# FAULT DETECTION AND DIAGNOSIS IN A HEAT EXCHANGER

Juan C. Tudon Martinez, Ruben Morales-Menendez and Luis E. Garza Castañon
*Tecnológico de Monterrey, Av. E. Garza Sada 2501, 64849, Monterrey N. L., México*

Keywords: Dynamic principal component analysis, Diagnostic observers, Fault detection, Fault diagnosis.

Abstract: A comparison between the Dynamic Principal Component Analysis (DPCA) method and a bank of Diagnostic Observers (DO) under the same experimental data from a shell and tube industrial heat exchanger is presented. The comparative analysis shows the performance of both methods when sensors and/or actuators fail. Different metrics are discussed (i.e. robustness, quick detection, isolability capacity, explanation facility, false alarm rates and multiple faults identifiability). DO showed quicker detection for sensor and actuator faults with lower false alarm rate. Also, DO can isolate multiple faults. DPCA required a minor training effort; however, it cannot identify two or more sequential faults.

## 1 INTRODUCTION

Early detection and diagnosis of abnormal events in industrial processes represent economic, social and environmental profits. Generally, the measuring and actuating elements of a control system fail causing abnormal events. Thus, when the process has a great quantity of sensors or actuators, the Fault Detection and Isolation (FDI) task is very difficult.

Most of the existing FDI approaches for Heat Exchangers (HE), are based on quantitative model-based methods. In (Ballé et al., 1997), fuzzy models are used to generate residuals; since each fault has an unique residual incidence, it is possible the fault isolation. Similarly, a residual generator is proposed to create fault signatures in (Krishnan and Pappa, 2005). Generalized Likelihood Ratio is frequently used to estimate the fault magnitude from a residual generation (Aitouche et al., 1998). On the other hand, a particle filtering approach for predicting the probability distribution of different heat exchanger states (faults) is proposed in (Morales-Menendez et al., 2003).

A comparative analysis between two FDI systems in an industrial HE is proposed. One of them is based on the Dynamic Principal Component Analysis (DPCA) and another one on a bank of Diagnostic Observers (DO).

Some researches are related to this work. Recently, DPCA and correspondence analysis (CA) have been compared (Detroja et al., 2005). CA shows a greater efficiency of fault detection in terms of the shorter detection delay and lower false alarm rates; however, CA needs a greater computational effort. An adaptive standardization of the DPCA has been proposed for MIMO systems (Mina and Verde, 2007); simulation results allow to detect faults and avoid normal variations in process signals.

An adaptive observer of a nonlinear discrete-time system with actuator faults is proposed in (Caccavale and Villani, 2004). Using process linear models, a dynamic observer detects malfunctions caused by measurement and modeling errors (Simmani and Patton, 2008). In order to detect multiple faults in a process, a set of unknown input-observers can be used, each one of them is sensitive to a fault while insensitive to the remaining faults (Verde, 2001).

The aforementioned works were implemented under different types of faults and processes; then, a comparison under same experimental data in an industrial HE is considered.

This paper is organized as follows: Section 2 formulates the DPCA approach. Section 3 describes the steps for designing a set of DO. Section 4 describes the experimental system. Section 5 discusses the results. Finally, conclusions of this work are presented.

## 2 DPCA

Let $\mathbb{X}$ be a matrix of $m$ observations and $n$ variables collected from a real process. This data set represents the normal operating conditions. $\underline{\mathbb{X}}$ is the scaled data matrix and $\bar{x}$ is a vector containing mean ($\mu$)



C. Tudon Martinez J., Morales-Menendez R. and E. Garza Castañón L. (2009).
FAULT DETECTION AND DIAGNOSIS IN A HEAT EXCHANGER.
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Intelligent Control Systems and Optimization*, pages 265-270
DOI: 10.5220/0002216402650270

of each variable. Such that $\bar{x} = (\frac{1}{m})\mathbb{X}^T\mathbf{1}$ and $\bar{\mathbb{X}} = (\mathbb{X} - \mathbf{1}\bar{x}^T)\mathbb{D}^{-1}$ where $\mathbb{D}$ is a diagonal matrix containing standard deviation ($\sigma$) of each variable and $\mathbf{1}$ is a vector of elements equal to 1.

