Preprints of the 2012 20th Mediterranean Conference on Control & Automation (MED), Barcelona, Spain, July 3-6, 2012

Application of a Mahalanobis-based Pattern Recognition technique for Fault Diagnosis on a chemical process

Silvia M. Zanoli, Giacomo Astolfi

Abstract— **The paper proposes a Fault Detection and Isolation (FDI) procedure based on a model-free approach and the use of pattern recognition techniques. In particular this paper aims to improve the isolation performance of a Fuzzy Faults Classifier (FFC) previously proposed by the author by modifications of the fuzzification module and by the use of the Mahalanobis distance as metric for identifying the most probable fault. In the paper faults due to the wear and tear of the thrust bearing and to fouling of the compressor stage of an industrial multishaft centrifugal compressor are considered. The presented results show the goodness of the overall procedure in the detections of single as well as multiple faults and its promptness in terms of faults isolation.**

I. INTRODUCTION

In order to enhance the efficiency level and the profitability of oil refineries as well as the commitment to meet precise production standards, an increasing level of system automation is required. Several processes of a refinery plant involve the use of centrifugal compressor and their overall efficiency can be greatly influenced by compressors working conditions.

Centrifugal machines are critical equipments, their essential characteristics been the large pressure rises and flow rates involved. Centrifugal compressors realize compression by transferring momentum to the fluid and the subsequent diffusion to convert the kinetic energy into pressure. The momentum transfer takes place at the doubly curved blades of the impeller that is mounted on a rotating shaft. Diffusion takes place in the annular channel of increasing radius around the impeller, usually referred to as diffuser.

Furthermore, the existence of rigorous environmental standards together with the need to operate in high safety conditions contribute to the implementation of automatic systems of rising complexity devoted to supervision, control and prevention of possible frequent malfunctions and potential faults. Typical control problems of interest when dealing with centrifugal compressor are speed regulation [1], robust stabilization [2] and surge avoiding [3],[4], [5], while for fault diagnosis purposes most of the works are based on the study of vibration signals [6], [7], [8]., A recent study of the authors on centrifugal compressors has concerned a Fault Detection and Isolation (FDI) data-based approach based on Principal Component Analysis (PCA), Cluster Analysis (CA) and Pattern Recognition (PR) algorithm [9], [10]. For other centrifugal equipments like centrifugal pumps in literature application of Fault Diagnosis model-based techniques can be found [11], [12], and13].

In the present paper, isolation performances of a Fault Detection and Isolation system presented in a previous work are improved [14]. At each sample time the FDI system suggests the most probable faulty (unfaulty) condition. In particular, the promptness of the system in term of faults isolation has been enhanced by a more effective procedure by the introduction of a different metric.

The paper is organized as follows: the process description is summarized in section II. In section III the developed fault detection and isolation system is presented. The improvements introduced by the modification of the membership function employed in the internal fuzzification module and by the use of the Mahalanobis distance are described. In Section IV the results concerning the detection and the isolation of a real fault of the centrifugal compressor are presented. Finally, concluding remarks are reported in section V.

II. PROCESS DESCRIPTION

A multishaft centrifugal compressor, employed for the nitrogen (N_2) compression in the dilution of a particular synthetic gas, called *syngas*, is considered in the present paper. The compressor is located in the Air Separation Unit (ASU) plant of an Integrated Gasification and Combined Cycle (IGCC) section of a refinery plant. The combined cycle plants, like the IGCC, enable for the thermal energy recovery by re-using it for other industrial processes, such as the steam production. The centrifugal is composed of five coupled compressor stages; in Figure 1 the coupling between two of the five stages can be observed.

Given the importance and the central role played by centrifugal compressors in many industrial processes, it is natural to focus much attention on the prevention of possible frequent malfunctions and potential faults which may cause inactivity or even their complete break. In the present work faults concerning the third stage of the compressor are considered. The approach adopted is based on multivariable data-driven, model free approach. Specifically the developed Fault Diagnosis system is based on the PCA (Principal Component Analysis) technique. The main reason for choosing a model free technique rather than

S. M. Zanoli is with the Department of Information Engineering (*DII*), Polytechnic University of Ancona, Italy (e-mail: s.zanoli@univpm.it).

