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

# Multiple Fault Detection and Diagnosis in a Gas Turbine using Principal Component Analysis and Structured Residuals

Amilcar Rincón

*Abstract*— **This paper presents the study of the detection and diagnosis of multiple faults in a Gas Turbine using principal component analysis and structured residuals method. The study includes developing a mathematical model and implementation of algorithms for detection and diagnosis of multiple faults in the system using principal component analysis and structured residuals.**

## I. INTRODUCTION

The multiple fault detection and diagnosis is an increasing research domain. The detection and diagnosis of faults in devices and processes has been a field of development and has been approached from multiple perspectives: control engineering, artificial intelligence and statistics, among others [1]. Within each with a multitude of techniques: expert systems, neural networks, case-based reasoning, signal analysis, observers, equations of analytical redundancy, consistency-based diagnosis, but considering the presence of single faults and not multiples [2].

Multiple fault diagnosis is a challenging problem because the number of candidates grows exponentially in the number of faults. In addition, multiple faults in dynamic systems may be hard to detect, because they can mask or compensate each other's effects. The multiple faults problem is important, since the single fault assumption can lead to incorrect or failed diagnoses when multiple faults occur [3].

If a system is diagnose every so often (once a day), the faults may have occurred during that period of time an outbreak of non-simultaneous faults, sequentially, when you want to analyze the system has more than one fault. Here the traditional methods to diagnose single faults can't be used.

Another reason why it is important to study multiple failures is that there are systems that when you want to analyze for the first time, and has long occurred simultaneously or not, remain dormant fault when this happens, it cannot apply methods that consider single faults because there may be multiple faults or latent desire to diagnose.

Numerous works are devoted to the analysis of the methods of detection and diagnosis of single faults and very few with focus on multiple faults, simultaneous or sequential. Some of the most popular include extended diagnostic matrices [4], [5], set covering theory [6], [7], consistency-based reasoning [8], [9], [3], logical causal graphs [3] and many others [10], [11], [16], [17]. In the area of fault detection and isolation, emerging from the classical theory of automatic control, these problems are discussed in [18], [19] and [17].

This paper is organized as follows: a brief system description of the gas turbine SR-30 and the mathematical model is presented in section 2. The section 3 describes the fundamental of principal component analysis and structured residuals method used for diagnosis. In section 4, the algorithm for fault detection and diagnosis using PCA and structured residuals method is presented. In section 5 the results and discussion are exposed. Finally, conclusions are summarized in section 6.

## II. GAS TURBINE MODEL

### A. System Description

The Gas turbines are one of the main movers. The aim of using a gas turbine can be double: either generating thrust force for an aircraft, or generating power for any other machine, for example a generator, propeller and so on [20].

The main parts of a gas turbine include the inlet duct, the compressor, the combustion chamber, the turbine and the nozzle. The operation of gas turbines is basically the same. The air is drawn into the engine through the inlet duct by the compressor, which compresses it and then delivers it to the combustion chamber. Within the combustion chamber the air is mixed with fuel and the mixture is ignited, producing a rise in temperature and hence an expansion of the gases. These gases are exhausted through the engine nozzle, but first pass through the turbine, which is designed to extract sufficient energy from them to keep the compressor rotating, so that the engine is self-sustaining [21].

The main parts of a gas turbine are shown schematically in figure 1. A real SR-30 gas turbine is used for our studies. The equipment is installed in the Inter American University

---

A. Rincón is with Mechanical Engineering Department, Inter American University of Puerto Rico, Bayamon Campus, 500 Dr. John Will Harris Road, Bayamon, Puerto Rico 00957. USA. Phone: 787-279-1912; fax: 787-279-2205; (e-mail: arincon@ bayamon.inter.edu).

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

of Puerto Rico, Bayamon Campus, and Department of Mechanical Engineering.

fig.1 SR-30 Gas turbine system

### B. Mathematical Model

The nonlinear state equations are derived from first engineering principles. Dynamic conservation balance equations are constructed for the overall mass $m$, the internal energy $U$ and the mechanical energy $E$ [20]. The notation list is given separately in the Appendix A. These dynamic equations have to be transformed to intensive variable form to contain the measurable quantities.

