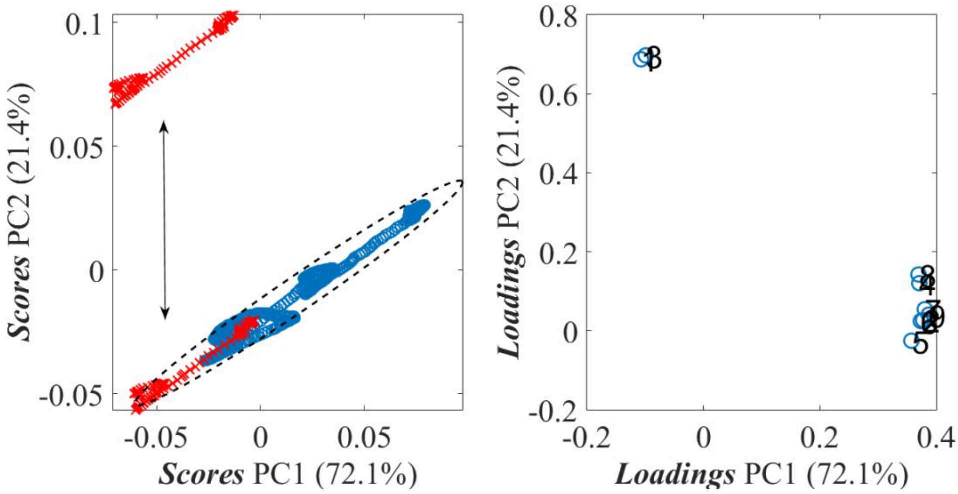

Figure 9 – Scores (left) and loadings (right) plots at 3000 seconds for the simulated faulty condition D.

Case E test with the algorithm resulted in the plots shown by Figure 10. The algorithm in this situation was able to identify faults. It is possible to observe three groupings of scores: the normal data set and two groups corresponding to faults of different nature. The red group in the upper right corner can be explained by the loadings of detector 1, while the red group in the lower right corner was due to the loadings of the other detectors. It is concluded that all detectors have failed, and the fault in the detector 1 has a different nature from the faults in the other detectors, which is in line with what was expected.

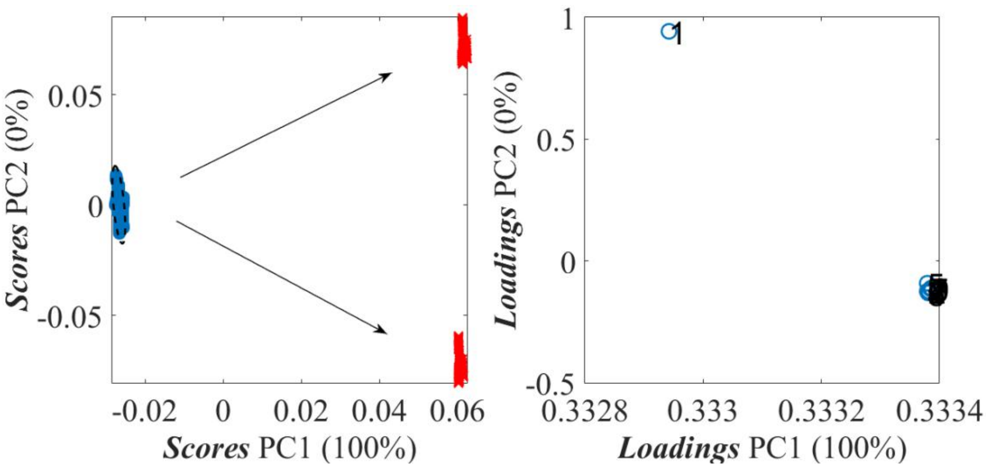

Figure 10 – Scores (left) and loadings (right) plots at 3000 seconds for the simulated faulty condition E.

Figure 11 shows the results obtained for Case F. It is possible again to observe in the scores plot the occurrence of failures of different nature. In this situation, we conclude that detector 1 was responsible for the fault represented by the set of scores in the upper left corner. The fault represented by the set of scores in the lower left corner, in turn, can be explained by the loadings of the other detectors. In this case, the identification of the fault occurring in detector 3 is confused with the faults that occur in the other detectors, due to the proximity of their temperature variations, compared to the temperature measured by the detector 1. Thus, the fault that occurred in detector 3 is not differentiated from the others with only two principal components. Although these two components are enough to explain the variability of the system, it is suggested to use one more principal component to differentiate the fault that occurred in detector 3 (Figure 12).

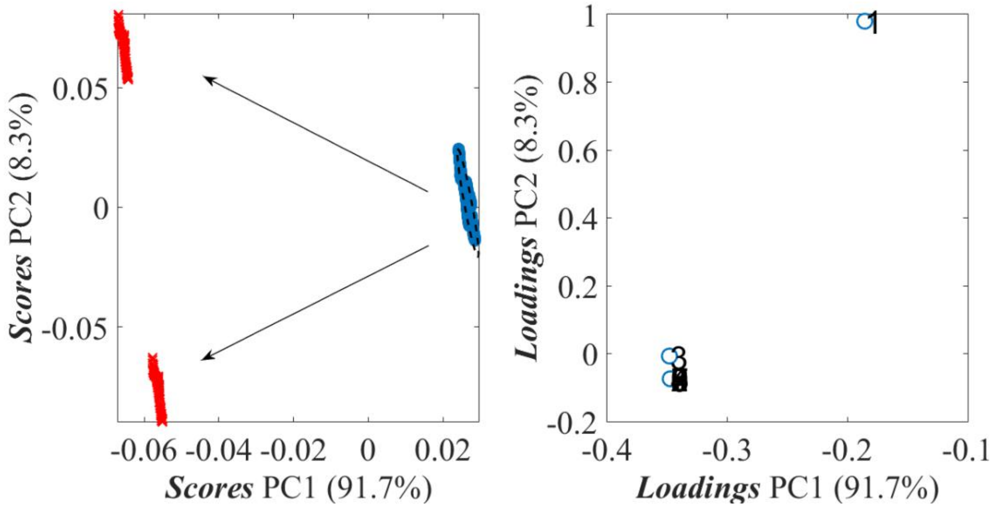

Figure 11 – Scores (left) and loadings (right) plots at 3000 seconds for the simulated faulty condition F.