G. Astolfi is with the Department of Information Engineering (*DII*), Polytechnic University of Ancona, Italy (e-mail: g.astolfi@univpm.it).

978-1-4673-2529-5/12/$31.00 ©2012 IEEE

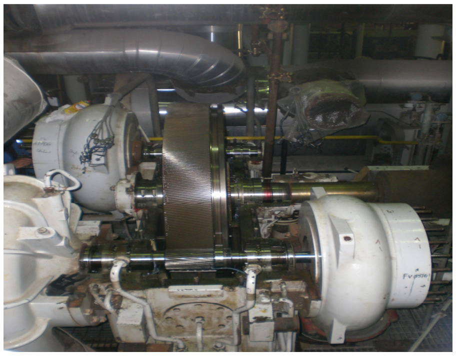

Fig. 1. The multishaft centrifugal compressor employed for the nitrogen (N_2) compression: view of the coupling between two stages of the compressor

Fig.2. Fouling of a compressor stage: as it can be see, due to sea water leakage in the exchanger, sea salt can accumulate in the compressor frame.

a model-based one, is that developing detailed physical models for industrial plant is generally a very demanding task; moreover difficulties arise given that typically the resulting models are of high dimensions and nonlinear. On the other hand, thanks to its simplicity and efficiency in processing huge amount of process data, PCA is recognized as a powerful tool of statistical process monitoring and it is widely used in the process industry for fault detection and diagnosis purpose (see for example [15], [16])

Typical faults in industrial process concern sensors and actuators but more complex process faults that influence several variables may affect the process.

In this work, a fault database consisting of three different faulty/unfaulty conditions have been considered which are expressive for the main system conditions (see table I). From a historical analysis of the compressor fouling (contamination) of compression stages (fault code #2) and breaks of thrust bearing (fault code #1) resulted to be the most frequent causes of faults. Fouling is typically due to sea water leakage in the heat exchanger: water from the tube enters the shell side and mixes with the process gas. In Normal Operative Conditions (NOC), the nitrogen pressure which is higher than the water pressure prevents the sea water to enter the shell side of the exchange. However, during compressor shut-down, gas pressure is reduced and water may spurt into the shell side. As a consequence sea salt accumulates in the compressor frame (see Figure 2). If this fault is not promptly identified, the fouling will cause a decrease in the efficiency, an increase of horizontal and vertical shaft vibrations and an increase in the temperature of thrust bearing.

The analysed system conditions are summarized in Table I and comprehend single (fault code #2) and multiple variables (fault code#3) faults.

TABLE I
FAULT PROTOTYPES CODE NUMBER AND DESCRIPTION

FAULT PROTOTYPE CODE	DESCRIPTION
(1)	Absence of faults.
(2)	Breaking of the thrust bearing.
(3)	Fouling (contamination) of the compressor stage.

For the actual implementation of the system, a larger fault database that comprises other process faults and many single and multiple actuators\sensors faults has been developed.

Visual inspection of critical ASU plant signal is regularly performed at the scope to detect anomalous system behaviors and warn for possible faults. Nevertheless it is not always simple to associate symptoms to the correct fault. This is the case when considering breaks of the thrust bearing: an increment of the frequency of the horizontal vibration may be associated to different faults (like break of the thrust bearing, fouling, fault in the sensor, etc.,) so that, fault isolation from simple visual inspection of the signal trend, is not feasible.

Diversely, the proposed approach was successful in the prompt and correct isolation of the break of the thrust bearing as it will be shown in the following.

III. Fault Detection and Isolation

To approach the problem of the fault detection and isolation of the centrifugal compressor at issue, a data-based approach centered on Principal Component Analysis (PCA), Cluster Analysis (CA) and Pattern Recognition (PR) algorithm has been developed. The first problem to be solved when approaching fault diagnosis problems with PCA techniques is the choice of a proper dimension for the Principal Component subspace. This technique is based on a mathematical procedure that transforms a number of possibly correlated variables of the process, into a smaller number of uncorrelated variables called Principal Components (PCs).