Therefore the set of transformed differential balances include the dynamic mass balance for the combustion chamber, the pressure form of the state equation derived from the internal energy balance for the combustion chamber and the intensive form of the overall mechanical energy balance expressed for the number of revolutions $n$. Thus 3 independent balance equations can be constructed; therefore the gas turbine can be described by only 3 state variables [20]. The form of the nonlinear dynamic model equations is the following:

$$\dot{m}_{comb} = \dot{m}_c + \dot{m}_{fuel} - \dot{m}_t \tag{1}$$

$$\dot{p}_3 = \frac{R}{C_v V_{comb}}[(\dot{m}_c C_p T_2 - \dot{m}_t C_p T_3 + LHV.\eta_{comb}\dot{m}_{fuel}] \tag{2}$$

$$\dot{N} = \frac{1}{4\pi^2\theta N}[\dot{m}_t C_p (T_3 - T_4)\eta_{mech} - \dot{m}_c C_p (T_2 - T_1) - 2\pi\frac{3}{50}NM_{load}] \tag{3}$$

In order to complete the model, constitutive algebraic equations are also needed. These equations describe the static behavior of the gas turbine in various operating points, and all of them can be substituted into the dynamic equations.

$$T_2 = T_1(1 + \frac{1}{\eta_c}[(\frac{p_2}{p_1})^{\frac{k-1}{k}} - 1]) \tag{4}$$

$$T_3 = \frac{p_3 V_{comb}}{m_{comb} R} \tag{5}$$

$$T_4 = T_3(1 - \eta_t[1 - \frac{1}{(\frac{p_3}{p_4})^{\frac{k-1}{k}}}]) \tag{6}$$

$$p_2 = \frac{p_3}{\sigma_{comb}} \tag{7}$$

$$p_4 = \frac{p_1}{\sigma_I \sigma_N} \tag{8}$$

$$\dot{m}_c = \beta A_1 \frac{p_1}{\sqrt{T_1}}[\alpha_1 \frac{N}{\sqrt{\frac{T_1}{288.15}}} \frac{p_3}{p_1 \sigma_{comb}} + \alpha_2 \frac{N}{\sqrt{\frac{T_1}{288.15}}} + \alpha_3 \frac{p_3}{p_1 \sigma_{comb}} + \alpha_4]$$

$$\dot{m}_t = \beta A_3 \frac{p_3}{\sqrt{\frac{p_3 V_{comb}}{m_{comb} R}}}[\gamma_1 \frac{\pi n}{\sqrt{\frac{p_3 V_{comb}}{m_{comb} R}}} \frac{p_3 \sigma_I \sigma_N}{p_1} + \gamma_2 \frac{\pi n}{\sqrt{\frac{p_3 V_{comb}}{m_{comb} R}}} + \gamma_3 \frac{p_3 \sigma_I \sigma_N}{p_1} + \gamma_4]$$

The parameters, the constants of these functions can be determined with the help of the results of the measurements, the compressor and turbine characteristic at different speeds (55.000, 60.000, 65.000, 70.000, 75.000 and 78.000 rpm).

The constant of these functions can be calculated with the help of the least squares method. The resulted model consist of 3 independent dynamic equations, therefore the gas turbine can be described by only 3 state variables. The figure 2 shown the regression between the speed, mass compressor and pressure ratio.

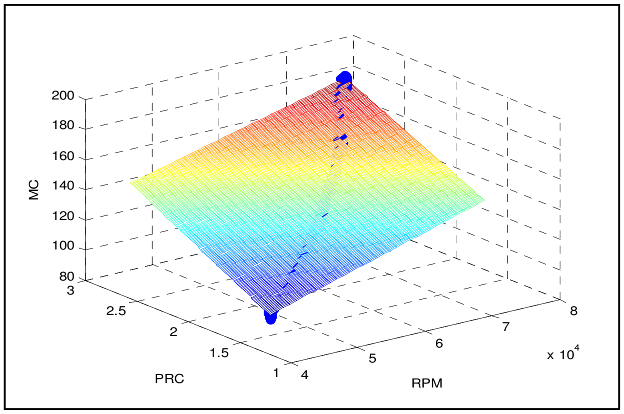

fig. 2. Regression between variables.

A Simulink model was developed and adjusted with real data from real turbine to simulate faults in temperature, pressure, speed and torque sensors. See figure 3.