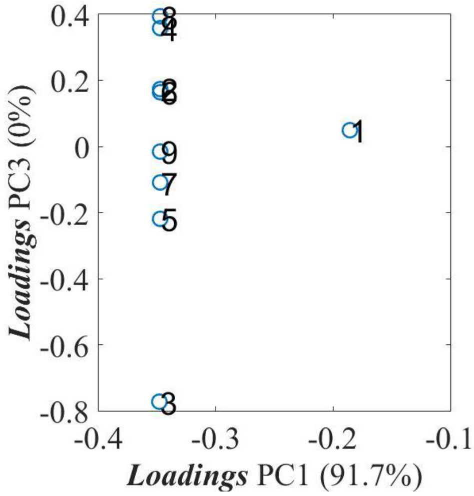

Figure 12 – Loadings plot of the first principal component (PC1) against the third (PC3) at 3000 seconds for the simulated faulty condition F.

4. CONCLUSIONS

Results showed that the developed algorithm was capable of identifying and diagnosing failures effectively. The statistic used was sensitive to abrupt changes in measured temperatures, whereas for the gradual fault tested there was a delay in recognizing fault. In spite of the restricted set of variables used, the developed algorithm, due to its simplicity and ease of implementation, has great potential to assist in the identification of more complex problems, such as those found in industries.

REFERENCES

CHEN, T. On Reducing False Alarms in Multivariate Statistical Process Control. **Chemical Engineering Research and Design**, v.88, p. 430-436, 2010.

FEITAL, T. S. **Monitoramento da condição de processos químicos industriais**. 2011. 182 f. Tese (Doutorado em Engenharia Química) - COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

GEMPERLINE, P. **Practical Guide to Chemometrics**. 2 ed. Boca Raton, London e New York: CRC Press, 2006.

HERNANDEZ-VARGAS, M.; CABAL-YEPEZ, E.; GARCIA-PEREZ, A. Real-time SVD-based detection of multiple combined faults in induction motors. **Computers and Electrical Engineering**, v. 40, p. 2193–2203, 2014.

MACGREGOR, J.; CINAR, A. Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods. **Computers and Chemical Engineering**, v.47, p. 111-120, 2012.

MILLER, J. N; MILLER, J. C. **Statistics and Chemometrics for Analytical Chemistry**. 5 ed. Harlow: Pearson Prentice Hall, 2005.

MONTGOMERY, D. C. **Introduction to statistical quality control**. 5 ed., New York: John Wiley & Sons, 2005.

QIN, S.J. Statistical process monitoring: basics and beyond. **Journal of Chemometrics**, 17, 480-502, 2003.

QIN, S.J. Survey on data-driven industrial process monitoring and diagnosis. **Annual Reviews in Control**, v. 36, p. 220-234, 2012.

SARTORI, I.; AMARO, C. A.; DE SOUZA, M. B. J.; EMBIRUÇU, M. Detecção, Diagnóstico e Correção de Falhas: Uma Proposição Consistente de Definições e Terminologias. **Science & Engineering Journal**, v. 21, p. 41–53, 2012.

SILVA, J. G. B. **Aplicação da Análise de Componentes Principais (PCA) no Diagnóstico de Defeitos em Rolamentos através da Assinatura Elétrica de Motores de Indução**. 2008. 109 f. Dissertação (Mestrado em Engenharia Elétrica), Universidade Federal de Itajubá, Itajubá, 2008.

SOARES, V. B. **Detecção e diagnóstico de falhas em plantas industriais com base em padrões de alarmes**. 2016. 204 f. Tese (Doutorado em Engenharia Química)-COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

SOARES, V. B.; PINTO, J. C.; DE SOUZA JR., M. B. Alarm management practices in natural gas processing plants. **Control Engineering Practice**, v.55, p. 185-196, 2016.

TEÓFILO, R. F. **Métodos Quimiométricos em Estudos Eletroquímicos de Fenóis sobre Filmes de Diamante Dopado com Boro**. 2007. 329 f. Tese (Doutorado em Química) - Instituto de Química, Universidade Estadual de Campinas, Campinas, 2007.

VENKATASUBRAMANIAN, V.; RENGASWAMY, R.; YIN, K.; KAVURI, S. N. A review of process fault detection and diagnosis Part I: Quantitative model-based methods**. Computers and Chemical Engineering**, v. 27, p. 293-311, 2003a.

VENKATASUBRAMANIAN, V.; RENGASWAMY, R.; YIN, K.; KAVURI, S. N. A review of process fault detection and diagnosis Part III: Process history based methods. **Computers and Chemical Engineering**, v. 27, p. 327-346, 2003b.

VILLEGAS, T.; FUENTE, M. J.; RODRÍGUEZ, M. Principal Component Analysis for Fault Detection and Diagnosis. Experience with a Pilot Plant. In: Proceedings of the 9th WSEAS international conference on computational intelligence, man-machine systems and cybernetics, Wisconsin, USA, p. 147-152, 2010.

YUE, H.; QIN, S. J. Reconstruction based fault identification using a combined index. **Industrial & Engineering Chemistry Research**, vol. 40, p. 4403–4414, 2001.