At this regard, a rigorous objective procedure that resulted reliable in real contexts, assuring uniqueness of the proposed solution, has been presented in [17] and here discussed.

For accomplishing the primary function of isolation, in the present paper the so called structured residual approach [16] has been adopted which has been integrated by an original statistical methodology for the selection of the l most significant eigenvalues, proposed by the authors in [10].

To overcome to the growth of complexity in the analysis of process faults that typically involve many variables, an automatic procedure for the isolation of the principal known faults has been developed. Complexity is related to the high number of signals (i.e. SPE: Square Prediction Errors), that have to be investigated. This is particularly verified in case of process faults that involve many variables. Furthermore some of the variables involved in the considered process faults may exhibit a slightly different behavior from the training case due to different environmental and/or operative conditions. To solve this problem an automatic procedure for the inspection of all the SPEs, the Fuzzy Faults Classifier (FFC), has been developed.

The Fuzzy Faults Classifier (FFC) has been introduced by the authors in a [14]. The system, based on Cluster Analysis and Pattern Recognition of fuzzified SPEs values, suggests the most likely faults within a record of known faults. Historical data concerning known system faulty/unfaulty operative conditions are processed offline and proper fault prototypes are generated. As depicted in Figure 3 the offline procedure is based on three main preliminary steps: computation of the SPEs which is performed by the PCA algorithm, fuzzification of the SPEs values and finally, the construction of the faults prototype vectors (denoted F_{SPEk}) performed by the Cluster Analysis.

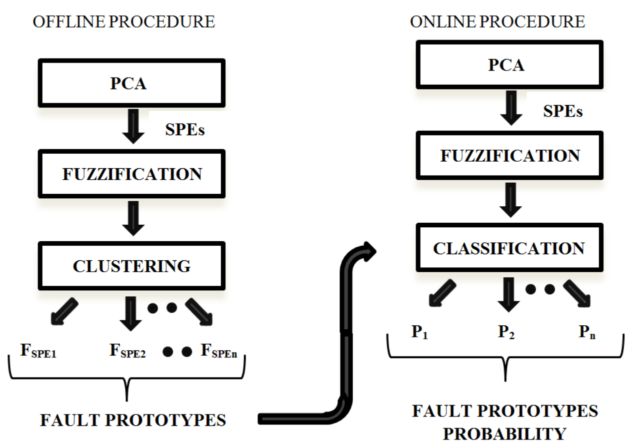

Fig. 3. Offline and online FDI procedure

Online (see figure 3), the fuzzification of the currently computed SPEs values is performed by the Fuzzy Faults Classifier and the vector F_{SPE} is computed.

To accomplish the classification task, that is assigning the degree of membership of the current F_{SPE} with respect to each of the stored fault prototype, the Euclidean distance between the computed F_{SPE} configuration and the stored F_{SPEk} configurations is computed. Finally, the current probability of each fault prototype is given as output.

More specifically, the probabilities are the values of the membership function U used in the Fuzzy C-Means (FCM) algorithm. At each sample time j, the elements μ_{ij} of the matrix U is computed which denotes the membership degree of the current F_{SPE} elements j from the centroid i:

$$\mu_{ij} = \frac{1}{\left(\frac{d_{ij}^2}{d_{1j}^2}\right)^{(1/(m-1))} + \ldots + \left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{(1/(m-1))}} \qquad (1)$$

where k is the number of the centroids, m >1 is the fuzziness factor and d_{ij} is the Euclidean distance. The common choice of the parameter m is 2. More specifically the highest probability is associated to the minimal distance. In order to avoid false alarms and to avoid "chattering", i.e. the possible high frequency fault turnover, in the score ranking a specific logic has been developed.

In order to improve the isolation capability, the Fuzzy Faults Classifier has been first enhanced by modification of the membership function for the fuzzification of the SPEs value. The scope of the Fuzzification module is to perform the matching of SPEs values to their thresholds in a fuzzy, not crispy way. The main motivation for this is the fact that an accurate setting of SPEs thresholds values related to process faults where many variables are involved in cascade is quite critical. A detailed inspection of the SPEs components values provided clear evidence that a characterization in term of the exceeding of their threshold could also be considered. The original and the modified membership function are shown in Figure 1.