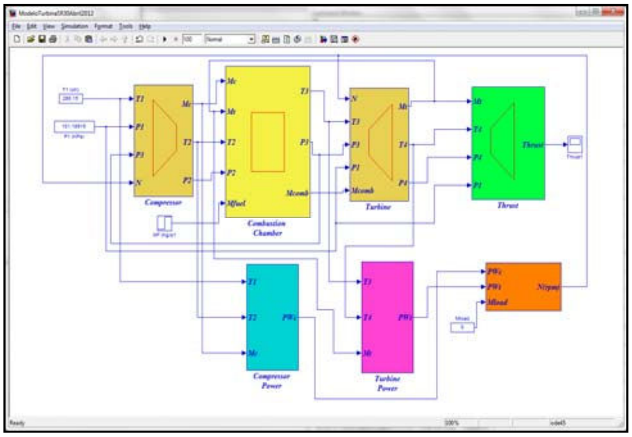

fig. 3. SR-30 Gas turbine Simulink model

The figures 4, 5 and 6 shown the temperature, pressure and speed sensor signals for the model versus real data measured with a real gas turbine SR-30.

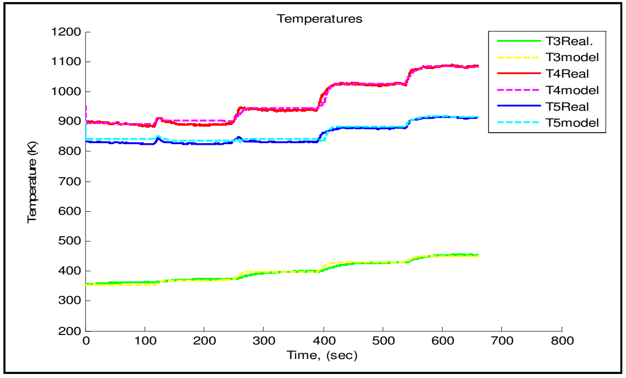

fig. 4 Temperature signal model versus real data

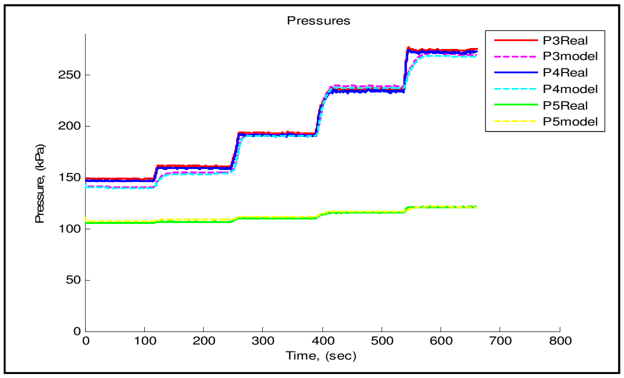

fig. 5 Pressure signal model versus real data

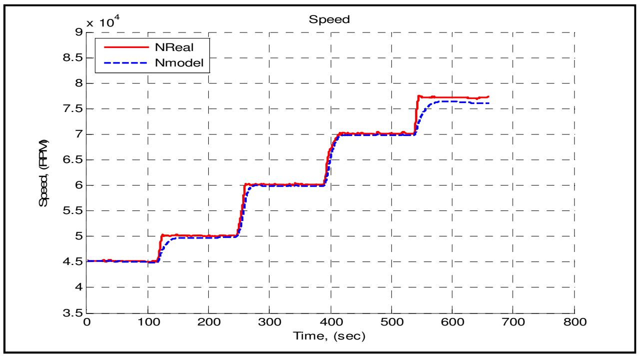

fig. 6 Pressure signal model versus real data

Although the actual data of temperature, pressure and speed contain noise and non-linearities, the results obtained have a 5.1% error. In general, the results approximate very well the behavior of the gas turbine. There are methodologies that would fit the model even further and is currently working on them.

## III. PRELIMINARY CONCEPTS

### A. *Principal Component Analysis*

Principal component analysis (PCA) is a multivariate analysis technique which was first introduced by Pearson in 1901 and developed independently by Hotelling in 1933.

Mathematically, PCA can be generally considered as a subspace decomposition technique where the process measurement space is divided into two orthogonal subspaces, that is, the principal component (PC) subspace and residual subspace.

