8th International Conference of Modeling and Simulation - MOSIM'10 - May 10-12, 2010 - Hammamet - Tunisia
"Evaluation and optimization of innovative production systems of goods and services"

NEW PCA-BASED METHODOLOGY FOR SENSOR FAULT DETECTION AND LOCALIZATION

Anissa BENAICHA[1], Mohamed GUERFEL[2] Nasreddine BOUGUILA[1], Kamel BENOTHMAN[2]

[1] ATSI
National Engineering School of Monastir
Rue Ibn El Jazzar, 5019 Monastir - Tunisia
{ ben.aicha_anissa, nasreddinebouguila}@yahoo.fr

[2] LARA Automatique
National Engineering School of Tunis
BP 37, le Belvédère, 1002 Tunis - Tunisia
{guerfel_mohamed, kamelbenothman}@yahoo.fr

ABSTRACT: *This work proposes a new methodology for sensor fault detection and localization using principal component analysis (PCA). A new index is proposed in order to detect simple and multiple faults affecting the dependent and independent process variables. A new iterative selection method of principal component number is presented. This method determines a model allowing the detection of faults without a priori knowledge of their natures. The fault localization is carried out using hierarchical contribution plots applied to the proposed detection index. The performance of this approach becomes poor for a bad partitioning of the variables into blocs. A new partitioning method is proposed to identify correctly all the faults affecting the process. The whole proposed results were applied to a non linear noisy system subjected to simple and multiple faults.*

KEYWORDS: *PCA, principal component number, sensor fault, detection, localization, contribution plots.*

1 INTRODUCTION

The use of the multi statistical process control methods (MSPC) became frequent for the diagnosis of complex and over instrumented physical processes (chemical engineering, micro-electronics..., see (Venkatsubramanian et al., 2003), (Qin, 2003)). These methods are based on the construction of models obtained from the system historic in nominal operating mode. Principal Component Analysis (PCA) is among the most popular statistical methods. It was successfully used as a tool for sensor fault detection and localization (Qin, 2003), (Harkat et al., 2006), (Tharrault et al., 2008a), (Tamura and Tsujita, 2007), (Guerfel et al., 2009). The detection stage is related to the generation of residuals, also known as detection indices, which are signals that reveal the fault presence. Those indices are obtained from the analysis of the difference between the process measurments and their estimations obtained from the PCA model. There is a large variety of these indices (Ben Aicha, 2008) but most of them is insensitive to the independent variables faults (Tharrault, 2008b). In order to mitigate this disadvantage, a new index allowing the detection of faults affecting the dependent and independent variables is proposed in this work. The PCA based fault detection stage depends closely on the retained number of principal components. Most of the methods permitting the choice of this number (Valle, 1999) do not take into account the fault effect on the computed PCA model (Tamura and Tsujita, 2007). Inspired from the work of (Harkat et al., 2006), a new iterative method is proposed for the determination of the principal component number to retain in the PCA model. This method uses conjointly nominal operating data to identify the PCA model and faulty data in order to fix its structure. The fault localization can be carried out using many methods (Harkat et al., 2002). The contribution plots is a widely used localization method (Qin et al., 2001), (Kourti and MacGregor, 1996), (Westerhuis et al., 2000), (Qin, 2003), (Alcala and Qin, 2009). However, it can give wrong localizations for the simple and multiple fault case (Yue and Qin, 2001), (Harkat et al., 2002). In order to minimize the wrong localizations, a pertinent approach called hierarchical contribution plots was proposed (MacGregor et al., 1994). This approach consists in dividing the process variables into blocs. The computation of the contributions permits the localization of the faulty bloc(s). Those blocs are analysed in order to identify the faulty variable(s). The performance of this approach becomes poor for a bad partitioning of the variables into blocs. To avoid this risk, a new method of variable partitioning is proposed. It determines the blocs according to the variables contributions in the nominal operating mode. The application of this method on the proposed index permits to identify correctly the simple and multiple faults affecting the dependent and independent variables. The paper is organised as follows. Section 2 is a brief recall of the PCA principle. Section 3 is devoted to the proposition of a new fault detection index as well as its detectability conditions. The proposed criterion for the PCA model fixation is presented in the section 4. Section 5 presents the localization via the classical and hierarchical contribution plots in the case of the proposed index. A new method permitting the partitioning of the system variables into blocs is also presented. The last section illustrates the application of the previous sections results on a noisy non linear system affected with simple and multiple faults.