When the system has a dynamic behavior, the data present a serial and cross-correlation among the variables. This violates the assumption of normality and statistical independence of the samples. To overcome these limitations, the column space of the data matrix $\mathbb{X}$ must be augmented with a few past observations for generating a static context of dynamic relations.

$$\bar{\mathbb{X}}_{\mathbb{D}} = \begin{array}{l}[X_1(t)X_1(t-1),\ldots,X_1(t-w),\ldots \\ X_n(t)X_n(t-1),\ldots,X_n(t-w)]\end{array} \quad (1)$$

where $w$ represents the quantity of time delays. By performing PCA on the augmented data matrix, a multivariate auto regressive model is extracted directly from the data (Ku et al., 1995). For a multivariate system, the process variables can have different ranges of values, thus the data matrix $\mathbb{X}_{\mathbb{D}[m\times(n\times w)]}$ must be standardized. With the scaled data matrix, a set of a smaller number ($r < n$) of variables is searched through the process of decomposing the variance in the data. $r$ must preserve most of the information given in these variances and covariances.

The dimensionality reduction is obtained by a set of orthogonal vectors, called *loading vectors* ($p$), which are obtained by solving an optimization problem involving maximization of the explained variance in the data matrix by each direction ($j$) with $t_j = \bar{\mathbb{X}} p_j$; the maximal variance of $t_j$ must be computed from:

$$max(t_j^T t_j) = max(p_j^T \bar{\mathbb{X}}^T \bar{\mathbb{X}} p_j) = max(p_j^T \mathbb{A} p_j) \quad (2)$$

Such that $p_j^T p_j = 1$. Solving the optimization problem through the Singular Value Decomposition (SVD), the eigenvalues $\lambda_j$ of the matrix $\mathbb{A}$ are computed from,

$$(\mathbb{A} - \lambda_j \mathbb{I})p_j = 0 \quad \text{for j=1,...,n} \quad (3)$$

where, $\mathbb{A}$ represents the correlation matrix of the data matrix $\bar{\mathbb{X}}$, and $\mathbb{I}$ is a $n \times n$ identity matrix. Using the new orthogonal coordinate system, the data matrix $\bar{\mathbb{X}}$ can be transformed into a new smaller data matrix $\mathbb{T}$, called *scores matrix*.

$$\mathbb{T}_{[m\times r]} = \bar{\mathbb{X}}_{[m\times n]} \mathbb{P}_{[n\times r]} \quad (4)$$

where, $\mathbb{P}$ represents the obtained loading vectors of the SVD with the most significant eigenvalues $\lambda_j$. As this transformation is a rotation matrix, it holds $\mathbb{P}^T\mathbb{P} = \mathbb{I}$. Therefore also $\bar{\mathbb{X}} = \mathbb{T}\mathbb{P}^T$ is valid. Thus, PCA decomposes the matrix $\bar{\mathbb{X}}$ as,

$$\bar{\mathbb{X}} = t_1 p_1^T + t_2 p_2^T + \ldots + t_r p_r^T \quad (5)$$

The matrix $\mathbb{T}$ can be back-transformed into the original data coordination system as,

$$\mathbb{X}^*{}_{[m\times n]} = \mathbb{T}_{[m\times r]} \mathbb{P}^T_{[r\times n]} \quad (6)$$

## 2.1 FDI using DPCA

The normal operating conditions can be characterized by $T^2$-statistic (Hotelling, 1993). Equation (7) allows to generate online the $T^2$-statistic based on the first $r$ loading vectors (*principal components*).

$$T^2 = x^T_{[1\times n]} \mathbb{P}_{[n\times r]} \Lambda^{-1}_{[r\times r]} \mathbb{P}^T_{[r\times n]} x_{[n\times 1]} \quad (7)$$

where, $x$ is a new measurement vector taken online and $\Lambda$ is a diagonal matrix which contains first $r$ eigenvalues of the correlation matrix ($\mathbb{A}$). If the value of $T^2$-statistic stays within its control limit then, the status of the process is considered normal (Ku et al., 1995). Thus, a fault occurs, when a value of $T^2$-statistic is greater than its control limit ($T^2_\alpha$).

$$T^2_\alpha = \frac{(m-1)r}{(m-r)} F_\alpha(r, m-r) \quad (8)$$

where, $F_\alpha(r,m-r)$ is the $F$-distribution with $r$ and $m-r$ degrees of freedom with 100$\alpha$% of confidence.