Consider a normalized data matrix $X(m \times n)$ composed of $m$ sample vectors and n process variables collected under normal operation conditions. PCA transforms the matrix X to a linear combination of variables as follows

$$X = TP^T + \hat{T}\hat{P}^T = TP^T + E \qquad (9)$$

Where $P(n \times k)$ are the first $k(<n)$ principal component loadings, $T(m \times k)$ are corresponding scores. Matrices $\hat{P}$ and $\hat{T}$ consist of the last *n-k* column vector of loading and row vector of scores, respectively. Matrix $E(m \times n)$ is correspond to the residual subspace consisted by the abandoned information in PC subspace.

The decomposition of data matrix X in eq. 6 can be implemented by singular value decomposition of correlation matrix R as:

$$R = U\Lambda^{1/2}U^T \qquad (10)$$

Where:
$R = (X^T X)/(n-1)$, $[P^T \quad \hat{P}^T] = U$ And $[T \quad \hat{T}] = U\Lambda^{1/2}$,
and matrix $\Lambda$ is a diagonal one composed of the eigenvalues of the matrix **R**.

Process monitoring and fault diagnosis can then be carried out based upon statistical hypothesis tests in these two subspaces. Two indices, the Hotelling-$T^2$ and the Q statistics are used to describe process behavior, and subsequently detect process change and faults. In PC subspace, the hypothesis is done by using $T^2$ statistic, which is defined as:

$$T^2 = t\Lambda_k{}^{-1}t^T \qquad (11)$$

Where $\Lambda_k{}^{-1}$ is a diagonal matrix consisted by reciprocals of the $k$ largest eigenvalues of matrix **R**. In residual subspace, the Q statistic is used for statistical test, it is defined as follows:

$$Q = ee^T = x(I - \hat{P}\hat{P}^T)x^T \tag{12}$$

The $T^2$ statistic in conventional PCA corresponding to the PC subspace mainly describes behavior of process variables that have significant correlation with PCs. While the $Q$ statistic corresponding to the residual subspace that is related with all monitored variables.

### *A. Structured Residuals Method*

Among the various Fault Detection and Isolation (FDI) schemes, the structured residual approach proposed by J. Gertler, M. Staroswiecki and M. Shen [22] is powerful in isolating faults. Structured residual approach proposed by Gertler [23], [24] is considered for this paper for multi fault detection and isolation for a Gas Turbine system.

The structured residual approach involves two steps 1) generation of Primary Residual Vector (PRV) for fault detection and 2) transformation of PRV into structured residual vector (SRV) for fault isolation.

The implementation procedure of the proposed FDI scheme is illustrated in Figure 7. When there is a fault in the process, its output differs with model output. This difference is termed as residual. By simply monitoring the residuals one can say that something is going wrong. But it is not possible to identify the location of the fault. So the residual has to be processed to enhance isolation.

Fault Detection Using Residual Generator Residuals are generated from the observable variable of the monitored plant, that is, from the command values of the inputs and the outputs [25]. Ideally, the residuals should only be affected by the faults. However, the presence of disturbances, noise and modeling errors also causes the residuals to become nonzero and thus interferes with the detection of faults.

Therefore the residual generator needs to be designed so that it is maximally unaffected by these nuisance inputs, which means that it is robust in the face of disturbance, noise and model errors. Structured residual are so designed that each residual responds to a different subset of faults and insensitive to the others. When a particular fault occurs, some of the residuals do respond and others do not. Then the pattern of the response set, the fault signature or fault code, is characteristic of the particular fault.

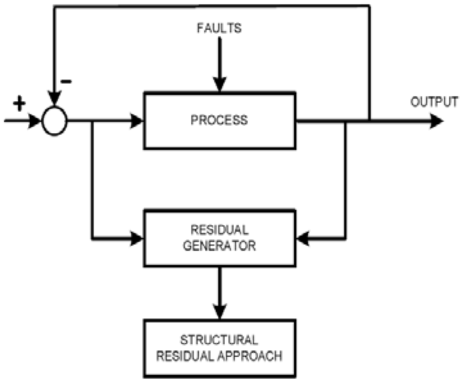

fig. 7 Block diagram representation of proposed FDI scheme

## IV. PROPOSED DETECTION AND ISOLATION METHODOLOGY USING PCA AND STRUCTURED RESIDUALS

### *A. Algorithm for multiple fault detection using PCA.*

The algorithm for multiple fault detection using PCA consists of the following steps:

1 - Acquire original training data matrix $\mathbf{X}_o$ which represents normal process operations.
2 - Scale $\mathbf{X}_o$ to $\mathbf{X}$ using its mean and standard deviation. Then carry out singular value decomposition (SVD) or use PCA toolbox to obtain loading matrix $\mathbf{P}$ and the score matrix $\mathbf{T}$ of the PCA model.
3 - Choose $k$ principal components based on cumulative percentage variance explained. (i.e. first $k$ principal components capture more than 80% variance).
4 - Obtain data sample and scale it to $\mathbf{x}$ (dimension mxn).
5 - Compute the $T^2$ statistic is the squared norm of the current sample from the center of the normal operation region.
6 - Compute the square prediction error (SPE), also called $Q$ statistic.
7- If currently the process is in normal status, check whether $Q$ or $T^2$ exceeds its control limit. If $Q$ or $T^2$ exceeds its control limit, go to step 8), otherwise go back to step 4) for next sample.
8 - A fault is detected and an alarm is set, indicating that the process is in abnormal status.
9 - Plot the variable contribution chart for violation of $Q$ or $T_2$ to isolate dominating process variables for root cause diagnosis.
10 - Go back to step 4 for next samples.

Figure 8 shows the PCA used for the detection of two abrupt simultaneous faults for the $T_3$ temperature sensors and pressure $P_3$ at 500 seconds of simulation.

Figure 9 shows the Square prediction error for two abrupt simultaneous faults in $T_3$ Temperature and $P_3$ pressure sensor at 500 seconds of simulation.

Figure 10 shows the Hotelling statistic $T^2$ for simultaneous fault in $T_3$ Temperature and $P_3$ pressure sensor at 500 seconds of simulation.

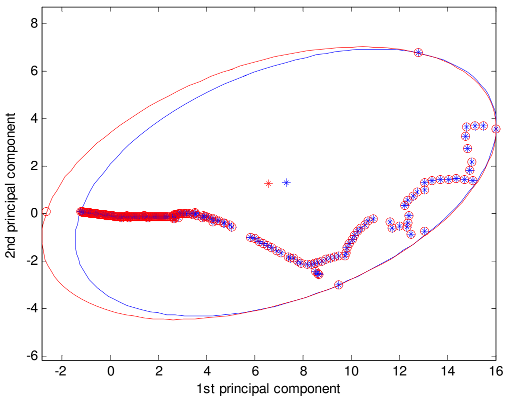

Fig. 8 Principal Component for simultaneous fault in $T_3$ and $P_3$ sensors.

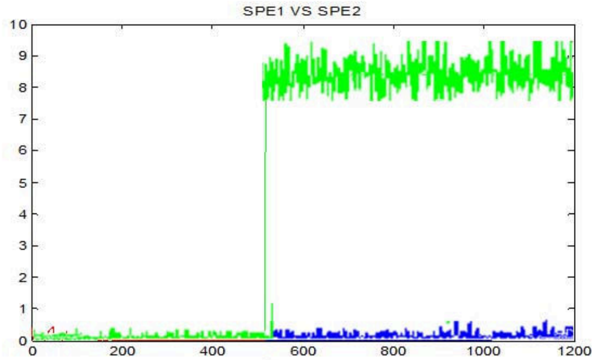

Fig. 9 Square Prediction Error for simultaneous fault in $T_3$ and $P_3$ sensors.

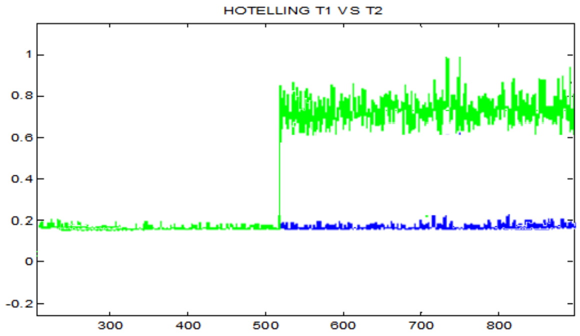

Fig. 10 Hotelling $T^2$ statistic for simultaneous fault in $T_3$ and $P_3$ sensors.