2 Principal Component Analysis

Let $x(k) = [x_1(k) \ldots\ x_m(k)]^T$ be a normalized vector, scaled to zero mean and unity variance, containing the m observed variables (inputs and/or outputs) in the instant k of the process in nominal operating mode. Modeling a process via PCA consists in seeking an optimal linear transformation (with respect to a variance criterion) of the vector $x(k)$ into a new one called $t\,(k)$ and defined as follows :

$$t\,(k) = P^T x\,(k) = \begin{bmatrix} t_1(k) & \ldots & t_m(k) \end{bmatrix}^T \qquad (1)$$

The quantities t_j; $j \in \{1, \ldots, m\}$, called principal components are uncorrelated and arranged in the decreasing variance order. The column vectors p_j of the matrix $P \in \mathbb{R}^{m\times m}$ represent the eigenvectors corresponding to the eigenvalues λ_j obtained from the diagonalization of the correlation matrix Σ of a data matrix X :

$$\Sigma = P\Lambda P^T \quad \text{with} \quad PP^T = P^T P = I_m \qquad (2)$$

The data matrix X is formed from the juxtaposition of $x(k)$ in different instants. It is defined as follows :

$$X = \begin{bmatrix} x(1) & \cdots & x(N) \end{bmatrix}^T \qquad (3)$$

The eigenvectors and eigenvalues matrices Λ and P are divided into two parts. The first part corresponds to the retained process variations and the second part is associated to the neglected process variations :

$$\Lambda = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & \tilde{\Lambda} \end{bmatrix} \qquad P = \begin{bmatrix} \hat{P} \mid \tilde{P} \end{bmatrix} \qquad (4)$$

The principal component vector $t\,(k)$ is also divided into two parts :

$$t\,(k) = \begin{bmatrix} \hat{t}(k) & \mid & \tilde{t}(k) \end{bmatrix}^T \qquad (5)$$
$$\text{with} \quad \hat{t}\,(k) = \hat{P}^T x\,(k) \;\; ; \;\; \tilde{t}\,(k) = \tilde{P}^T x\,(k)$$

where $\hat{t}(k) \in \mathbb{R}^{m-i}$ contains the significative process variations while $\tilde{t}(k) \in \mathbb{R}^i$ represents the quasi null linear combinations between the process variables $x_j(k)$. Thus, the data vector $x(k)$ is decomposed :

$$x\,(k) = \hat{x}\,(k) + e\,(k) \qquad (6)$$
$$\text{with} \quad \hat{x}\,(k) = \hat{C}x\,(k) \;\; ; \;\; e\,(k) = \tilde{C}x\,(k)$$

The vectors $\hat{x}\,(k)$ and $e\,(k)$ represent respectively the estimate and the error vectors obtained from the PCA model. The matrices $\hat{C} = \hat{P}\hat{P}^t$ and $\tilde{C} = I_m - \hat{C}$ form the PCA model of the process. Thus, this model divides the data space into two orthogonal subspaces : the principal subspace of representation formed by the $m-i$ first eigenvectors and the residual subspace formed by the i last eigenvectors.

3 Fault detection via PCA

The detection stage is based on the analysis of the residuals also called indicators or detection indices. Those indicators are obtained via the analysis of the error vector $e(k)$ or via the analysis of the estimates vector $\hat{x}(k)$. The process is declared in failure mode if the detection index is higher than its threshold. Among these indices, one finds the Hotelling T^2 statistic (Hotelling, 1933) and the squared prediction error SPE (Box, 1954). The index T^2 can only detect the shifts in the mean of the process variables (Guerfel et al., 2008). This statistic is able to detect faults affecting the independent variables and not able to detect weak magnitude faults that affect the dependent variables (Ben Aicha, 2008). The index SPE can only detect the changes in the correlations between the process variables (Harkat et al., 2006). Consequently, this index is not able to detect the faults on the independent variables but can detect the small magnitude faults affecting the dependent variables. This index is very sensitive to modeling errors and its performance closely depends on the choice of i. This work proposes a new detection index allowing the detection of dependent and independent faulty variables, which allows a good robustness with respect to the choice of i (Guerfel et al., 2008).

3.1 Proposed fault detection index

The proposed fault detection index G_i; $i \in \{1, \ldots, m\}$ is based on the sum of the squares of the i last principal components weighted by the inverse of their variances. Its expression is :

$$G_i(k) = x^T(k)Hx(k) = \left\| H^{1/2}x(k) \right\|^2 \qquad (7a)$$
$$= \sum_{h=m-i+1}^{m} \frac{t_h^2(k)}{\lambda_h} \qquad (7b)$$

where $H = \tilde{P}\tilde{\Lambda}^{-1}\tilde{P}^T \in \mathbb{R}^{m\times m}$ and $\|\cdot\|$ is the Frobenius norm.
The index G_i follows theoritically a *Chi-square* distribution with i liberty degrees $\chi^2_{i,\alpha}$, α represents the confidence limit. A fault is detected in an instant k, if :

$$G_i(k) > \chi^2_{i,\alpha} \qquad (8)$$

The index G_i represents an SWE (Oxby and Shah, 1998) computed with a PCA model formed by $m-i$

principal components. The SWE index is robust regarding the choice of i and permits the detection of weak magnitude faults affecting the process (Oxby and Shah, 1998). In the case where i is equal to m, the proposed index corresponds to a Mahalanobis distance D (Qin, 2003). This distance allows the detection of independent variable faults (Thaurrault, 2008b). Thus, the significant feature of the proposed index resides in the fact that it can be extended to cover different subspaces in the data space. Consequently, this index has the aptitude to detect any type of fault affecting the process. In some cases, it is not necessary to exploit all the data space because the augmentation of the detection threshold can prevent the detection of weak magnitude faults. Thus, an adequate selection of i adapted to the type of fault to detect can solve this compromise.