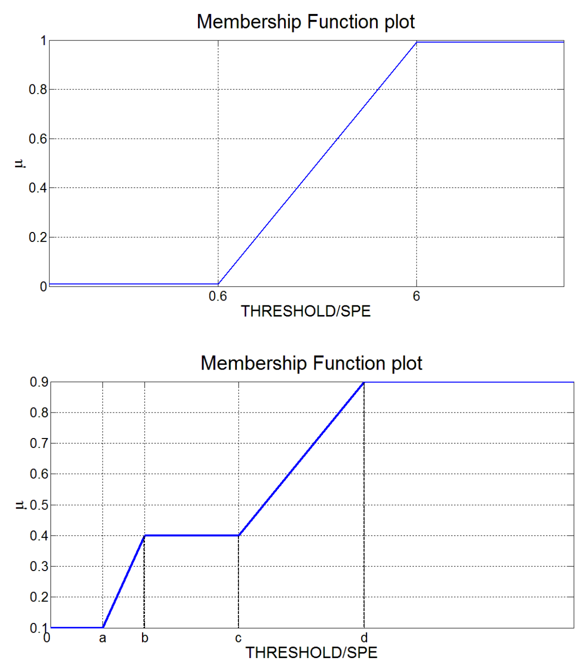

Fig. 4 Membership function for the fuzzification of the SPE signal: the old membership function (upper chart) and the new proposed one (bottom chart).

From a careful inspection of the output results of the developed FDI procedure, it was observed that adopting the Euclidean distance as classification metric, in some cases a poor discerning capability was attained. Thus in order to achieve a better characterization of each fault cluster, a different metric than the Euclidean has been considered so to include correlation data information. The authors propose the use of the Mahalanobis distance as metric for comparing the current operative condition from the stored faults. This solution is motivated by the desire to include in the analysis more discriminant data about each cluster as explained in the following.

A. *Mahalanobis Distance*

The benefits introduced when adopting, in the FCC, the Mahalanobis distance [18], [19] with respect to others non statistical distances (Euclidean distance, Manhattan distance, Minkowski distance, etc) are exemplified in the following.
Consider, for example, different data clusters and assume to be interested to compute the distance of a new FSPE point to the stored F_{SPEk} cluster. It is clear that the introduction of a metric which accounts for statistical indexes of the clusters themselves is of particular relevance.
In Figure 5 an exemplification of the features introduced by the Mahalanobis metric is reported. The example refers to a two-dimensional data space, but the extension at n-dimensional case is trivial. Here, the two points, (x1,y1) and (x2,y2) represent the projection of two different fault prototype clusters, F_{SPE1} and F_{SPE1}. The point F_R is the current fuzzified F_{SPE} vector computed online (as described in Figure 3)
The distance of the two points, (x_1,y_1) and (x_2,y_2) from the point F_R is computed in each of the two metrics. These values, reported in Figure 5, show that the Euclidean distances computed for the two chosen points, have the same value, as it can be clearly evinced by visual inspection of Figure 4 while the Mahalanobis distances result different. In fact, the latter metric highlights that when considering the dispersion of the points in each cluster, point F_R is much more in "agreement" with point (x_2,y_2) than with the point (x_1,y_1).

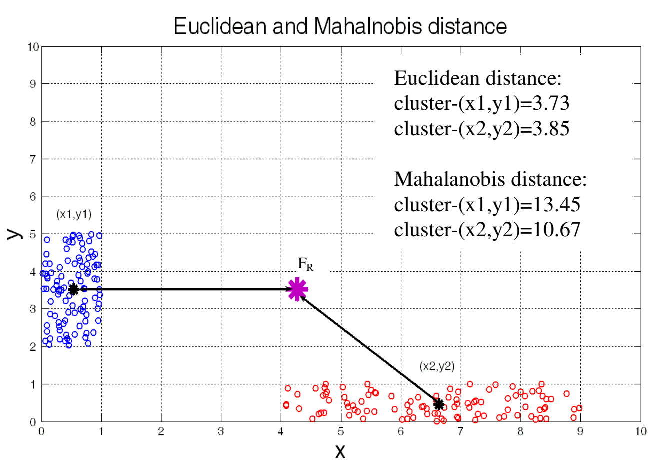

Fig. 5. Difference between the Euclidean and Mahalanobis distance.