### B. *Isolation for multiple fault using structured residuals*

While sensor and actuator faults may be handled by a simple extension of the fault-free system model, process faults require explicit modeling. This may be done on the physical system, by emulating faults, or in our case using the SR-30 simulink model. In either case, a fault $f_j$ is introduced (with known location and size) and the resulting $\mathbf{x}(t/f_j)$ vector is observed (centered with the average of the fault-free training data) [26].

In the following, we will make use of the transformed sensitivity vectors $S_{\bullet j} = [S_{1j} \ldots S_{(n-m+k)j}]'$, $j=1 \ldots k$, where:

$$S_{ij} = q_i' x(t|f_j)/f_j \qquad \text{i =1…n-m+k, j=1…k} \quad (13)$$

The hypothetical faults considered are described in Table1.

Table 1. Fault Description.

| **Fault** | **Description** |
|---|---|
| Fault (1) MC | Air leak in compresor |
| Fault (2) MT | Leak in the combustion chamber |
| Fault (3) EC | Mechanical fault in compresor |
| Fault (4) ET | Mechanical fault in turbine |
| Fault (5) WF | Blockage of the fuel valve |

The residuals considered are described in table 2.

Table 2. Residual description

| Residual | Sensor Name |
|---|---|
| 1 | Temperature Compressor Inlet ($T_2$) |
| 2 | Temperature Compressor Outlet ($T_3$) |
| 3 | Temperature Turbine Inlet ($T_4$) |
| 4 | Temperature Turbine Outlet ($T_5$) |
| 5 | Pressure Compressor Inlet ($T_2$) |
| 6 | Pressure Compressor Outlet ($T_3$) |
| 7 | Pressure Turbine Inlet ($T_4$) |
| 8 | Pressure Turbine Outlet ($T_5$) |
| 9 | Revolutions (N) |
| 10 | Torque ($F_G$) |

In the monitoring phase, measurements x(t) are taken from the Gas Turbine Simulink model system (and centered with the average of the fault-free training data-set). These centered measurements are transformed into the representation space, as p(t). These latter are then used as residuals in fault detection and isolation.

In a structured framework, the fact that the dimension of the residual vector p(t) equals the number of faults allows for the design of a diagonal structure. Define S = [S1 … Sk], then the transformed residual:

$$r(t) = S^{-1} p(t) \qquad (6)$$

Obeys a diagonal structure each residual responds to one specific fault. Such structure allows for the isolation of multiple simultaneous faults.

Due to the diagonal structure of r (t), the elements of the residual vector are estimates of the individual fault sizes. These estimates, however, beyond being noisy, are also subject to an error due to the linearization of a nonlinear relationship.

## V. RESULTS AND DISCUSSION

The figure 11 shows the isolation for a single fault using structured residual. The figure 12 show the isolation for two simultaneous faults and the figure 13 is for the case with three simultaneous faults which were created with the Simulink model.

The S matriz will be used to detect small faults. In the operating point selected, is obtained as:

$$s=\begin{bmatrix} -0.0033 & -0.1041 & 0.1622 & 0.2753 & 0.2065 \\ -0.1347 & -0.0273 & 0.0256 & 0.0728 & 0.2354 \\ 0.7122 & -0.0903 & 0.2190 & 0.2371 & 0.2069 \\ -0.0071 & -0.0920 & 0.2301 & 0.2414 & 0.1798 \\ 0.0833 & 0.0665 & 0.2682 & 0.2771 & 0.2591 \end{bmatrix}$$