Due $T^2$-statistic only detects variation in the direction of the first $r$ principal components, Jackson *et al.* (Jackson and Mudholkar, 1979) propose to monitor the variation in the residual space (components associated with the smallest singular values) using $Q$-statistic for helping to fault detection. Both statistics must detect a fault, however they have not the same resolution in the deviation when the fault occurs. Similarly to $T^2$-statistic, when a value of $Q$-statistic is greater than its threshold ($Q_\alpha$) indicates the occurrence of a fault. The values of $Q$-statistic and its control limit can be calculated through the equations:

$$Q = [(\mathbb{I} - \mathbb{P}\mathbb{P}^T)x]^T[(\mathbb{I} - \mathbb{P}\mathbb{P}^T)x] \quad (9)$$

$$Q_\alpha = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2}\right]^{\frac{1}{h_0}} \quad (10)$$

where, $\theta_i = \sum_{j=r+1}^{n} (\lambda_j)^{2i}$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$ and $c_\alpha$ is the normal deviation corresponding to $(1-\alpha)$ percentile.

Once a fault is detected, the next step is the isolation. In order to determine which variable is the most relevant to cause the fault, the use of contribution plots has been proposed (Miller et al., 1998). Contribution plots quantify the error of each process variable when the process is not in normal operating conditions. The variable which shows the highest contribution ($Con_i$) to the error is isolated and associated as the most relevant to the fault which has occurred.

$$Con_i = \frac{R_i^2}{\sum_{j=1}^{r} R_j^2} \quad (11)$$

where, $R_i$ represents the residue in the residuals space (Isermann, 2006). The residue $\mathbb{R}$ can be calculated by subtracting the back-transformation data (equation 6) to scaled data matrix ($\underline{\mathbb{X}}$),

$$\mathbb{R}_{[m\times n]} = \overline{\mathbb{X}}_{[m\times n]} - \mathbb{T}_{[m\times r]}\mathbb{P}^T_{[r\times n]} \quad (12)$$

where $\mathbb{P}$ contains the loading vectors corresponding to components with the smallest singular values.

## 3 DESIGN OF A BANK OF DO

As the state observer compute the error between the process states and adjustable model states, it can be used as a further alternative for model-based fault detection. The discrete state space model which can describe the process dynamic is,

$$\begin{aligned} x_p(k+1) &= \mathbb{G}x_p(k) + \mathbb{H}u(k) \\ y(k) &= \mathbb{C}x_p(k) \end{aligned} \quad (13)$$

A state observer for unmeasurable state variables can be represented as

$$\begin{aligned} \tilde{x}_o(k+1) &= \mathbb{G}\tilde{x}_o(k) + \mathbb{H}u(k) + \mathbb{K}_e[y(k) - \hat{y}(k)] \\ \hat{y} &= \mathbb{C}\tilde{x}_o(k) \end{aligned} \quad (14)$$

where, $\mathbb{K}_e$ is the observer feedback matrix.

### 3.1 FDI using a Bank of DO

The error of the observer can be computed as:

$$x_p(k+1) - \tilde{x}_o(k+1) = (\mathbb{G} - \mathbb{K}_e\mathbb{C})[x_p - \tilde{x}_o] \quad (15)$$

Defining $e(k) = x_p - \tilde{x}_o$ as the error vector, the predicted error can be calculated as

$$e(k+1) = (\mathbb{G} - \mathbb{K}_e\mathbb{C})e(k) \quad (16)$$

The dynamic behavior of the error $e(k)$ is determined by the eigenvalues of $\mathbb{G}$-$\mathbb{K}_e\mathbb{C}$. If the matrix $\mathbb{G}$-$\mathbb{K}_e\mathbb{C}$ is a stable matrix, the error vector will converge to zero for any initial error $e(0)$.

When an unknown input (fault) changes the process normal operation, the error signal called *residual*, should be different to zero. Therefore, if the residual is close to zero (i.e. noise with $\mu = 0$ and $\sigma = 1$), the process variable is into its normal operating condition, called nominal behavior.