Thus by the use of the Mahalanobis distance, the recognition of the current F_{SPE} configuration of the compressor becomes more robust. During the online data acquisition, as described previously in this section (see Figure 3), the system performs the fuzzification of the current SPEs value and evaluates the distance between the current configuration (F_{SPE}) and the stored ones; the distance is computed as:

$$\Delta^2 = \left(c_i - x\right)^T \Sigma^{-1} \left(c_i - x\right) \quad (2)$$

TABLE II
TAG AND DESCRIPTION OF THE CONSIDERED VARIABLES

Variable	Description
TI_89707	Measurement of N_2 temperature at the inlet of the 3^{th} stage of the BLNC.
PI_89706	Measurement of N_2 pressure at the inlet of the 3^{th} stage of the BLNC.
PCA_89708	Measurement of N_2 pressure at the exit of the heat exchanger used in the 3^{th} stage of the BLNC.
TV_89709	Measurement of N_2 temperature at the exit of the heat exchanger used in the 3^{th} stage of the BLNC.
TIA_89720	Measurement of the thrust bearing temperature of the shaft.
VXIA_89720	Measurement of the horizontal vibrations of the 3^{rd} shaft of BLNC.
VYIA_89720	Measurement of the vertical vibrations of the 3^{rd} shaft of BLNC.
TIA_89720	Measurement of the thrust bearing temperature of the shaft.
TIA_89725	Measurement of the thrust bearing temperature of the shaft.
TIA_89710A	Measurement of N_2 temperature at the exit of the 3^{th} stage of the BLNC.
90PAB12CT001	Measurement of H_2O temperature at the exit of the heat exchanger used in the 3^{th} stage of the BLNC.
ε	Polytropic efficiency of the 3^{th} of the BLNC.

where c_i is the cluster that represents an F_{SPE} relative to a fault condition of the compressor, x is a vector which indicates the actual F_{SPE} configuration and Σ is the covariance matrix. In order to calculate the previous expression it is necessary to know the covariance matrix of the data associated to each cluster. Thus, during the clusters generation step, the dispersion of the data around the stored cluster has been computed.
The benefits introduced by the use of the Mahalanobis distance in the new FFC are demonstrated in the following section and the detection and the isolation of a real fault, i.e. the break of the thrust bearing on the centrifugal compressor, is shown.

IV. Euclidean and Mahalanobis FFC results

The dataset analyzed for the FDI system has been chosen to contain 12 variables, concerning sensors, actuators and performance parameters (like polytropic efficiency) as summarized in Table II. Data are sampled at a rate of 10 minutes.
The first step of the PCA approach is the selection of the proper number of the Principal Component. The method used at this purpose is the statistical test ANOVA [20]. The ANalysis Of VAriance (ANOVA) is a statistical test, which is adopted to verify the consistence of different hypotheses. In the present context hypotheses are formulated about the parameters of the model, in particular about the eigenvalues of the correlation matrix. In this way, in successive iterations, it is checked whether additional information is gathered when adding a new eigenvalue, and in conclusion, the correct order of the system is detected. The suitable dimension of the Principal Component subspace for the case at issue has been computed to be five: this represents a good compromise between faithful data reconstruction and fine

noise filtering.

Following the off-line procedure described in section III, the three clusters prototypes for the faulty (unfaulty) conditions, reported in Table I, are generated. Online, after the computation of the fuzzified values F_{SPE} has been performed, the FFC executes a matching with the stored clusters in order to indicate the one to which the current configuration most probably belongs.