3.2 Faults detectability conditions

In the presence of n_d faults of magnitudes equal to $F \in \mathbb{R}^{n_d}; n_d \geq 1$, acting in the directions Ξ_d and affecting a subset d (of cardinal $= n_d$) of variables, the measurement vector $x(k)$ is written :

$$x(k) = x^*(k) + \Xi_d F(k) \tag{9}$$

where $x^*(k)$ designates the non faulty mesurement vector. The orthonormal matrix $\Xi_d \in \mathbb{R}^{m \times n_d}$ is formed with 0 to indicate a non faulty variables (respectively with 1 for faulty variables).
By replacing the equation (9) in (7a), one obtains :

$$G_i(k) = \left\| H^{1/2}(x^*(k) + \Xi_d F(k)) \right\|^2 \tag{10}$$

In the case where $\tilde{\Xi}_d = H^{1/2}\Xi_d = 0$, the faults are not detectable. Thus, the necessary detectability condition is :

$$\left\| \tilde{\Xi}_d \right\| = \sigma_{\max}(\tilde{\Xi}_d) \neq 0 \tag{11}$$

where $\sigma_{\max}(*)$ désignates the maximal singular value of the matrix $*$.
The application of the triangular inequality on the equation (10) leads to the following result :

$$\left\| H^{1/2}(x^*(k) + \Xi_d F(k)) \right\| \geq \left\| \tilde{\Xi}_d F(k) \right\| - \left\| H^{1/2} x^*(k) \right\| \tag{12}$$

As $\left\| H^{1/2} x^*(k) \right\| \leq \chi_{i,\alpha}$ and $G_i(k) > \chi^2_{i,\alpha}$, the sufficient fault detectability condition is :

$$\sigma_{\max}(\tilde{\Xi}_d F(k)) \geq 2\chi_{i,\alpha} \tag{13}$$

4 Proposed method for the fixation of the PCA model structure

The number of the retained prinicipal component (equal to $m - i$) has a significatif impact on the fault detection stage (Kano et al., 2002). Several criteria were proposed for the selection of this number (Valle et al., 1999), (Qin and Dunia, 2000). *Valle et al.* have demonstrated in (Valle et al., 1999) that the *VNR* criterion (Qin and Dunia, 2000) is the most interesting because it takes into account the redundancies between the process variables. However, this criterion does not take into account the influence of the faults on the choice of i and gives an average number i for all kind of faults affecting the process. To take into account the influence of faults on the choice of i, the authors in (Tamura and Tsujita, 2007) propose to determine this number according to the fault directions. However, the generalization of this method to multiple faults is unrealizable due to the great number of cases to take into account. All the methods cited previously use nominal operating data and determine off-line the number i. Inspired from the work of (Harkat et al., 2006), one proposes an iterative selection method. It uses conjointly nominal operating data to identify the PCA model and faulty data to fix i in an on-line mode. The principle of this method, illustrated on figure 1, is explained as follows :

1. Initialization $i = 1$.
2. Acquire nominal operation data. Compute X, Σ, Λ and P.
3. Compute $G_i(k)$, $\chi^2_{i,\alpha}$ from data in failure mode.
4. If $G_i(k) > \chi^2_{i,\alpha}$, go to 6. Else go to 5.
5. If $i < m$ then $i = i + 1$, go to 3. Else fault non detectable, go to 7.
6. Localization via contribution plots using $G_i(k)$.
7. End algorithm.

The proposed method can be applied in the simple and multiple fault case. This method determines the smallest number i which enables the fault detection for the first time without a priori knowledge on the fault(s), their direction(s) and their type(s). Its disadvantage lies in the use of faulty process data to ensure the choice of the PCA model structure.
In order to supress false alarms, the process is considered in failure mode ($G_i(k) > \chi^2_{i,\alpha}$), if $G_i(k)$ has shown six succeeding values larger than $\chi^2_{i,\alpha}$. The value "six" is determined in an empirical way and must be adjusted according to the treated application.