If the process is affected by several faults, it is possible to use a bank of DO for identification of different faults. All DO are designed from different fault models and they are sensitive to any fault except the used fault for their design.

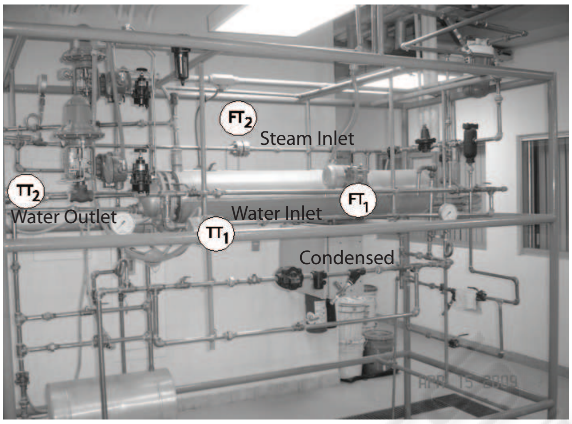


Figure 1: Experimental System.

## 4 EXPERIMENTAL SYSTEM

An industrial shell-tube heat exchanger is used, whose characteristics of non-linearity and slow transient response are the most relevant, see Figure 1.

Faults in sensors and actuators, called *soft faults*, have been implemented in additive form. Also, the process always was free of disturbances.

DPCA used 1 second as sample time delay; and 1900 measurement data of each sensor were taken.

$$x(t) = [FT_2(t)\ FT_1(t)\ TT_1(t)\ TT_2(t)] \quad (17)$$

where, $FT_1$ and $FT_2$ are flow transmitters and $TT_1$ and $TT_2$ are temperature transmitters.

In case of diagnostic observers, 5 seconds of sample time are used to obtain the state space models for each faulty condition. The observer feedback matrix in each observer is designed via pole placement with closed loop poles close to origin in the discrete space.

Four types of additive soft faults will be implemented: abrupt fault in sensors, gradual fault in sensors, abrupt faults in actuators and multiple faults in sensors, Table 1.

Table 1: Types of faults in the sensors.

| Sensor | Abrupt fault | Gradual fault (slope) |
|---|---|---|
| $FT_1$ | 6% (5σ) | $0.1\%/sec$ |
| $FT_2$ | 8% (5σ) | $0.1\%/sec$ |
| $TT_1$ | 2°C (8σ) | $0.1°C/sec$ |
| $TT_2$ | 2°C (8σ) | $0.1°C/sec$ |

Five types of faults were implemented in the steam and water control valves, Table 2.

Table 2: Types of faults in actuators.

| Case | Status of the steam valve | Status of the water valve |
|---|---|---|
| 0 | normal (70%) | normal (38%) |
| 1 | low pressure (60%) | normal (38%) |
| 2 | high pressure (80%) | normal (38%) |
| 3 | normal (70%) | low pressure (28%) |
| 4 | normal (70%) | high pressure (48%) |

# 5 RESULTS

## 5.1 DPCA Approach

Taking one sample time delay of each measurement, it is possible to explain a high quantity of variance including the possible auto and cross correlations. The normal operating conditions can be explained with 5 principal components (99.95%).

When an abrupt fault was implemented in the $TT_2$ sensor at time 105, the Figure 2(left plot) shows that $Q$ and $T^2$ statistics clearly overshoot their control limits. Figure 2(right lot) shows how the contribution plot helps correctly with the fault isolation. The 78% of total error corresponds to outlet temperature signal.

Figure 3 (left plot) shows a gradual fault in the $TT_2$ sensor at time 200. $Q$ and $T^2$ statistics overshoot their control limits and indicate the fault detection after 14 and 10 seconds respectively once the fault has occurred. Figure 3(right plot) shows that 64% of total error corresponds to outlet temperature signal.

For actuator faults, independently if the bias is positive or negative, there is a reaction in both statistics. When $T^2$ and $Q$ statistics overshoot their control limits, the fault is detected (Figure 4).

Using contribution plots, for the cases 1 and 2, the steam flow signal has the greatest error contribution followed by the outlet temperature signal (Figure 5). This result is right because the faults are associated to changes in the pressure of steam valve (negative and positive respectively). Similarly, the water flow signal has the greatest contribution to the error when the water valve is affected by a pressure change.