Results of the FFC module when considering the Euclidean distance as metric are shown in Figure 5. In this case the FFC correctly recognizes the break of the thrust bearing (fault code #2) at sample 852 (approximately just after the symptom on the horizontal vibration has appeared) assigning to it a probability of about 55%. At the same sample, the probability associated to the Fouling (fault code #3) is about 35% and the probability associated to the unfaulty case (fault code #1) is 8%. At sample 1080, i.e. one day and a half later, the probability relative to the correct fault increases up to 81%.

From these results it can be stated that the FFC system exhibits good detecting capabilities and appreciable isolating performances. Nevertheless improvements in the isolating performances may be preferred and, as the following data will confirm, these can be achieved by the introduction of the Mahalanobis distance.

On the same process data, new computations are performed by the adoption of the proposed metric and new distances associated to the current F_{SPE} configuration are obtained. The results are shown in Figure 6.

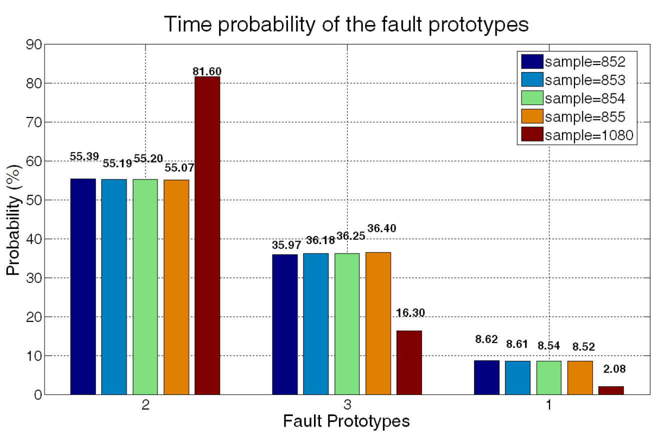

Fig. 6. Trend of the fault prototypes probabilities computed by the use of the Euclidean distance.

At sample 852, when the first symptoms are detected the distance from the cluster associated to the absence of faults increases. Accordingly, the distances from the others centroids decrease and the one relative to the break of thrust bearing is the lowest one (see Figure 7). This situation remains unchanged for all the remaining data since no correction action was performed.

The output of the FFC when considering the new metric approach is depicted in Figure 8.

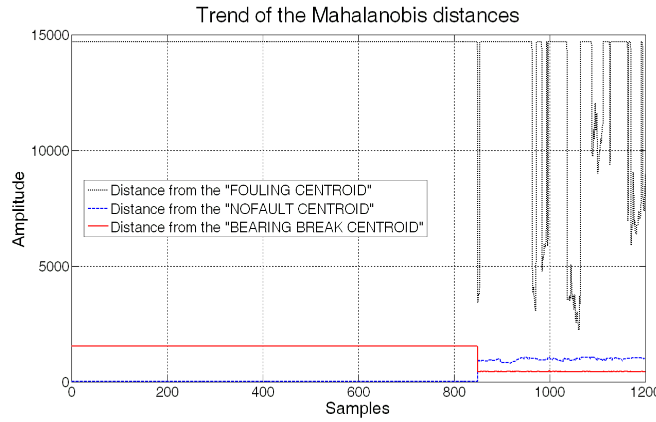

Fig. 7. The Mahalanobis distances of the analysed data from the stored fault clusters.

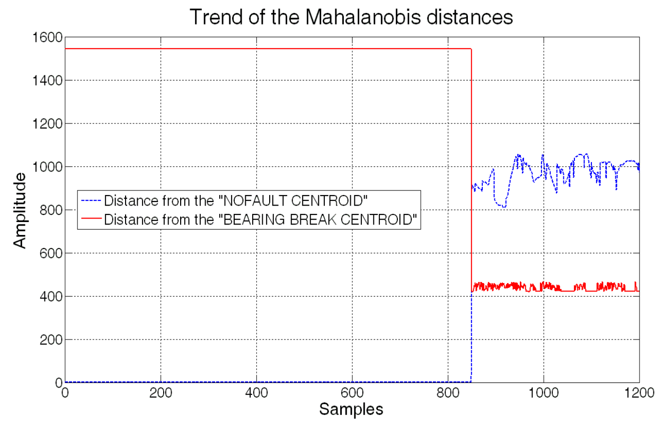

Fig. 8. Zoom of the Mahalanobis distances of the analysed data from the stored fault centroids.