5 Localization via contribution plots

Three approaches can be used for fault localization via PCA. The first is based on the residual structuration (Gertler and Cao, 2004). The second uses a bench of models sensitive to a particular subset of faults (Huang et al., 2000), (stork et al., 1997), (Dunia and Qin, 1998). The third approach, treated

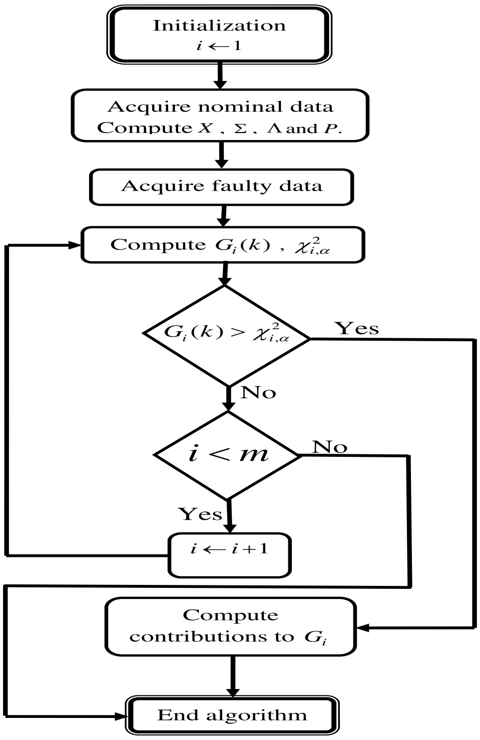

Figure 1: Algorithm of the proposed sensor fault detection and isolation method

in this work, is based on the computation of the contribution of different variables to the detection index (Westerhuis et al., 2000), (Qin, 2003), (MacGregor el al., 1994). The variables with the greatest contributions are suspected to be faulty. This method presents many inconvenients. On one hand, most of the works are based on the definition of the approximative contributions of the variables to the statistics SPE and T^2 (Kourti and MacGregor, 1996), (Harkat et al., 2002), (Qin et al., 2001). On the other hand, those contributions are sensitive to the variables amplitudes (Yue and Qin, 2001). The variables with the higher values have great probability to be suspected as faulty. To mitigate the disadvantages mentioned above, a pertinent approach called hierarchical contribution plots was proposed (MacGregor el al., 1994), (Qin et al., 2001). This approach divides the variables into multiple blocs based on the knowledge of the process. The hierarchical contribution plots gives better localization results compared to the classical approach (Qin, 2003). This section defines the classical and hierarchical contribution plots in the case of the proposed index G_i.

5.1 Classical contribution plots

Using the definition of *Alcala and Qin* (Alcala and Qin, 2009), the equation (7a) can be written :

$$G_i(k) = \sum_{j=1}^{m} (\xi_j^T H^{1/2} x(k))^2 = \sum_{j=1}^{m} cont_j(k) \quad (14)$$

The classical contribution of the j^{th} variable to the index G_i is :

$$cont_j(k) = (\xi_j^T H^{1/2} x(k))^2 \quad (15)$$

where ξ_j designates a column vector having 1 in the j^{th} position and 0 elsewhere.
The j^{th} variable is considered faulty, if :

$$\frac{cont_j(k)}{\tau_j^2} > 1 \quad (16)$$

The quantity τ_j^2 represents the threshold of $cont_j(k)$, the results of (Box, 1954) permit the determination of its coefficient r_j and its liberty degree b_j :

$$\tau_j^2 \equiv r_j \chi^2_{b_j,\alpha} \quad (17)$$

with

$$r_j = \frac{\xi_j^T H \Sigma H \xi_j}{\xi_j^T H \xi_j} \quad ; \quad b_j = 1 \quad (18)$$

5.2 Hierarchical contribution plots

The hierarchical contribution plots divides the m variables in n blocs containing m_b variables, $b \in \{1, \dots, n\}$. The $G_i(k)$ index can be written :

$$G_i(k) = \left\| \sum_{b=1}^{n} H_b^{1/2} x_b(k) \right\|^2 \quad (19)$$

The notation $x_b(k) = \begin{bmatrix} x_{b,1} & \dots & x_{b,m_b} \end{bmatrix}^T \in \mathbb{R}^{m_b}$ designates a vector containing the variables of the b^{th} bloc. The matrix H_b is given as follows :

$$H_b = \tilde{P}_b \tilde{\Lambda}^{-1} \tilde{P}_b^T \in \mathbb{R}^{m_b \times m_b} \quad (20)$$

where $\tilde{P}_b^T = \begin{bmatrix} p_{b,1} & \dots & p_{b,m_b} \end{bmatrix} \in \mathbb{R}^{i \times m_b}$ is the matrix formed via the juxtaposition of the eigenvectors $p_{b,s}; s \in \{1, \dots, m_b\}$ associated to the b^{th} bloc.
The expression of the contribution of the b^{th} bloc is :

$$cont_b(k) = \left\| H_b^{1/2} x_b(k) \right\|^2 = \left\| \sum_{s=1}^{m_b} H_{b,s}^{1/2} x_{b,s}(k) \right\|^2 \quad (21)$$

where $x_{b,s}(k)$ and $H_{b,s}$ represent respectively the s^{th} component of $x_b(k)$ and the s^{th} line of H_b.
The b^{th} bloc is considered faulty, if :