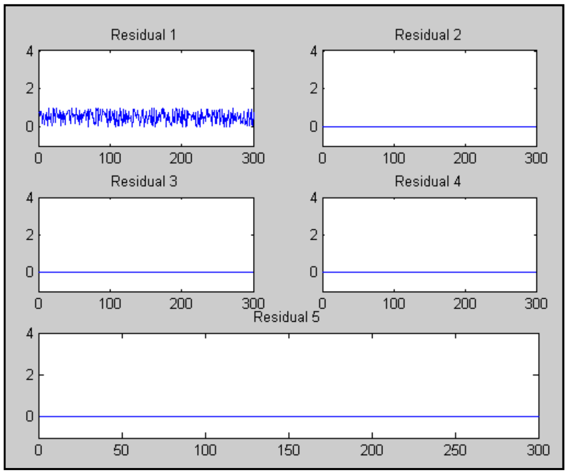

Fig.11. Fault 1% magnitude in Mass Compressor

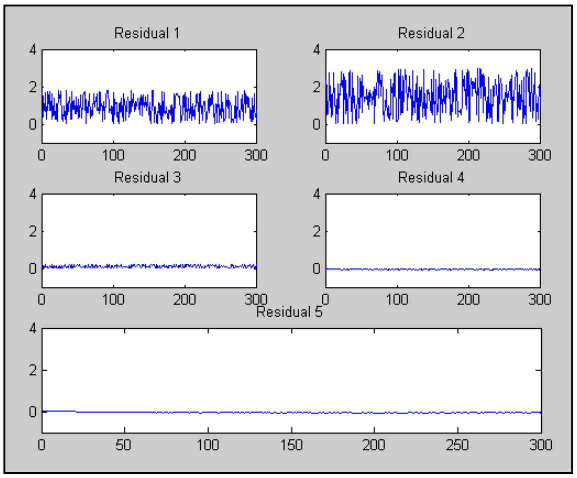

Fig.12. Simultaneous fault for 2% magnitude in mass compressor and 3% magnitude in mass turbine

Finally the filtered diagonally structured residuals are compared to the respective thresholds and determine the magnitudes for each fault.

This method can detect faults in the turbine components including noise and disturbance.

In Figures 11, 12 and 13 each residue is related to a fault. See table 1.

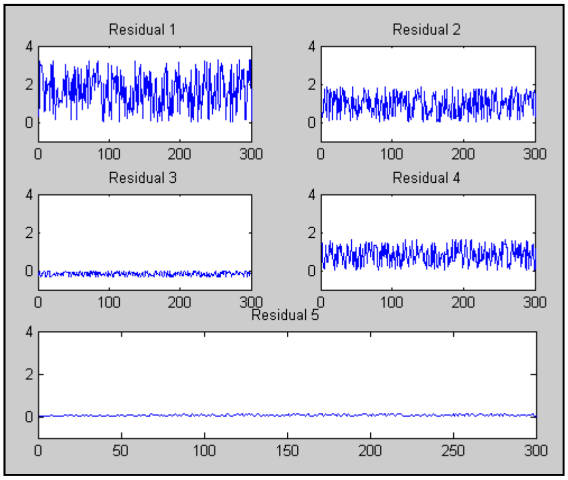

Fig.13. Simultaneous fault for 3% magnitude in compressor mass and 1.5% magnitude in turbine efficiency and 2% turbine mass

## VI. CONCLUSIONS

In general, PCA is a good tool for multiple fault detection, simultaneous and sequential faults, but it was necessary to complement it with the structured residuals methodology. Additional, PCA is a linear dimensionality reduction technique, which ignore the nonlinearities that exist in the process data.

The methodology presented here has been applied to the gas turbine can successfully diagnose many multiple faults and appears promising extension to other industrial processes with possible multiple faults.

## APPENDIX

Appendix A. Nomenclature for Variables and Constants

| | |
|---|---|
| $\dot{m}_{comb}$ | Mass flow in the combustion chamber |
| $\dot{m}_c$ | Mass flow in the compressor |
| $\dot{m}_{fuel}$ | Mass flow of fuel |
| $\dot{m}_t$ | Mass flow in the turbine |
| $p_1$ | Pressure in the compressor inlet |
| $p_2$ | Pressure in the compressor outlet |
| $p_3$ | Pressure in the turbine inlet |
| $p_4$ | Pressure in the turbine outlet |
| $T_1$ | Temperature in the compressor inlet |
| $T_2$ | Temperature in the compressor outlet |
| $T_3$ | Temperature in the turbine inlet |
| $T_4$ | Temperature in the turbine outlet |
| $N$ | Rotational speed in the turbine shaft |
| V | Volume combustion chamber |

# REFERENCES

[1] M. Daigle, X. Koutsoukos y G. Biswas. Multiple Diagnosis in Complex Physical Systems. Institute for Software Integrated Systems (ISIS). 2007.

[2] V. Venkatasubramanian. A review of process fault detection and diagnosis. El Sevier, Computer and Chemical Engineering 27 (2003) 327-346.

[3] J. Ligeza, J. Koscielny. A New Approach to Multiple Diagnosis: A Combination of Diagnostic matrices, Graphs, Algebraic and Rule-based Models. The case of two layer models. Int. Journal Appl. Math. Comput. Sci., 2008, Vol. 18, No.4, 465-476.