Thus for what concerns the detection capability of the proposed FDI system, from the analysis of the results, it is possible to conclude that for the considered case the two metrics are comparable: at the same sample time 852 the faulty situation is detected. On the other hand, the experimental results shows that the Mahalanobis metric offers better results with respect to the Euclidean metric, when the ability in the faults isolation is considered.

In fact, from the application of the two metrics, different values of the fault probabilities are computed by (2). In particular, from inspection of Figure 6 it can be noted that, at the time the fault is firstly detected and correctly isolated the gap between the faults #3 is not so remarkable and a significant increase can be observed only several samples (225 samples, i.e. about 38 hours) after the first detection. Differently by the use of the Mahalanobis distance, since the first detection of the fault, it is possible to associate the fault to the break of the thrust bearing. This result can be explained by the fact that the Mahalanobis distance taking into account, for each cluster, the correlation between the data enhances their distinguish ability and this, ultimately, considerably reduce the time from fault occurrence to its isolation.

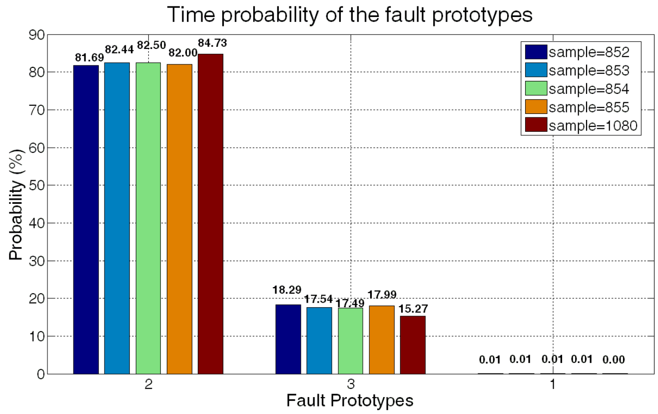

Fig. 9. Trend of the fault prototypes probabilities computed by the use of the Mahalanobis distance

TABLE III
SAMPLE PROBABILITY OF THE FAULT PROTOTYPES

	EUCLIDEAN DISTANCE			**MAHALANOBIS DISTANCE**		
	Fault prototype			**Fault prototype**		
SAMPLE	#2	3	1	2	3	1
852	55.39	35.97	8.62	81.69	18.29	0.01
853	55.19	36.18	8.61	82.44	17.54	0.01
854	55.20	36.25	8.54	82.50	17.49	0.01
855	55.07	36.40	8.52	82.00	17.09	0.01
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1080	81.60	16.30	2.08	84.73	15.27	0.00

V. CONCLUSION

In this work, the results concerning the detection and the isolation of a break of the thrust bearing in a multishaft centrifugal compressor are shown. From the inspection of historical data, there is clear evidence that without the use of a diagnostic tool that assists the process engineer operations, to recognize the true fault it is not trivial unless its effects are not appreciable. This delayed detection/isolation may generate considerable economic loss damage and/or cause undesirables wear of the system. In the present research work fault detection is approached by a PCA model-free technique, while a Fuzzy Faults Classifier (FFC), as proposed by the authors in a previous work, performs fault isolation. The most effective enhancements achieved are related to the modification of the membership function employed in the fuzzification module and the use of the Mahalanobis distance as metric for the fault recognition purpose. The main benefit of this metric is the possibility of taking into account the correlation between the data and this sensitively increases the performances of the overall system. The improvements are mainly associated to the faster response for the true fault isolation with respect to the previous FFC module. The previous module required about one day and a half for the true fault isolation while with the introduction of the new metric in less than one hour from the appearance of the symptom, the fault can be correctly detected and isolated.

ACKNOWLEDGMENT

The authors wish to extend warm thanks to the api refinery of Ancona staff (in particular Luca Barboni for his invaluable research assistance) and Syas S.r.l. team for their insights and support.