$$\frac{cont_b(k)}{\tau_b^2} > 1 \quad (22)$$

where τ_b^2 designates the threshold of $cont_b(k)$, it is given as follows :

$$\tau_b^2 \equiv r_b \chi^2_{b_b,\alpha} \quad (23)$$

with

$$r_b = \frac{trace\left\{(\Sigma_b H_b)^2\right\}}{trace\left\{\Sigma_b H_b\right\}} \; ; \; b_b = \frac{\left\{trace(\Sigma_b H_b)\right\}^2}{trace\left\{(\Sigma_b H_b)^2\right\}} \quad (24)$$

and $\Sigma_b = \mathrm{cov}(x_b^*(k))$.
If the b^{th} bloc is declared to be in failure, a contribution plots computation of its different variables is necessary in order to determine the origin of the fault. The s^{th} variable contribution to the b^{th} bloc is :

$$cont_{b,s}(k) = \left\| H_{b,s}^{1/2} x_{b,s}(k) \right\|^2 \quad (25a)$$

$$= \frac{\sum_{h=m-i+1}^{m} p_{b,hs}^2 x_{b,s}^2(k)}{\lambda_h} \quad (25b)$$

where $p_{b,hs}$ designates the h^{th} component of the vector $p_{b,s}$. The s^{th} variable of the b^{th} bloc is faulty, if :

$$\frac{cont_{b,s}(k)}{\tau_{b,s}^2} > 1 \quad (26)$$

The threshold $\tau_{b,s}^2$ is computed similarly to τ_b^2.

Remark The thresholds computed in a theoritical manner can be inadequate in the case of the classical and hierarchical contribution plots. For this reason, they can be adapted by training on nominal data.

5.3 Proposed method for the choice of blocs

The choice of the blocs and their variables has a major impact on the performance of the localization results. *Qin* (Qin, 2003) subdivides the different variables according to the physical knowledge of the process. In this work, one proposes a new partitioning method of the variables into blocs.
Let Z be a vector defined as follows :

$$Z = \frac{1}{l} \sum_{k=1}^{l} M(k) y(k) \quad (27a)$$

$$= \left[\ \overline{cont_1^*} \quad \ldots \quad \overline{cont_m^*}\ \right]^T \quad (27b)$$

where $y(k) = H^{1/2} x^*(k) \in \mathbb{R}^m$.
The notation $M(k) = diag(y_1(k) \ldots y_m(k))$ designates a diagonal matrix formed with the j^{th} components of $y(k)$.
The quantity $\overline{cont_j^*}$ represents the mean computed on l samples of the contribution of the j^{th} variable in nominal operating mode.
The partitioning of the m variables into n blocs is determined in a heuristic way according to the magnitude of the vector Z components (27b). The variables x_j associated to the great coefficients $\overline{cont_j^*}$ are gathered into a bloc, the variables corresponding to the average coefficients are gathered into another bloc and so on.
This partitioning method brings back the average contributions of variables of every bloc to the same order of magnitude in nominal operation. In the presence of faults in the b^{th} bloc, the contributions of all its variables increase with respect to their nominal values. The most significant augmentation is that of the faulty variables. This augmentation can be masked by the variables having the greatest contributions in nominal operation in the case where this partitioning is not respected. Thus, this partitioning method minimizes the probability of erroneous localizations.

6 Application

In order to test the efficiency of the new methodology for the detection and localization of simple and multiple faults affecting the dependent and independent variables, one considers the system described by the following equations :

$$\begin{cases} z_1^b(k) = \sin(k/4) + v(k)^2 + 1 \\ z_2^b(k) = \cos(k/4)^3 \exp(-k/N) \\ x_1^b(k) = z_1^b(k) + \varepsilon_1(k) \\ x_2^b(k) = z_2^b(k) + \varepsilon_2(k) \\ x_3^b(k) = z_2^b(k)^3 + \varepsilon_3(k) \\ x_4^b(k) = z_1^b(k) + z_2^b(k) + \varepsilon_4(k) \\ x_5^b(k) = z_2^b(k) - z_1^b(k) + \varepsilon_5(k) \\ x_6^b(k) = 2z_1^b(k) + z_2^b(k) + \varepsilon_6(k) \\ x_7^b(k) = z_1^b(k) + z_3^b(k) + \varepsilon_7(k) \\ x_8^b(k) = v_1(k) + \varepsilon_8(k) \end{cases} \quad (28)$$

where $x_j^b(k)\,;\ j \in \{1, \ldots, 8\}$ designate the system measurements (inputs or outputs). $z_1^b(k)$ and $z_2^b(k)$ designate the real system inputs. The quantities $v(k)$ and $v_1(k)$ designate random variables following a reduced centered normal law. The variables $\varepsilon_j(k)$ represent the measurements noise. They are obtained via realizations of random variables following centered normal law with steady deviation equal to 0.095. The system (28) presents linear and non linear redundancy relations as well as an independent variable ($x_8^b(k)$). The system is simulated a first time for $N = 500$ observations. The evolution of its measurements is illustrated on the figure 2.