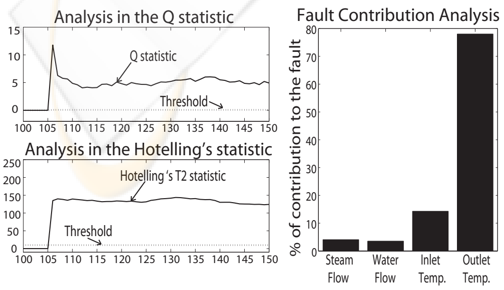


Figure 2: FDI analysis for an abrupt fault in the outlet temperature sensor using DPCA.

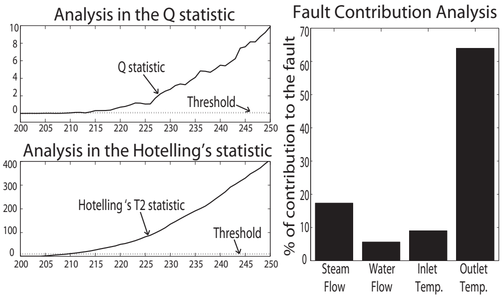


Figure 3: FDI analysis for a gradual fault in the outlet temperature sensor using DPCA.

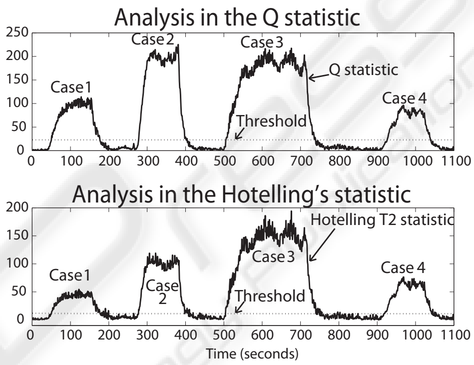


Figure 4: Fault detection for actuator faults using DPCA.

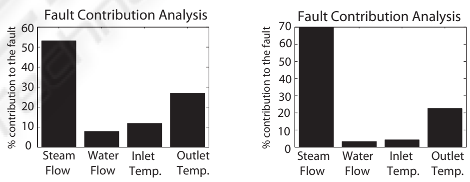


Figure 5: Results in actuators: case 1(left), case 2(right).

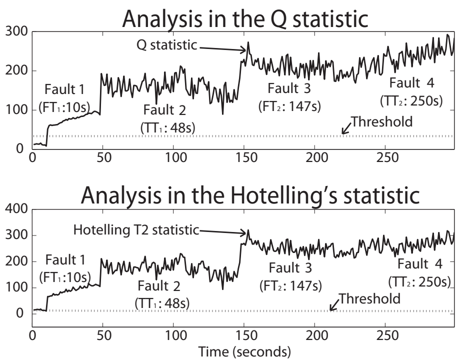


Figure 6: Fault detection using DPCA under multiple faults.

Finally, multiple faults have been activated sequentially at different times. Figure 6 shows the performance of DPCA; each fault presents its activation time. Both statistics overshoot their control limits when the fault 1 has occurred at time 10. When the remainder of the faults were introduced, the statistics stay inside their control limits; however, they move more away from their thresholds. None of the statistics comes back to its normal status since none of the faults was deactivated. Figure 7 shows that it is not possible to isolate multiple faults since contribution plots can not associate the error to a specific variable.

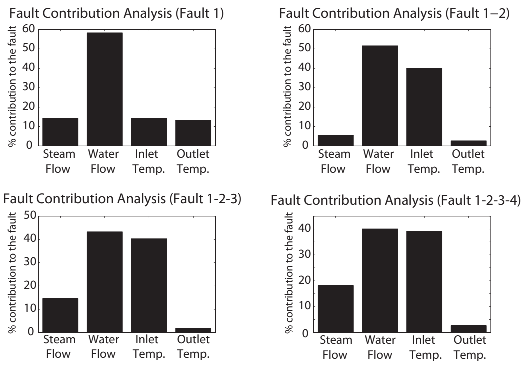


Figure 7: Diagnostic result for multiple faults in all sensors.