[4] J. Koscielny. Methodology of process diagnosis, *in* J. Korbicz, J. Koscielny, Z. Kowalczuk and W. Cholewa *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer-Verlag, Berlin, pp. 57–114. 2004a.

[5] J. Koscielny. Models in process diagnosis, *in* J. Korbicz, J. Koscielny, Z. Kowalczuk and W. Cholewa, *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer-Verlag, Berlin, pp. 29–43. 2004b.

[6] J. Reggia, D. Nau and P. Wang. Diagnostic expert system based on a set covering model, *International Journal on Man-Machine Studies* **19**(5): 437–460. 1983.

[7] J. Reggia, D. Nau. and P. Wang. (1985). A formal model of diagnostic inference. Problem formulation and decomposition, *Information Sciences* **37**(1–3): 227–256. 1985.

[8] R. Reiter. A theory of diagnosis from first principles, *Artificial Intelligence* **32**(1): 57–95. 1987.Benoff. "Reliability Centered Maintenance Gives a new Meaning to the Phrase: If it can't Broke don't fix it," *Business & Commercial Aviation Magazine*, Oct 2000.

[9] W. Hamscher, L. Console and J. de Kleer. *Readings in Model-Based Diagnosis*, Morgan Kaufmann, San Mateo, CA. 1992.

[10] S. Tzafestas. System fault diagnosis using the knowledge-based methodology, *in* Tzafestas S. *Knowledge-Based System Diagnosis, Supervision and Control*, Plenum Press, New York, NY/London, 1989a.

[11] S. Tzafestas. *Knowledge-Based System Diagnosis, Supervision and Control*, Plenum Press, New York, NY/London. 1989b.

[12] L. Blázquez, L. Miguel. Diagnóstico Automático de Fallos para Sistemas Dinámicos No Lineales. Publicación del Departamento de Ingeniería Eléctrica y Electrónica de la Universidad de León. 2002.

[13] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. Diagnosis and Fault Tolerant Control. Springer, 2003.

[14] O. Camacho. Fault Diagnostic base on multivariate statistical techniques. Revista técnica de la Universidad de Zulia. Vol. 30, No 3, 253-262, 2007.

[15] A. Casavola, D. Famularo y G. Franze. Robust Multiple-fault detection and isolation: A gradient flow approach. Proceeding of the 2004 American Control Conference. 2004.

[16] R. Davis and W. Hamscher. Model-based reasoning: Troubleshooting, *in* W. Hamscher, L. Console and J. de Kleer (Eds.), *Readings in Model-Based Diagnosis*, Morgan Kaufmann Publishers, San Mateo, CA, pp. 3–24. 1992.

[17] J. Korbicz, J. Koscielny, Z. Kowalczuk. and W. Cholewa. *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer-Verlag, Berlin. 2004.

[18] P. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – A survey and some new results, *Automática 1990* **26**(3): 459–474.

[19] P. Frank. Analytical and qualitative model-based fault diagnosis – A survey and some new results, *European Journal of Control* **2**(1): 6–28. 1996.

[20] A. Piroska. Nonlinear model building of a low power gas turbine. Periodica Polytechnica SER. Transp. Eng. Vol. 29. No 1-2, PP 117-135. 2001.

[21] T. Witkowski, S. White and O. Dueñas. Characterizing the performance of the SR-30 Turbojet engine. Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition, American Society for Engineering Education 2003.

[22] J. Gertler, M. Staroswiecki, M. Shen: Direct Design of structured residuals for fault diagnosis in linear systems, American Control Conference, 2002.

[23] J. Gertler: Residual Generation in Model Based Fault Diagnosis, control-Theory and Advanced Technology, Vol. 9, pp. 259-285, March 1993.

[24] J. Gertler: A Numerical-Structural Approach to Failure Detection and Isolation in Complex Plants, Proceedings of the IEEE Conference on Decision and Control, Athens, Greece, pp. 1576-1580, 1986.

[25] J. Gertler. Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, Inc. New York- Basel- Hong Kong. 1998.

[26] J. Gertler, J. Romera, V. Puig y J. Quevedo. Leak Detection and Isolation in water distribution network using Principal Component Analysis and Structured Residuals. Conference on Control and Fault Tolerant Systems. 2010.