After the centering and the reduction of the measurements, they are used to build the data matrix X via (3). The diagonalization of the correlation matrix of X permits the identification of its eigenvalues and its eigenvectors.

6.1 Simple fault case

The system (28) is simulated a second time on 500 samples and a fault is added to the dependent variable x_6^b since the sample 300 till the end of this simulation. This fault is represented by a constant bias of amplitude equal to 6% of the variation domain of x_6^b. The application of the proposed method for PCA model selection gives $i = 1$ (the fault is detected via G_1). The evolution of this index is illustrated on the figure 3. In order to identify the faulty variable, one applies the classical contribution plots on the index

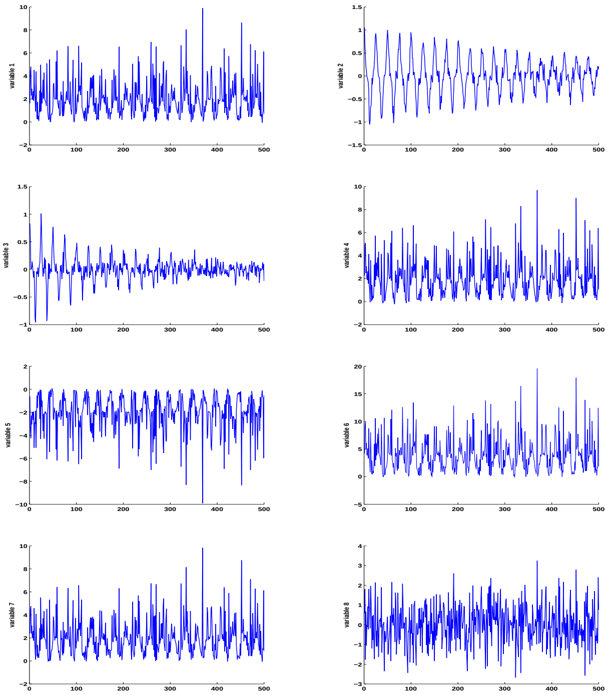

Figure 2: Evolution of the variables x_j^b

G_1. The left hand side of figure 4 shows the classical contribution of variables weighted by the inverse of their thresholds at the sample 400. This figure shows that the variables x_1, x_4, x_6 and x_7 are faulty but only x_6 is really responsible of the fault. The analysis of the coefficients of the matrix Z (27b), computed with $l = 20$ samples in nominal operation mode, permits the division of the system variables into three blocs. A bloc A containing the variables x_1, x_2 and x_3, a bloc B gathering the variables x_4, x_5 and x_6 and a bloc C formed by the variables x_7 and x_8. The right hand side of figure 4 illustrates the contribution of these three blocs weighted by the inverse of their thresholds at the sample 400. Only the contribution of the bloc B exceeds 1 which indicates that this bloc is responsible of the fault. The figure 5 presents the values of the normalized contribution of the different variables of the bloc B at the sample 400. This figure reveals that only x_6 is responsible of the fault. Thus, the hierarchical contribution plots permits to identify correctly the faulty variable contrary to the classical contribution plots.

6.1.1 Multiple fault case

A third simulation is carried out on 500 samples and three bias type faults affecting the dependent variables x_1^b, x_4^b (of amplitude equal to 10% of their variation domain) and the independent variable x_8^b (of amplitude equal to 150% of its variation domain) are simultaneously introduced from the sample 300 till the end. The application of the proposed method for

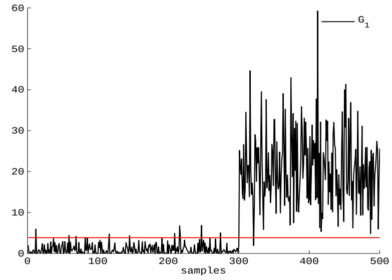

Figure 3: Evolution of G_1 in the second simulation case

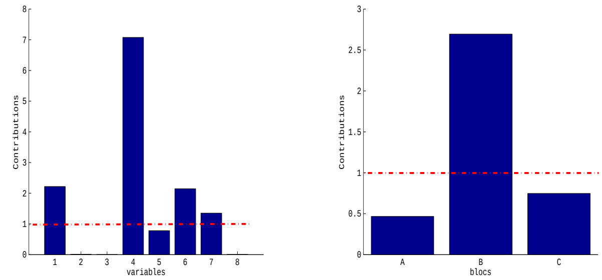

Figure 4: Classical and hierarchical contribution plots at the sample 400 (second simulation case)

PCA model selection gives $i = 1$ (the fault is detected via G_1). The evolution of this index is illustrated on the figure 6. The left hand side of figure 7 illustrates the value of the normalized classical contibutions obtained at the sample 400. This figure shows that the variables x_4, x_5 and x_6 are faulty. The bloc partitioning used in this case for the hierarchical contribution is the same as the one used in the second simulation. The values of the normalized bloc contributions at the sample 400 are shown on the right hand side of the figure 7. This figure shows that the blocs A, B and C contain faulty variables. The figure 8 illustrates the normalized contributions of each bloc variables. From this figure, one can conclude that the faulty variables are x_4, x_6 and x_8.