## 5.2 DO Approach

In order to distinguish different fault conditions, a bank of four DO was designed (i.e. water flow, steam flow, outlet temperature and inlet temperature).

When an abrupt fault is implemented in the $TT_2$ sensor, the outlet temperature residue is the unique signal which does not change its nominal behavior whereas the remainder of the residues are deviated negatively 1.5 units at time 10 when the fault is activated, Figure 8(top plot). Thus, it is possible to associate the fault to the $TT_2$ sensor. Same FDI result is obtained when a gradual fault is implemented in the $TT_2$ sensor. Figure 8(bottom plot) shows the fault detection after 5 seconds once the fault has occurred.

Figure 9 shows the performance of DO for faults in actuators Table 2. When is implemented a fault in the water control valve, independently of the bias direction, the water flow residue does not change its behavior from its nominal value; whereas, the remainder of the residues are deviated. Similarly, when is implemented a fault in the steam control valve, the steam flow residue does not change its behavior.

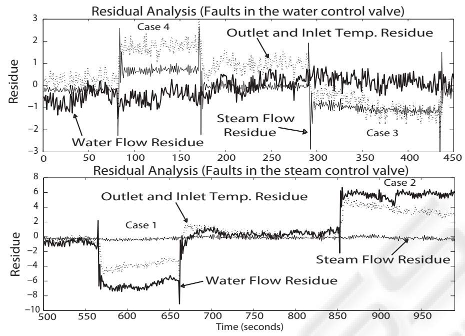


Figure 9: FDI analysis for actuator faults using the DO.

Figure 10 shows the FDD result using a set of DO when multiple faults have been activated sequentially at different time instants. It is important to note that only one signal is not deviated from its behavior when is introduced any abrupt sensor fault. The residual signal which does not change its behavior is associated to the occurred fault.

**Comparison of the Methods.** According to the Table 3, DO shows a quicker detection than DPCA when is implemented a gradual fault in a sensor signal. In this work, the gradual faults are added to a signal and only the deviations about the normal operating point are analyzed as residuals. In all fault cases, it is easy to explain the fault propagation using both FDI methods, i.e. the explanation facility metric is achieved.

Contribution plots indicate which variables are hypothetically more associated to the fault since it is possible that more fault cases are involved. On the other hand, a set of DO can correctly isolate a fault if all fault models and the model of the normal operating condition are known with high reliability. For faults in actuators, the normal operating conditions change in more than two sensors and the diagnosis task can

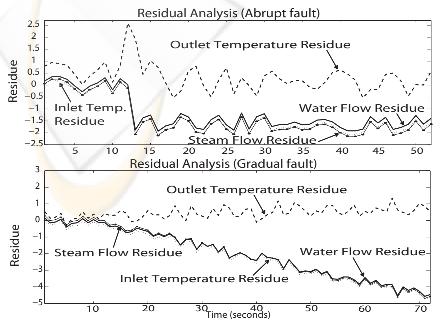


Figure 8: FDI analysis for an abrupt fault (top plot) and gradual fault (bottom plot) in the $TT_2$ sensor using DO.

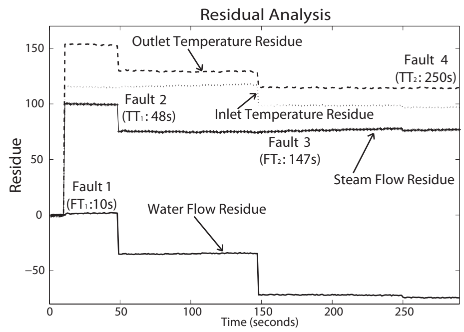


Figure 10: FDI result using DO under multiple faults.

Table 3: Comparison of DPCA and DO approaches.

| | DPCA | | | | DO | | | |
|---|---|---|---|---|---|---|---|---|
| Metrics | Abrupt Fault($TT_2$) | Gradual Fault($TT_2$) | Actuator faults | Multiple faults | Abrupt Fault($TT_2$) | Gradual Fault($TT_2$) | Actuator faults | Multiple faults |
| Detection (s) | 0 | $10-14$ | $9-18$ | 0 | 0 | 5 | $5-8$ | 0 |
| Isolation | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ |
| Explanation | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ |
| False alarm (%) | 0 | 0 | 14.13 | 0 | 0 | 0 | 9.94 | 0 |

be complicated. In this work, DO shows a quicker detection (i.e. almost the half of detection time) than DPCA when faults in both actuators are implemented at different times. For these faults, DO presents a lower false alarm rate than DPCA (Table 3).