REFERENCES

[1] D.-C. Liaw, C-C. Song and J-T. Huang, “Robust Stabilization of a Centrifugal Compressor With Spool Dynamics”, in *IEEE Trans. Contr. Syst. Technol.*, vol. 12, no. 6, 2004.
[2] Epstein, A. H., Ffowcs Williams, J. E., and Greitzer, E. M., “Active suppression of aerodynamic instabilities in turbomachines”, in *Proc. 10th AIAA Aeroacoustics Conf.*, Seattle, WA, USA. Paper 86–1994
[3] Gravdahl, J. T. and Egeland, O., *Compressor Surge and Rotating Stall: Modeling and Control*. Springer-Verlag, London, 1999.
[4] De Jager, B., “Rotating stall and surge control: A survey”, in *Proc. 34th IEEE Conf. on Decision and Control*, 1857–1862, New Orleans, LA, USA, 1995.
[5] Liao, H., HUANG, S., “The fault diagnosis for centrifugal compressor based on time series analysis with neural network”, in *International Conference on Advanced Computer Theory and Engineering (ICACTE)* Chengdu, China, 2010.
[6] Yay, Z., Guo, H., Duan, L., “New method to establish fault diagnostic standard of centrifugal compressor”, in Oil Field Equipment, China, 2010.
[7] Yu, X., Zhang, L., Wang, Z., “Fault diagnosis of refrigerator compressor on the vibrating spectral analysis”, in *Oil Field Equipment*, China, 2005.
[8] Dalton, T. and Patton, R., “Model-based fault diagnosis of a two-pump systems”, in *IFAC World Congress*, pages 79–84, San Francisco, CA, USA, 1996.
[9] Zanoli, S., M., Astolfi, G., Barboni, L., “Applications of Fault Diagnosis Techniques for a Multishaft Centrifugal Compressor”, in *18th Mediterranean Conference on Control and Automation (MED10)*, Marrakech, Morocco.
[10] Zanoli, S., M., Astolfi, G., Barboni, L., “FDI of Process Faults based on PCA and Cluster Analysis”, *in Conference on Control and Fault Tolerant System* (*SysTol10*), Nice, France, 2010.
[11] Isermann, R., *Fault-Diagnosis Applications. Model-based condition monitoring: Actuators, drives, machinery, plants sensors, and fault-tolerant systems*. Springer, Berlin, 2011.
[12] Higham, E. and Perovic, S, “Predictive maintenance of pumps based on signal analysis of pressure and differential pressure (flow) measurements”, *Trans. Of the Institute of Measurement and Control*, 23(4):226–248, 2001.
[13] Nold, S. and Isermann, R., “Model-based fault detection for centrifugal pumps and AC drives”, in *11th IMEKO World Congress*, Houston, TX, USA, 1998.
[14] Zanoli S.M., L. Barboni, G. Astolfi, “Principal Component Analysis based on ANOVA Test for Multishaft Centrifugal Compressor Fault Detection and Diagnosis”, in *4th IFAC Symposium on System, Structure and Control (SSCC)*, Ancona, Italy, 2010.
[15] Y. Tharrault, G. Mourot, J. Ragot and D. Maquin, “Fault detection and isolation with robust principal component analysis”, in *International Application Mathematical Computer Science.*, 2008, vol. 18, No. 4, pp. 429–442.
[16] Bezergianni, S. and Kalogianni, A., “Application of Principal Component Analysis for Monitoring and Disturbance Detection of a Hydrotreating Process”, in *Ind. Eng. Chem. Res.*, No. 47, page(s) 6972–6982, 2008.
[17] Zanoli, S., M., Astolfi, G., Barboni, L., “Applications of Fault Diagnosis Techniques for a Multishaft Centrifugal Compressor”, in *9th European Workshop on Advanced Control and Diagnosis (ACD 2011)*, Budapest, Hungary, 2011.
[18] Mahalanobis, P. C., “A statistical study of the Chinese head”, in *Man In India (8),* 107-122, 1928.
[19] McLachlan, G., “Mahalanobis Distance”, in *RESONANCE,* 20-26, 1999.
[20] G. R. Miller, Beyond ANOVA – Basics of Applied Statistics, Chapman&Hall/CRC, 1997.