The performance of the hierarchical contribution

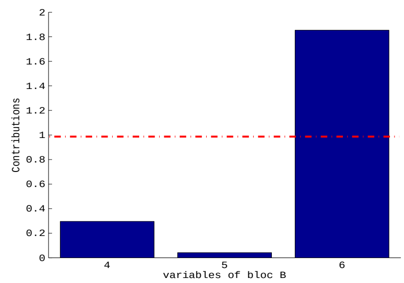

Figure 5: Contribution of the variables of the bloc B at the sample 400 (second simulation case)

plots can be degraded for a bad partitioning of the system variables into blocs. One considers the case where the system variables are splitted into three blocs as follows : a boc A_1 containing the variables x_1, x_5 and x_7, a bloc B_1 formed with x_2, x_3 and x_8 and a bloc C_1 gathering x_4 and x_6. The value of the normalized contribution of these blocs, in the third simulation case, at the sample 400 is shown on the figure 9. This figure shows that the blocs A_1 and C_1 are not responsible of the fault whereas A_1 contains the faulty variable x_1 and C_1 contains the faulty variable x_4.

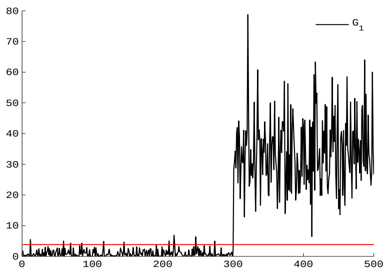

Figure 6: Evolution of G_1 in the third simulation case

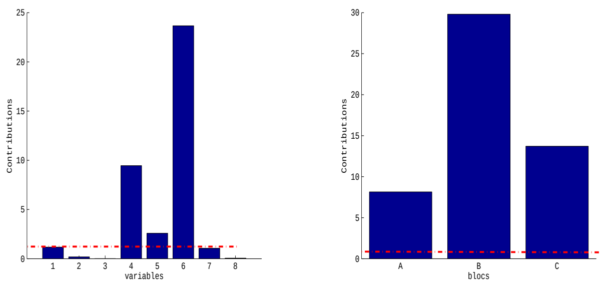

Figure 7: Classical and hierarchical contribution plots at the sample 400 (third simulation case)

7 Conclusion

The proposed index is able to detect the faults affecting the dependent and independent variables. A new iterative method is proposed to determine the structure of the PCA model. Unlike existing methods, this one determines the principal components number in an on-line mode. This modeling method allows the detection of the faults without a priori knowledge on their natures. The localization method, adopted in this work, uses the contribution plots applied to the proposed index. The hierarchical contribution plots is more efficient than the classical one in fault localization. However, the performance of the hierarchical contribution plots depends closely on the partitioning of the blocs and its variables. A partitioning, computed from the suggested method, permits the correct

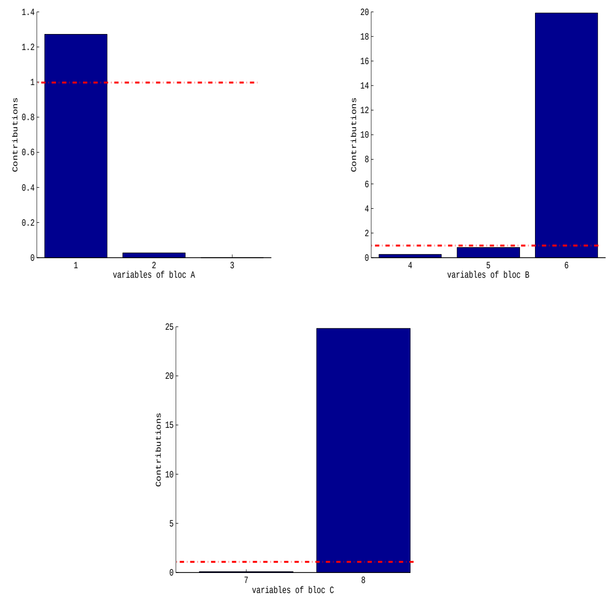

Figure 8: Contribution of the variables of the blocs A, B and C (third simulation case)

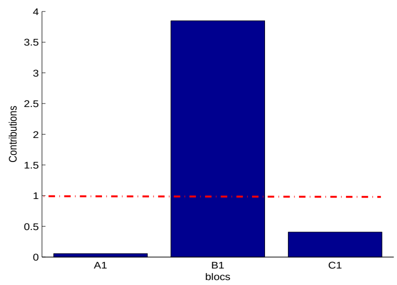

Figure 9: Contribution of the blocs A_1, B_1 and C_1 (third simulation case)

localization of simple and multiple faults affecting the dependent and independent variables.