On the other hand, both FDI methods can detect multiple faults which are implemented in all sensors. However, DPCA can not isolate correctly when several faults have been implemented. According to computational requirements, the design of DO needs greater computational resources. The training stage of this method is more complicated than the DPCA training; DO requires firstly a reliable ARX model which must be translated to a state space model. Furthermore, each fault case must be modeled in a particular state space model. Once the fault model is known with high reliability, is designed a state observer; particularly in this work all models (fault cases and normal operating) are obtained in parallel. On the other hand, the DPCA training is quickly executed once historical data of the normal operating point are known.

## 6 CONCLUSIONS

A comparison between the Dynamic Principal Component Analysis (DPCA) and a set of Diagnostic Observers (DO) under same experimental data from an industrial Heat Exchanger (HE) is presented. DPCA do very well on fast detection of abnormal situations, it is easier to implement in industrial applications. A process model was not required; however, a broad acquisition of the historical measurements is needed. Respect to false alarm rate, DPCA showed 42% more of false alarms than DO for actuator faults.

DO presents a quicker detection than DPCA ([4 – 10] seconds lower), DO requires an accurate state space model of the process. Furthermore, each fault case must be modeled. If the model is not reliable, DO can not detect a fault correctly. Due to HE is inherently a nonlinear system, it is more difficult to implement a FDI method based on quantitative models. Finally, DPCA can not identify multiple faults whereas DO can.

## REFERENCES


Aitouche, A., Maquin, D., and Busson, F. (1998). Multiple Sensor Fault Detection in Heat Exchanger Systems . In *Proc. of Int. Conf. on Ctrl. Appl.*, pages 741–745, Trieste, Italy.

Ballé, P., Fischer, M., Füssel, D., and Isermann, R. (1997). Integrated Control, Diagnosis and Reconfiguration of a Heat Exchanger . In *American Control Conference*, pages 922–926, Albuquerque, New Mexico.

Caccavale, F. and Villani, L. (2004). An Adaptive Observer for Fault Diagnosis Nonlinear Discrete-Time Systems. In *American Control Conference*, pages 2463–2468, Boston, Massachusetts.

Detroja, K., Gudi, R., and Patwardhan, S. (2005). Plant-wide Detection and Diagnosis using Correspondece Analysis. *Control Engineering Practice*.

Hotelling, H. (1993). Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.*, 24.

Isermann, R. (2006). *Fault-Diagnosis Systems*. Springer, Germany, $1^{st}$ edition.

Jackson, J. and Mudholkar, G. (1979). Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, 21:341–349.

Krishnan, R. and Pappa, N. (2005). Real Time Fault Diagnosis for a Heat Exchanger A Model Based Approach. In *IEEE Indicon Conference*, pages 78–82, Chennai, India.

Ku, W., Storer, R., and Georgakis, C. (1995). Disturbance Detection and Isolation by Dinamic Principal Components Analysis. *Chemometrics and Intelligent. Lab. Syst.*, 30:179–196.

Miller, P., Swanson, R., and Heckler, C. (1998). Contribution Plots: A Missing Link in Multivariate Quality Control. *Appl. Math. and Comp. Sci.*, 4(8):775–792.

Mina, J. and Verde, C. (2007). Fault Detection for MIMO Systems Integrating Multivariate Statistical Analysis and Identification Methods . In *American Control Conference*, pages 3234–3239, New York City, USA.

Morales-Menendez, R., Freitas, N. D., and Poole, D. (2003). State Estimation and Control of Industrial Processes using Particles Filters. In *American Control Conference*, pages 579–584, Denver, Colorado, USA.

Simmani, S. and Patton, R. (2008). Fault Diagnosis of an Industrial Gas Turbine prototype Using a System Identification Approach. *Ctrl. Eng. Prac.*, (16):769–786.

Verde, C. (2001). Multi-leak Detection and Isolation in Fluid Pipelines . *Control Eng. Practice*, 9:673–682.