REFERENCES

Alcala C. F. and Qin S. J., 2009. Reconstruction-based contribution for process monitoring. *Automatica*, vol. 45, p. 1593-1600.

Ben Aicha A., Juillet 2008. Diagnostic des systèmes par analyse en composantes principales ACP linéaire. *Master thesis of National Engineering School of Monastir*, Tunisia.

Box G. E. P., 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems: Effect of inequality of variance in one-way classification. *The Annals of Mathematical Statistics*, vol. 25, p. 290-302.

Dunia R. and Qin S.J., 1998, Subspace approach to multidimentional identification and reconstruction. *AIChE Journal*, vol. 44, p. 1813-1831.

Gertler J. and Cao J., 2004. PCA based fault diagnosis in the presence of control and dynamics. *AIChE Journal*, vol. 53, p. 388-402.

Guerfel M., Mourot G., Ragot J., Benrejeb M. and Benothman K., 2008. Comparaison des indices de détection de changements des modes de fonctionnement par ACP. Cas des indices basés sur l'estimation d'état. *1st International Workshop on Systems Engineering Design & Applications, SENDA'08*, Monastir, Tunisia 24-26 October.

Guerfel M., Ben Aicha A, Ben Othman K. and Benrejeb M., 2009. An improved principal component analysis scheme for sensor fault detection and isolation : application to a three tanks system. *17th Mediterranean Conference on Control and Automation Makedonia Palace*, Thessaloniki, Greece.

Harkat M. F., Mourot G. and Ragot J., 2002. Différentes méthodes de localisation de défauts basées sur les dernières composantes principales. *Conférence Internationale Francophone d'Automatique CIFA*, Nantes, France.

Harkat M. F., Mourot G. and Ragot J., 2006. An improved PCA scheme for sensor. FDI: application to an air quality monitoring network. *Journal of Process Control*, vol. 16, p. 625-634.

Huang Y., Gertler J. and McAvoy T., 2000, Sensor and actuator fault isolation by structured partial PCA with nonlinear extensions, *Journal of Process Control*, vol. 10, p. 444-459.

Hotelling H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, vol. 24, p. 417-441.

Kano M., Nagao K., Hasebe S., Hashimoto I., Ohno H., Strauss R. et al, 2002. Comparison of multivariate statistical process control monitoring methods with applications to the Eastman challenge problem. *Computers and Chemical Engineering*, vol. 26, p. 161-174.

Kourti T. and MacGregor J. F., 1996. Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, vol. 28, p. 409-428.

MacGregor J.F., Jaeckle C., Kiparissides C and Koutoudi M., 1994. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, vol. 40, p. 826-828.

Oxby P. and Shah S., 1998. A critique of the use of PCA for fault detection and diagnosis, *Technical report*, Matrikon, Edmonton, Alberta, Canada.

Qin S.J. and Dunia R., 2000. Determining the number of principal components for best reconstruction. *Journal of Process Control*, vol. 10, p. 245-250.

Qin S. J., Valle S. and Piovoso M., 2001. On unifying multiblock analysis with applications to decentralized process monitoring. *Journal of Chemometrics*, vol. 15, p. 715-742.

Qin S. J., 2003. Statistical process monitoring: basis and beyond. *Journal of Chemometrics*, vol. 17, p. 480-502.

Stork C. L., Veltkamp D. J. and Kowalski B. R., 1997. Identification of multiple sensor disturbances during process monitoring. *Analytical Chemistry*, vol. 69, p. 5031-5036.

Tamura M. and Tsujita S., 2007. A study on the number of principal components and sensitivity of fault detection using PCA. *Journal of Computers and chemical Engineering*, vol. 31, p. 1035-1046.

Tharrault Y., Mourot G., Ragot J. et Maquin D., 2008a. Fault detection and isolation with robust principal component analysis. *International Journal of Applied Mathematics and Computer Science AMCS*, vol. 18, p. 429-442.

Tharrault Y., 2008b. Diagnostic de fonctionnement par analyse en composantes principales : Application à une station de traitement des eaux usées. *PhD Thesis of National Polytechnic Institute of Lorraine.*

Valle S., Weihua L. and Qin S. J., 1999. Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. *Industrial and Engineering Chemistry Research*, vol. 38, p. 4389-4401.

Venkatsubramanian V., Rengaswamy R., Yin K., Kavuri S. N., 2003. A review of fault detection and diagnosis; part I: quantitative model-based methods. *Computers and Chemical Engineering*, vol. 27, p. 293-311.

Westerhuis J. A., Gurden S. and Smilde A., 2000. Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, vol. 51, p. 95-114.

Yue H. and Qin S. J., 2001. Reconstruction based fault identification using a combined index. *Industrial and Engineering Chemistry Research*, vol. 40, p. 4403-4414.